А К А Д Е М И Я Н А У К С С С Р
УРАЛЬСКИЙ ФИЛИАЛ

ВЫП. 37 ТРУДЫ ИНСТИТУТА БИОЛОГИИ 1963

Н. В. ЛУЧНИК

СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРОБЛЕМЫ АМИНОКИСЛОТНОГО КОДА

СВЕРДЛОВСК

А К А Д Е М И Я Н А У К С С С Р
УРАЛЬСКИЙ ФИЛИАЛ

ВЫП. 37 ТРУДЫ ИНСТИТУТА БИОЛОГИИ 1963

Н. В. ЛУЧНИК

СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРОБЛЕМЫ АМИНОКИСЛОТНОГО КОДА

СВЕРДЛОВСК

Печатается по постановлению
Редакционно-издательского совета
Уральского филиала АН СССР

ОТВЕТСТВЕННЫЙ РЕДАКТОР
доктор биол. наук проф. **С. С. Шварц**

ПРЕДИСЛОВИЕ

В конце 1961—начале 1962 гг. было опубликовано несколько работ (выполненных соединенными усилиями английских, американских, немецких и японских ученых), впервые проливших свет на один из важнейших вопросов биологии — на то, каким образом определяется первичная структура белков при их биосинтезе[1]. Интерес, вызванный этими работами среди самых широких кругов ученых, так же как и высказывавшееся мнение о том, что скоро эта проблема будет полностью решена, вполне понятны. Так, газета «Нью-Йорк Таймс» посвятила в номере от 3 февраля 1962 г. этим работам целую страницу, снабдив ее шапкой: «Успехи в биологии заставляют надеяться, что химические загадки наследственности будут разгаданы в этом году». Но год прошел, а окончательная разгадка «аминокислотного кода», как часто обозначают эту проблему, еще далека. Виттман, один из главных участников этих работ, в статье, опубликованной в феврале этого года [343], был вынужден констатировать, что «несмотря на оптимистические предсказания, высказывавшиеся год назад, мы еще довольно далеки от окончательного решения этой проблемы, и совершенно неизвестно, когда удастся полностью расшифровать код».

Автору этих строк, давно интересовавшемуся проблемой кода, бросалось в глаза, что во многих работах речь шла о выяснении конкретных частных вопросов без анализа и точного определения тех понятий, с которыми оперировали, и что при интерпретации результатов нередко принимали за достоверные некоторые положения, являвшиеся довольно гипотетичными. Благодаря этому, с одной стороны, выводы зачастую оказывались неточными, с другой — из экспериментальных данных извлекалась не вся заключенная в них информация. В частности, почти не применялись статистические методы, использование которых при решении подобных вопросов часто оказывается перспективным. Это побудило автора к ревизии принятых по-

---

[1] Основные из работ переведены на русский язык и изданы в виде отдельного сборника [44].

нятий и имеющихся данных, причем удалось укрепить некоторые из уже делавшихся утверждений, а также прийти к новым выводам об общих свойствах кода и его конкретной расшифровке. Описанию этих работ и посвящена настоящая небольшая монография. К сожалению, в ней нет новых экспериментальных данных. Это обусловлено тем, что автор был лишен какой бы то ни было возможности ставить опыты в этой области, в связи с оторванностью от приспособленных к этому лабораторий, и, хотя в ряде случаев постановка тех или иных опытов напрашивалась сама собой, они оставались неосуществленными (некоторые из них уже выполнены независимо другими авторами).

Поскольку современной биологией и, в частности, проблемой кода в настоящее время интересуются очень многие, автор старался сделать книгу понятной для возможно большего круга читателей. Для этого были написаны две первые главы, дающие краткий обзор современного состояния вопросов о механизме биосинтеза белка и об аминокислотном коде. Читатель, знакомый с этими вопросами хотя бы по обзорным статьям [1, 2, 25, 34, 43, 44, 58], может без ущерба для дальнейшего эти главы опустить. После подробного описания собственных работ, в последней главе дается их сжатое изложение, по которому, не теряя общей нити, легко проследить за ходом рассуждений и получаемыми выводами.

Развитие молекулярной биологии происходит так быстро, что не исключена возможность того, что еще до выхода этой книги в свет проблема кода будет полностью решена, может быть, и совершенно неожиданным образом. Нужно заметить, что за время подготовки этой работы появилось несколько статей, в которых высказывались некоторые сходные мысли, или приводились данные, подтверждавшие наши выводы, что заставляло делать соответствующие изменения и добавления. Некоторая возникавшая в такой обстановке спешка не могла не вызвать недочетов, которых можно было бы избежать. Автор заранее благодарит за указание ошибок и другие критические замечания.

Считаю приятным долгом поблагодарить П. С. Зырянова, Г. Г. Талуца и Н. В. Тимофеева-Ресовского за ценные советы, С. И. Ворошилина, Н. В. Глотова, В. Мархасина, А. П. Никифорова, А. А. Позолотина и Н. А. Порядкову за активную и бескорыстную помощь в подготовке рукописи к печати, а последнюю также за изготовление иллюстраций и постоянную моральную поддержку.

ГЛАВА 1

СОВРЕМЕННЫЕ ПРЕДСТАВЛЕНИЯ О МЕХАНИЗМЕ БИОСИНТЕЗА БЕЛКА

ПРОБЛЕМА БИОСИНТЕЗА БЕЛКА

Значение белков в явлениях жизни общеизвестно. Они являются основными веществами, из которых построены живые тела; например, в организме человека белки составляют около 45% сухого веса, а в отдельных органах их содержание превышает 80% [23]. Но что еще более существенно — все биологически активные вещества либо сами по себе являются белками (такие гормоны, как инсулин, вазопрессин, такие ферменты, как рибонуклеаза, трипсин и многие другие являются чистыми белками), либо их синтез происходит под контролем белковых ферментов. Развитие биологической науки полностью подтвердило положение Ф. Энгельса о том, что «жизнь есть способ существования белковых тел, и этот способ существования состоит по своему существу в постоянном самообновлении химических составных частей этих тел» (Ф. Энгельс, Анти-Дюринг, Госполитиздат, 1950, стр. 77). Можно без преувеличения сказать, что проблема изучения физико-химических основ жизненных явлений в первую очередь сводится к вопросу о том, каким образом происходит самообновление белковых тел, или говоря более специальным языком, каковы механизмы биосинтеза специфических белковых молекул.

Строение белков. Белки представляют собой высокополимерные вещества с молекулярным весом, как правило, порядка от десятков тысяч до нескольких миллионов. В отличие от других биополимеров (крахмал, гликоген, целлюлоза), являющихся гомополимерами, т. е. состоящих из одинаковых элементарных единиц[1], или периодических гетерополимеров, в которых составные части правильно чередуются, белки являются апериодическими гетерополимерами, что и обуславливает их исключительное разнообразие. Элементарной единицей бел-

[1] Например, при гидролизе перечисленных веществ получается единственный продукт — глюкоза.

ка являются α-аминокислоты, т. е. органические кислоты, у которых к одному и тому же углеродному атому присоединены аминная и карбоксильная группы. Число различных аминокислот, участвующих в построении белков, равно 20. Во всех организмах и почти во всех достаточно крупных белках обнаружи-

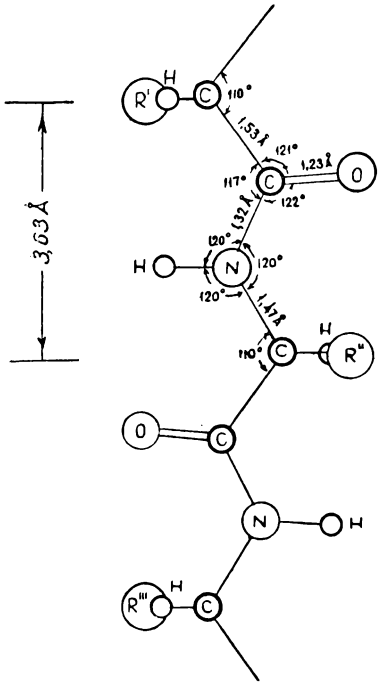

Рис. 1. Схема полностью растянутой полипептидной цепи. Показаны межатомные расстояния и валентные углы, вычисленные на основании данных рентгеноструктурного анализа. По Анфинсену [1].

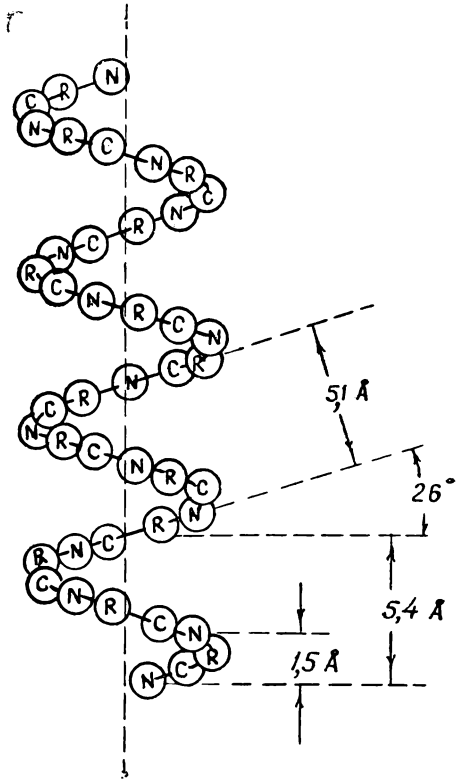

Рис. 2. Схема α-спирали белка. Для большей ясности показаны не все атомы. По Анфинсену [1].

ваются: аланин, аргинин, аспарагиновая кислота, аспарагин, валин, гистидин, глицин, глютаминовая кислота, глютамин, изолейцин, лейцин, лизин, метионин, пролин, серин, тирозин, треонин, триптофан, фенилаланин и цистеин.

В действительности, число аминокислот, встречаемых в природе, в том числе входящих в состав белков, больше 20. Не все из них входят в состав белков; например, такие промежуточные

соединения как орнитин и цитруллин никогда не обнаруживаются в составе белков. Такие, как цистин, оксилизин, оксипролин, фосфосерин, тироксин и другие, несомненно, входят в состав белков. Но если перечисленные выше 20 аминокислот входят в состав большинства белков, остальные встречаются (кроме цистина) только в немногих. Считают, что эти «лишние» аминокислоты возникают из перечисленных двадцати в результате их окисления, фосфорилирования и т. п., уже после их включения в белковую цепь, что подтверждается убедительными экспериментами. Так, в опытах с мечеными атомами было показано, что предшественником окисилизина в белках является лизин [275], тогда как сам оксилизин непосредственно в белок не включается [274]. Таким образом, хотя число аминокислот, входящих в белки, больше двадцати, но в построении белковых молекул участвует, по-видимому, только 20 аминокислот. Подробно этот вопрос обсужден в обзоре Ичаса [25].

По предложению Линденстром-Ланга [226 а], в белках различают первичную, вторичную и третичную структуры. Под первичной структурой понимают порядок чередования аминокислот в полипептидной цепи. Схематически такая полностью растянутая цепь изображена на рис. 1, где показаны основные размеры и углы. В связи с закономерным распределением амидных звеньев (CONH), между которыми образуются водородные связи, полипептидная цепь принимает спиральную конфигурацию (255, 256), образуя так называемую α-спираль (рис. 2). Эта спираль носит название вторичной структуры. Наконец, белковые цепи могут иметь третичную структуру, давая определенное и подчас очень сложное пространственное расположение полипептидных цепей. В качестве примера такой третичной структуры на рис. 3 изображена пространственная модель миоглобина. Как вторичная структура связана с закономерностями первичной, точно так же и третичная структура, вероятно, в большинстве случаев определяется порядком чередования аминокислот в первичной цепи. Так, локализация дисульфидных мостиков между спиралями определяется положением остатков цистеина, водородные связи между ними — тирозином, углы поворота — пролином, гистидин обеспечивает связь с простетической группой или имеет другое значение для ферментативной активности [128а]. Поэтому вопрос о биосинтезе белков в первую очередь сводится к выяснению механизмов, определяющих их первичную структуру. При выяснении этого механизма всегда нужно иметь в виду, что структура белковой молекулы воспроизводится с исключительной степенью точности. Все молекулы белка данного сорта, насколько это позволяют установить современные методы анализа, имеют не только совершенно одинаковый состав, но и идентичное чередование аминокислотных остатков. Различия между разными белками

иногда оказываются минимальными. Так, большинство мутантных форм гемоглобина отличается от нормального гемоглобина замещением лишь единственной аминокислоты [1, 24, 65].

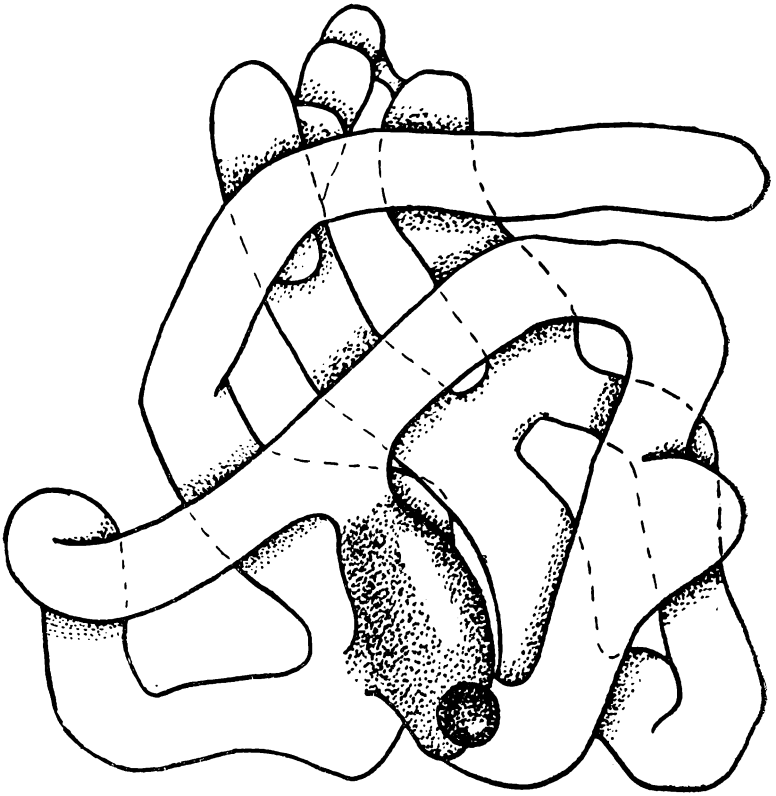

Рис. 3. Трехмерная модель молекулы миоглобина. Темная часть модели — простетическая группа. По Анфинсену [1].

Межвидовые различия белков также, как правило, незначительны: например, в молекуле инсулина, состоящей из 51 аминокислоты, межвидовые различия заключены лишь в небольшом участке между восьмым и десятым аминокислотными остатками А-цепи [1, 287].

Проблема биосинтеза белка. При образовании неорганических веществ результат реакции определяется термодинамическими условиями: получаются продукты, которым в данных условиях соответствует минимум свободной энергии. При синтезе мелких органических молекул, в связи с их большим разнообразием, приобретает исключительно важную роль катализатор (фермент), благодаря которому из нескольких

энергетически одинаково возможных веществ синтезируется лишь одно вполне определенное. Таким же образом происходит и синтез гомополимеров. Что же касается белков, то их чрезвычайная сложность и разнообразие, а также их центральная роль в процессах обмена делают невероятным предположение о том, что их синтез следует только что перечисленным механизмам. Действительно, с энергетической точки зрения, любые последовательности аминокислот равновероятны. Если бы первичная структура белков определялась специальными ферментами такого же рода, как те, которые определяют синтез других веществ, то это потребовало бы существования для каждого белка большого числа ферментов, имеющих белковую природу, специфика которых также должна была бы определяться большим числом ферментов второго порядка и т. д. Идя по этому пути, мы неизбежно приходим к абсурду, так как при этом микроскопически малая клетка должна была бы вмещать в себя бесконечно большое число разных ферментов.

Логично предположить, что в основе биосинтеза белка лежит явление ауторепродукции, т. е. либо белковые молекулы сами обладают свойством строить точно такие же молекулы, либо этим свойством обладают какие-то другие структуры, определяющие, в конечном счете, первичную структуру белков. Необходимость существования в клетке ауторепродуцирующихся структур постулировалась давно и главным образом в связи с явлениями наследственности. Впервые эта проблема в ее полном виде была сформулирована, и притом с поразительной ясностью, еще в 1927 г. Н. К. Кольцовым [28, 29]. Схема, которая была дана в его «Физико-химических основах морфологии», оказалась в своей основе совершенно правильной, если не говорить о второстепенных деталях, о выяснении которых в те времена, конечно, не могло быть речи. Приоритет генетиков в постановке и принципиальном решении этой проблемы, конечно, не случаен. Уже первые попытки конкретизации понятия гена и известные в то время факты из области цитогенетики с неизбежностью должны были привести к представлению об ауторепродукцирующихся структурах, локализованных в хромосомах клеточного ядра, с которыми причинно связаны все фенотипические признаки. Правда, должно было казаться странным, что гены локализованы в клеточном ядре, в то время как синтез белка происходит в цитоплазматических гранулах. Но это затруднение не привлекало серьезного внимания до тех пор, пока генетические исследования ограничивались анализом наследования чисто морфологических признаков. Вероятно, именно поэтому большинство биохимиков долгое время мало интересовалось генетикой. Однако с появлением биохимической генетики стало ясно, что биохимические признаки подчинены тем же самым законам наследственности, что и любые другие [11]. Это

оказалось справедливым и для строения белков. Например, подробное исследование одного из наследственных заболеваний крови — серповидноклеточной анемии, — с одной стороны, показало, что эта болезнь является менделирующим признаком, связанным с изменением одного гена, а с другой, что при ней происходит изменение первичной структуры белка гемоглобина. Интересно заметить, что это изменение, приводящее к столь драматическому эффекту (в семьях, больных этой формой анемии, четвертая часть детей оказывается нежизнеспособной), связано с тончайшим изменением белковой структуры [65]. При этой болезни оказывается измененным один единственный из 287 аминокислотных остатков гемоглобина: на шестом месте в β-цепи вместо остатка глютаминовой кислоты находится остаток валина [191, 195, 196].

В последнее время проблема хранения и передачи наследственной информации о первичной структуре белков стала привлекать к себе внимание биохимиков не меньше, чем генетиков. Вместе с тем, в связи с возникновением кибернетики — науки об управляющих системах в технике и живой природе [12, 26, 49, 66] — этой проблемой заинтересовались многочисленные круги математиков и физиков, близких к кибернетике, так как система хранения, передачи и использования внутриклеточной информации, по-видимому, является наиболее совершенным из известных до сих пор кибернетических устройств. Поэтому к этой проблеме нередко подходят с точки зрения теории информации [15, 43, 345].

Роль нуклеиновых кислот в синтезе белка. То что нуклеиновые кислоты представляют собой обязательную составную часть клетки, стало известно еще в конце прошлого века [238]. Но их биологическая роль до недавнего времени оставалась неясной, хотя в косвенных указаниях на их участие в белковом синтезе не было недостатка. Так, еще в конце 30-х — начале 40-х годов Касперссон [111, 112] и Браше [92, 93, 94] на основании цитохимических исследований заключили, что содержание рибонуклеиновой кислоты в клетках коррелирует с интенсивностью в них белкового синтеза. Позже стали появляться и более прямые указания. Была выяснена химическая природа фактора трансформации пневмококков. Еще в 1928 г. было обнаружено, что вытяжка из бактерий одного штамма вызывает массовое наследственное изменение другого. Долгое время механизм этого явления был непонятен, пока в 1944 г. не выяснилось, что этот эффект обусловлен присутствовавшей в вытяжках дезоксирибонуклеиновой кислотой [76]. При исследовании механизма заражения бактерий бактериофагом установили, что при этом внутрь бактериальной клетки проникает только нуклеиновая кислота, в результате чего строятся новые частицы фага, содержащие и свой специфичный белок [183].

Сходные результаты получались также на растительных и животных вирусах [161]: введение чистой нуклеиновой кислоты приводило к развитию инфекции и появлению вирусных частиц со своим специфичным белком. Правда, что касается всех этих и ряда других подобных опытов, то у скептиков всегда оставался простор для возражений, так как ни в одном случае нельзя было полностью исключить возможность небольших субаналитических загрязнений нуклеиновой кислоты белком, которому можно было при желании приписывать основную роль. Однако в 1961 г. М. Ниренберг и Г. Маттеи [46, 47, 250, 251] провели опыты, не допускающие возможности других толкований. Была создана в пробирке система, синтезирующая белковые вещества в присутствии полимерной нуклеиновой кислоты. Здесь особенно важно подчеркнуть два обстоятельства: первое — специфика синтезируемого продукта определяется только спецификой присутствующей в системе нуклеиновой кислоты; так, внесение нуклеиновой кислоты вируса мозаичной болезни табака в систему, созданную из разрушенных клеток кишечной бактерии, ведет к образованию продукта, подобного вирусному белку [309]. Второй важный факт — то, что эта реакция синтеза идет в присутствии не только естественных, но и искусственных нуклеиновых кислот, чистоту и состав которых можно контролировать вполне точно, и которые, конечно, вполне свободны от специфического белка. Вряд ли после этих опытов, подтвержденных и развитых дальше другими авторами [75, 102, 146, 219, 232, 328], может оставаться сомнение в том, что нуклеиновые кислоты должны играть очень важную роль в определении первичной структуры белка.

Состав нуклеиновых кислот. Упомянутые выше нуклеиновые кислоты являются, подобно белкам, высокополимерными веществами, молекулярный вес которых может достигать нескольких миллионов. Так же, как и белки, нуклеиновые кислоты являются гетерополимерами. Но если белки состоят из 20 различных аминокислот, то каждая нуклеиновая кислота построена только из четырех разных нуклеотидов. Каждый нуклеотид состоит из остатка фосфорной кислоты, пентозного сахара и азотистого основания пуринового или пиримидинового типа. Различают два основных типа нуклеиновых кислот — дезоксирибонуклеиновую (ДНК) и рибонуклеиновую (РНК). В клетках, имеющих ядро, почти вся ДНК локализована в хромосомах клеточного ядра, а РНК — преимущественно в ядрышке и цитоплазме. Состав ДНК и РНК сходен — они отличаются характером входящего в них сахара (в ДНК это — дезоксирибоза, а в РНК — d-рибоза) и одним из азотистых оснований. Из азотистых оснований в дезоксирибонуклеотиды входят аденин (А), гуанин (Г), цитозин (Ц) и тимин (Т), в рибонуклеотидах тоже присутствует аденин, гуанин и цитозин,

но место тимина в них занимает урацил (У). Четыре типа оснований содержатся в нуклеиновых кислотах в близких количествах; кроме них в значительно меньших количествах присутствуют некоторые другие основания (5-метилцитозин, псевдоуридин и др.), но так как об их роли пока ничего достоверного неизвестно, мы о них говорить не будем.

Относительно строения нуклеиновых кислот долгое время господствовала тетрануклеотидная теория [171], предполагавшая, что нуклеиновая кислота построена из тетрануклеотидов, в каждый из которых входят четыре разных нуклеотида. Таким образом не оставалось места для разнообразия и специфичности веществ этой группы; считали, что все ДНК одинаковы, так же как и все РНК. Однако с уточнением аналитических методов было показано, что соотношение оснований в нуклеиновых кислотах достоверно отличается от отношения 1 : 1 : 1 : 1, требуемого тетрануклеотидной теорией. В результате многочисленных работ, проведенных в лабораториях А. Н. Белозерского [3, 5, 6] и Е. Чаргафа [115, 116, 117, 317], было установлено, что состав нуклеиновых кислот является специфичным для разных видов и, вместе с тем, очень константным. После этого, в связи с очень высоким молекулярным весом нуклеиновых кислот, стало ясно, что в них мы имеем дело с классом соединений, разнообразие и специфичность которых вполне сравнимы с разнообразием и специфичностью белков.

Выше были приведены только самые элементарные сведения о составе нуклеиновых кислот. Для интересующихся можно указать на ряд обзоров и монографий по химии нуклеиновых кислот [4, 6, 20, 52, 61].

Комплементарная модель ДНК. Давно казалось ясным, что в основе явлений жизни должны лежать управляющие системы, обладающие способностью к самовоспроизведению. Вначале думали, что такой способностью должны обладать белки [28]. Но поиски ауторепродуцирующихся структур среди белков остались бесплодными. Иногда создавалось впечатление, что белки обладают способностью к ауторепродукции. Так, после того, как выяснилось, что пищеварительный фермент трипсин образуется из своего предшественника трипсиногена только в присутствии готовых молекул трипсина, складывалось впечатление, что новые молекулы строятся по образцу старой из более мелких субъединиц. Однако дальнейшее исследование показало, что молекула трипсиногена больше молекулы трипсина и ее активация происходит в результате частичного разрушения, причем видовая специфичность продукта определяется природой трипсиногена, а не присутствующего трипсина [45, 249].

Но с 1953 г. появились основания для того, чтобы считать, что способностью к ауторепродукции могут обладать макромо-

лекулы ДНК. Дж. Уотсон и Ф. Крик [33, 126, 326] в результате анализа рентгеноструктурных снимков ДНК [151, 336] выяснили, что ее молекулы являются двойными спиралями, причем каждая из спиралей представляет собой сахаро-фосфатную нить, а азотистые основания обращены внутрь структуры и связаны друг с другом водородными связями, за счет которых и соединены друг с другом две части спирали. К аналогичным выводам приводят также данные по инфракрасной спектроскопии ДНК [296]. Эта модель схематически изображена на рис. 4. Сопоставление размеров спирали с величиной пуринов и пиримидинов, а также расположение в последних атомов азота и кислорода приводят к выводу о том, что водородные мостики могут образоваться только между А и Т и между Ц и Г. Таким образом, последовательности нуклеотидов в каждой из двух спиралей должны быть комплементарными и любая одна из этих последовательностей однозначно определяет другую.

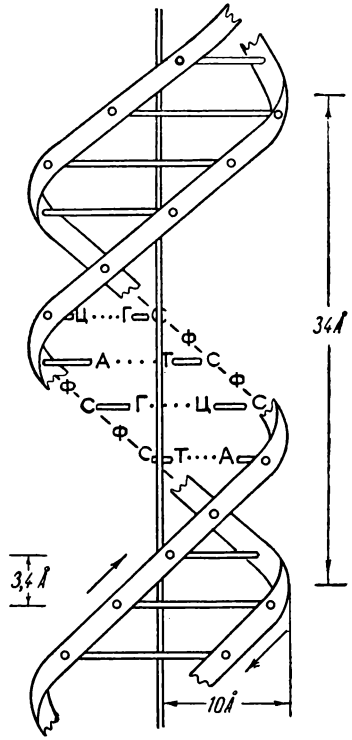

Рис. 4. Схема комплементарной модели макромолекулы ДНК, по Уотсону и Крику [326].

Это приводит к интересной гипотезе, высказанной тогда же Уотсоном и Криком. Если представить себе, что полинуклеотидные нити, образующие двойную спираль ДНК, разъединились, и возле каждой из них начала строиться новая нить, то после того, как этот процесс закончится, должны возникнуть две двойные спирали, совершенно идентичные исходной. Таким образом, казалось весьма вероятным, что новые молекулы ДНК возникают путем ауторепродукции с помощью матричного механизма. Это казалось тем более правдоподобным, что хромосомы, в состав которых ДНК входит в качестве одной из главных составных частей, являются внутриклеточными органеллами, для которых способность к ауторепродукции казалась ясной уже давно [129]. Поэтому ауторепродукцию ДНК можно рассматривать как молекулярную основу ауторепродукции хромосом.

Сразу же вслед за выяснением макромолекулярной структуры ДНК было высказано предположение о том, что последо-

вательность нуклеотидов в ДНК может определять последовательность аминокислот в белках. Г. Гамов [153, 154] обратил внимание на то, что расстояние между соседними нуклеотидами в ДНК равно расстоянию между соседними аминокислотами в полипептидной цепи. Он высказал гипотезу, согласно которой аминокислота определяется четырьмя нуклеотидами, в число которых входит пара А — Т или Г — Ц и по одному нуклеотиду из верхней и нижней пары (рис. 5). При этом получается ровно 20 разных «ромбов», что точно соответствует числу подлежащих кодированию аминокислот. Хотя этот так называемый ромбический код был вскоре опровергнут [99], так же как и мысль, что последовательность нуклеотидов в ДНК непосредственно определяет последовательность аминокислот, но основная идея о том, что информация о первичной структуре белков закодирована в виде последовательности нуклеотидов ДНК, оказалась очень плодотворной и, как будет ясно из дальнейшего, может в настоящее время считаться доказанной.

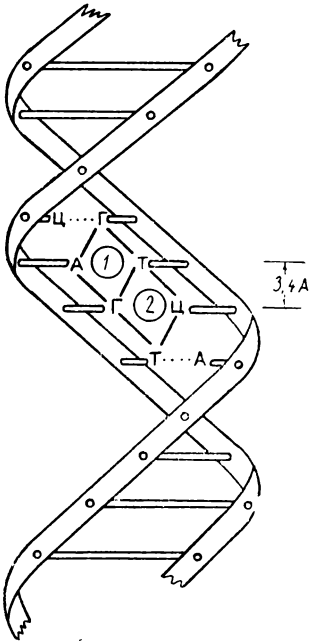

Рис. 5. Схема «ромбического кода». Показаны два «ромба», определяющие включение аминокислот *1* и *2*.

Расшифровка макромолекулярной структуры ДНК и возникшие в связи с ней гипотезы открыли широкое поле для экспериментирования, в результате чего механизм биосинтеза белка был в общих чертах выяснен. Прежде чем вдаваться в детали происходящих при этом процессов, мы кратко нарисуем общую схему биосинтеза белка в том виде, в каком она представляется в настоящее время, после чего будет описан соответствующий фактический материал и рассмотрены некоторые подробности.

Общая схема биосинтеза белка. В принципе механизм биосинтеза белка сводится к следующему. Информация о порядке аминокислот в белках закодирована в виде последовательности оснований ДНК. От клетки к клетке эта информация передается в процессе ауторепродукции ДНК и хромосом, предшествующих клеточному делению. Внутри клетки информация передается через посредство так называемой информационной РНК (и-РНК), синтезируемой при участии ДНК; порядок оснований в и-РНК в точности соответствует таковому

ДНК. Образовавшаяся в ядре и-РНК мигрирует в цитоплазму, где присоединяется к специальным внутриклеточным гранулам, так называемым рибосомам, где и происходит синтез полипептидных цепей.

Полипептидные цепи синтезируются не непосредственно из аминокислот; так как для синтеза пептидной связи нужна довольно большая энергия, аминокислоты должны быть предва-

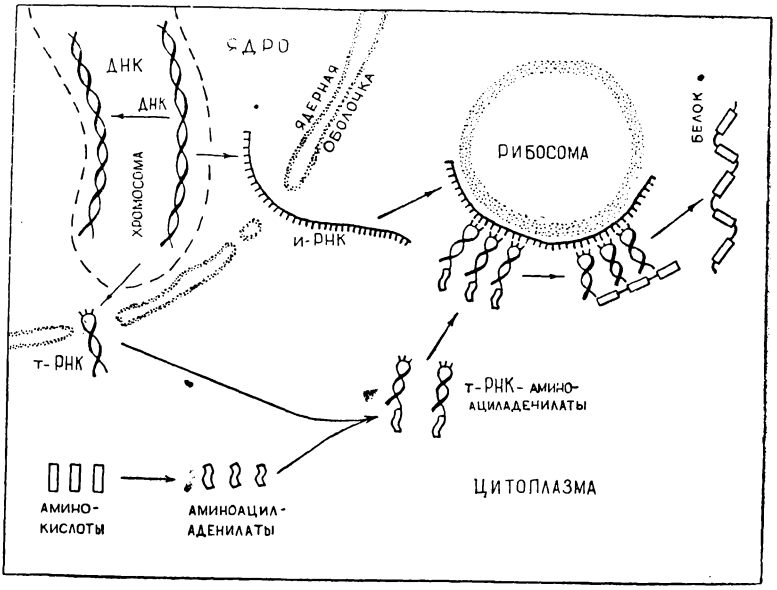

Рис. 6. Общая схема биосинтеза белка.

рительно активированы. Это активирование осуществляется специальными ферментами, присоединяющими к аминокислотам фосфорные группы, богатые энергией. Те же ферменты специфически присоединяют активированные аминокислоты к специальным молекулам транспортной РНК (т-РНК), которая также синтезируется под контролем ДНК и обладает специфическим сродством к определенным участкам и-РНК. Комплексы т-РНК с активированными аминокислотами мигрируют к рибосомам, к которым присоединена и-РНК, несущая информацию о первичной структуре белков. Благодаря упомянутому химическому сродству участков и-РНК и т-РНК, последние, взаимодействуя с и-РНК, располагаются вполне определенным образом и располагают тем самым присоединенные к ним аминокислоты в заданную последовательность, в соответствии с той информацией, которая была первоначально записана в

ДНК. После соединения аминокислот друг с другом и отделения полипептидной цепи от матрицы интересующий нас процесс заканчивается.

Участки (или молекулы?) ДНК, в которых записана информация о первичной структуре белков, называют структурными генами. Интенсивность работы структурных генов или их полное выключение контролируются генетически с помощью специальных генов-операторов; регуляция эта происходит по принципу отрицательной обратной связи. Описанные процессы для большей наглядности схематически изображены на рис. 6.

Прежде чем перейти к более подробному рассмотрению отдельных из этих процессов, мы хотим обратить внимание на следующее обстоятельство. Иногда на основании популярных статей складывается неправильное впечатление, что современная генетика ставит хромосомы и ДНК в изолированное и автономное положение. Как должно быть ясно из рассмотренной схемы, хотя ДНК играет вполне определенную и очень важную роль, она является только одним из звеньев сложнейшего процесса, в который вовлечены разные части клетки и разные биологически активные вещества. К этому нужно добавить, что те процессы, которые были рассмотрены, теснейшим образом связаны с остальными процессами внутриклеточного обмена, рассмотрение которых здесь не представляется возможным.

МЕХАНИЗМ БИОСИНТЕЗА БЕЛКА

Рассмотрим кратко перечисленные выше процессы более подробно, с точки зрения их экспериментального обоснования, и укажем на некоторые из наиболее важных деталей.

Ауторепродукция ДНК. Если вначале представление об ауторепродукции ДНК было только гипотезой, напрашивающейся на основании комплементарности молекул ДНК, то в настоящее время ее можно считать доказанной [58]. Особенно важную роль при этом сыграли работы А. Корнберга и сотрудников [31, 213], выделивших из бактериальных клеток фермент, осуществляющий ин витро синтез ДНК из нуклеозидтрифосфатов. Позже тот же фермент был обнаружен и в клетках других организмов [73, 74, 91, 166]. При этом выяснилось, что синтез новой ДНК происходит только в присутствии затравки в виде полимерной ДНК. Получающийся продукт оказывается во всех отношениях идентичным исходной ДНК, включая соотношение оснований [218] и частоту их ближайших соседств [202]. Теоретически рассуждая, можно представить, что ауторепродукция идет по одному из трех механизмов [18]: консервативному, при котором исходная молекула остается интактной, полуконсервативному, при котором каждая из двух молекул, имеющихся после редупликации, состоит из одной новой и одной старой нити и дисперсному, при котором после редупликации каждая

нить двойной спирали включает в себя новый и старый материал. Все имеющиеся данные говорят в пользу того, что редупликация следует полуконсервативному механизму. Так, если бактерии, содержащие только тяжелый азот, перенести на один цикл деления в нормальную среду, а затем выделить из них ДНК и разделить ее на отдельные спирали (что достигается тепловой денатурацией), то оказывается, что каждая из нитей содержит либо только нормальный, либо только тяжелый изотоп азота [235, 236, 237]. Очень интересны опыты с ДНК бактериофага φX, которая представляет собой не двойные, а одинарные нити. Если при искусственном синтезе ДНК брать в качестве затравки ДНК от фага φX, то в результате синтезируется ДНК, не идентичная исходной, а комплементарная ей [297]. Эти и многие другие данные убеждают в том, что молекулы ДНК возникают в результате ауторепродукции и притом по полуконсервативному механизму, хотя некоторые авторы и придерживаются других взглядов [19, 88].

Синтез РНК. Что касается синтеза молекул РНК, то он может не требовать никакой затравки. Очоа и сотрудники [48, 170] открыли фермент, осуществляющий синтез полирибонуклеотидов в отсутствии каких-либо затравок (именно эти искусственные РНК и оказывают большую службу в расшифровке аминокислотного кода, о чем подробно будет говориться в следующей главе). Однако синтез *специфичной* РНК должен требовать информации о последовательности нуклеотидов, и здесь следует ожидать осуществления матричного механизма. И действительно, удалось доказать существование синтеза РНК, зависимого либо от ДНК, либо от самой РНК. Существуют вирусы и фаги, содержащие РНК и не содержащие ДНК. Хотя механизм их редупликации нельзя считать выясненным, но уже имеются указания на осуществление синтеза РНК вируса табачной мозаики ин витро [121, 205, 209] и о том, что синтез РНК бактериофага f2 (не содержащего ДНК) происходит ин виво и при полностью выключенном синтезе ДНК бактерии-хозяина [228]. Сходные результаты были получены и на животных вирусах [273]. Ряду авторов удалось выделить ферменты, катализирующие синтез РНК, зависимый от РНК [106, 109, 137, 211, 239].

В клетках, содержащих РНК и ДНК, по-видимому, нормальным является синтез специфичной РНК, зависящий от ДНК [198]. Ферменты, осуществляющие такой синтез ин витро, были выделены из клеток печени [331], вилочковой железы [70], из растительных клеток [190] и из микроорганизмов [107]. Важным подспорьем для выяснения места синтеза РНК явилось открытие метода получения «гибридных» молекул ДНК и РНК [178, 258, 271, 332]. Этот метод основан на том, что при нагревании двойных спиралей ДНК они разъединяются на отдельные нити, которые в случае медленного охлаждения снова образуют двойные

спирали. Если при этом присутствуют нити РНК, то, в случае их комплементарности к ДНК, может возникать «гибридная» спираль, одна нить которой состоит из дезоксирибонуклеотидов, а другая — из рибонуклеотидов. Такие гибридные молекулы оказываются устойчивыми к действию рибонуклеазы. Факт образования гибридных спиралей между определенными ДНК и РНК говорит об их комплементарности, что можно объяснить тем, что соответствующие молекулы РНК синтезируются под контролем ДНК. Важно отметить, что подобная комплементарность была показана как для и-РНК, так и для т-РНК. В специальных опытах было показано, что состав таких молекул РНК действительно является идентичным ДНК [332].

В опытах по синтезу РНК, зависимому от ДНК, синтез происходил как на одноцепочечной, так и на двухцепочечной ДНК. Каким образом он происходит в естественных условиях ин виво — несут ли при этом матричную функцию обе нити ДНК или только одна из них — пока неизвестно. Хотя этот и некоторые другие важные вопросы, связанные с механизмом ауторепродукции, еще ждут своего разрешения, но тот факт, что новые молекулы ДНК возникают путем ауторепродукции и что, по крайней мере, некоторые фракции РНК возникают сходным образом—на матрицах, представляющих собой ДНК или РНК — в настоящее время уже не подлежит сомнению. Впрочем, здесь нужно упомянуть об интересных работах Г. Шрамма и сотрудников [269], нашедших способ чисто химического получения полинуклеотидов, без использования ферментов.

Информационная РНК. Хотя и-РНК впервые была экспериментально обнаружена только в 1956 г., о ее существовании подозревали значительно раньше. Еще в упоминавшихся работах Н. К. Кольцова 20-х годов постулируется существование генных продуктов, переносящих генетическую информацию в цитоплазму. Позже, после выяснения роли нуклеиновых кислот в биосинтезе белка, ряд авторов высказывал гипотезу о том, что посредником при передаче генетической информации должна служить РНК [16, 32, 60, 123]. Однако химический состав РНК оказывался значительно менее специфичным, чем таковой ДНК, и не наблюдалось той корреляции между их составом, которую следовало ожидать [3, 5, 6]. Поэтому можно было думать, что переносчиком информации является только небольшая доля всей РНК, содержащейся в клетке. Впервые и-РНК была обнаружена в бактериальных клетках, зараженных фагом [56, 318]; ее состав соответствовал составу ДНК фага. Эти результаты были подтверждены другими авторами [167, 168, 349], позже такая же РНК была найдена и в клетках высших организмов [184, 188, 267, 268, 272]. Она составляет всего от 1 до 3% всей РНК клетки и характеризуется очень высокой скоростью обновления. Так, в клетках Bac. subtilis средняя продолжительность жизни моле-

кул и-РНК составляет всего около 2 мин [222]. Было показано, что бактериальная и-РНК оказывается очень нестойкой также в опытах ин витро [301]. Правда, есть основания считать, что и-РНК другого, не бактериального происхождения, оказывается значительно более стабильной [10, 95, 150, 172, 189, 230]. Возможно, что такая лабильность свойственна бактериальным и-РНК, в связи с исключительной способностью бактерий к биохимической адаптации. По размерам молекулы и-РНК очень гетерогенны, но в общем довольно велики: их константы осаждения лежат в пределах от 8 до 30 единиц Сведберга [119, 210, 241, 254]. В клетках и-РНК связывается лабильно с рибосомами, причем одновременно связанной оказывается только небольшая часть всей наличной и-РНК [246, 300]. Имеются указания на то, что одна молекула и-РНК может одновременно взаимодействовать с несколькими рибосомами [323]. Сравнение состава ДНК и и-РНК и опыты по их молекулярной гибридизации [171 а] показывают, что и-РНК действительно представляют собой копии ДНК. Синтезируется и-РНК специальным ферментом из нуклеозидтрифосфатов в присутствии полимерной ДНК [194, 214, 324]. Все перечисленные свойства полностью соответствуют тем требованиям, которым должен отвечать переносчик генетической информации о структуре белков. Более подробные сведения об и-РНК приводятся в обзоре А. Спирина и В. Смирнова [53].

Рибосомы. Рибосомы — цитоплазматические гранулы, на которых происходит при участии присоединяющихся к ним и-РНК синтез белка — очень богаты и собственной, так называемой рибосомной, РНК: 85—90% всей РНК, находящейся в клетке, локализовано в рибосомах. Эта РНК, в отличие от и-РНК, довольно гомогенна по размерам [120, 333] и составу [178, 227]. Белок, входящий в состав рибосом, богат основными аминокислотами, причем у всех исследованных организмов аминокислотный состав белка рибосом оказывается довольно сходным [304]. Таким образом, ни белок, ни РНК рибосом не показывают специфичности, необходимой для матричных структур, хотя раньше представление об их матричной роли было довольно популярным. Большая часть рибосом имеет диаметр порядка 10—20 *ммк* и константу седиментации около 70—80 единиц Сведберга. В отсутствии ионов магния рибосомы распадаются на более мелкие частицы (с константами седиментации 50 и 30 единиц Свердберга), причем становятся совершенно неактивными [302].

Как уже говорилось, синтез белка происходит на рибосомах, соединенных с и-РНК. Из всех рибосом, находящихся в данное время в клетке, только около 10% способны соединяться с и-РНК [177, 259]. Хотя участие рибосом в синтезе белка кажется необходимым, они играют при этом неспецифическую

роль. Так, в опытах с фагом Т2 было показано, что при заражении бактерий не образуется новых рибосом, а фаговая и-РНК присоединяется к рибосомам, существовавшим до инфекции [101].

В опытах ин витро было показано, что внесение в бактериальную систему РНК вируса табачной мозаики приводит к синтезу белка, подобного вирусному [309]. Синтезируется рибосомная РНК также, по-видимому, на ДНК, однако участок ДНК, комплементарный рибосомной РНК, составляет не более одной десятитысячной части генома [346], хотя два ее компонента синтезируются разными цистронами [347]. Биологическая роль рибосомной РНК остается пока неясной.

Активирование аминокислот. Так как образование пептидной связи требует энергии, то аминокислоты, участвующие в синтезе белка, должны быть активизированы. Как и в большинстве процессов такого рода, для этого служат богатые энергией («макроэргические») фосфорные связи аденозинтрифосфорной кислоты [АТФ]. Активирование осуществляется специальными ферментами, имеющими белковую природу. Эти ферменты специфичны: каждой аминокислоте соответствует свой особый фермент; кроме того, эти ферменты обладают известной видовой специфичностью [85]. Они катализируют реакцию между АТФ и аминокислотой, причем образуется аминоациладенилат (смешанный ангидрид между карбоксильной группой аминокислоты и фосфатом 5′-АМФ) с отщеплением пирофосфата. Тот же фермент катализирует и следующий шаг реакции — образование комплекса между аминоациладенилатом и соответствующей т-РНК.

Интересно отметить, что концевой участок т-РНК, к которому присоединяется аминоациладенилат, одинаков у всех т-РНК. Поэтому очевидно, что специфичное присоединение активированных аминокислот к т-РНК осуществляется не за счет сродства между ними. Скорее здесь может идти речь о сродстве фермента, с одной стороны, к аминокислоте, с другой — к соответствующей т-РНК. В этой связи хочется упомянуть об изящных опытах по определению видовой специфичности т-РНК [164]. В этих опытах было показано, что т-РНК образуют гибридные молекулы с ДНК того же вида (этим, кстати, доказывается хромосомное происхождение т-РНК); если же т-РНК и ДНК брались от разных видов, то степень их химического сродства хорошо коррелировала со степенью систематической близости тех видов, которым они принадлежат. Что же касается видовой специфичности активирующего фермента, то она была выяснена раньше; так, фермент из печени быка реагирует с т-РНК из животных тканей, но не реагирует с т-РНК бактерий и дрожжей [176]. Более подробно процессы активации аминокислот описаны в обзорной статье Ф. Липманна [36].

Транспортные РНК. Уже упоминавшиеся выше т-РНК (вначале носившие также название растворимая РНК) составляют около 10% всей РНК, содержащейся в клетке. Это сравнительно небольшие молекулы с числом нуклеотидов 70—80. Каждой аминокислоте соответствует не меньше одной т-РНК. Для некоторых же доказано существование нескольких [295, 330]. Концы всех т-РНК одинаковы. На 3′-конце во всех т-РНК расположена тройка оснований ЦЦА, а на 5′-конце, по крайней мере у большинства, находится гуанозин — 5′-фосфат [215, 276, 351]. В отличие от других типов РНК, т-РНК содержит необычные для РНК основания (такие как тимин и псевдоуридин), роль которых неясна [130, 134, 227, 240, 253]. Недавно было показано, что т-РНК дают вторичную структуру, спирализуясь сами с собой [104, 122, 229, 288]; предполагают, что в месте перегиба т-РНК находится группа нуклеотидов, взаимодействующая с и-РНК. Хотя с аминокислотой взаимодействует 3′-конец т-РНК, но в связи с тем, что концы всех т-РНК одинаковы, ясно, что распознавание аминокислоты происходит не за счет участка, к которому последняя присоединяется. О том, что аминокислота взаимодействует с и-РНК только через посредство т-РНК, говорят следующие остроумные опыты [114]. Цистеин, находившийся в комплексе с соответствующей т-РНК, подвергали на никелевом катализаторе окислительному десульфированию, в результате чего к специфичной для цистеина т-РНК оказывается присоединен аланин. После этого аланин включается в белки не на место аланина, а на место цистеина. Эти опыты ясно показывают, что включение аминокислот в белок происходит через посредство т-РНК. Подробная сводка современных данных о т-РНК дается в обзорной статье Л. Л. Киселева [27].

Заключительные стадии биосинтеза белка. После того как аминокислоты расположены в должном порядке, они должны соединиться, а получившаяся полипептидная цепь — отделиться от т-РНК и рибосомы. К сожалению, об этих заключительных этапах биосинтеза белка известно пока очень мало. Известно, что соединение аминокислот в полипептидную цепь катализируется энзиматически [87, 131, 169, 245] и требует присутствия гуанозинтрифосфата [207, 327], но до сих пор, например, остается неясным вопрос, идет ли здесь речь о ряде ферментов, специфичных для отдельных аминокислот, или об одном неспецифичном ферменте. Также неясен вопрос об отделении уже готовой цепи. Имеются сообщения о существовании специального фермента [216, 329], тогда как в других случаях его обнаружить не удавалось [140]. Хотя эти процессы и не имеют принципиального значения в нашей связи, но выяснение их механизма, конечно, должно являться одной из важных ближайших задач.

Регуляция белкового синтеза. Последний вопрос, который необходимо рассмотреть в связи с общей проблемой биосинтеза белка, это вопрос о том, каким образом регулируется передача генетической информации. Действительно, давно известно, что все клетки организма несут одну и ту же генетическую информацию; можно сказать, что в геноме каждой клетки содержится информация о структуре всех белков организма. Между тем, разные клетки продуцируют разные белки, более того, они дают количественно и качественно разную продукцию в зависимости от стадии развития и сопутствующих условий. До недавнего времени вопрос о механизме регуляции этих процессов оставался совершенно неясен. Но в последние годы был достигнут значительный прогресс в познании механизма индукции и репрессии у микроорганизмов, которые можно рассматривать как модель биохимических координаций у высших организмов.

Этот вопрос подробно рассмотрен в прекрасной статье Ф. Жакоба и Ж. Моно [22], и мы здесь ограничимся только приведением основных выводов. У бактерий давно известен эффект индукции, сводящийся к тому, что внесение в среду некоторого вещества влечет за собой, и притом очень быстро, синтез фермента, который до этого не синтезировался. Например, кишечная палочка на обычной среде не продуцирует фермент β-галактозидазу, если же в среду внести β-галактозу — сахар, сбраживаемый этим ферментом, то уже через несколько минут можно наблюдать интенсивный синтез β-галактозидазы. Эффект репрессии обратен эффекту индукции: он сводится к тому, что внесение некоторого вещества специфически подавляет синтез определенного фермента. В ряде тончайших биохимических и генетических опытов было показано, что наряду со структурными генами, определяющими первичную структуру белка (т. е. наряду с участками ДНК, продуцирующими и-РНК), геном содержит еще так называемые гены-регуляторы и гены-операторы. Ген-регулятор продуцирует специальное вещество — репрессор, которое при соединении с геном-оператором блокирует его, благодаря чему выключается группа структурных генов, контролируемых данным оператором.

Таким образом, механизм регуляции носит отрицательный характер: когда он включен, соответствующие структурные гены выключены. Новый метаболит, попадая в среду, либо связывает репрессор (явление индукции), либо активирует его (явление репрессии). Следует заметить, что, хотя перенос этого механизма с бактериальных систем на все остальные является, несмотря на его заманчивость, гипотетичным, для изученных систем он доказан вполне убедительно.

Заканчивая обзор современных представлений о механизме биосинтеза белковых веществ, мы хотим подчеркнуть еще раз,

что он был предпринят с единственной целью облегчить чтение последующих глав для читателя не биохимика. Поэтому мы были вынуждены воздержаться от сравнительного критического рассмотрения аналогичных данных разных авторов, несколько упрощали некоторые положения и опускали описание многих очень доказательных опытов, ведя изложение иногда в почти что аксиоматической форме. С другой стороны, мы старались везде, где это возможно, указывать на спорность некоторых утверждений и невыясненность иных существенных вопросов, а также давали указания на более подробные обзоры, где приведен больший фактический материал и имеется обширная библиография.

ГЛАВА 2

СОВРЕМЕННОЕ СОСТОЯНИЕ ПРОБЛЕМЫ АМИНОКИСЛОТНОГО КОДА

Один из важнейших вопросов в общей проблеме биосинтеза белка — выяснение общих свойств аминокислотного кода и его расшифровка, т. е. установление характера соответствия между нуклеотидами и-РНК и аминокислотами белка. Поскольку проблема аминокислотного кода составляет нашу главную задачу, ее современное состояние должно быть рассмотрено более подробно. Однако мы при этом воздержимся от разбора всех многочисленных работ, касающихся проблемы кода, многие из которых приводили, как впоследствии выяснилось, к неверным выводам. Мы ограничимся рассмотрением основных направлений поисков и более подробно остановимся лишь на работах, имеющих бо́льшее значение. Для читателей, интересующихся более подробно историей вопроса, можно указать на несколько специальных обзоров [16, 25, 125, 341, 348].

ПЕРВЫЕ ПОДХОДЫ К РЕШЕНИЮ ПРОБЛЕМЫ

Криптографические методы. В первое время, пока еще не были найдены экспериментальные подходы к расшифровке кода, предпринимался ряд попыток использовать для нее криптографические методы, исходящие из некоторых численных соответствий и статистических закономерностей. О самой первой гипотезе такого рода мы уже упоминали в предыдущей главе. Это — гипотеза «ромбического кода», предложенная Г. Гамовым [153, 154], которая была пояснена на рис. 5. Нетрудно заметить, что, поскольку в кодировании двух соседних аминокислот участвуют два общих нуклеотида, в белках должны быть запрещенные соседства. Поэтому, в случае правильности этого кода, его можно было бы расшифровать, изучив статистику полипептидных цепей. Хотя этот код подкупал хорошими численными совпадениями (равенство расстояний между соседними парами нуклеотидов в ДНК и между аминокислотами в белках, одинаковое число разных «ромбов» и подлежащих кодированию аминокислот), но он противоречил имевшимся

данным о месте белкового синтеза: ДНК локализована в хромосомах клеточного ядра, тогда как белки синтезируются в рибосомах. Поэтому вскоре был предложен новый код для одноцепочечной матрицы, известный под названием «треугольного» [16]. Предполагалось, что ДНК передает информацию на одноцепочечную РНК, которая и служит матрицей. Принцип «тре-

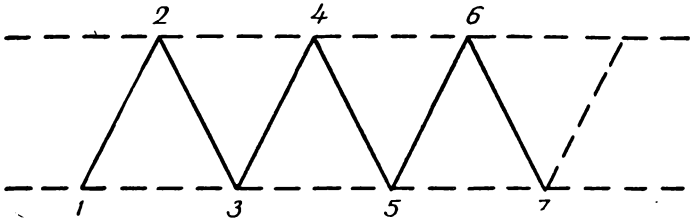

Рис. 7. Схема «треугольного кода». Объяснение в тексте.

угольного кода» пояснен на рис. 7. Каждая аминокислота определяется треугольником, в вершинах которого находятся основания. Возможны два варианта этого кода. В «плотном» треугольном коде первая аминокислота кодируется треугольником 123, вторая — 234, третья — 345 и т. д. В «неплотном»: первая — 123, вторая — 345, третья — 567 и т. д. В обоих вариантах треугольного кода в кодировании соседних аминокислот участвуют, так же как и в ромбическом коде, общие основания (в таких случаях код называют перекрывающимся), следовательно, и здесь нужно ожидать запрещенных соседств в белках.

В то время, когда создавались первые гипотезы об общих свойствах аминокислотного кода, не было ничего известно о последовательности нуклеотидов в нуклеиновых кислотах и почти ничего о последовательности аминокислот в белках. Если бы обе задачи были с самого начала доступны экспериментальному исследованию, не было бы надобности в тех многочисленных гипотезах, которые высказывались в течение истекших десяти лет. Достаточно было бы сравнить порядок аминокислот в нескольких белках с последовательностью нуклеотидов в соответствующих им матрицах. Но, как известно, эта задача еще не решена. Помимо того, что до сих пор не удалось выделить ни одной и-РНК, соответствующей определенному белку, последовательность нуклеотидов в ДНК или РНК еще не поддается расшифровке. Правда, в этой области ведутся работы в разнообразных направлениях. Помимо химических [8 а] и биохимических [201а, 202] попыток определения последовательностей нуклеотидов применяют и другие остроумные методы. Так, предлагалось применить для этой цели анализ обратных мутаций, вызванных в генетически хорошо исследованном участке, [113] и даже метить с помощью тяжелых атомов определенные

нуклеотиды с последующим исследованием под электронным микроскопом [81].

Первые перекрывающиеся коды были предложены в то время, когда об аминокислотных последовательностях в белках не было почти ничего известно, так что вопрос о существовании запрещенных соседств оставался открытым. Но уже к 1957 г. накопилось достаточно данных для того, чтобы отвергнуть возможность существования перекрывающихся кодов [99]: слишком большим оказалось наблюдаемое число соседей для большинства аминокислот. Анализ частоты соседств в белках неоднократно повторялся при получении бо́льшего числа данных [25, 62, 287, 337], причем первоначальный вывод об отсутствии запрещенных соседств только подтверждался. Иногда создавалось впечатление, что распределение частоты соседств отклоняется от случайного; так, Ф. Шорм [287] утверждал, что для ряда пар аминокислот один порядок наблюдается значительно чаще другого. Однако статистический анализ и накопление большого материала показали, что это и другие отклонения от случайного распределения — кажущиеся [337]. В табл. 1 мы даем сводку по частоте соседств аминокислот в белках, составленную на основании новейших данных. Хотя заполнены еще не все клетки таблицы, не может оставаться сомнений в том, что запрещенных соседств не существует, так как в имеющемся материале отсутствуют соседства в основном для аминокислот, встречающихся редко, по которым собрано еще недостаточно данных. Частота соседей не зависит от их порядка и пропорциональна встречаемости соответствующих аминокислот в белках [25, 337].

Таким образом, анализ аминокислотных последовательностей исключает возможность существования простейших перекрывающихся кодов. Результаты этого анализа — не единственный довод против того, что код перекрывающийся. В настоящее время проанализировано уже довольно большое число мутантных белков (сводка соответствующих данных будет приведена в главе 3). Для перекрывающегося кода следовало бы ожидать, что при мутациях довольно часто должны быть изменены две аминокислоты, стоящие рядом, — в тех случаях, когда изменен нуклеотид, участвующий в кодировании обеих соседних аминокислот. Однако до сих пор не найдено ни одного подобного случая. Чаще всего измененной оказывается одна единственная аминокислота; в относительно редких случаях, когда изменены две аминокислоты, они никогда не стоят рядом. Правда, эти доводы справедливы только для простейших перекрывающихся кодов, в первую очередь для плотного треугольного кода. Недавно Р. Уолл [322] рассмотрел все доводы, приводимые против перекрывающихся кодов, и показал, что они недостаточны; он предложил перекрывающийся код, которому не противоречат

Частота соседств аминокислот в неструктурных белках Таблица 1

N-концевые аминокислоты	С-концевые аминокислоты																			
	ала	арг	асп	асп-NH_2	вал	гис	глиц	глют	глю-NH_2	илей	лей	лиз	мет	про	сер	тир	тре	трип	фала	цис
ала	8	2	7	3	7	9	7	4	6	5	10	3	1	3	8	3	7	2	3	3
арг	2	3	+	2	3	+	2	2	2	1	2	3	1	2	3	2	1	2	3	1
асп	8	5	5	+	4	2	3	4	2	2	6	3	1	4	4	1	+	—	4	1
асп-NH_2	7	2	—	+	4	1	1	1	4	+	5	3	—	4	+	3	1	+	1	+
вал	9	4	5	6	2	5	8	7	6	2	6	3	1	1	1	4	5	2	1	4
гис	4	—	1	—	3	1	4	2	—	1	8	3	1	2	2	—	1	—	6	2
глиц	8	3	6	5	3	4	6	6	1	4	6	10	—	3	6	2	5	—	6	+
глют	7	4	8	2	3	3	4	4	1	1	7	4	1	—	5	2	4	1	3	—
глю-NH_2	5	2	2	+	6	1	1	—	1	—	3	4	2	+	1	—	2	1	3	3
илей	2	2	2	1	4	2	2	2	1	3	1	3	+	2	1	+	5	—	2	+
лей	7	1	5	+	9	4	9	8	1	6	5	5	2	2	11	2	8	1	4	2
лиз	9	2	4	3	8	4	+	6	—	2	3	8	3	3	3	5	7	—	3	1
мет	+	—	—	1	1	—	2	4	+	1	1	2	1	1	1	—	+	—	1	+
про	5	+	3	3	7	1	2	4	2	2	3	6	—	3	3	2	3	1	1	+
сер	9	3	6	2	3	5	5	4	4	3	6	4	1	3	10	4	6	1	3	+
тир	—	2	1	2	2	1	3	+	3	1	4	4	—	3	5	—	2	—	1	2
тре	5	2	2	6	7	—	4	6	2	1	7	4	1	5	7	2	4	1	3	+
трип	1	—	—	—	+	—	3	—	1	—	1	2	—	—	2	—	2	—	—	—
фала	2	4	4	+	5	1	4	3	1	+	4	5	—	5	3	2	4	—	2	+
цис	2	1	—	3	1	1	2	1	+	+	1	2	+	+	2	2	2	—	+	1

Примечание. Цифры — частота соседств данного сорта в белках с полностью расшифрованной последовательностью, кресты — соседства, наблюдавшиеся в белках, аминокислотные последовательности которых изучены не полностью. При составлении таблицы использованы данные о полных аминокислотных последовательностях в следующих белках: рибонуклеаза быка [187], глюкагон [103], гипертензин [143], цитохром С из сердца лошади [55], меланотропин и кортикотропины [82, 160, 226], А и В пептиды кофибрина [148, 149], гемоглобин и миоглобин [42, 183, 191, 192, 193, 212, 242, 270, 325], инсулин разных видов [64, 263], вирус табачной мозаики [71, 305, 342]. Данные о соседствах в неполных последовательностях взяты из работ [25, 110, 287, 337]. В случае сходных белков (например, цепи гемоглобина) или мутаций в таблицу включались только не совпадающие соседства. (Таблицу составил А. А. Позолотин).

экспериментальные данные . В этом коде каждой аминокислоте соответствует четыре основания, причем у соседних квадруплетов имеется по одному общему нуклеотиду, которым должен быть либо урацил, либо гуанин.

Проблема «запятых». Признание положения о неперекрываемости кода приносит новые затруднения. Коды типа ромбического и плотного треугольного обладают двумя очень удобными свойствами. Во-первых, расстояние между соседними кодирующими группами численно равно расстоянию между соседними аминокислотами в полипептидной цепи. Во-вторых, любая последовательность нуклеотидов может быть прочитана только одним способом. Неперекрывающиеся коды этими свойствами не обладают. Так как кодированию подлежит 20 аминокислот, то если считать, что вся информация об их последовательности заключена в матричной РНК, окажется, что каждая аминокислота должна кодироваться последовательностью более чем из двух нуклеотидов, так как 4 разные буквы могут дать только $4^2 = 16$ двухбуквенных слов. Именно поэтому с самого начала чаще всего принималось, что код является триплетным. Однако принятие неперекрывающегося триплетного кода вносит геометрические трудности, так как расстояние между соседними триплетами втрое больше, чем между соседними аминокислотами; совершенно ясно, что в случае непосредственного взаимодействия аминокислот с матрицей белок не может быть построен. Кроме того, в случае неперекрывающегося кода произвольная нуклеотидная последовательность не может быть прочитана однозначно, если не ставить дополнительных условий [15]. Так, в последовательности 1 — 2 — 3 — 4 — 5 —... за кодирующий триплет можно принять и 1 — 2 — 3, и 2 — 3 — 4, и 3 — 4 — 5.

Оба эти затруднения были рассмотрены в работе Ф. Крика, И. Грифица и И. Орджела [127]. Было высказано предположение, что с матрицей взаимодействует не аминокислота, а посредник, представляющий собой вытянутую молекулу, один конец которой специфически взаимодействует с матрицей, а другой несет соответствующую аминокислоту. Как мы уже знаем, эта гипотеза подтвердилась: была открыта транспортная РНК, обладающая именно теми свойствами, которые приписывались гипотетическому «посреднику».

В той же работе была предложена гипотеза «кода без запятых». Идея ее сводится к следующему. Предполагают, что для кодирования аминокислот служат только такие триплеты, при любом соседстве которых не может получиться из их частей осмысленных комбинаций. Так, если АБА и АББ кодируют какие-то аминокислоты, то триплеты БАА, ААБ, ББА, БАБ не должны ничего кодировать. Интересно, что составленная в соответствии с этим требованием группа содержит 20 триплетов из 64 (см. также [162, 163]). Получение «магического числа» 20

казалось бы говорило в пользу правильности этой гипотезы. Но дальнейшие исследования показали, что она неверна.

Кроме этой гипотезы, можно, для объяснения однозначности считывания, предложить еще две других: либо считать, что кодовые группы отделены друг от друга какими-то добавочными, специальными раздельными группами («запятыми»), либо что триплеты считываются по порядку, начиная с некоторой фиксированной точки. В очень изящных опытах Ф. Крика и сотрудников [128], с мутациями бактериофага Т4, которые будут специально рассмотрены несколько позже, было довольно убедительно показано, что правильным является последнее из сделанных предположений. Сущность доказательства сводится к следующему: если в матрицу вставить один лишний нуклеотид или удалить один из нуклеотидов, то в случае двух первых гипотез неизбежно должны возникнуть бессмысленные триплеты и соответствующий белок не будет строиться. В случае же последней гипотезы можно ожидать, что в ряде случаев молекула белка (конечно, с частично измененными аминокислотами) будет строиться. Опыты показали, что сдвиг довольно большого участка матрицы на один нуклеотид вправо или влево в значительном числе случаев не препятствует синтезу белка со свойствами, близкими к нормальному.

Некоторые гипотезы. Только что шла речь о некоторых из общих свойств кода. Что касается его конкретной расшифровки, то первые попытки начали предприниматься уже довольно давно. Г. Гамов и М. Ичас [155] обратили внимание на то, что имеется ровно 20 триплетов, отличающихся друг от друга составом (без учета порядка нуклеотидов), поэтому можно было высказать гипотезу о том, что код триплетный, неперекрывающийся и что все триплеты одинакового состава кодируют одну и ту же аминокислоту. Далее, если предположить, что последовательность нуклеотидов в РНК близка к случайной, можно вычислить вероятность встречаемости триплетов разного состава и сравнить ее со встречаемостью аминокислот в соответствующем белке. Из соответствия частот в обоих распределениях можно сделать заключение о том, какой из триплетов кодирует какую аминокислоту. Такой расчет был сделан для некоторых растительных вирусов, в результате чего была дана первая расшифровка кода. Результат этой расшифровки оказался очень далеким от истины, точно так же, как и те предположения об общих свойствах кода, которые были положены в его основу.

Как мы видели, все ранние попытки так или иначе вращались вокруг «магического числа» — 20. Если гипотеза давала 20 кодовых слов, то она считалась правдоподобной. Однако, как мы уже видели, это число получали при целом ряде гипотез, которые так или иначе были опровергнуты. Между тем, подле-

жит обсуждению сама правомерность требования того, чтобы число кодирующих групп было равно 20. Это число должно получаться только в том случае, если каждой аминокислоте соответствует только одна кодирующая группа.

На самом деле такое соответствие не обязательно должно иметь место. С одной стороны, всем или некоторым аминокислотам может соответствовать более одного кодового слова (т. е. словарь может иметь синонимы); в таких случаях говорят о вырожденности кода. С другой стороны, одна и та же кодовая группа может соответствовать более чем одной аминокислоте, т. е. код может быть неоднозначным (неоднозначность может быть снята в том случае, если информация о первичной структуре белков заключена не только в последовательности нуклеотидов и-РНК). Между тем, вопрос о вырожденности или однозначности кода не может быть решен априорно. Это обстоятельство не учитывалось ни в одной из работ раннего периода.

Кроме рассмотренных выше подходов к проблеме кодирования, предпринималось большое число других попыток, не сыгравших существенной роли в решении проблемы кода и представляющих теперь только исторический интерес. Кратко перечислим некоторые из них.

Чтобы в рамках перекрывающегося кода объяснить тот факт, что при мутациях, как правило, бывает изменена только одна аминокислота, Орджел и др. предложили так называемый «код старший — младший». Предположили, что каждой аминокислоте соответствует некоторый нуклеотид, являющийся «главным детерминантом». Естественно, что один и тот же главный детерминат должен соответствовать нескольким аминокислотам. Поэтому однозначность в считывании текста вносится за счет влияния прилегающих нуклеотидов, которые играют подчиненную роль [16]. Этот код также должен давать запрещенные соседства, и противоречие его аминокислотным последовательностям было обнаружено самими авторами. Сходную гипотезу высказывал также А. Даунс [132, 133].

Хотя представление о перекрывающемся коде было популярно преимущественно в начальный период расшифровки кода, но попытки его «реабилитации» не прекращаются до сих пор. Кроме уже упоминавшейся работы Р. Уолла [322], недавно появилась работа Ж. Медведева [233], автор которой предлагает другой вариант перекрывающегося кода, не противоречащий экспериментальным данным. Он предполагает, что нуклеотиды, кодирующие одну аминокислоту, находятся не рядом, а на некотором расстоянии, так что, например, в последовательности нуклеотидов

АБВ... КЛМ... СТУ...

первая аминокислота может кодироваться триплетом АКС, вторая — БЛТ, третья — ВМУ и т. д. Хотя так же как и для кода Уолла здесь трудно привести данные, опровергающие эту схему, но сама она кажется очень искусственной, и нет данных, говорящих сколько-нибудь серьезно в ее пользу. С доводами в пользу перекрывающегося кода выступил недавно также Эк [135, 136], изучавший статистику полипептидных цепей. Однако, о чем речь будет ниже, его данные скорее объясняются тем, что вырожденность кода не является случайной.

Выше уже говорилось, что «магическое число» — 20, считавшееся критерием правдоподобности кода, было получено разными способами. Кроме работ, упоминавшихся выше, это число или близкие к нему получались и другими способами. Так, В. Чавчанидзе [59], рассматривавший, как и большинство других авторов, триплетный код, ввел два постулата, согласно первому из которых замена А на Т, Т на А, Г на Ц и Ц на Г не меняет информации, заключенной в триплете, а согласно второму «зеркальные» триплеты кодируют одну и ту же аминокислоту, получив ровно 20 комбинаций, дающих разную информацию. С. Везе [345] получает необходимое количество триплетов, несущих разную информацию, сделав предположение, что при считывании нуклеотидного текста имеет значение только наличие замещений при втором или четвертом атоме углерода пуриновых и пиримидиновых колец. Н. В. Лучник [38] получил 22 значащих триплета, ограничив рассмотрение урацилсодержащими триплетами и предположив, что два зеркальных триплета не могут кодировать две разные аминокислоты.

Трудно представить себе какое-либо из общих свойств кода, о котором разные авторы не делали бы разных предположений. Даже представление о четырехбуквенном «алфавите» нуклеиновых кислот не является единственным. Некоторые авторы приводили доводы в пользу того, что этот алфавит состоит всего из двух букв. Так, в некоторых работах [278, 294] на основании того, что у бактерий сильно варьирует процентное содержание Г + Ц в их ДНК при значительном сходстве продуцируемых ими белков, было высказано предположение о том, что смысл имеют только 6-амино- и 6-кето-группы. Таким образом, А имеет то же значение, что Ц, а Г — такое же, как У.

Э. Теллер предложил «код последовательного ряда», в котором каждая аминокислота кодируется двумя основаниями и предыдущей аминокислотой [16]. Это — не единственная гипотеза, согласно которой в кодировании участвуют не только азотистые основания. Так, еще в 1950 г. предлагался диплетный код [108], элементами которого являются, кроме пуриновых и пиримидиновых оснований, также фосфорибозные остатки.

Комплементарные нити РНК способны образовывать двойные спирали. Если такую вторичную структуру дает и-РНК, то

становится невозможным взаимодействие с ней т-РНК [277]. Исходя из этих соображений, В. Г. Туманян и сотрудники [54] предложили такой порядок оснований в триплетах, при котором сводится к минимуму возможность образования и-РНК вторичной структуры.

Н. Суеока [293] исследовал зависимость содержания аминокислот в белках разных видов бактерий от содержания пары Г + Ц в их ДНК. Положительная связь говорила за преобладание в кодирующих словах Г и Ц над А и У, отрицательная — наоборот. Но для большого числа аминокислот корреляции с составом ДНК не было обнаружено. Из этого можно сделать вывод, что кодирующие группы для этих аминокислот содержат одинаковое число Г + Ц и А + У, что может быть только при четном числе нуклеотидов в группе. На этом основании Суеока обсуждает возможность того, что кодовое число равно четырем или шести. Кодовое число, близкое к шести, получали и другие авторы [21, 199, 291].

В последнее время открылись значительно более широкие и глубокие основания для расшифровки кода, в результате использования которых выяснился ряд свойств кода и многие гипотетические представления сразу смогли быть отброшены. Сделанный выше краткий обзор предлагавшихся гипотез (в который вошли не все работы) имел целью показать обилие делавшихся предположений, для каждого из которых были большие или меньшие основания. Уже в 1956 г. М. Ичас в одной из своих статей [25, 350] вынужден был констатировать, что «в настоящее время главная трудность состоит не в нахождении принципа кодирования, объясняющего известные факты, а в выборе между многими возможными». К этому можно добавить, что против некоторых из высказывавшихся, в том числе и малоправдоподобных, гипотез и сейчас, при накоплении значительно большего фактического материала, нельзя привести уничтожающих доводов и что некоторые подобные гипотезы высказывались даже в самый разгар экспериментальных исследований проблемы кода и с полным знанием всех имеющихся данных. Это обстоятельство заставляет с осторожностью относиться к казалось бы общепринятым и вполне правдоподобным представлениям и стараться искать новые методы для их проверки.

ЭКСПЕРИМЕНТАЛЬНЫЕ ПОДХОДЫ К РАСШИФРОВКЕ КОДА

Сравнительно недавно стали намечаться гораздо более плодотворные экспериментальные подходы к разрешению проблемы аминокислотного кода. В 1960 и 1961 г. появились три новых метода, которые уже сейчас дали существенные результаты. Поскольку нам в дальнейшем часто придется иметь с ними дело, нужно дать представление об этих методах.

Опыты с мутациями фага Т4, вызванными профлавином. В настоящее время лучше всего изученным генетически на молекулярном уровне является участок rII бактериофага Т4, паразитирующего на E. coli. Существуют методы точнейшей локализации мутаций в этом участке, и очень большое число мутаций, полученных с помощью разных агентов, точно локализовано и подробно исследовано. Методике работы с этим объектом и обзору полученных результатов посвящен ряд статей [7, 8, 84]. Из исследованных мутагенных агентов довольно своеобразно действовал профлавин (одно из производных акридина): при воздействии на профлавиновые мутации другими агентами, про которые было известно, что они вызывают изменение азотистых оснований, не наблюдалось реверсий (обратных мутаций) к нормальному дикому типу [57, 152]. Поэтому можно было думать, что профлавин оказывает свой эффект не за счет изменения оснований, а за счет образования делеций или вставок [100]. Это подтверждают также результаты модельных опытов [220, 221], показавших, что при взаимодействии с ДНК молекулы профлавина вклиниваются между соседними парами нуклеотидов. В пользу такой интерпретации говорят и другие данные [147, 200, 291]. Эти выводы использовали Ф. Крик и сотрудники [35, 128], проведшие ряд остроумных опытов, давших важную информацию об общих свойствах генетического кода.

При воздействии профлавином на мутантные формы, возникшие под действием профлавина же, нередко возникают реверсии к дикому типу. Оказалось, что эти реверсии возникают не путем обратного изменения, а путем новой мутации, находящейся на некотором расстоянии от первой. Если эта новая мутация находится в геноме изолированно, то она дает тот же фенотипический эффект, что и первая; если же они присутствуют вместе, то фаг неотличим от дикой формы.

Такой результат можно объяснить тем, что первая мутация состояла во вставке одного нуклеотида, а вторая — в делеции одного нуклеотида (или наоборот). В соответствии с этим все мутации можно разбить на два класса: мутации (+)-типа и мутации (—)-типа. На большом материале было показано, что совмещение в геноме двух мутаций (+)-типа или двух мутаций (—)-типа всегда дает мутантный фенотип, а совмещение одной мутации (+)-типа и одной мутации (—)-типа дает (за исключением некоторых точно предсказываемых случаев) возврат к дикому типу. Отсюда, кроме подтверждения высказанной выше гипотезы о механизме мутагенного действия профлавина, следует важный вывод о том, что код не имеет запятых, о чем мы уже говорили выше.

Для рассмотрения дальнейших опытов представим себе вначале, чего следует ожидать, если на некотором участке будет

сосредоточено столько вставок (или делеций), сколько нуклеотидов входит в одно кодовое слово. Очевидно, что между первой и последней вставкой первоначальный текст должен быть нарушен, а за их пределами он будет читаться нормально, поэтому если в нарушенном участке нет мест, необходимых для нормального функционирования белка, то должен наблюдаться нормальный фенотип. Опыты показали, что совмещение в одном геноме трех мутаций (+) или (—)-типа приводит к реверсиям. Отсюда следует вывод о том, что кодовое отношение равно трем, т. е. каждой аминокислоте соответствует 3 нуклеотида в и-РНК.

В специальных опытах было показано, что текст считывается по порядку, начиная с некоторой точки. Для этого была взята делеция, затрагивавшая два цистрона — конец цистрона А и начало цистрона Б. Обычно делеции, происшедшие в цистроне А, не влияют на функцию цистрона Б. Если же такие делеции вводились в только что описанный штамм, то они в ряде случаев подавляли и функционирование цистрона Б. Эти результаты можно объяснить тем, что текст всегда читается с начала цистрона, а в данном штамме выпало начало цистрона Б и потому текст считывается с начала цистрона А до конца цистрона Б.

Из этих опытов следует также вывод о том, что код должен быть сильно вырожденным. Поскольку при совмещении в геноме делеции и вставки продуцируется жизнеспособный белок, несмотря на то, что большинство триплетов, заключенных между ними, должны быть измененными, значит почти любое сочетание трех нуклеотидов кодирует какую-либо из аминокислот. Более подробно эти опыты описаны в работах [34, 35, 128].

Опыты с нитритными мутациями вируса табачной мозаики. Нуклеиновые кислоты РНК-содержащих вирусов, к которым относится вирус табачной мозаики, действуют непосредственно в качестве и-РНК [309]. Поэтому замены нуклеотидов в вирусной РНК (точнее в той ее части, которая соответствует структурным генам) должны приводить к замещению аминокислот в вирусном белке. Поскольку не только аминокислотный состав, но и последовательность аминокислот в белке в настоящее время доступны точному анализу, можно в каждом случае выяснить характер замещения, из чего могут делаться выводы о некоторых общих свойствах кода и о составе кодовых слов. Особенно большую информацию дают опыты, в которых мутагенным агентом служит азотистая кислота [243]. Дело в том, что единственная реакция, в которую это вещество вступает с РНК — дезаминирование, в связи с чем под действием азотистой кислоты возможно только два превращения оснований: цитозин превращается в урацил, а аденин (через гипоксантин) в гуанин [17, 63, 80]. Поэтому, если на достаточно

большом материале окажется, что, например, фенилаланин замещает серин и лейцин, но сам не замещается ни одной аминокислотой, то отсюда следует, что кодовое слово, соответствующее фенилаланину, может состоять только из У и Г. Обратно, если какая-нибудь аминокислота дает замещения на другие, но сама не замещает других аминокислот, то ее кодовое слово должно состоять из А и Ц. Наконец, если некоторая кислота и замещает и замещается, то ее кодовое слово состоит из нуклеотидов обеих групп. Один только анализ нитритных замещений не может сам по себе дать расшифровку ни одного кодового слова. Но если имеется другая информация, то нитритные замещения могут сильно помочь расшифровке. Так, если в рассмотренном выше примере (замещение серина и лейцина на фенилаланин), известно, что фенилаланину соответствует триплет УУУ, то из этих результатов следует, что кодовые слова для серина и лейцина должны иметь состав 2У, 1Ц. (Порядок букв остается неизвестным).

Анализ нитритных мутаций дает также сведения о вырожденности кода. Так, мы только что установили, что серин и лейцин имеют триплеты одинакового состава: 2У, 1Ц. Но под влиянием азотистой кислоты наблюдалось замещение серина на лейцин. Поскольку для триплетов указанного состава такое замещение невозможно, следует предполагать, что, по крайней мере, одна из этих аминокислот имеет дополнительный триплет (например, серин имеет кроме 2У, 1Ц, еще триплет состава 1У, 2Ц). Кроме того, опыты с нитритными мутациями дают сведения и о других общих свойствах кода. Так, в опытах Виттмана [342] было исследовано 117 мутаций. Из них у 82 состав белка не отличался от исходной формы, у 29 мутаций было замещено по одной аминокислоте, у шести — по две, причем ни в одном случае замещенные аминокислоты не стояли рядом. Мутации с тремя и более замещениями не наблюдались ни разу. Такой результат говорит о том, что код неперекрывающийся. Более подробно методика и результаты описаны в работах [13, 14, 341, 342, 343].

Синтез белков под влиянием искусственных полинуклеотидов. В первой главе уже говорилось о том, что М. Ниренбергу и Г. Маттеи [46, 47, 250, 251] удалось создать бесклеточную систему, в которой происходил синтез полипептидов, в частности при использовании в качестве и-РНК искусственных полинуклеотидов. Известен фермент—полинуклеотидфосфорилаза, синтезирующий полинуклеотиды из нуклеозид-5′-дифосфатов в отсутствие затравки [48, 170]. Поэтому могут быть синтезированы искусственные РНК с заранее заданным составом нуклеотидов (считают, что чередование их в цепи случайно). Этот полинуклеотид вносится в систему в качестве и-РНК и инкубируется со смесью аминокислот, одна из которых

радиоактивно маркирована. После инкубации исследуют осадок, нерастворимый в горячей трихлоруксусной кислоте, на наличие радиоактивности, по чему судят о степени включения данной аминокислоты в белок.

Первый полученный таким образом результат состоял в том, что полиуридиловая кислота (т. е. РНК состава УУУУУУ...) стимулировала включение только одной аминокислоты из 20, а именно фенилаланина. Отсюда можно сделать вывод о том, что, в случае триплетности кода, фенилаланин кодируется «словом» УУУ. Остальные полинуклеотиды, состоявшие из нуклеотидов одного только типа, в первых опытах не стимулировали включения аминокислот. Зато все кополимеры, состоявшие из нуклеотидов двух или трех сортов, из которых один был уридиловой кислотой, активно включали по нескольку аминокислот. Так, поли-УЦ включала фенилаланин, серин, лейцин и пролин. Отсюда, поскольку уже известно, что фенилаланин кодируется триплетом УУУ, а триплет ЦЦЦ ничего не кодирует (последний вывод, как выяснилось позже, оказался неверным), делается заключение о том, что серин, лейцин и пролин кодируются словами, в состав которых входят и У, и Ц.

Анализируя результаты количественно, можно указать и на соотношение У и Ц в словах. Так, в поли-УЦ с соотношением У к Ц 5 : 1 триплеты состава 2У, 1Ц должны встречаться в 5 раз, а состава 1У, 2Ц — в 25 раз реже, чем триплет УУУ. Опыты показывают, что экспериментально наблюдаемые выходы близки к вычисленным, а именно, серина и лейцина включается примерно в 5 раз, а пролина в 25 раз меньше, чем фенилаланина [51, 231, 289]. Отсюда следует, что первым двум аминокислотам соответствуют триплеты состава 2У, 1Ц, а последней — 1У, 2Ц. В подобных опытах удалось определить уже в начале 1962 г. состав триплетов для 19 аминокислот из 20 [51, 219, 219а, 231, 289, 290]. Для оставшейся аминокислоты (глютамин) состав триплета был предсказан на основании нитритного замещения у вируса табачной мозаики [290].

Для некоторых аминокислот было найдено по два и даже по три триплета, что указывало, в согласии с другими данными, приводившимися выше, на частичную вырожденность кода. В табл. 2 дается сводка результатов этих опытов. Подробно применявшаяся методика описана в статьях [46, 47]. Впервые бесклеточная система, синтезирующая белок, была создана из продуктов, выделенных из кишечной палочки (все данные, приведенные в табл. 2, взяты именно из этих опытов). Позже удалось создать такие же системы из материала разных клеток млекопитающих [75, 146, 232, 328].

Применение в этих системах искусственных полинуклеотидов приводит к тому же составу триплетов, который получен на бактериальной системе.

Таблица 2

Состав кодирующих триплетов на основании опытов с урацилсодержащими СПРН

Аминокислота	Состав триплетов	
	данные группы С. Очоа [290]	данные группы М. Ниренберга [231].
Аланин	У, Г, Ц	(У, Г, Ц)
Аргинин	У, Г, Ц	(У, Г, Ц)
Аспарагиновая кислота	У, А, Г	—
Аспарагин	1У, 2А+У, А, Ц	—
Валин	2У, 1Г	2У, 1Г
Гистидин	У, А, Ц	—
Глицин	1У, 2Г	1У, 2Г
Глютаминовая кислота	У, А, Г	У, А, Г
Глютамин	(У, Г, Ц)	—
Изолейцин	2У, 1А	2У, 1А
Лейцин	2У, 1Ц+2У, 1Г+2У, 1А	2У, 1Ц+2У, 1Г
Лизин	1У, 2А	1У, 2А
Метионин	У, А, Г	У, А, Г
Пролин	1У, 2Ц	1У, 2Ц
Серин	2У, 1Ц	2У, 1Ц+У, Г, Ц
Тирозин	2У, 1А	2У, 1А
Треонин	У, А, Ц+(1У, 2Ц)	—
Триптофан	1У, 2Г	1У, 2Г
Фенилаланин	УУУ	УУУ
Цистеин	2У, 1Г	2У, 1Г или 1У, 2Г

Примечания. Триплет для глютамина не был найден в опытах с СПРН и предсказан на основании нитритного замещения глю-NH_2→вал. В скобках приведены те триплеты, от которых авторы отказались в последующих публикациях [201, 320].

Основные работы трех описанных направлений были опубликованы в конце 1961 — начале 1962 гг. и вызвали шумную сенсацию: казалось, что проблема генетического кода в основном решена и остается найти последовательность нуклеотидов в уже известных триплетах. Как выяснилось, расшифровать порядок нуклеотидов в триплетах, в рамках изложенных работ, оказалось не так трудно, и это было с большим или меньшим успехом сделано независимо друг от друга несколькими авторами [38, 203, 260, 262, 280]. Но критическое рассмотрение сделанных работ возбудило ряд вопросов, заставляющих с большой осторожностью относиться к делавшимся выводам; с другой стороны, появились новые данные, сильно осложнившие проблему расшифровки.

Основным источником информации для расшифровки кода служили опыты со стимуляцией включения аминокислот синтетическими полинуклеотидами. Их выводы отнюдь не бесспорны. Так, Г. Виттман в своей последней статье [343] приводит целый десяток возражений, заставляющих если не усомниться,

то, по крайней мере, очень критически подходить к делавшимся выводам. Эти возражения сводятся к следующему.

1. В ряде работ состав применявшихся синтетических полинуклеотидов не определялся, и делалось предположение, что их состав соответствует пропорции нуклеотидов, взятых для синтеза. Однако было показано, что между пропорцией нуклеотидов, взятых для синтеза, и соотношением оснований в образовавшемся полимере могут быть очень большие расхождения [102].

2. Предположение о чисто статистическом распределении оснований вдоль полинуклеотидной цепи, являющееся необходимой предпосылкой для расшифровки результатов, совершенно не проверено. О том, что предположение о статистическом распределении не является единственно возможным, говорят как результаты опытов по синтезу ин витро полидезоксирибонуклеотидов, где, в случае, например, поли- (dAT), получается правильная последовательность ... ATATAT... [266], так и результаты, упомянутые в предыдущем пункте.

3. Вторичная структура, которую могут принимать полинуклеотиды, может препятствовать включению аминокислот даже при наличии кодирующей последовательности нуклеотидов [277].

4. Нельзя считать доказанным, что в опытах, о которых идет речь, действительно синтезируются белки [352].

5. Методика расчета, применяемая для расшифровки, пригодна, строго говоря, только для невырожденного кода [343].

6. Зачастую расчеты состава триплетов основываются на очень малых выходах, достоверность которых подвергается сомнению [125, 201].

7. Кажется весьма вероятным, что до сих пор улавливается только часть синтезируемых продуктов [290].

8. Неизвестно, как влияют «нонсенсы» на считывание информации ин витро [343].

9. Степень включения определенной аминокислоты данным полинуклеотидом зависит от присутствия и отсутствия других аминокислот. Так, чистая поли-У, включающая в присутствии всех аминокислот практически только фенилаланин, в его отсутствие включает довольно значительные количества лейцина [230а].

10. Полученные до сих пор триплеты требуют очень высокого содержания У в и-РНК, что противоречит фактам [118, 298]. К этим возражениям можно, при желании, прибавить и ряд других. Ни одно из этих возражений, правда, не опровергает делавшихся выводов, а только заставляет в них сомневаться, так как в большинстве случаев речь идет не о противоречии фактам, а о недостаточной обоснованности исходных положений. Наиболее серьезными возражениями являются, на наш

взгляд, приведенные в пунктах 5 и 10, так как в них уже идет речь о противоречии фактам.

Последнее возражение — несоответственно высокое содержание урацила в кодовых словах, по сравнению с истинным содержанием его в РНК, недавно отпало после того, как была усовершенствована методика, позволившая экспериментировать и с полинуклеотидами, не содержащими урацила [102, 156, 201, 277, 320]. В этих опытах было найдено довольно большое число значащих безурацильных триплетов (табл. 3). Однако эти

Таблица 3

Дополнения к списку кодирующих триплетов, полученные в последних опытах с СПРН

Аминокислота	Состав триплетов	
	данные группы С. Очоа [156, 320]	данные группы М. Ниренберга [201, 277]
Аланин	А, Г, Ц+1Г, 2Ц	1Г, 2Ц
Аргинин	2А, 1Г+1Г, 2Ц	1Г, 2Ц
Аспарагиновая к-та	А, Г, Ц	2А, 1Ц
Аспарагин	2А, 1Ц	
Валин	—	—
Гистидин	1А, 2Ц	1А, 2Ц
Глицин	1А, 2Г+2Г, 1Ц	—
Глютаминовая к-та	2А, 1Г	2А, 1Г+2А, 1Ц
Глютамин	1А, 2Г+2А, 1Ц	
Изолейцин	2А, 1У	—
Лейцин	—	УУУ
Лизин	ААА	ААА+2А, 1Г+2А, 1Ц
Метионин	—	—
Пролин	ЦЦЦ+1А, 2Ц	ЦЦЦ+1А, 2Ц+1Г,2Ц
Серин	А, Г, Ц	—
Тирозин	—	—
Треонин	2А, 1Ц+2Ц, 1Г	2А, 1Ц+1А, 2Ц
Триптофан	—	—
Фенилаланин	—	—
Цистеин	—	—

результаты еще более усиливают возражение 5, так как показывают, что код в действительности является довольно сильно вырожденным. В связи с этим становится еще более важным, чем раньше, найти метод определения состава кодовых слов, полностью независимый от результатов опытов с синтетическими полинуклеотидами.

Остается кратко рассказать об основных работах, вышедших в свет за самое последнее время.

Вполне естественно, что многих авторов занимал вопрос, решение которого казалось ближайшей задачей: определение порядка нуклеотидов в триплетах. Во всех работах такого рода производилось сопоставление состава кодирующих трипле-

тов, полученного на основании опытов с синтетическими полинуклеотидами и замещений аминокислот в мутантных белках. Если не ставить дополнительных условий, то на основании только этих предположений установить абсолютную последовательность нуклеотидов в триплетах невозможно. Но можно попытаться установить их относительную последовательность.

По этому пути пошли И. Рыхлик и Ф. Шорм [262, 283] и Э. Смит [280], давшие по шесть вариантов кода, соответственно шести возможным способам, которыми можно расположить буквы в том триплете, с которого начинается расшифровка. При этом получается любопытный результат: в 16 триплетах урацил занимает постоянное место. Это позволило И. Рыхлику и Ф. Шорму [262] и Р. Робертсу [260] отбросить этот общий урацил и рассматривать диплетный код, который дает два эквивалентных варианта. Н. В. Лучник [38], обратив внимание на некоторые закономерности в использовании для кодирования урацилсодержащих триплетов, наложил на код ограничения, позволившие из шести вариантов триплетного кода оставить только два.

Важной для расшифровки порядка нуклеотидов в триплетах оказалась работа с синтетическими полинуклеотидами, в которой использовалась в качестве и-РНК смесь полинуклеотидов АУУУУ... и ААУУУ... [319]. Эти полимеры стимулировали, кроме фенилаланина, включение тирозина, не влияя на включение изолейцина, аспарагина и лизина. Аналогично смесь ГУУУУ... и ГГУУУ... включала, наряду с фенилаланином, небольшие количества цистеина и не включала валина, глицина и триптофана. Отсюда следует, что тирозин кодируется триплетом АУУ, а цистеин — ГУУ. После этого можно выбрать правильный вариант кода из тех шести, которые были первоначально возможны, что и было сделано в одной из работ [203]. Все эти работы должны были привести к безусловно правильным результатам, если бы код был однозначным, т. е. если бы каждой аминокислоте соответствовало только по одному кодирующему триплету и если бы все замещения аминокислот действительно являлись результатом замены одного нуклеотида. Однако с полной определенностью можно сказать, что ни одно из этих условий не выполняется в действительности. В одной из работ [180] даже ставится под сомнение допустимость использования данных по замещениям в связи с возможностью чисто случайных совпадений. В работах Н. В. Лучника [38] и Т. Юкса [203] эти обстоятельства учитываются: рассматривается слабо вырожденный код и отбираются наиболее достоверные замещения, но это сделано в совершенно недостаточной мере.

В табл. 4 приведены все коды с определенной последовательностью нуклеотидов. В тех из них, где в основу расшифровки брался состав триплетов из работ [231, 290], коды преобра-

Таблица 4

Коды с определенной последовательностью букв

Аминокислота	И. Рыхлик и Ф. Шорм [262]	Р. Робертс [261]	Н. В. Лучник [38]	Т. Юкс [203]	Г. Зьюби и Х. Квастлер [352]	Р. Везе [345]	Е. Смит [280]	В. Туманян и сотр. [54]
Аланин	ЦУГ	ЦУГ	ГУЦ	ЦУГ	УЦГ	ГУЦ	УЦГ	ЦУГ
Аргинин	ГУЦ	ГУЦ	ЦУГ	ГУЦ	УГЦ	ГЦУ	ГУЦ	ГЦУ
Аспарагиновая к-та	АУГ	ГУА	АУГ	ГУА	УЦА	ГУА	ГУА	ГУА
Аспарагин	АУА	ГУА	АУЦ+ +УАА	ЦУА+УАА	УЦА	АЦУ+ААУ	ЦУА	ААУ+АЦУ
Валин	УУГ	УУГ	УУГ	УУГ	УУЦ	УУГ	УУГ	УУГ
Гистидин	АУЦ	АУЦ	ЦУА	АУЦ	УГУ	АУЦ	АУЦ	АУЦ
Глицин	ГУГ	ГУГ	ГУГ	ГУГ	УУГ	ГУГ	ГУГ	ГУГ
Глютаминовая к-та	АУГ	АУГ	ГУА	АУГ	УУА	АУГ	АУГ	АУГ
Глютамин	ЦУГ	АУГ	УЦГ	УЦГ	УУА	—	УУГ	ГУЦ
Изолейцин	УУА	УУА	УУА	УУА	УАЦ	УУА	УУА	УУА
Лейцин	УУЦ	УУЦ	ЦУУ	УАУ+ ЦУУ+УГУ	УЦУ	УЦУ	ЦУУ	УУЦ+ АУУ+ГУУ
Лизин	АУА	АУА	АУА	АУА	УГА	АУА	АУА	АУА
Метионин	ГУА	АУА	УГА	УГА	УАУ	УАГ	УГА	УАГ
Пролин	ЦУЦ	ЦУЦ	ЦУЦ	УЦЦ+ЦУЦ	УЦЦ	УЦЦ	ЦЦУ	УЦЦ
Серин	ЦУУ	ЦУУ	УУЦ	УУЦ+УЦУ	УГГ	УУЦ	УЦУ	УЦУ
Тирозин	АУУ	АУУ	АУУ	АУУ	?АУ	АУУ	АУУ	АУУ
Треонин	ЦУА	ЦУА	УЦЦ+ +УЦА	УЦА.	УАГ	УЦА	УЦА	УЦА+ЦЦУ
Триптофан	ГУГ	ГУГ	УГГ	УГГ	УАА	ГУУ	ГГУ или УГГ	ГГУ
Фенилаланин	УУУ	УУУ	УУУ	УУУ	УУУ	УУУ	УУУ	УУУ
Цистеин	ГУУ	ГУУ	ГУУ	ГУУ	?ЦГ	ГУУ	УГУ или ГУУ	ГУУ

Примечания. 1. Код Робертса изменен добавлением ко всем словам У. 2. Из кода Везе взяты только триплеты, соответствующие составу триплетов Очоа.

зованы таким образом, чтобы тирозин кодировался триплетом АУУ, а изолейцин — УУА. Как видно, между этими кодами имеются расхождения, что связано с уже указанными причинами. С другой стороны, сходство между разными кодами значительно больше, чем случайное, а это говорит за то, что установленный в обсуждаемых работах порядок нуклеотидов содержит какую-то долю истины.

Ряд работ был посвящен обсуждению противоречия урацилсодержащего кода с наблюдаемым в действительности составом РНК и попыткам примирить эти противоречия [217, 260, 345]. На обсуждении этих работ нет нужды останавливаться, так как после открытия значительного числа триплетов, не содержащих урацил, они потеряли свою актуальность. Упомянем в этой связи лишь о работах Р. Робертса [260, 261]. Этот автор обратил внимание на то, что в урацилсодержащем триплетном коде общий урацил не несет никакой информации и что основаные данные по включению аминокислот под действием синтетических полинуклеотидов и по замещениям аминокислот в белках также хорошо объясняются в рамках диплетного кода, который может быть получен из триплетного путем «сокращения» всех триплетов на общий урацил. Хотя диплетный код, как уже упоминалось, дает слишком мало кодовых слов, почему он никогда и не рассматривался всерьез, но, как показал Р. Робертс, он очень хорошо соответствует наблюдаемому соотношению нуклеотидов в РНК; кроме того, в этих работах приведены и некоторые другие данные, говорящие в пользу правдоподобности диплетного кода.

Интересна работа Г. Зьюби и Х. Квастлера [352], давших код (также приведенный в табл. 4) с использованием из работ с полинуклеотидами только данных о составе триплетов для фенилаланина и пролина. Хотя данная ими расшифровка сильно противоречит всем остальным, работа заслуживает специального внимания в связи с интересным методом, примененным в ней. Авторы располагают все возможные триплеты в виде трехмерной решетки таким образом, что триплеты, переходящие один в другой в результате замены одного нуклеотида, находятся на одной прямой. Эта решетка сильно облегчает расшифровку кода, и, как будет показано ниже, некоторое ее обобщение позволяет, кроме того, делать выводы и об общих свойствах кода.

Самым важным достижением последних месяцев является, конечно, открытие большого числа новых триплетов, не содержащих урацила, что удалось за счет ряда усовершенствований методики [102, 156, 201, 277, 320]. После этих работ отпали некоторые из существенных возражений, о которых говорилось выше, и стало окончательно ясно, что код должен быть довольно сильно вырожден.

Как уже говорилось, наибольшую информацию об общих свойствах кода дают работы трех направлений: генетические опыты с мутациями фага Т4, вызванными профлавином, анализ белков в мутациях и использование синтетических полинуклеотидов в качестве и-РНК. Однако этим не исчерпываются источники информации, имеющие значение для проблемы кода. Кратко перечислим их.

1. *Статистика аминокислот в белках.* Как уже говорилось, в случае справедливости перекрывающихся кодов, порядок аминокислот в белках мог дать ключ к расшифровке кода. Этого не оказалось, но анализ аминокислотных соседств явился главным основанием для отказа от перекрывающихся кодов. Поэтому закономерности строения белков могут служить для проверки правильности тех или иных гипотез. Кроме анализа соседств аминокислот, может оказать службу также и анализ частоты встречаемости аминокислот в белках, так как не лишено оснований предположение о том, что встречаемость аминокислот коррелирует с числом кодовых слов, соответствующих определенным аминокислотам.

2. *Генетические рекомбинации на молекулярном уровне.* В последнее время удалось получить генетические рекомбинации, соответствующие обменам внутри кодирующих слов [182]. Эти данные, в принципе, могут дать не меньше оснований для определения порядка нуклеотидов в триплетах, чем замещения аминокислот. Остается только пожалеть, что число таких рекомбинаций, изученных до сих пор, можно пересчитать по пальцам одной руки.

3. *Корреляции в составе РНК и белков.* К этой категории данных относятся по сути дела обсужденные выше опыты с использованием искусственных полирибонуклеотидов в качестве и-РНК. В принципе сходным образом могли бы быть использованы и естественные РНК и белки. Но, к сожалению, до сих пор еще не выделено ни одной и-РНК, соответствующей определенному белку. Далеко не отвечают этой цели и вирусные РНК. Хотя таковые несут функцию и-РНК, но структурный ген, соответствующий исследуемому белку, составляет только небольшую ее часть [341]. Тем не менее сопоставление даже общего состава РНК с составом белков в рамках урацилсодержащего кода явилось одним из оснований для его критики, которая, как стало ясно после открытия безурацильных триплетов, оказалась правильной. Были попытки, уже упомянутые выше [155, 344] использовать корреляции в составе вирусной РНК и белка для определения состава кодовых слов.

4. *Корреляции в составе ДНК и белков.* Поскольку белки синтезируются на одноцепочечной и-РНК, речь может итти только о корреляции состава белков с парами нуклеотидов (А—Т и Г—Ц), а не с отдельными нуклеотидами ДНК. Та-

кое сопоставление было проведено Н. Суеока для суммарного белка и суммарной ДНК ряда бактерий [293]. Полученные зависимости, как это уже отмечалось [217, 348], хорошо коррелируют с составом кодовых слов, установленным в опытах с синтетическими нуклеотидами. Как будет показано позже, эти результаты можно успешно использовать и для независимого определения состава кодовых слов.

5. *Гетерогенность т-РНК*. В ряде работ [295, 330] было показано, что, по крайней мере, некоторым аминокислотам соответствует более одной т-РНК. При накоплении данных такого рода их можно было бы использовать для определения степени вырожденности кода.

Этим, в основном, исчерпываются предпринимавшиеся подходы к расшифровке аминокислотного кода и полученные результаты.

ГЛАВА 3

ОСНОВНЫЕ ПОНЯТИЯ И ПОСТАНОВКА ЗАДАЧИ

В настоящей главе мы дадим определение тем понятиям, которыми будем в дальнейшем оперировать, и критически рассмотрим на основании материала, приведенного в двух первых главах, обоснованность современных представлений об основных общих свойствах аминокислотного кода После этого будет сформулирована задача настоящей работы.

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Для большей ясности и, отчасти, краткости последующего изложения дадим определение основным понятиям и терминам, с которыми приходится иметь дело при рассмотрении проблемы кода. При этом мы, в первую очередь, стремимся к максимальной понятности терминологии и везде, где только возможно, будем применять слова, уже употреблявшиеся для обозначения рассматриваемых понятий и потому привычных. Нетрудно заметить, что мы отдаем предпочтение терминам, применяемым в аналогичном смысле в теории информации [26, 66]; это вполне естественно, так как проблема аминокислотного кода является в значительной степени кибернетической проблемой. Ряд терминов уже был пояснен выше, тем не менее здесь им также дается определение, иногда более четкое. Ниже следуют эти термины и их определения. Приводя их, мы не претендуем на то, чтобы эту систему считали наилучшей, и не стремимся предлагать их для всеобщего применения. Единственная цель этого раздела — сделать возможно более ясным дальнейшее изложение.

Код — характер соответствия групп нуклеотидов в и-РНК (или в ДНК) аминокислотам полипептидной цепи.

Однозначный код — такой код, в котором каждой аминокислоте соответствует только одна группа нуклеотидов, расположенных в определенном порядке, причем каждая такая группа кодирует только одну аминокислоту.

Неоднозначный код — такой код, в котором хотя бы одна группа нуклеотидов кодирует более чем одну аминокислоту.

Вырожденный код — такой код, в котором хотя бы одной аминокислоте соответствует более одной кодирующей группы нуклеотидов.

Перекрывающийся код — такой код, в котором один и тот же нуклеотид и-РНК (или ДНК) может входить в состав нескольких кодирующих групп.

Универсальный код — такой код, общие свойства которого одинаковы для всей живой природы, причем по крайней мере большинство аминокислот кодируются у разных организмов одинаковыми сочетаниями нуклеотидов.

Буква — единица кодирующей группы и-РНК или ДНК (т. е. один нуклеотид, входящий в состав кодирующей группы).

Алфавит — совокупность букв, участвующих в кодировании.

Слово — группа букв, расположенных в определенном порядке, кодирующая аминокислоту.

Кодовое число — число букв в словах.

Кодовое отношение — число букв, приходящееся в среднем на одну кодируемую аминокислоту; другими словами — отношение числа нуклеотидов и-РНК к числу аминокислот в синтезируемой на ней полипептидной цепи. Ясно, что в неперекрывающемся коде без запятых кодовое число и кодовое отношение равны друг другу.

Так как есть основания считать, что не все буквы слова могут быть необходимыми для кодирования аминокислоты, о чем речь будет в последующих главах, целесообразно уже сейчас ввести понятие облигатных и факультативных букв.

Облигатные буквы — такие буквы слова, которые не могут быть заменены на другие буквы без того, чтобы тем самым не изменился смысл слова.

Факультативные буквы — такие буквы слова, при замене которых на другие (любые или вполне определенные) буквы, слово сохраняет первоначальный смысл, т. е. кодирует ту же самую аминокислоту.

Синонимы — слова (в вырожденном коде), имеющие одинаковый смысл (т. е. кодирующие одну и ту же аминокислоту).

Изокоды — слова (в неоднозначном коде), кодирующие более одной аминокислоты.

Знаки препинания — группы букв, не являющиеся словами, но участвующие в кодировании. Можно предположительно говорить о запятых, отделяющих слова друг от друга и о точках, кодирующих начало или конец аминокислотной цепи.

Словарь — совокупность всех слов и знаков препинания.

Нонсенс — группа букв, отсутствующая в словаре.

Текст — совокупность слов, кодирующих некоторый полипептид.

Замещения — различия в отдельных аминокислотах, стоящих на определенных местах, обнаруживаемые при анализе белков, измененных в результате мутаций, или при сравнении белков, близких по составу и строению.

Молекулярные механизмы возникновения замещений могут быть, вообще говоря, очень различными. Дадим определения только тем из них, о которых будет идти речь ниже и которые, согласно имеющимся данным, являются наиболее распространенными в природе.

Транзиция — замена одного пурина на пурин или одного пиримидина на пиримидин, т. е. такое изменение слова, при котором происходит одна из следующих замен букв А → Г, Г → А, Ц → Т(У), Т(У) → Ц.

Трансверзия — замена одного пурина на пиримидин, или одного пиримидина на пурин, т. е. такое изменение слова, которое произошло в результате одной из следующих замен букв: А → Ц, Ц → А, Ц → Г, Г → Ц, Г → Т(У), Т(У), → Г, А → Т(У), Т(У) → А.

Делеция — выпадение одного или нескольких стоящих рядом букв из текста.

Вставка — процесс, обратный делеции.

Разрешенными замещениями мы будем называть такие замещения аминокислот, причиной которых является замена одной буквы в слове (т. е. транзиции и трансверзии). Соответственно, запрещенными замещениями будем называть замещения, происшедшие в результате других механизмов, например, в результате замены нескольких букв в слове, делеций, вставок или других, не упомянутых выше, механизмов.

СОКРАЩЕННЫЕ ОБОЗНАЧЕНИЯ

Здесь же уместно дать сводку употребляемых сокращений (некоторые из них также уже были введены в первых двух главах).

Нуклеиновые кислоты и их компоненты. А — адениловая кислота, Г — гуаниловая кислота, Ц — цитидиловая кислота, У — уридиловая кислота, Т — тимидиловая кислота. Для соответствующих оснований будут применяться те же сокращения (например, А — аденин и т. д.).

ДНК — дезоксирибонуклеиновая кислота,

РНК — рибонуклеиновая кислота,

и-РНК — информационная («матричная») РНК,

т-РНК — транспортная («растворимая») РНК.

СПРН — синтетические полирибонуклеотиды. Состав СПРН обозначает с помощью сокращений для соответствующих нуклеотидов; например, поли-У — полиуридиловая кислота, поли-УЦ—СПРН, содержащий только У и Ц в определенной пропорции.

При обозначении порядка нуклеотидов в цепи слева стоит нуклеотид, имеющий свободную 5′-гидроксильную группу, справа 3′-группу. В случаях, когда порядок нуклеотидов неизвестен, они отделяются друг от друга запятыми. Если известен относительный порядок двух нуклеотидов из трех, то последний отделяется от первых двух точкой.

Аминокислоты и белки. Ала — аланин, арг — аргинин, асп — аспарагиновая кислота, асп-NH_2 — аспарагин, вал — валин, гис — гистидин, глиц — глицин, глют — глютаминовая кислота, глю-NH_2 — глютамин, илей — изолейцин, лей — лейцин, лиз — лизин, мет — метионин, про — пролин, сер — серин, тир — тирозин, тре — треонин, трип — триптофан, фала — фенилаланин, цис — цистеин. Асп*х* и глу*х* будут употребляться в тех случаях, когда неизвестно, идет ли речь об аспарагиновой или глютаминовой кислоте или соответствующих амидах.

При написании аминокислотных последовательностей в полипептидах слева ставится N-концевая аминокислота, справа С-концевая.

Замещения. В тех случаях, когда известно, какая из двух аминокислот является замещенной, а какая — заместившей, обозначаются сокращенными названиями аминокислот, соединенными стрелкой, показывающей направление изменения. Если направление замещения неизвестно, или в данной связи оно не имеет значения, то аминокислоты располагаются по алфавиту и соединяются черточкой.

Прочие сокращения. ВТМ — вирус табачной мозаики. Белок А — белок А-триптофансинтетазы кишечной палочки, АТФ — аденозинтрифосфат, ТХУ — трихлоруксусная кислота.

ОБОСНОВАННОСТЬ ПРЕДСТАВЛЕНИЙ ОБ ОБЩИХ СВОЙСТВАХ КОДА

Для того чтобы ясней сформулировать стоящие перед нами задачи, подытожим, на основании фактического материала, приведенного в двух первых главах, основные современные представления об общих свойствах кода и посмотрим, насколько они твердо обоснованы. Из этого будет ясно, из каких положений можно будет в дальнейшем исходить как из достаточно правдоподобных, а какие нуждаются в проверке или подтверждении. Речь идет о шести вопросах: является ли код перекрывающимся, чему равно кодовое число, каково кодовое отношение, считывается ли текст последовательно, или есть необходимость в существовании запятых, универсален ли код, и, наконец, является ли код однозначным, неоднозначным или вырожденным; в последних двух случаях необходимо знать степень неоднозначности или вырожденности. На эти вопросы в настоящее время дают обычно следующие ответы.

1. Код неперекрывающийся. Этот вывод делается главным образом на основании того, что аминокислоты распределены вдоль полипептидных цепей статистически случайно [99, 337] и потому, что при мутациях не наблюдается одновременного изменения двух смежных аминокислот [24, 342]. Однако предлагались модели перекрывающихся кодов, не противоречащие этим данным [322].

2. Кодовое отношение равно трем. Единственный известный нам ясный экспериментальный довод — результаты опытов с профлавиновыми мутациями фага Т4, описанные во второй главе [35, 128]. Некоторые другие методы определения кодового отношения, такие как стехиометрия включения фенилаланина под действием поли-У [231] или сопоставления, делаемые на основании анализа нитритных мутаций у ВТМ [342], приводят к значительно менее ясным результатам. Но эти результаты не противоречат утверждению о том, что кодовое отношение равно трем.

3. Кодовое число равно трем, т. е. код является триплетным. Хотя это положение почти общепринято, но пока нет ни одного прямого указания на его справедливость. С самого начала триплетный код казался наиболее вероятным, так как диплетный может дать только 16 слов [16]. Однако это не исключает возможности диплетного кода с дополнительным источником информации, вроде упомянутого выше «кода последовательного ряда», а тем более возможности кодового числа, большего трех. Выводы на основании опытов Ф. Крика с фагом Т4 являются довольно косвенными, так как они касаются не кодового числа, а кодового отношения, и перенесение этого вывода на кодовое число требует выполнения двух условий: неперекрываемости и отсутствия запятых, однако эти положения вполне строго доказанными считать нельзя. В пользу того, что код триплетный, пытались приводить также результаты опытов с СПРН, где оказывалось, что в ряде случаев для включения аминокислоты необходимо присутствие в матрице трех разных букв, но во всех случаях трех букв достаточно [231]. Эти соображения, конечно, не могут быть убедительными, тем более, что в настоящее время фактические данные изменились: для 19 из 20 аминокислот оказывается достаточным присутствия лишь двух разных букв и лишь для одного метионина необходимы три [201].

4. Текст считывается последовательно, начиная с некоторой фиксированной точки. Этот вывод делается опять-таки на основании изучения профлавиновых мутаций фага Т4 [128]. Тот же вывод следует из исследований механизма синтеза полипептидных цепей; опыты на ретикулоцитах и на E. coli показывают, что рост цепи идет постепенно, начиная с N-концевой аминокислоты [131а, 161а, 248, 353]. Из этого де-

лают заключение, что код не имеет запятых, так как при последовательном считывании в них нет надобности.

5. Код является универсальным. В пользу этого положения говорит довольно большое число данных. Состав кодовых слов, полученный в опытах с СПРН и с нитритными мутациями ВТМ, оказывается одинаковым [342]. РНК вируса табачной мозаики способна играть роль матрицы в бесклеточной бактериальной системе, причем синтезируется белок, подобный вирусному [309]. Замещения аминокислот при мутациях гемоглобина человека не противоречат составу слов, установленному на основании опытов с СПРН в бактериальной системе [279], то же справедливо и для замещений аминокислот вообще [38]. Опыты с СПРН в бесклеточных системах из тканей млекопитающих приводят к тому же составу кодовых слов, что и опыты с бактериальными системами [75, 146, 232, 328]. В пользу положения об универсальности кода говорят и некоторые другие соображения. Однако опыты с фагом Т4 [86] дают результаты, которые можно интерпретировать таким образом, что одна и та же группа нуклеотидов при развитии фага на одном штамме бактерий является кодирующим словом, а на другом — нонсенсом. Было бы интересно подтверждение этих результатов на другом материале. Но и в случае подтверждения этих данных они не будут противоречить представлению о том, что код универсален в рамках данного выше определения.

6. Вырожденность кода. Этот вопрос оказывается наименее ясным из всех. Разные авторы считали код сильно вырожденным [35, 128], слабо вырожденным [38, 203], не вырожденным однозначным [280] и даже неоднозначным [260, 261]. Ни одна из этих точек зрения не была до сих пор достаточно твердо обоснована. Поэтому Ф. Крик в одной из своих последних работ был вынужден прийти к выводу от том, что в настоящее время степень вырожденности кода — вопрос точки зрения [124].

Почти с полной уверенностью можно утверждать, что код в какой-то степени вырожден. В пользу этого говорит достаточное число данных, которые будут перечислены ниже, а авторы работ, в которых принимается не вырожденный код, не делают никаких попыток для того, чтобы объяснить эти данные иначе.

В пользу того, что код вырожден, говорили уже первые результаты по определению состава слов в опытах с СПРН. Для некоторых аминокислот было обнаружено по два или даже по три разных кодирующих слова [51, 231, 290]. Еще большая степень вырожденности следует из данных последних работ [156, 201]. Однако о степени вырожденности по этим данным судить нельзя. С одной стороны, нет гарантии, что найдены уже все кодирующие слова, с другой — в ряде случаев нельзя быть уверенным в достоверности выводов о составе слов. Выше уже

говорилось о тех возражениях, которые могут быть сделаны по отношению к заключениям, получаемым на основании опытов с СПРН, и о том, что подчас авторы сами отказываются от некоторых расшифровок, делавшихся в более ранних публикациях. Кроме того, в пользу вырожденности кода говорят данные Ф. Крика и сотрудников, полученные в опытах с профлавиновыми мутациями фага Т4 [35, 128], данные о гетерогенности т-РНК [295, 330] и другие доводы [118, 298]. Однако сколько-нибудь определенных высказываний о степени вырожденности, т. е. об объеме словаря, до сих пор сделать было нельзя. Из упомянутых опытов Ф. Крика и сотрудников [35, 128] как будто следует вывод о том, что почти любая случайная комбинация из трех букв является словом, но поскольку этот вывод нельзя не считать очень косвенным, одних этих данных для окончательного суждения об объеме словаря, конечно, недостаточно.

Как видно из этого рассмотрения, прочность, с которой обоснованы разные из перечисленных положений, очень различна. Так, о том, что считывание кода происходит последовательно, и о том, что код, по крайней мере, в значительной своей части является универсальным, говорят разные данные, а данных или гипотез, которые противоречили бы этим представлениям, нет, поэтому эти утверждения кажутся близкими к достоверным. Положения о том, что код неперекрывающийся и что кодовое отношение равно трем, также подтверждаются вполне определенными данными, которые однако допускают и другие, хотя и менее вероятные толкования. Наконец, положение о том, что код является триплетным, и вопрос о степени вырожденности кода являются наиболее неясными и слабо обоснованными. Странным образом оказывается, что наименее обоснованы как раз те свойства кода, знание которых особенно необходимо для его расшифровки, и те, которые на основании популярных статей наиболее известны. Кроме того, нетрудно заметить, что разные свойства кода тесно связаны друг с другом. Так, если код не перекрывающийся и не имеет запятых, то кодовое число должно быть равно кодовому отношению; о вырожденном коде речь может итти только в том случае, если кодовое число больше двух и т. д.

Интересно также отметить, что хотя число работ, посвященных проблеме кода, уже очень велико, вопросу об общих свойствах кода посвящена всерьез только одна работа [128], из которой и следует большинство из перечисленных выводов.

ПОСТАНОВКА ЗАДАЧИ

Излишне говорить о тех успехах в области проблемы кода, которые были достигнуты в течение последних полутора — двух лет; они очевидны. Однако совершенно очевидны также имею-

щиеся пробелы и недостатки. Прежде всего, недостаточно обоснованы представления об общих свойствах кода и, в частности, о таких важнейших, как кодовое число и объем словаря. Что же касается состава отдельных слов и порядка букв в них, то все проведенные до сих пор работы не считались с довольно значительной вырожденностью кода и потому их выводы не могут считаться достаточно достоверными.

При просмотре литературы последних лет, касающейся проблемы кода, бросается в глаза почти полное игнорирование статистических методов. Те же работы, где эти методы применяются, носят преимущественно негативный характер. Например, С. Бреннер [99] применяет статистические методы для доказательства невозможности простых перекрывающихся кодов, Р. Хендлер [180] для того, чтобы показать, что наблюдающееся соответствие между составом слов, полученным в опытах с СПРН, и замещениями может объясняться чистой случайностью. Между тем, статистические методы могут успешно применяться и для выяснения общих свойств кода и для расшифровки словаря. Попыткой такого применения и является настоящая работа.

Общий ход анализа проблемы кода, после сказанного выше, напрашивается сам собой. Прежде всего должны быть выяснены общие свойства кода. Поскольку эти свойства взаимосвязаны, целесообразно начать с того из них, которое в настоящее время является наиболее неясным — с определения степени вырожденности кода. Лишь после того, как установлены общие свойства кода — кодовое число, объем словаря, возможные ограничения в выборе букв для составления слов или ограничения в выборе слов для кодирования аминокислот — можно приступать к расшифровке смысла отдельных слов.

Кратко поясним характер тех путей, которые могут использоваться при решении поставленной задачи. Основной материал, которым мы будем пользоваться вначале, — замещения аминокислот. В принципе, замещение — типичная статистическая величина, так как она может принимать разные «значения», причем эти «значения» наблюдаются с различной частотой. Совершенно ясно, что набор этих значений (т. е. набор возможных замещений) и частота, с которой они происходят, должны зависеть от свойств кода. Ясно также, что число возможных замещений определенного класса (например, «разрешенных» замещений) теснейшим образом связано с объемом словаря. Так как нет никакой гарантии в том, что все разрешенные замещения уже наблюдались, то их число может быть определено статистически, т. е. речь идет об одной из обычнейших статистических задач: об оценке, на основании статистических характеристик некоторой случайной выборки, объема генеральной совокупности.

Принцип такой оценки логически очень прост. Представим себе, что в ящике находится некоторое (неизвестное) число шаров разного цвета. Будем наудачу вынимать из ящика шары, записывать их цвет и возвращать обратно. Через некоторое время исследуем полученную выборку. Если каждый наблюдавшийся цвет был обнаружен по нескольку раз, то можно быть почти уверенным в том, что число необнаруженных шаров очень невелико; если же каждый цвет наблюдался лишь однократно, то очевидно, что довольно большое число шаров не наблюдалось ни разу. Это рассмотрение носит сугубо качественный характер. Однако математическая статистика дает методы для достаточно точного количественного решения подобных задач. Если бы действительная ситуация вполне соответствовала только что рассмотренной модели, т. е. если бы наблюдались только «разрешенные» замещения и вероятность их обнаружения была одинаковой, то задача была бы очень простой. Однако заранее можно сказать, что это довольно сомнительно. Но степень влияния этих и других факторов также может быть количественно оценена путем статистического анализа материала, в результате чего можно довольно точно оценить число разрешенных замещений.

Если бы все остальные свойства кода были известны, то теперь уже можно было бы определить объем словаря. Так, если, например, было бы известно, что код триплетный, не имеет ограничений в выборе букв и что слова для кодирования аминокислот «выбираются» случайным образом, то после оценки числа разрешенных замещений можно было бы оценить и объем словаря. Но и эти общие свойства кода также не вполне ясны. Поэтому можно пойти обратным путем — использовать уже известное число разрешенных замещений для проверки разных гипотез о свойствах кода, так как для любого конкретного кода нетрудно вычислить теоретическое число разрешенных замещений.

Расшифровка конкретного смысла отдельных слов, производимая на основании сведений об общих свойствах кода, также требует применения статистического анализа. Фактические данные, которые могут быть использованы для расшифровки, не точны. Не только сами по себе количественные данные опытов с СПРН несут статистическую ошибку, но и любая их интерпретация сознательно или бессознательно базируется на определенных статистических представлениях (случайный порядок букв в СПРН, случайное или не случайное соответствие между словами и аминокислотами). То же справедливо и для данных о замещениях: сколько-нибудь определенные выводы можно делать только на основании их совокупности; про любое же отдельно взятое замещение нельзя с уверенностью сказать, возникло ли оно под действием данного агента или спонтанно,

является ли разрешенным или запрещенным. Поэтому расшифровка кода должна идти по пути сравнения совокупности экспериментальных данных с предсказаниями всех гипотез, возможных в рамках определенных представлений о его общих свойствах.

Поскольку основной фактический материал, которым нам предстоит пользоваться, — замещения аминокислот, в следующей главе будет дана сводка имеющихся данных, после чего будет проведен статистический анализ проблемы кода. Две главы будут посвящены общим свойствам кода: вначале будет с помощью нескольких методов оценено число разрешенных замещений, а затем, также с помощью нескольких методов, будут рассмотрены другие свойства кода. После этого, также в двух главах будет предпринята попытка расшифровки словаря, которая будет проводиться, исходя из полученных выводов об общих свойствах кода.

В связи с тем, что общие свойства кода нельзя считать окончательно выясненными, постараемся исходить с самого начала из минимального числа допущений. Единственное действительно необходимое допущение сводится к тому, что код существует, т. е. что информация о последовательности аминокислот в белке записана в виде последовательности нуклеотидов в нуклеиновой кислоте. Однако в предыдущих главах было приведено достаточно аргументов в пользу справедливости этого положения и здесь нет нужды к ним возвращаться. Кроме того, мы будем считать, что код универсален, в пользу чего также имеется достаточно доводов.

ГЛАВА 4

СТАТИСТИКА АМИНОКИСЛОТНЫХ ЗАМЕЩЕНИЙ

Прежде чем приступить к разрешению задачи, сформулированной в предыдущей главе, необходимо рассмотреть материал по замещениям, имеющийся в нашем распоряжении, а также подвергнуть его предварительному анализу для выяснения вопроса, насколько он соответствует предъявляемым к нему требованиям.

Замещения аминокислот анализировались уже в нашей работе, касавшейся определения последовательности нуклеотидов в урацилсодержащих триплетах [38], где и была дана сводка соответствующих данных. Этот материал был разбит по предположительной степени достоверности на три группы: в первую были отнесены одиночные замещения, по обе стороны которых в обоих сравниваемых пептидах совпадает, по крайней мере, по две аминокислоты, во вторую — одиночные замещения, по обе стороны от которых совпадает только по одной аминокислоте, в третью попали замещения, при которых изменены две аминокислоты, стоящие рядом; случаи, когда измененными оказывалось больше двух стоящих рядом аминокислот, не рассматривались.

Основой для анализа замещений служит предположение о том, что они являются результатом замены одной буквы в кодовом слове. Ряд данных говорит о том, что изменение отдельного нуклеотида действительно часто является причиной мутации; по-видимому, этот механизм можно считать наиболее распространенным [13, 14, 25, 57, 152, 279, 341]. Потому и была избрана принятая нами классификация. Проведенный в цитированной работе анализ показал, что замещения третьей группы довольно плохо соответствуют предположению о том, что они связаны с заменой одного нуклеотида, тогда как для двух первых групп соответствие оказывается довольно хорошим. Поэтому в настоящей работе мы не будем делать различий между замещениями первой и второй группы, а третью вообще исключим из рассмотрения.

За время, прошедшее со времени сдачи в печать упомянутой работы [38], появилось довольно много новых данных о замещениях, а некоторые из старых данных были уточнены. Кроме того, в ряде случаев раньше мы не имели возможности использовать оригинальные публикации. Поэтому ниже мы даем новую сводку аминокислотных замещений. Она охватывает материал, опубликованный до конца 1962 г., причем почти все данные проверены по первоисточникам. В эту сводку не вошла важная работа Х. Виттмана [342], вышедшая уже после окончания проведенного ниже анализа. Поскольку ее данные не могут ни в какой мере повлиять на сделанные выводы, мы не стали повторять трудоемких расчетов. Однако ниже, при расшифровке кода, новые данные Х. Виттмана будут использованы в полной мере. Не все данные по замещениям являются одинаково надежными. Наиболее достоверны, конечно, замещения, полученные в специальных мутационных опытах, где всегда точно известно, какая аминокислота находилась на данном месте вначале и какая ее заменила. На второе место, вероятно, можно поставить мутантные гемоглобины человека, так как в этих случаях идет речь об изменении лишь одной единственной аминокислоты из нескольких сотен. Что же касается замещений, наблюдаемых при сравнении сходных полипептидных цепей (например α и β-цепи гемоглобина) или аналогичных полипептидов у разных видов (например α-цепь гемоглобина человека и лошади), то здесь, конечно, замещение может являться следствием нескольких независимых мутаций, а также неизвестно направление мутационного изменения. Однако эти данные также пригодны для анализа, так как и в этих случаях одиночные замещения обычно не противоречат предположению о том, что в их основе лежит замена одного нуклеотида.

В табл. 5 приведены данные по замещениям в белке оболочки вируса табачной мозаики (ВТМ). Данные сгруппированы по авторам и по применявшимся воздействиям. К этому нужно сделать следующие замечания. За наиболее достоверный источник мы принимали последние оригинальные публикации авторов [13, 14, 305, 306, 308, 341]. Однако в некоторых случаях имеются противоречия в публикациях самих экспериментаторов, а иногда другие авторы, имевшие доступ к дополнительным неопубликованным данным, приводят другие замещения. Такие случаи выделены нами в отдельную группу (во всех случаях это касается мутаций, вызванных азотистой кислотой). Основания для выделения были следующие. А. Цуджита приводит в работе [305] данные дважды — в таблице и на рисунке, — причем относительно рисунка пишет, что на нем приведены все данные, полученные в лаборатории. Между тем в таблице значится замещение арг → лиз, которого нет на рисунке, а на рисунке приведено замещение тир → фала, которо-

Таблица 5

Замещения в мутациях вируса табачной мозаики

Действующий агент	Замещения	Литература
Азотистая к-та	асп*x*→ала (4), асп*x*→глиц (4), асп*x*→сер (2), глу*x*→вал (2), глу*x*→глиц (2), илей→вал (1), лей→фала (1), про→сер (3), про→лей (1), сер→фала (2), тре→илей (7), тре→мет (3), сер→лей (1)	[13, 14, 338, 339, 340, 341]
Азотистая к-та	асп*x*→ала (2), асп*x*→сер (1), арг→глиц (2), про→лей (2), сер→фала (3), тре→сер (2), глют→глиц (1)	[305, 307, 308]
Азотистая к-та	асп→лиз (?), асп-NH_2→арг (?), асп-NH_2→лиз (?), асп*x*→сер (1), арг→лиз (1), глю-NH_2→вал (1), илей→мет (?), илей→тре (?), тир→фала (1),	[203, 290, 305]
N-бромосукцинимид	асп*x*→сер (4), арг→глиц (1), про→лей (4), илей→тре (1)	[305]
N-бромосукцинимид (из сложных)	ала→глиц (2) (сравн. с G-TAMV) лей→илей (1) (сравн. с Y-TAMV)	[305, 306]
Диметилсульфат	арг→глиц (1), про→лей (3). сер→фала (1)	[305]
Спонтанные	асп-NH_2→арг (1), асп-NH_2→лиз (1), асп*x*→ала (2), илей→тре (1), лей→фала (1), сер→фала (1)	[13, 14, 339, 341]
Спонтанные	арг→глиц (1), асп*x*→сер (1) (сравн. с Y-TAMV)	[234, 305, 306]

го нет в таблице. И. Спейер и др. [290], приводя сводку по нитритным замещениям и ссылаясь, наряду с опубликованными работами, на устное сообщение А. Цуджиты, указывает на еще одно замещение асп*x* → сер и приводит к замещению глю-NH_2→ → вал. Эти четыре случая, по которым в литературе имеются противоречия, мы будем рассматривать как действительно происшедшие под действием азотистой кислоты. Т. Юкс [203] приводит в своей сводке, ссылаясь на опубликованную работу и личное сообщение Х. Виттмана, следующие новые нитритные мутации: асп-NH_2 → лиз, илей → тре, асп-NH_2 → арг,

асп → лиз и илей → мет (без указания их частоты). Поскольку три первых упомянуты в работе Х. Виттмана как спонтанные, тем более, что илей → тре противоречит тре → илей, мы склонны считать, что их указание вызвано недосмотром, что же касается мутаций асп → лиз и илей → мет, не приведенных в опубликованных работах, то мы будем их учитывать , для избежания возможных ошибок, как спонтанные. Кроме того, Т. Юкс там же уточняет асп*x* → глиц как асп → глиц и асп*x* → сер как асп-NH_2 → сер; с этим мы в дальнейшем будем условно считаться[1].

Среди мутаций ВТМ было обнаружено три, вызванные N-бромосукцинимидом [306], и одна спонтанная [234], отличающиеся от исходной формы 16 — 18 аминокислотами. Исследование показало, что они очень близки к естественным штаммам ВТМ — G-TAMV и Y-TAMV и отличаются от них в каждом случае всего лишь на одну аминокислоту [306]. В связи со столь большим сходством мы также будем принимать эти мутации во внимание, сравнивая их, однако, не с исходным штаммом, а с только что упомянутыми (см. табл. 5).

В табл. 6 приведены в соответствии с номенклатурой, предложенной В. Ингрэмом, [24, 159], мутации гемоглобина челове-

Таблица 6

Замещения аминокислот в мутациях гемоглобина человека

Обозначение мутации	Местоположение мутации	Замещение	Литература
Hb—I	$\alpha_2^{16\,асп}\ \beta_2^{A}$	лиз → асп	[244]
Hb—G (Гонолулу)	$\alpha_2^{30\,глю\text{-}NH_2}\ \beta_2^{A}$	глют → глю-NH_2	[354[
Hb—(Норфольк)	$\alpha_2^{57\,асп}\ \beta_2^{A}$	глиц → асп	[77]
Hb—M (Бостон)	$\alpha_2^{58\,тир}\ \beta_2^{A}$	гис → тир	[158]
Hb—G (Филадельфия) . .	$\alpha_2^{68\,лиз}\ \beta_2^{A}$	асп-NH_2 → лиз	[78]
Hb—S	$\alpha_2^{A}\ \beta_2^{6\,вал}$	глют → вал	[191, 192, 195]
Hb—C	$\alpha_2^{A}\ \beta_2^{6\,лиз}$	глют → лиз	[191, 192]
Hb—G (Сан-Хозе)	$\alpha_2^{A}\ \beta_2^{7\,глиц}$	глют → глиц	[185, 186]
Hb—E	$\alpha_2^{A}\ \beta_2^{26\,лиз}$	глют → лиз	[193]
Hb—(Цюрих)	$\alpha_2^{A}\ \beta_2^{63\,арг}$	гис → арг	[242]
Hb—M (Саскатун)	$\alpha_2^{A}\ \beta_2^{63\,тир}$	гис → тир	[158]
Hb—M (Милуоки)	$\alpha_2^{A}\ \beta_2^{67\,глют}$	вал → глют	[158]
Hb—D β (Пенджаб)	$\alpha_2^{A}\ \beta_2^{125\,глю\text{-}NH_2}$	глют → глю=NH_2	[24]

[1] Появившаяся позже и уже упоминавшаяся нами работа Х. Виттмана [342] разъясняет эти сомнения. Все только что написанное оказалось верным, за исключением того, что замещение илей → мет произошло под действием азотистой кислоты, а замещение асп → лиз вообще не наблюдалось.

ка. Приведены только те мутации, в которых изменена только одна единственная аминокислота. Такие аномальные гемоглобины, как Hb A_2 [197] или Hb-L_{holl}→[79], в которых изменено по нескольку аминокислот и которые трудно рассматривать как результат одного мутационного изменения, в таблицу не включены.

В табл. 7 приведены данные о замещениях в белке А триптофансинтетазы кишечной палочки. Наряду с мутациями, вы-

Таблица 7

Замещения в белке А триптофансинтетазы E. coli

Действующий агент	Замещения	Литература
Ультрафиолетовые лучи	глиц→арг (2), глиц→глют (1)	[179, 181]
Спонтанные	арг→глиц (2), арг→сер (1) глют→ала (1), глют→вал (1) глют→глиц (1)	[182]
Рекомбинации	арг→глиц (3), арг→сер (2) вал→глиц (1), вал→сер (2) глют→глиц (2)	[182]
(не указан)	тир—цис (?)	[203]

званными ультрафиолетовыми лучами, и спонтанными обратными мутациями приведены также результаты опытов по генетической рекомбинации между мутантными штаммами, так как в этих опытах даже увереннее, чем в любых других можно считать (в случае диплетного или триплетного кода), что изменение аминокислоты связано с заменой одного нуклеотида. Замещения, приведенные Т. Юксом [203] со ссылкой на личное сообщение, приведены отдельно, так как способ их получения неизвестен.

Приводимые Т. Юксом [203] замещения в лактоглобулинах [165, 204, 257] мы в группу мутаций не вносим, так как не считаем их достаточно хорошо изученными.

Для того чтобы сделать данные более обозримыми, в табл. 8 объединены все данные по замещениям аминокислот при мутациях. Малочисленные группы суммированы. В тех случаях, когда была неизвестна частота мутирования (данные, заимствованные у Т. Юкса) считали, что замещение наблюдалось однократно. В отношении аспарагиновой и глютаминовой кислот и их амидов в ряде случаев неизвестно, о чем идет речь.

Таблица 8

Общая сводка данных по замещениям аминокислот при мутациях

	Объект	Замещения
Вирус табачной мозаики	Нитритные мутации	асп → ала (4+2), асп → глиц (4), асп-NH_2 → сер (2+2), арг→глиц (2), арг→лиз (1), глю-NH_2→ →вал (2+1), глют→глиц (3), илей→вал (1), лей→фала (1), про→лей (3), про→сер (3), сер→лей (1), сер→фала (5), тир→фала (1), тре→илей (7), тре→мет (3), тре→сер (2)
	Прочие химические мутации	ала→глиц (2), арг→глиц (2), асп-NH_2→сер (0+4), илей→тре (1), лей→илей (1), про→лей (7), сер→фала (1)
	Спонтанные мутации	асп→ала (2), асп-NH_2→сер (0+1), асп→лиз (1), асп-NH_2→арг (1), асп-NH_2→лиз (1), арг→глиц (1), илей→мет (1), илей→тре (1), лей→фала (1), сер→фала (1)
Гемоглобин человека		асп-NH_2→лиз (1), вал→глют (1), гис→арг (1), гис→тир (2), глиц→асп (1), глют→вал (1), глют→глиц (1), глют→глю-NH_2 (2), глют→лиз (2), лиз→асп (1)
Триптофансинтетаза		арг→глиц (5), арг→сер (3), вал→глиц (1), вал→ →сер (2), глиц→арг (2), глиц→глют (1), глют→ →ала, глют→вал (1), глют→глиц (3), тир—цис (1)
Сумма		ала→асп (8), ала→глиц (2), ала→глют (1), арг→ →асп-NH_2 (1), арг→гис (1), арг→глиц (10), арг→лиз (1), арг→сер (2), асп→глиц (5), асп→лиз (2), асп-NH_2→лиз 2) асп-NH_2→сер (9), вал→глиц (1), вал→глют (3), вал→глю-NH_2 (3), вал→илей (1), вал→сер (1), гис→тир (2), глиц→глют (8), глют→глю-NH_2 (2), глю-NH_2→лиз (2), илей→лей (1), илей→мет (1), илей→тре (9), лей→про (10), лей→сер (1), лей →фала (2), мет→тре (3), про→сер (3), сер→тре (2), сер→фала (7), тир→фала (1), тир—цис (1)

Если хотя бы один случай известен точно, то неидентифицированные случаи включены в ту же группу. При этом даются две цифры. Например асп → ала (4 + 2) означает, что в четырех случаях речь идет о асп → ала, а в двух об асп*x* → ала.

В табл. 9, 10 и 11 приведены данные о замещениях в естественных белках. Как уже указывалось, учитывались только те замещения, по обе стороны от которых в обоих сравниваемых белках находятся одни и те же аминокислоты. Так как генетические связи между сравниваемыми белками или полипептидными цепями ни в одном случае не известны достоверно, то

указывалась во всех случаях минимальная встречаемость. Например, при сравнении гормонов передней доли гипофиза. [68, 69, 315, 316], имеющих следующие последовательности аминокислот: *вазотоцин:* Цис. тир. илей, глю-NH_2. асп-NH_2. цис. про. арг. глю-NH_2. *вазопрессин:* Цис. тир. фала. глю-NH_2. асп-NH_2. цис. про. арг. глю-NH_2. *окситоцин:* Цис. тир. илей. глю-NH_2. асп-NH_2. цис. про. лей. глю-NH_2; замещения илей — фала и арг — илей учтены каждое только по одному разу. В табл. 9

Таблица 9

Замещения аминокислот при сравнении цепей гемоглобина человека

Сравниваемые цепи	Замещения	Литература
α-цепь и β-цепь	ала—гис (1), ала—глиц (2), ала—лей (1), ала—про (1), арг—гис (2), арг—лиз (1), асп—глют (1), вал—глют (I), вал—лей (1), вал—тир (1), гис—глю-NH_2 (1), глиц—сер (1), лиз—тре (1), сер—тре (1), тре—трип (1)	[9, 30, 96, 97]
β- цепь и γ-цепь	ала—сер (1), ала—тре (2), арг—лиз (1), асп—асп-NH_2 (2), асп—глют (2), асп-NH_2—сер (1), вал—глиц (1), вал—илей (1), вал—мет (1), гис—илей (1), глют—про (2), глю-NH_2-тре (1), лей—фала (1), тре—цис (1)	[9, 97, 270]
α-цепь и γ-цепь (не повторяющиеся выше)	ала—глют (1), ала—илей (1), ала—тре (1), арг—лиз (1), сер—тре (1)	[9, 96, 260]
β-цепь и δ-цепь	ала—глют (1), арг—вал (1), асп—сер (1), сер—тре (2)	[9, 97, 292]

приведены замещения из сравнений α-, β-, γ-, и δ-цепей гемоглобина человека; они выделены потому, что представляют собой довольно большую группу, пригодную для самостоятельного анализа. В табл. 10 приведены замещения, обнаруживаемые

Таблица 10

Замещения аминокислот в сходных по составу полипептидных цепях

Сравниваемые белки	Замещения	Литература
Вазопрессин и окситоцин	арг—лей (1), илей—фала (1)	[68, 315, 316]
Гемоглобин и миоглобин	ала—гис (1), ала—фала (1), арг—лей (1), арг—сер (1), асп—глют (1), вал—глиц(1), гис—фала (1), глют—тре (1), глют—трип (1), лей—лиз (1), лиз—сер (1), лиз—трип (1)	[9, 96, 97, 138, 208]

Таблица 10 (*продолжение*)

Сравниваемые белки	Замещения	Литература
А- и В-цепи инсулина	ала—глиц (1), глиц—тре (1)	[264, 265]
Клупеины	ала—про (1), ала—сер (1)	[72, 145, 303, 321]
Кортикотропин и меланотропин	про—сер (1)	[160, 174, 224, 334]
А- и В-β лактоглобулины	ала—вал (1), асп—глиц (1)	[165, 204, 257]
α- и β-меланотропин	лиз—сер (2)	[160, 173, 174]
А- и В-фибринопептиды	ала—вал (1), глиц—лей (1)	[89, 90]

при сравнении других белков, близких по составу. Наконец, в табл. 11 приведены замещения, получаемые при сравнении аналогичных белков у разных видов.

Мы совершенно не используем довольно широко известные

Таблица 11

Замещения аминокислот в аналогичных белках разных видов

Белок	Замещения	Литератур
Вазопрессин (бык, свинья и др.)	арг—лиз (1)	[67, 68, 315, 316]
Вирусный белок (разные штаммы ВТМ)	ала—тре (1)	[16]
Гемоглобин (лошадь, человек)	ала—асп (1), ала—вал (1), ала—глиц (4), ала—про (1), ала—сер (1), асп-NH_2—глиц (1), асп-NH_2—лей (1), вал—лей (1), вал—сер (1), гис—глю-NH_2 (1), глиц—сер (2), лей—мет (1)	[9, 42, 50, 96, 97, 98]
Гипертензин (лошадь, бык)	вал—илей (1)	[141, 142]
Глюкопептиды (бык, кролик, человек)	глют—глю-NH_2 (1)	[252]
Инсулин (бык, свинья, лошадь, полосатик и др.)	ала—тре (1), глиц—сер (1), вал—илей (1), илей—тре (1)	[1, 105, 175, 264, 265]
Клупеин (*Clupea harengus C. pallasii*)	ала—арг (1), арг—сер (1)	[72, 145, 303, 321]
Кортикотропин (бык, овца и др.)	ала—асп (1), глют—глю-NH_2 (1)	[83, 223, 224, 334]
Меланотропин (бык, свинья)	глют—сер (1)	[160, 173, 174]
Протамины (*Salmo iridea*, *S. trutta*)	вал—илей (1)	[144]

Таблица 11 (*продолжение*)

Белок	Замещения	Литература
Соматотропин (бык, обезьяна, человек и др)	ала—сер (1), глиц—фала (1)	[225, 247]
Сывороточный альбумин (человек, бык и др.)	ала—тре (1)	[299, 335]
Фибринопептиды (человек, бык, свинья, кролик)	ала—вал (1), ала—глют (1), ала—тре (1), асп—лей (1), вал—глют (1), про—сер (1)	[89, 90]
Цитохром С (бык, шелкопряд и др.)	ала—сер (1), арг—лиз (1)	[139, 310, 311,312, 313 314]

Примечание. Указаны, как правило, только те виды, данных по которым достаточно для обнаружения приведенных замещений.

данные Ф. Шорма и его сотрудников [62, 206, 281, 282, 284, 285, 286, 287] о структурных корреляциях внутри белковых молекул. Поскольку замещений в настоящее время известно уже довольно много, то скорее нужно стремиться к отбору наиболее достоверных, чем к некритическому увеличению их числа любой ценой. Что же касается внутримолекулярных корреляций, то само существование их пока является только предположением. Выводы школы Ф. Шорма не подкрепляются статистическим анализом, а статистический анализ, проведенный другими авторами [337], делает очень вероятным, что таких корреляций в действительности не существует.

В табл. 12, так же, как это было сделано выше для мутаций, приведены в более компактной форме, чем в табл. 9, 10 и

Таблица 12

Общая сводка данных по замещениям аминокислот в естественных белках

Объекты	Замещения
Цепи гемоглобина человека	ала—гис (1), ала—глиц (2), ала—глют (2), ала—илей (1), ала—лей (1), ала—про (1), ала—сер (1), ала—тре (3), арг—вал (1), арг—гис (2), арг—лиз (3), асп—асп-NH_2 (2), асп—глют (3), асп—сер (1), асп-NH_2—сер (1), вал—глиц (1), вал—глют (1), вал—илей (1), вал—лей (1), вал—мет (1), вал—тир (1), гис—глю-NH_2 (1), гис—илей (1), глиц—сер (1), глют—про (2), глю-NH_2—тре (1), лей—фала (1), лиз—тре (1), сер—тре (4), тре—трип (1), тре—цис (1)

Таблица 12 (*продолжение*)

Объекты	Замещения
Сходные полипептидные цепи	ала—вал (2), ала—гис (1), ала—глиц (1), ала—про (1), ала—сер (1), ала—фала (1), арг—лей (2), арг—сер (1), асп—глиц (1), асп—глют (1), вал—глиц (1), гис—фала (1), глиц—лей (1), глиц—тре (1), глют—тре (1), глют—трип (1), илей—фала (1), лей—лиз (1), лиз—сер (3), лиз—трип (1), про—сер (1)
Аналогичные белки разных видов	ала—арг (1), ала—асп (2), ала—вал (2), ала—глиц (4) ала—глют (1), ала—про (1), ала—сер (4), ала—тре (4), арг—лиз (2), арг—сер (1), асп—лей (1), асп-NH_2—глиц (1), асп-NH_2—лей (1), вал—глют (1), вал—илей (3), вал—лей (1), вал—сер (1), гис—глю-NH_2 (1), глиц—сер (3), глиц—фала (1), глют—глю-NH_2 (2), глют—лиз (1), глют—сер (1), илей—тре (1), лей—мет (1), про—сер (1), сер—тре (1)
Сумма	ала—арг (1), ала—асп (2), ала—вал (4), ала—гис (2), ала—глиц (7), ала—глют (3), ала—илей (1), ала—лей (1), ала—про (3), ала—сер (6), ала—тре (7), ала—фала (1), арг—вал (1), арг—гис (2), арг—лей (2), арг—лиз (5), арг—сер (2), асп—асп-NH_2 (2), асп—глиц (1), асп—глют (4), асп—лей (1), асп—сер (1), асп-NH_2—глиц (1), асп-NH_2—лей (1), асп-NH_2—сер (1), вал—глиц (2), вал—глют (2), вал—илей (4), вал—лей (2), вал—мет (1), вал—сер (1), вал—тир (1), гис—глю-NH_2 (2), гис—илей (1), гис—фала (1) глиц—лей (1), глиц—сер (4), глиц—тре (1), глиц—фала (1), глют—глю-NH_2 (2), глют—лиз (1), глют—про (2), глют—сер (1), глют—тре (1), глют—трип (1), глю-NH_2—тре (1), илей—тре (1), илей—фала (1), лей—лиз (1), лей—мет (1), лей—фала (1), лиз—сер (3), лиз—тре (1), лиз—трип (1), про—сер (2), сер—тре (5), тре—трип(1), тре—цис (1)

11 все данные по замещениям, обнаруживаемым при сравнении естественных белков.

Прежде чем использовать приведенный материал для оценки числа разрешенных замещений, необходимо выяснить, насколько соответствует истине предположение о равновероятности всех замещений, и, в случае плохого соответствия, найти факторы, наиболее сильно влияющие на частоту встречаемости замещений.

Конечно, выводы статистического анализа должны быть тем более достоверными, чем на большем материале они основаны. Однако некритическое объединение всего вышеприведенного материала может привести и к неверным выводам. Если какая-либо группа опытов ставилась таким образом, что примененная

методика позволяла улавливать только небольшую часть всех разрешенных замещений, и данные этой группы составляют процентуально большую часть всего материала, то это приведет к преувеличению вероятности одних и к занижению вероятности других замещений. А в нашем материале это вполне возможно. В опытах по вызыванию мутаций азотистой кислотой из двенадцати возможных превращений нуклеотидов (каждый из четырех может быть, в принципе, заменен любым из трех остальных) осуществляется только два (А → Г и Ц → У); поэтому в этих опытах должна обнаруживаться только примерно одна шестая часть всех возможных замещений. Еще меньшая доля всех замещений может быть обнаружена в опытах с мутациями в белке А, где исследовались изменения только одного определенного аминокислотного «локуса».

Рассмотрение имеющихся данных показывает, что среди мутаций химические мутации ВТМ составляют количественно значительно большую группу, чем все остальные. Это, конечно, может при объединении данных исказить истинную частоту отдельных замещений. Для того чтобы избежать этого, мы решили из всех химических мутаций ВТМ взять только часть примерно такого объема, как и другие мутационные данные. Такими являются нитритные мутации, описанные в работе [305]. Позже (в главе 5) мы сможем показать, что такой отбор действительно нужен и достигает своей цели. В табл. 13 приведен материал, который будет в дальнейшем подвергаться анализу.

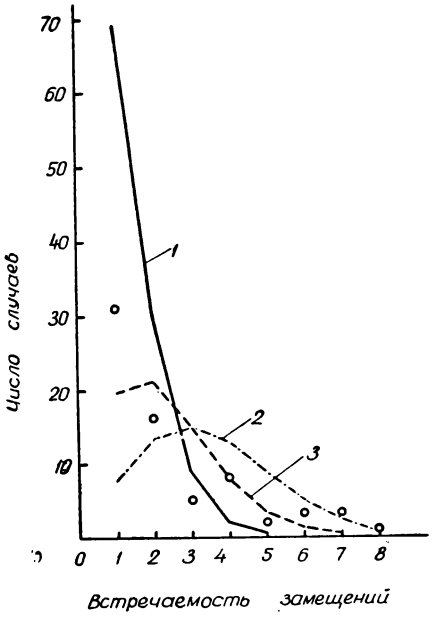

Рис. 8. Распределение замещений по их встречаемости.

Точки — наблюдаемые цифры. *1* — теоретическая кривая Пуассоновского распределения для предположения, что разрешены все замещения; *2* — то же, для предположения, что разрешены только наблюдавшиеся замещения; *3* — теоретическая кривая для вырожденного распределения Пуассона.

Первый вопрос, который должен нас интересовать — одинакова ли вероятность появления всех замещений. Для этого распределение замещений из табл. 13 по их частоте было сравнено с теоретическими распределениями. Так как замещения представляют собой редкие независимые события, то, в случае их равновероятности, их появление должно было бы следовать

Частота отдель

Аминокислота	ала	арг	асп	асп-NH_2	вал	гис	глиц	глют
Аланин	×	1	6	—	4	2	7	4
Аргинин	1	×	—	1	1	3	8	—
Аспарагиновая к-та	10	—	×	2	—	—	2	4
Аспарагин	—	1	2	×	—	—	1	—
Валин	4	1	—	—	×	—	3	5
Гистидин	2	3	—	—	—	×	—	—
Глицин	9	10	6	1	3	—	×	6
Глютаминовая к-та	4	—	4	—	5	—	8	×
Глютамин	—	—	—	—	3	2	—	4
Изолейцин	1	—	—	—	5	1	—	—
Лейцин	1	2	1	1	2	—	1	—
Лизин	—	6	2	2	—	—	—	1
Метионин	—	—	—	—	1	—	—	—
Пролин	3	—	—	—	—	—	—	2
Серин	6	4	1	10	2	—	4	1
Тирозин	—	—	—	—	1	2	—	—
Треонин	7	—	—	—	—	—	1	1
Триптофан	—	—	—	—	—	—	—	1
Фенилаланин	1	—	—	—	—	1	1	—
Цистеин	—	—	—	—	—	—	—	—

Примечание. В левой нижней части—весь материал; в правой верхней из

распределению Пуассона. Мы произвели два таких сравнения: первое при предположении, что все 190 пар аминокислот могут давать замещения и второе — считая, что разрешенными замещениями являются те и только те замещения, которые приведены в табл. 11. На рис. 8 (две первые кривые) наблюдаемое распределение сравнивается с теоретическими для тех же средних. Этот рисунок ясно показывает, что обе теоретические кривые плохо соответствуют наблюдаемому распределению. В том же убеждает и результат статистического анализа. Сравнение наблюденного и теоретического распределения с помощью критерия χ^2 дает для обеих гипотез значения P меньше 10^{-10}. Несоответствие наблюдаемого распределения теоретическим указывает на то, что замещения не являются равновероятными.

Наиболее очевидный фактор, который не может не влиять на частоту замещений и влияние которого легче всего оценить, это различная встречаемость разных аминокислот. Естественно, что чаще должны наблюдаться замещения с участием распространенных аминокислот, чем с участием более редких. Поэтому проверим, как зависит средняя частота замещений с участием данной аминокислоты от ее встречаемости в белках.

Таблица 13

ных замещений

глю-NH_2	илей	лей	лиз	мет	про	сер	тир	тре	трип	фала	цис
—	1	1	—	—	3	6	—	7	—	1	—
—	—	2	5	—	—	4	—	—	—	—	—
—	—	1	2	—	—	1	—	—	—	—	—
—	—	1	2	—	—	3	—	—	—	—	—
—	4	2	—	1	—	2	1	—	—	—	—
2	1	—	—	—	—	—	2	—	—	1	—
—	—	1	—	—	—	4	—	1	—	1	—
4	—	—	1	—	2	1	—	1	1	—	—
×	—	—	2	—	—	—	—	1	—	—	—
—	×	—	—	1	—	—	—	2	—	1	—
—	1	×	1	1	2	—	—	—	—	2	—
2	—	1	×	—	—	3	—	1	1	—	—
—	1	1	—	×	—	—	—	—	—	—	—
—	—	10	—	—	×	2	—	—	—	—	—
—	—	1	3	—	5	×	—	7	—	4	—
—	—	—	—	—	—	—	×	—	—	—	1
1	10	—	1	3	—	7	—	×	1	—	1
—	—	—	1	—	—	—	1	1	×	—	—
—	1	3	—	—	—	7	1	—	—	×	—
—	—	—	—	—	—	—	1	1	—	—	×

химических мутаций ВТМ включены только нитритные мутации из работы [305].

Данные по встречаемости аминокислот в белках мы взяли из работы И. Уильямса, И. Клегга и М. Матч [337], а данные о средней встречаемости замещений получили на основании материала, приведенного в табл. 13. Эти исходные данные сведены в табл. 14. На рис. 9 изображена графически зависимость средней частоты замещений от встречаемости аминокислот. Из этого рисунка видно, что действительно, средняя частота замещений растет с возрастанием встречаемости аминокислоты. Прямая, описывающая данную зависимость, вычисленная методом наименьших квадратов, выражается уравнением

$$y = 0{,}87 + 0{,}257\ x, \qquad (1)$$

где x — встречаемость аминокислоты, а y — средняя частота замещений для этой аминокислоты. Прямая, проведенная на рисунке, соответствует этому уравнению. Коэффициент регрессии $b = 0{,}257 + 0{,}052$ вполне достоверен ($t = 4{,}99$). Как видно из уравнения, ордината нулевой точки (0,87) близка к единице, как этого и следовало бы ожидать; разница ($0{,}13 \pm 0{,}28$) статистически недостоверна ($t = 0{,}46$). В последней колонке табл. 12 приведено для каждой аминокислоты ожидаемое число заме-

Таблица 14

Зависимость средней частоты замещений от встречаемости аминокислот в белках

Аминокислота	Встречаемость в белках, %	Общее число замещений	Число сортов замещений	Средняя частота	Ожидаемое число замещений
Серин	9,1	37	11	3,4	35,2
Аланин	8,6	43	12	3,6	37,2
Валин	8,4	23	9	2,6	27,0
Глицин	8,0	34	10	3,4	29,0
Лейцин	7,1	14	10	1,4	27,0
Аспарагиновая к-та	6,7	18	7	2,6	18,2
Глутаминовая к-та	5,8	29	10	2,9	24,0
Треонин	5,8	22	9	2,4	21,6
Лизин	5,7	18	9	2,0	20,7
Аргинин	4,8	25	8	3,1	16,8
Пролин	4,5	9	4	2,3	8,0
Фенилаланин	4,2	10	6	1,7	11,4
Цистеин	3,9	2	2	1,0	3,8
Тирозин	3,7	4	3	1,3	5,4
Аспарагин	3,6	10	6	1,7	10,8
Глутамин	3,5	9	4	2,3	7,2
Изолейцин	3,0	10	6	1,7	9,6
Гистидин	2,2	11	6	1,8	8,4
Метионин	0,8	3	3	1,0	3,3
Триптофан	0,8	3	3	1,0	3,3

щений, вычисленное на основании уравнения (1). Так как для разных аминокислот число разрешенных замещений может быть различным, то ожидаемые частоты получались путем умножения ожидаемой средней частоты на число сортов замещений, которые наблюдались для данной аминокислоты. Сравнение вычисленных частот с наблюдаемыми с помощью критерия χ^2 показало, что наблюдающиеся между ними различия нельзя считать статистически достоверными ($P = 0{,}95$).

Достоверный коэффициент регрессии для зависимости средней частоты замещений от встречаемости аминокислоты доказывает, что встречаемость аминокислот в белках существенно влияет на частоту замещений и поэтому с ней нужно считаться. Вместе с тем хорошее соответствие общего числа замещений для каждой аминокислоты с ожидаемым, вычисленным с учетом распространенности аминокислот, указывает на то, что влияние других факторов должно быть несущественным. Нужно заметить, что последний вывод не является вполне строгим. Поскольку расчет касался только средней частоты замещений для каждой аминокислоты, можно говорить только о малости влияния факторов, играющих разную роль для разных аминокислот.

Однако наиболее разумно предположение о существовании именно таких факторов (например, предположение о том, что замещение аминокислот, играющих важную роль в определении третичной структуры или биологической активности, чаще дол-

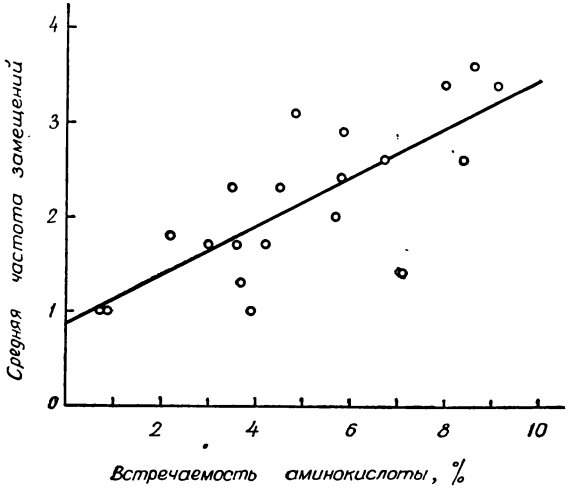

Рис. 9. Зависимость среднего числа замещений от встречаемости аминокислоты в белках:

Точки — результаты опытов, прямая соответствует уравнению (1).

жно вызывать фенотипический или летальный эффект, чем замещение «нейтральной» аминокислоты). Поэтому рассмотренным материалом можно, учитывая встречаемость аминокислот, пользоваться для нашей цели. Это тем более законно, что несколько позже мы сможем проверить влияние остальных факторов и более точно.

ГЛАВА 5

О СТЕПЕНИ ВЫРОЖДЕННОСТИ КОДА

Исходя из поставленной задачи, начнем анализ с оценки числа разрешенных замещений. В связи с тем, что для этого могут быть использованы только косвенные методы, эта оценка будет произведена пятью разными способами, причем подробно будет разобран только один из них, а остальные будут применены только приближенно для контроля результатов, полученных первым способом. В конце этой главы будет обсуждено значение полученного результата для определения степени вырожденности.

ЧИСЛО РАЗРЕШЕННЫХ ЗАМЕЩЕНИЙ

Первый метод. Принципиальная основа этого метода уже была пояснена в главе 3. Теперь сформулируем задачу количественно. Так как всего имеется 20 аминокислот, то максимально мыслимое число замещений равно числу возможных пар аминокислот (без учета их порядка), то есть

$$\frac{20 \cdot (20-1)}{2} = 190.$$

Число замещений разных сортов, наблюдавшихся до сих пор, меньше 190. Остальные замещения могли не наблюдаться по двум причинам — либо потому, что они вообще запрещены, либо потому, что хотя они и разрешены, но не были обнаружены случайно в связи с недостаточно большим числом опытов. Требуется определить, сколько из не наблюдавшихся замещений являются запрещенными (число их может оказаться и равным нулю, в случае максимального вырождения), а сколько разрешенными, но не наблюдавшимися. Будем вначале для простоты считать, что все наблюдавшиеся замещения разрешенные и что все разрешенные замещения обнаруживаются с одинаковой вероятностью. Так как здесь мы имеем дело с редкими независимыми событиями, то частота отдельных сортов замещений должна следовать распределению Пуассона: если средняя частота разрешен-

ных сортов замещений равна μ, то вероятность того, что среди разрешенных замещений некоторый сорт замещений произойдет ровно 0,1,2,.. n,... раз, равна

$$e^{-\mu}\left[1, \mu, \frac{\mu^2}{2}, \ldots, \frac{\mu^n}{n!}, \ldots\right].$$

Если из всех возможных пар аминокислот только некоторая доля p может давать разрешенные замещения, то вероятность того, что среди всех пар какая-либо даст ровно 1, 2,... n, ...замещений равна, очевидно,

$$p \cdot e^{-\mu} \cdot \left[\mu, \frac{\mu^2}{2}, \ldots, \frac{\mu^n}{n!}, \ldots\right],$$

а вероятность того, что некоторая пара не даст ни одного замещения, равна сумме вероятностей запрещенного замещения $q = 1 - p$ и вероятности того, что разрешенный сорт не произойдет ни разу:

$$q + (1-q) \cdot e^{-\mu} = e^{-\mu} + q(1 - e^{-\mu}).$$

Таким образом, распределение частоты замещений будет следовать не распределению Пуассона, а тому, что мы назвали в свое время вырожденным распределением Пуассона [39]. Интересно отметить, что это распределение было обнаружено и изучено в связи с совершенно иным явлением — восстановлением первичных радиационно-цитогенетических повреждений [37, 40, 41].

После этого можно предложить способ для оценки общего числа разрешенных замещений. Из сказанного ясно, что средняя частота **наблюдаемых** замещений R должна определяться уравнением

$$R = \frac{\mu}{1 - e^{-\mu}}, \tag{2}$$

так как форма вырожденного распределения отличается от обычного распределения Пуассона только относительной величиной нулевого класса, который при этом исключается из рассмотрения. Действительно, в вырожденном распределении Пуассона средняя частота разрешенных замещений равна $p \cdot \mu$, а число сортов замещений, происходящих не менее одного раза, $p \cdot (1 - e^{-\mu})$. Обозначив долю разрешенных замещений, происходящих при средней частоте μ через $P = 1 - e^{-\mu}$, находим

$$\mu = -\ln(1 - P),$$

после чего уравнение (2) превращается в

$$R = \frac{-\ln(1-P)}{P}. \tag{3}$$

Уравнения (2) и (3) не решаются (относительно μ и, соответственно, P) аналитически. Поэтому мы составили табл. 15 для решения уравнения (3).

Таблица 15

Зависимость средней встречаемости замещений от процента обнаруженных замещений

	0	1	2	3	4	5	6	7	8	9
0	—	1,002	1,007	1,013	1,018	1,023	1,029	1,034	1,039	1,045	1,051
10	1,051	1,056	1,062	1,068	1,074	1,080	1,087	1,093	1,099	1,106	1,112
20	1,112	1,119	1,126	1,133	1,140	1,147	1,154	1,162	1,169	1,177	1,185
30	1,185	1,193	1,201	1,209	1,218	1,226	1,235	1,244	1,253	1,263	1,272
40	1,272	1,282	1,292	1,302	1,312	1,323	1,334	1,345	1,356	1,368	1,380
50	1,380	1,392	1,405	1,418	1,431	1,445	1,459	1,473	1,488	1,503	1,519
60	1,519	1,535	1,552	1,569	1,587	1,606	1,625	1,645	1,665	1,686	1,709
70	1,709	1,732	1,756	1,781	1,807	1 24	1,863	1,893	1,925	1,958	1,993
80	1,993	2,031	2,070	2,113	2,158	2,206	2,259	2,315	2,377	2,444	2,518
90	2,518	2,601	2,694	2,800	2,923	3,069	3,247	3,474	3,783	4,264	—

Проиллюстрируем применение этого метода на всей совокупности данных, представленных в табл. 8 и 12 (включая все нитритные замещения). Так как всего наблюдалось 222 замещения 74 разных сортов, то $R = {}^{222}/_{74} = 3{,}0 \pm 0{,}20$. По табл. 15 находим, что отношение 3,0 заключено между табличными значениями 2,923 и 3,069, следовательно, соответствует 94%. Можно определить и доверительные пределы для этого процента, исходя из стандартной ошибки найденной частоты. Так как доверительные пределы, соответствующие вероятности 0,05, равны 2,05 и 3,95 ($3{,}0 \pm 0{,}20 \times 1{,}96$), то доверительные пределы для процента равны 81 и 98%. Для того же, чтобы определить число разрешенных сортов замещений, делим наблюдавшееся число сортов (74) на 0,94. Таким образом, число разрешенных замещений N равно 79 (76—91). Если повторить тот же расчет для материала с меньшим числом нитритных мутаций (ср. табл. 13), то получается очень близкий результат: $N = 78$ замещений.

Если бы действительная ситуация вполне отвечала сделанным вначале упрощающим предположениям, то задача была бы решена. Однако это не так. Поэтому ниже мы постараемся повторить этот расчет, принимая во внимание влияние всех возможных факторов.

Но сначала вернемся немного назад. В конце предыдущей главы мы сравнивали наблюдающееся распределение частоты замещений с теоретическим, получив очень плохое соответствие. Повторим то же для вырожденного распределения Пуассона, считая, что число разрешенных замещений равно 79. Соответствующая кривая также приведена на рис. 8 и дает уже значительно лучшее соответствие. Правда, расхождение ее с эмпири-

ческим распределением все еще остается статистически достоверным ($P = 0{,}005$), чему, конечно, не приходится удивляться в связи с предварительным характером произведенной оценки. Там же говорилось предположительно о том, что в мутационных опытах должна улавливаться только небольшая часть всех разрешенных замещений. Теперь это можно проверить. Расчет, который был проведен для всех замещений, мы повторили для отдельных их групп. Результаты, приведенные в табл. 16, пока-

Таблица 16

Оценка числа разрешенных замещений на основании разных данных

Исходные данные	Общее число замещений	Число сортов замещений	R	P	N
Данные по мутациям					
Вирус табачной мозаики, спонтанные	11	10	1,10	18	56
Вирус табачной мозаики, нитритные	50	17	2,94	94	18
Вирус табачной мозаики, другие химические	18	7	2,57	90	8
Триптофансинтетаза кишечной палочки	16	9	2,00	80	10
Аномальные гемоглобины человека	13	9	1,44	54	17
Данные по естественным белкам					
Цепи гемоглобина человека	45	31	1,45	55	56
Другие сходные полипептиды	25	21	1,19	30	70
Аналогичные белки разных видов	44	27	1,63	66	41

зывают, что вычисленное число разрешенных замещений для отдельных групп мутаций оказывается меньше, чем для замещений в естественных белках. Поэтому сделанное уравнивание числа нитритных мутаций с остальными действительно имело основания.

Поскольку уже известно, что встречаемость аминокислот в белках влияет на частоту замещений с их участием, причем этот фактор имеет, по-видимому, наибольшее значение, начнем с того, что учтем его влияние. Для этого вычислим тем же методом число разрешенных замещений отдельно для каждой аминокислоты. Так как каждое замещение будет в этих расчетах фигурировать дважды, то полусумма чисел, полученных для отдельных аминокислот, даст общее число разрешенных замещений. Необходимые для этого расчеты приведены в табл. 17.

Для трех аминокислот, давших все замещения только по одному разу, формальный расчет приводит к абсурдному результату (к бесконечности). Поэтому для соответствующих аминокислот (триптофан, метионин и цистеин) было предположено,

Таблица 17

Число разрешенных замещений для отдельных аминокислот

Аминокислота	Общее число замещений	Число сортов замещений	R	P	N
Аланин	43	12	3,58	97	12
Серин	37	11	3,36	96	11
Глицин	34	10	3,40	96	10
Глютаминовая к-та	29	10	2,90	93	11
Аргинин	25	8	3,13	95	8
Валин	23	9	2,56	90	10
Треонин	22	9	2,44	88	10
Аспарагиновая к-та	18	7	2,57	90	8
Лизин	18	9	2,00	80	11
Лейцин	14	10	1,40	51	20
Гистидин	11	6	1,83	75	8
Аспарагин	10	6	1,67	68	9
Изолейцин	10	6	1,67	68	9
Фенилаланин	10	6	1,67	68	9
Пролин	9	4	2,25	85	5
Глютамин	9	4	2,25	85	5
Тирозин	4	3	1,33	45	7
Триптофан	3	3	1,00	—	(5)
Метионин	3	3	1,00	—	(5)
Цистеин	2	2	1,00	—	(5)
Сумма	—	—	—	—	178

что они дают по пять замещений, т. е. столько же, сколько и другие наименее распространенные аминокислоты. После этого сумма разрешенных замещений для всех аминокислот оказывается равной 178. Отсюда общее число разрешенных замещений равно $178 : 2 = 89$.

На правильность только что полученной цифры могут влиять два фактора. Во-первых, вычисление производилось в молчаливом предположении, что все наблюдавшиеся замещения являются разрешенными, что само по себе требует специального анализа. Во-вторых, предполагалось, что вероятности обнаружения всех замещений с участием данной аминокислоты одинаковы. На самом деле это не так. Если вероятность обнаружить замещение какой-либо определенной аминокислоты на любую другую и можно полагать постоянной, то этого, строго говоря, нельзя делать для вероятности замещения *с участием* данной аминокислоты. А в нашем материале замещения в противоположных направлениях (например глют → глиц и глиц → глют) по неизбежности были объединены. Пусть вероятности всех замещений без учета встречаемости аминокислот одинаковы и равны p. Тогда математическое ожидание для числа замещений А → Б с учетом

их встречаемости должно быть равно pN_a, где N_a — встречаемость аминокислоты А. Точно так же математическое ожидание для числа замещений Б →А равно $pN_б$. Математическое ожидание для суммы обоих чисел, естественно, будет равно сумме математических ожиданий, т. е. $p \cdot (N_a + N_б)$. Поэтому частоты отдельных замещений должны быть пропорциональны сумме встречаемостей участвующих в них аминокислот. Хотя для каждой аминокислоты эти суммы не должны варьировать очень сильно, целесообразно постараться учесть это обстоятельство.

Это можно сделать следующим образом. Сгруппируем замещения в соответствии с суммой встречаемостей участвующих в них пар аминокислот. Так как в табл. 17 аминокислоты расположены в порядке уменьшения общего числа замещений, которое коррелирует со встречаемостью аминокислот, то мы разбили их на две категории: от аланина до лейцина и от гистидина до цистеина, и определили число замещений для трех групп: к первой группе были отнесены такие замещения, в которых участвуют обе аминокислоты из числа распространенных, ко второй — одна распространенная, другая — редкая, к третьей — обе редкие. Результаты расчетов, приведенные в табл. 18, показывают, что

Таблица 18

Число разрешенных замещений для разных групп аминокислотных пар

Пары аминокислот	Общее число замещений	Число сортов замещений	R	P	N
Распространенные с распространенными	105	33	3,18	95	34,7
Распространенные с редкими	53	29	1,83	75	38,7
Редкие с редкими	9	7	1,29	41	17,1
Сумма	—	—	—	—	90,5

сумма трех чисел, полученных для отдельных групп, дает такой же итог, который был получен при вычислениях по отдельным аминокислотам. Отсюда следует вывод, что число разрешенных замещений с учетом влияния встречаемости аминокислот, но без учета влияния других факторов должно быть около 90.

Что касается влияния других факторов, то их можно разбить на две группы. Во-первых, не все наблюдавшиеся замещения могут быть разрешенными, т. е. являться результатом замены одного нуклеотида; во-вторых, под влиянием тех или иных причин (летальный эффект, отсутствие фенотипического эффекта при вызывании мутаций и т. д.) некоторые разрешенные заме-

Сравнение частот отдельных

Аминокислоты	ала	сер	глиц	глют	арг	вал	тре	асп
Аланин	×	6,9/6	7,0/7	6,2/4	6,7/1	5,9/4	5,8/7	5,8/6
Серин	0,12	×	6,8/4	6,0/1	6,5/4	5,7/2	5,6/7	5,6/1
Глицин	0	1,16	×	6,1/6	6,5/8	5,7/3	5,6/1	5,7/2
Глютаминовая к-та	0,78	4,17	0	×	5,8/0	4,9/5	4,8/1	4,9/4
Аргинин	4,85	0,96	0,35	5,80	×	5,4/1	5,3/0	5,4/0
Валин	0,4	2,40	1,26	0	3,60	×	4,5/0	4,6/0
Треонин	0,25	0,35	3,79	3,02	5,32	4,49	×	4,5/0
Аспарагиновая к-та	0,01	3,80	2,39	0,17	5,40	4,60	4,50	×
Лизин	5,22	0,80	5,00	2,54	0,01	3,90	2,06	0,93
Лейцин	2,54	4,13	2,34	3,30	0,85	0,33	2,90	1,33
Гистидин	1,80	4,70	4,80	4,00	0,50	3,70	3,60	3,60
Аспарагин	4,70	0,50	2,74	3,90	2,46	3,40	3,30	0,58
Изолейцин	1,57	4,50	4,50	3,90	4,20	0,58	0,51	3,40
Фенилаланин	1,57	0,06	2,74	3,90	4,20	3,40	3,30	3,40
Пролин	1,07	1,98	5,20	1,10	4,90	4,10	4,00	4,10
лотамин	5,40	5,20	5,20	0,04	4,90	4,10	2,25	4,10
Тирозин	4,30	4,10	4,20	3,40	3,90	1,43	3,00	3,00
Триптофан	4,10	3,90	3,90	1,42	3,60	2,80	1,07	2,80
Метионин	4,10	3,90	3,90	3,10	3,60	1,16	2,70	2,80
Цистеин	4,10	3,90	3,90	3,10	3,60	2,80	1,07	2,80

Примечание. В правой верхней части таблицы в виде дроби приведе в знаменателе — наблюдавшаяся. В левой нижней части — соответствующие каж

замещений с теоретически ожидаемыми

Таблица 19

лиз	лей	гис	асп-NH_2	илей	фала	про	глю-NH_2	тир	трип	мет	цис
5,2/0	4,3/1	5,0/2	4,7/0	4,7/1	4,7/1	5,4/3	5,4/0	4,3/0	4,1/0	4,1/0	4,1/0
5,0/3	4,1/0	4,7/0	4,5/3	4,5/0	4,5/4	5,2/2	5,2/0	4,1/0	3,9/0	3,9/0	3,9/0
5,0/0	4,1/1	4,8/0	4,5/1	4,5/0	4,5/1	5,2/0	5,2/0	4,2/0	3,9/0	3,9/0	3,9/0
4,3/1	3,3/0	4,0/0	3,9/0	3,9/0	3,9/0	4,4/2	4,4/4	3,4/0	3,1/1	3,1/0	3,1/0
4,8/5	3,8/2	4,5/3	4,2/1	4,2/0	4,9/0	4,9/0	3,9/0	3,9/0	3,6/0	3,6/0	3,6/0
3,9/0	3,0/2	3,7/0	3,4/0	3,4/4	3,4/0	4,1/0	4,1/0	3,1/1	2,8/0	2,8/1	2,8/0
3,8/1	2,9/0	3,6/0	3,3/0	3,3/2	3,3/0	4,0/0	4,0/1	3,0/0	2,7/1	2,7/0	2,7/1
3,9/2	3,0/1	3,6/0	3,4/2	3,4/0	3,4/0	4,1/0	4,1/0	3,0/0	2,8/0	2,8/0	2,8/0
×	2,3/1	3,0/0	2,8/2	2,8/0	2,8/0	3,4/0	3,4/2	2,4/0	2,1/1	2,1/0	2,1/0
0,73	×	2,1/0	1,8/1	1,8/0	1,8/2	2,5/2	2,5/0	1,5/0	1,2/0	1,2/1	1,2/0
3,00	2,10	×	2,5/0	2,5/1	2,5/1	3,2/0	3,2/2	2,1/2	1,9/0	1,9/0	1,9/0
0,23	0,36	2,50	×	2,2/0	2,2/0	2,9/0	2,9/0	1,9/0	1,6/0	1,6/0	1,6/0
2,80	1,80	0,90	2,20	×	2,2/1	2,9/0	2,9/0	1,9/0	1,6/0	1,6/1	1,6/0
2,80	0,02	0,90	2,20	0,65	×	2,9/0	2,9/0	1,9/0	1,6/0	1,6/0	1,6/0
3,40	0,10	3,20	2,90	2,90	2,90	×	3,6/0	2,6/0	2,3/0	2,3/0	2,3/0
0,58	2,50	0,45	2,90	2,90	2,90	3,60	×	2,6/0	2,3/0	2,3/0	2,3/0
2,40	1,50	0	1,90	1,90	1,90	2,60	2,60	×	1,3/0	1,3/0	1,3/1
0,58	1,20	1,90	1,60	1,60	1,60	2,30	2,30	1,30	×	1,0/0	1,0/0
2,10	0,03	1,90	1,60	0,23	1,60	2,30	2,30	1,30	1,00	×	1,0/0
2,10	1,20	1,90	1,60	1,60	1,60	2,30	2,30	0,07	1,00	1,00	×

ны частоты: в числителе — ожидаемая частота (способ вычисления см. в тексте), дому сравнению индексы (нормированные квадраты отклонений).

щения могут обнаруживаться с пониженной вероятностью. В обоих случаях это будет вести к появлению замещений с пониженной частотой. Но в случае, если эти редкие замещения в действительности являются запрещенными, то число наблюдаемых разрешенных замещений окажется завышенным, что приведет и к завышению оценки их общего числа. Если же редкие замещения являются разрешенными, но встречаются с пониженной частотой, то это приведет к занижению средней частоты замещений, а следовательно, и оценки общего числа разрешенных замещений. Поэтому нашей ближайшей задачей должно явиться определение числа таких замещений и выяснение их природы. Для этого придется проанализировать частоту отдельных сортов замещений.

Сопоставим наблюдаемую частоту замещений с ожидаемой при предположении, что вероятность возникновения и обнаружения всех разрешенных замещений одинакова. Исходя из числа разрешенных замещений для каждой аминокислоты, приведенных в табл. 17, и общего числа замещений, нетрудно вычислить «теоретическую» среднюю частоту замещения с участием данной аминокислоты. Ожидаемое число замещений данного сорта должно быть равно, аналогично показанному выше, сумме этих средних для обеих аминокислот, участвующих в данном замещении. В табл. 19 приведены эти ожидаемые частоты для всех 190 пар аминокислот вместе с наблюдаемыми частотами. Поскольку нас интересует не общее соответствие наблюденных частот ожидаемым, а степень соответствия по отдельным парам аминокислот, не представляется возможным применить какой-либо из стандартных статистических методов. Для возможности количественного сравнения мы вычислили следующий индекс:

$$\frac{(\text{наблюдаемая частота} - \text{ожидаемая частота})^2}{\text{ожидаемая частота}}.$$

Значения этого индекса для всех пар аминокислот также приведены в таблице. (Хотя арифметически вычисление этого индекса соответствует вычислению χ^2, но эта величина в действительности χ^2 не является, так как число определяемых индексов превышает число степеней свободы для всей таблицы).

Более внимательное рассмотрение таблицы показывает, что в большом числе случаев наблюдаемые частоты очень мало отличаются от вычисленных, тогда как для других замещений отклонения довольно значительны, в то время как промежуточных отклонений сравнительно мало. Более объективно можно об этом судить, построив распределение индексов. В связи с тем, что для редких аминокислот ожидаемые частоты, а следовательно, и отклонения от них, оказываются слишком малыми, мы рассмотрели только данные до замещений с участием гистидина вклю-

чительно. Это дает 154 пары аминокислот из 190. Полученное распределение индексов изображено на рис. 10. Оно совершенно определенно является двувершинным, т. е. совокупность всех возможных пар аминокислот распадается на два дискретных класса: к одному принадлежат пары, дающие замещения с ча-

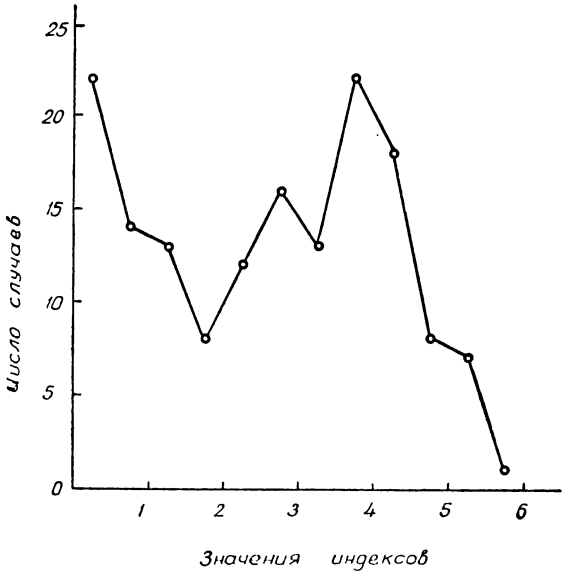

Рис. 10. Распределение индексов (нормированных квадратов отклонений) из табл. 19 по частоте.

стотой, близкой к ожидаемой, а к другому — те, для которых замещения наблюдаются значительно более редко.

Эти результаты позволяют сделать довольно определенные высказывания о характере факторов, обусловливающих появление замещений с низкой частотой. Если бы те или иные факторы, перечислявшиеся выше, существенно влияли на вероятность обнаружения некоторых сортов разрешенных замещений, то, поскольку рассматриваемый материал касается довольно разнообразных белков, эта вероятность должна была бы довольно плавно меняться от максимального значения до нуля. Если же замещения, наблюдающиеся редко, являются запрещенными, а разрешенные замещения обнаруживаются с более или менее равной вероятностью, то пары аминокислот должны распадаться на две группы. А поскольку распределение нормированных квадратов отклонений дало резко двухвершинную кривую, то второе предположение представляется более вероятным.

В пользу того, что редко встречающиеся замещения являются «запрещенными» можно привести и другой довод. В табл. 20

Таблица 20

Сопоставление редких замещений с предсказаниями некоторых кодов

Замещения	Индекс	Авторы сравниваемых кодов		
		И Спейер и др. [290]	Н. В. Лучник [38]	Т. Юкс [203]
ала—илей	1,57	—	—	—
ала—фала	1,57	—	—	—
лиз—тре	2,06	+	—	—
глют NH_2—тре	2,25	+	+	+
глиц—лей	2,34	+	—	—
арг—асп NH_2	2,46	+	—	—
ала—лей	2,54	+	—	+
глют—лиз	2,54	+	+	+
асп NH_2—глиц	2,76	—	—	—
глиц—фала	2,76	—	—	—
глют—тре	3,02	+	—	—
арг—вал	3,60	+	+	—
глиц—тре	3,79	—	—	—
асп—сер	3,80	—	—	—
глют—сер	4,17	—	—	—
ала—арг	4,85	—	—	—

Примечание. Приведены только замещения, в которых хотя бы одна из аминокислот не относится к числу редких. Знак (—) означает, что данное замещение запрещается соответствующим кодом, (+) — что оно разрешено.

выписаны замещения, встречавшиеся однократно, кроме тех случаев, когда обе аминокислоты, участвующие в замещении, относятся к редким. Там же показано, является ли данное замещение разрешенным или запрещенным с точки зрения кода С. Очоа [51, 290] и двух кодов с определенной последовательностью нуклеотидов [38, 203]. Из таблицы видно, что даже для кода с неопределенным порядком нуклеотидов в триплетах, при котором вероятность случайного совпадения довольно велика [180], половина этих замещений оказывается запрещенными. Что же касается кодов с определенным порядком нуклеотидов, то для них запрещенными оказываются почти все эти замещения.

Поэтому мы должны сделать вывод, что замещения, встречающиеся с пониженной против ожидания частотой, относятся большей частью к запрещенным. Следовательно, полученная выше оценка — 90 замещений — должна быть скорее завышенной, чем заниженной.

Поправку на наблюдаемые запрещенные замещения можно внести следующим образом. Если принять, что все замещения, встречающиеся реже, чем следовало ожидать, запрещенные (что, конечно, является преувеличением), то можно найти такое число однократно встреченных замещений, при отбрасывании которых распределение наблюдаемых замещений по частоте будет мень-

ше всего отклоняться от теоретического распределения Пуассона. Проведенный расчет показал, что χ^2 принимает минимальное значение при числе разрешенных замещений около 53. Поскольку далеко не все замещения, наблюдающиеся с пониженной частотой, являются запрещенными, эту цифру следует считать заниженной нижней границей. А так как верхняя граница была равна 91, то за наиболее вероятное значение для общего числа разрешенных замещений разумно принять среднее из двух, т. е. 72 замещения.

Следует заметить, что некоторые другие авторы обращали внимание на то, что одни замещения не наблюдались ни разу, в то время как другие происходили довольно часто, но давали этому неправильное толкование. Так Ф. Шорм [283] считает, что это говорит в пользу развиваемых им и его школой представлений о закономерностях строения полипептидных цепей, а Р. Экк [135, 136] полагает, что неслучайность замещений объясняется тем, что код перекрывающийся. Выше мы уже имели возможность говорить о том, что никаких закономерностей в чередовании аминокислот в белках нет и что код неперекрывающийся, приводя соответствующие доводы. Что же касается объема и устройства словаря, то он не может не влиять на число и частоту замещений. Интересно отметить, что Р. Экк также определяет на основании статистики замещений число пар аминокислот, способных давать замещения, пользуясь несколько иным методом. Полученная им цифра близка к нашей: $N = 67$ (52—87).

Таким образом, стоявшую перед нами задачу можно было бы считать выполненной. Однако, поскольку для ее решения был применен косвенный метод, могут быть резонные основания сомневаться в правильности полученного результата. Поэтому ниже мы попытаемся оценить число разрешенных замещений другими, тоже косвенными, методами, но более или менее независимыми от первого. Все они основаны на количественном анализе наблюдавшихся замещений, так как, на наш взгляд, в настоящее время только эти данные и могут служить для ответа на стоящий перед нами вопрос. Как уже говорилось, ряд других данных также дает возможность судить о степени вырожденности кода, но все эти данные могут служить основанием только для качественных высказываний, но не для количественных оценок. Кроме метода, подробно описанного выше, мы применили для оценки числа разрешенных замещений еще четыре других метода, к краткому описанию которых и переходим.

Второй метод. Представим себе некоторую совокупность, состоящую из объектов разных сортов, причем число сортов неизвестно. Из этой совокупности делаются две случайных выборки. Если сорта объектов в обеих выборках полностью совпали, то очевидно, что большого числа сортов, не попавших в эти выборки, не может существовать. Если же в двух выборках

вовсе нет совпадающих объектов, то очевидно, что совокупность имеет еще большое число объектов, не попавших в эти выборки. Эти соображения можно использовать для оценки числа разрешенных замещений.

Пусть из некоторой совокупности, содержащей N разных объектов, сделаны (с возвращением) две выборки объемом n_1 и n_2. Так как при однократном испытании вероятность обнаружить один определенный объект из N равна $1/N$, то вероятность обнаружить его, например, в выборке объема n_1 равна n_1/N. Вероятность обнаружить тот же объект в обеих независимых выборках по правилу умножения вероятностей равна

$$\frac{n_1 \cdot n_2}{N^2}.$$

А так как общее число объектов равно N, то математическое ожидание для числа объектов, обнаруживаемых в обеих выборках,

$$\text{МО} = \frac{N \cdot n_1 \cdot n_2}{N^2} = \frac{n_1 \cdot n_2}{N}.$$

Приравняв число объектов, обнаруженных в обеих выборках k, математическому ожиданию для него (М. О.), можно определить объем совокупности N из формулы

$$N = \frac{n_1 \cdot n_2}{k}.$$

В табл. 21 сопоставлены замещения (без учета их частоты) в двух количественно и качественно сравнимых группах: в цепях гемоглобина человека и в сходных белках разных видов. Из этой таблицы видно, что в гемоглобине наблюдалось всего 31 заме-

Таблица 21

Замещения, наблюдавшиеся при сравнении цепей гемоглобина человека и сходных белков у разных видов

Встречаемость замещений	Замещения
Только в гемоглобине человека	ала—гис, ала—илей, ала—лей, арг—вал, арг—гис, асп—асп-NH_2, асп—глют, асп—сер, асп-NH_2—сер, вал—глиц, вал—мет, вал—тир, гис—илей, глют—про, глют-NH_2—тре, лей—фала, лиз—тре, тре—трип, тре—цис
Только в белках разных видов	ала—арг, ала—асп, ала—вал, арг—сер, асп—лей, асп-NH_2—глиц, асп-NH_2—лей, вал—сер, глиц—фала, глют—глю-NH_2, глют—лиз, глют—сер, илей—тре, лей—мет, про—сер
Встречаются в обеих группах	ала—глиц, ала—глют, ала—про, ала—сер, ала—тре, арг—лиз, вал—глют, вал—илей, вал—лей, гис—глю-NH_2, глиц—сер, сер—тре

шение, а в сходных белках 27, причем 12 замещений наблюдались в обеих группах. Отсюда, используя только что приведенную формулу, можно оценить число разрешенных замещений:

$$N = \frac{31 \cdot 27}{12} = 69{,}7.$$

Эта цифра очень близка к полученной с помощью первого метода. Второй метод должен быть очень чувствителен к степени универсальности кода. Действительно, если у разных организмов одни и те же аминокислоты кодируются разными словами, то (в случае, если между этими словами нет структурной связи) разные группы данных должны давать разные замещения; число общих замещений будет резко уменьшаться, и второй метод даст сильно завышенную цифру числа замещений. Хорошее совпадение между результатами первого и второго метода дает новое доказательство в пользу универсальности кода.

К интересным выводам приводит такое же сопоставление замещений для двух групп, обладающих некоторым качественным различием. Рассмотрим данные по замещениям в мутациях и в естественных белках (в каждом случае взят весь материал, соответственно, из табл. 8 и 12). В мутационных опытах сравнение аминокислотного состава производится только в случае обнаружения фенотипического эффекта, т. е. тогда, когда функция белка заметным образом изменена. При сравнении же естественных белков в большом проценте случаев сравниваются белки из разных объектов, функционирующие одинаково. Поэтому можно ожидать некоторого различия в «спектре» замещений обеих групп, а следовательно, и завышенного результата при оценке общего числа разрешенных замещений вторым методом.

Таблица 22

Число сортов замещений, наблюдавшихся при мутациях и при сравнении естественных белков

Рассматриваемый материал	Число сортов замещений
Мутации	33
Сходные белки	58
Встречаются в обеих группах	17

Необходимые цифры приведены в табл. 22. Число «разрешенных замещений» оказывается равным

$$N = \frac{33 \cdot 58}{17} = 112{,}7,$$

что значительно выше результата, полученного вначале. Это подтверждает приведенные соображения о несколько разном характере замещений, улавливаемых в мутационных опытах и обнаруживаемых при сравнении естественных белков.

Третий метод. Для оценки общего числа разрешенных замещений можно применить также анализ индексов, приведенных в табл. 19. Правда, этот метод может дать только нижнюю границу для числа разрешенных замещений. Будем считать, что разрешенными являются только те замещения, частота которых дает небольшое отклонение от ожидаемой. Мы можем взять некоторую часть замещений, составляющую известную часть всего распределения разрешенных замещений, и по ее величине сделать заключение об общем числе замещений, составляющих это распределение. Затруднением является то, что теоретическая форма распределения индексов неизвестна. Удовлетворительный выход из положения может быть найден в том, что мы будем рассматривать замещения, для которых нормированный квадрат отклонения не больше единицы. Поскольку рассматриваемая величина близка к χ^2, то есть основания предполагать, что и распределена она будет сходно с χ^2. В пределах $\chi^2=1$ лежит 68,3% всех его значений. Но то же число значений лежит в пределах $\pm 1\sigma$ (и $\pm 1\sigma^2$) в нормальном распределении, являющемся предельным случаем ряда других. Поэтому при принятии указанной границы этот метод должен быть мало чувствительным к истинной форме распределения.

Как видно из рис. 10, 36 индексов имеют значение меньше единицы, но в распределение включены только 154 индекса из 190. Поэтому правильнее принять

$$N = \frac{36 \cdot 190}{154} = 44{,}4.$$

Как было показано, это число должно составлять 68,3% всей совокупности разрешенных замещений. Поэтому их общее число должно быть равно

$$44{,}4 \cdot 0{,}683 = 65,$$

что также близко к результатам, полученным с помощью двух первых методов.

Четвертый метод. Разобьем, как мы это делали и раньше (ср. табл. 18), аминокислоты на две равные группы, в соответствии с их встречаемостью. Тогда замещения распадутся на три группы: на замещения между двумя распространенными аминокислотами, между двумя редкими и на такие, в которых участвуют одна распространенная и одна редкая кислота. Если средняя частота разрешенных замещений для каждой из этих групп определена правильно, то, как это уже было показано при разборе первого метода, частота для третьей группы должна быть равна полусумме частот для двух первых. Поясним это на примере. Предположим, что разрешены все 190 мыслимых замещений. В первую и вторую группы попадают по 45 пар аминокислот, т. к. $(10 \cdot 10 - 10) : 2 = 45$, а в третью — 100. Наблюдавшееся число замещений, соответственно, было (см. табл. 18) —

105, 9 и 53. Отсюда средняя частота замещений для первой группы 105 : 45 = 2,33, для второй 0,20, для третьей 0,53. Если бы эти частоты соответствовали действительности, то последняя цифра (0,53) была бы близка к полусумме двух первых. Однако эта полусумма (2,33 + 0,20) : 2 = 1,27 оказывается в 2,4 раза больше, что говорит за то, что в действительности число замещений значительно меньше, чем это было принято при расчетах. Подобный расчет был повторен для нескольких гипотез о числе разрешенных замещений; результаты его представлены в табл. 23. Как

Таблица 23

Оценка средней частоты разных групп замещений при разных предположениях об их числе

Гипотеза	Число разрешенных замещений	Средняя частота замещений				
		1-й группы	2-й группы	3-й группы		
				вычисленная	наблюдаемая	вычисл. в % наблюд.
Разрешены замещения между любыми аминокислотами	190	2,33	0,20	1,27	0,53	240
Разрешены наблюдаемые и только наблюдаемые замещения . . .	69	3,18	1,28	2,23	1,83	122
Оценка первым методом на основании наблюдаемых замещений . .	91	3,00	0,52	1,76	1,36	129
Оценка первым методом при учете в первой и третьей группах только замещений, наблюдавшихся более одного раза	55	3,93	0,52	2,23	2,29	97

Примечание. К первой группе отнесены замещения десяти более распространенных аминокислот друг с другом, ко второй — десяти более редких друг с другом, к третьей — замещения, в которых участвует одна распространенная и одна редкая кислота.

видно из этой таблицы, 69 замещений дают результат выше, а 55 — ниже, чем наблюдаемые величины. Отсюда можно сделать заключение, что число разрешенных замещений должно быть около 60.

Пятый метод. Число разрешенных замещений можно оценить еще, если предположить, что все аминокислоты дают одинаковое число разрешенных замещений и что большая часть возможных замещений уже наблюдалась.

На рис. 11 показано распределение аминокислот по числу сортов даваемых ими замещений. При построении этого распределения однократно встреченные замещения с участием девяти наиболее распространенных аминокислот были отброшены, так как они с большой вероятностью не являются результатом замены одного нуклеотида. Распределение дает явный максимум на шести замещениях, при сильной асимметрии: вначале оно

медленно нарастает и, достигнув максимума около шести, дает затем резкий спад до нуля, за которым следует небольшой второй максимум. Такую картину можно интерпретировать таким образом, что большинство аминокислот дает по 6 замещений, из которых для ряда аминокислот наблюдались не все. Второй мак-

Рис. 11. Распределение аминокислот по числу даваемых ими замещений.

симум говорит о том, что некоторые (немногие) аминокислоты могут давать больше шести замещений (можно думать, что здесь идет речь об аминокислотах, по которым код вырожден). Если каждая аминокислота имеет по шесть разрешенных замещений, то общее их число должно быть равно $6 \cdot 20 : 2 = 60$. А так как некоторые аминокислоты дают большее число замещений, то эту цифру следует несколько увеличить.

О СТЕПЕНИ ВЫРОЖДЕННОСТИ КОДА

В табл. 24 сведены результаты, полученные с помощью всех пяти примененных методов. Хорошее согласие между этими результатами дает основание верить в их правильность и позволяет оценить число разрешенных замещений цифрой 70 с возможной погрешностью примерно в десяток замещений в ту и другую сторону.

Все эти расчеты были проведены с главной целью: через число разрешенных замещений (которое мы, наконец, выяснили) попытаться оценить число слов, т. е. выяснить степень вырожденности кода. Однако, хотя число замещений связано с числом слов, однозначного соответствия между ними нет, если не связывать себя заранее довольно жесткими предположениями об общих свойствах кода. Действительно, в случае словаря, состоящего из 16 слов, если каждое слово состоит из двух букв, то возможно 48 замещений, если из трех букв, то 63, а если из четырех букв, то 96 замещений. В случае вырожденного кода, число

Таблица 24

Число разрешенных замещений, полученное разными методами

Метод	Число разрешенных замещений
Предварительный расчет	79 (76—91)
I метод	70 (53—90)
II метод	70
III метод	больше 65
IV метод	60
V метод	больше 60
Наиболее вероятное значение	70 (60—80)

разрешенных замещений будет зависеть от того, какие слова будут соответствовать одной и той же аминокислоте. В самом деле, если несколько слов, кодирующих одну и ту же аминокислоту, различаются лишь одной буквой, то эти слова в результате замещений будут переходить друг в друга, не давая замещений аминокислот.

Кроме того, можно себе представить такой вырожденный код, в котором разным превращениям кодирующих групп будут соответствовать одни и те же замещения аминокислот. Это может быть, например, в следующем случае. Представим себе диплетный код; он разрешает 48 замещений. Если к каждому диплету добавить в одно и то же положение урацил, то получится триплетный код, разрешающий те же 48 замещений. Кроме того, вместо урацила мы сможем поставить любой другой нуклеотид без изменения смысла диплета, так что, скажем, УАА, ЦАА, ГАА и ААА кодируют одну и ту же аминокислоту. В таком случае получается крайне вырожденный код, разрешающий очень большое число изменений кодирующих слов, но разрешающий только 48 замещений аминокислот.

Поэтому определение числа разрешенных замещений не дает возможности однозначно определить степень вырожденности. Однако, зная это число, можно сразу отбросить некоторые гипотезы, которые иначе могли бы казаться приемлемыми, а также предложить несколько вариантов кода, для выбора между которыми потребуются дополнительные данные.

Из только что проведенного рассмотрения должно быть ясно, что нельзя ставить знака равенства между возможным числом изменений слов и числом возможных замещений аминокислот. Число замещений может быть равно числу возможных изменений слов (в некоторых вариантах однозначного кода) или быть меньше его. Это уменьшение может происходить за счет возникновения нонсенсов и за счет того, что разные изменения слов могут соответствовать одним и тем же замещениям аминокислот, а также за счет того, что некоторые изменения слов могут и не приводить к замещениям.

Для простоты будем предполагать, что слова в словаре данного типа подобраны таким образом, что их изменение дает минимальное число нонсенсов. Именно таким образом составлена табл. 25, в которой приведено число возможных изменений слов

Таблица 25

Число разрешенных изменений слов для разных словарей

Тип словаря	Число разрешенных изменений слов
Без ограничений в использовании букв	
Слова из:	
двух букв, словарь из 16 слов	48
трех букв, словарь из 20 слов	90
четырех букв, словарь из 20 слов	120
трех букв, словарь из 26 слов	117
Слова из трех букв, содержащие хотя бы одно У	
Словарь из:	
20 слов	58
24 слов	72
28 слов	90
31 слова	99
34 слов	111
37 слов	126

для обсуждавшихся до сих пор вариантов кода. В связи с тем, что в таблице нет очень сильно вырожденных кодов, число замещений аминокислот не должно сильно отличаться от числа изменений слов. Сопоставим числа этой таблицы с тем числом, которое мы получили для числа разрешенных замещений.

При диплетном коде возможно 48 замещений, что приближается к нижнему пределу для нашей оценки числа замещений. Однако диплетный код дает недостаточное число слов. Квадруплетный код, разрешающий даже без вырождения 120 замещений, также должен быть сразу отброшен, как дающий резкое несоответствие с полученной цифрой.

Рассмотрим более подробно триплетные коды. Как видно из таблицы, невырожденный триплетный код дает 90 замещений, что лежит уже около верхней границы действительного числа замещений. Однако ряд данных говорит о том, что код имеет какую-то степень вырожденности. Если принять 26 триплетов — число уже установленных триплетов в опытах с урацилсодержащими СПРН — то они дадут 117 замещений, что явно противоречит нашей оценке числа разрешенных замещений. Поэтому, хотя триплетный код представляется наиболее вероятным, но и он возможен только в случае определенных ограничений, накладываемых на выбор кодирующих триплетов.

Одно время казалось, что ограничением в выборе слов может являться требование того, чтобы в состав каждого из них вхо-

дил урацил. Рассмотрение табл. 25 показывает, что 72 замещения должен давать урацилсодержащий код при словаре, состоящем из 24 слов, что близко к действительному числу урацилсодержащих триплетов. Однако последние работы [156, 201, 277, 320] показывают, что существует значительное число слов, не содержащих У. Поэтому ограничения должны иметь какую-то другую природу. Так как ограничений в выборе букв не существует, то очевидно, что ограничения должны касаться выбора слов, кодирующих одни и те же аминокислоты.

Что касается объема словаря, то сколько-нибудь точно о нем сейчас ничего сказать нельзя. Однако следующие соображения говорят о том, что вырожденность должна быть довольно значительной. После статистического анализа, результаты которого были приведены в табл. 19, можно со значительно большей уверенностью, чем раньше, говорить о том, какие из замещений являются разрешенными, а какие запрещенными. В табл. 26 при-

Таблица 26

Сравнение разрешенных и запрещенных на основании статистического анализа замещений с предсказаниями урацилсодержащих кодов

Разрешены				Запрещенные			
Замещение	Индекс	Предсказание кода		Замещение	Индекс	Предсказание кода	
		Н. Лучника [38]	Т. Юкса [203]			Н. Лучника [38]	Т. Юкса [203]
ала—глиц	0	+	+	вал—тре	4,49	—	—
вал—глют	0	—	+	асп—тре	4,50	—	—
глиц—глют	0	+	+	глиц—илей	4,50	—	—
ала—асп	0,01	—	—	илей—сер	4,50	+	+
арг—лиз	0,01	—	—	асп—вал	4,60	+	—
глют—глю-NH_2	0,04	—	—	ала—асп-NH_2	4,70	+	+
сер—фала	0,06	+	+	гис—сер	4,70	—	+
ала—сер	0,12	+	—	гис—глиц	4,80	—	—
асп—глют	0,17	—	—	ала-арг	4,85	—	—
ала—тре	0,25	—	—	арг—глю-NH_2	4,90	—	—
арг—глиц	0,35	+	+	арг—про	4,90	+	+
сер—тре	0,35	+	+	глиц—лиз	5,00	—	—
асп-NH_2—сер	0,50	+	—	глиц—глю-NH_2	5,20	—	—
арг—гис	0,50	+	+	глиц—про	5,20	—	—
ала—вал	0,61	—	+	глю-NH_2—сер	5,20	—	+
ала—глют	0,78	+	+	ала—лиз	5,22	—	—
лиз—сер	0,80	—	—	арг—тре	5,32	—	—
асп—лиз	0,93	+	+	арг—асп	5,40	+	+
арг—сер	0,96	—	+	ала—глю-NH_2	5,40	—	—
ала—про	1,07	+	+	арг—глют	5,80	—	—

Примечание. Знак «плюс» означает, что замещение разрешено данным кодом, знак «минус»—что запрещено.

ведены 20 замещений с наименьшими и 20 с наибольшими индексами; про первые можно достаточно уверенно говорить, что они являются разрешенными, а про вторые, что они запрещены. В той же таблице показано, являются ли эти замещения запрещенными или разрешенными для двух вариантов урацилсодержащего кода [38, 203], а в табл. 27 наблюдающиеся несоответст-

Таблица 27

Не случайность несоответствия замещений кодам

Аминокислота	Число несоответствий коду	
	Н. Лучника [38]	Т. Юкса [203]
Аланин	4	4
Серин	2	6
Аргинин	4	3
Аспарагиновая к-та	4	2
Лизин	3	2
Глютаминовая к-та	3	2
Аспарагин	1	3
Валин	3	0
Глютамин	1	2
Треонин	1	1
Изолейцин	1	1
Пролин	1	1
Гистидин	0	1

вия сгруппированы по аминокислотам. Обращает на себя внимание то, что несоответствия носят не случайный характер, а касаются одних и тех же аминокислот. Однако совершенно ясно, что если каждой из этих аминокислот дать по дополнительному триплету, то это устранит только небольшую часть несоответствий, так как новый триплет, не содержащий урацила, может давать замещение только с одним из урацилсодержащих. Поэтому, хотя урацилсодержащий код оставляет необъясненным сравнительно небольшое число разрешенных замещений, для их объяснения требуется довольно большое число дополнительных триплетов.

Таким образом, результаты, полученные в этой главе, делают вероятным вывод о том, что *код является триплетным и сильно вырожденным, причем триплеты, кодирующие одну и ту же аминокислоту, должны обладать сходным строением и составом.*

ГЛАВА 6

ОБ ОБЩИХ СВОЙСТВАХ КОДА

В предыдущей главе было выяснено, что общее число разрешенных замещений около 70. Хотя эта оценка может быть и не совсем точной, число замещений, разрешаемых разными кодами, оказывается настолько различным, что некоторые из них, вроде простых диплетного и квадруплетного, или невырожденного урацилсодержащего триплетного, можно было сразу исключить, как явно противоречащие полученному результату. После этого становится возможным анализировать общие свойства кода, так как появился способ проверки правдоподобности этих свойств без расшифровки состава кодирующих слов.

В этой главе мы рассмотрим два новых метода исследования некоторых общих свойств кода, что дает основание, используя данные о числе разрешенных замещений, сформулировать гипотезу об общей природе аминокислотного кода.

Основные свойства кода, которые необходимо выяснить до того, как определять состав слов, следующие: кодовое число, (т. е. число букв в слове), ограничения в выборе букв для составления кодирующих слов (например, требование того, чтобы каждое слово содержало хотя бы одно У), число слов в словаре (другими словами — вырожденность кода) и ограничения в выборе слов для кодирования одной и той же аминокислоты (например, возможное требование того, чтобы эти слова отличались не более, чем одной буквой).

ЧИСЛО ПЕРЕХОДОВ И ВЕЛИЧИНА СЛОВ

Если замещение между некоторой парой аминокислот разрешено, то кодирующие их слова по самому определению разрешенного замещения должны отличаться лишь одной буквой. Если же замещение для некоторой пары аминокислот запрещено, то, соответственно, кодирующие их слова различаются больше, чем одной буквой. Однако, если имеется третья аминокислота, дающая замещения с каждой из двух первых, т. е. если

эти две кислоты связаны через два перехода, то значит, что они отличаются на две буквы. Это ясно из следующего. Пусть первой аминокислоте соответствует слово ААББББ..., второй — БББББ..., а третьей — АББББ... . Ясно, что две первых аминокислоты не могут дать замещения непосредственно, но они могут дать его в результате двух переходов, через слово, соответствующее третьей аминокислоте: ААБББ... → АББББ... → БББББ... Точно так же, если пара аминокислот связана через три перехода, то это значит, что кодирующие их слова отличаются на три буквы.

Как было показано в предыдущей главе, в настоящее время известно большинство разрешенных замещений, а запрещенные замещения наблюдаются сравнительно редко. Поэтому можно использовать имеющиеся данные для определения числа букв в словах.

В табл. 28, составленной на основании данных, приведенных

Таблица 28

Число переходов для разных пар аминокислот

	Ала	Сер	Глиц	Глют	Арг	Вал	Тре	Асп	Лиз	Лей	Гис	Асп-NH_2	Илей	Фала	Про	Глют-NH_2	Тир	Трип	Мет	Цис
Аланин	×	1	1	1	2	1	1	1	2	2	3	2	2	2	1	2	4	3	2	5
Серин	1	×	2	2	1	2	1	2	1	2	2	1	2	1	1	2	3	2	2	4
Глицин	1	1	×	1	1	2	2	1	2	2	2	2	3	3	2	2	3	3	3	4
Глютаминовая к-та	1	1	1	×	2	1	2	1	2	2	2	2	2	3	2	1	3	3	2	4
Аргинин	1	1	1	2	×	2	2	2	1	1	1	2	2	2	2	2	2	2	2	3
Валин	1	1	1	1	1	×	2	2	2	1	2	2	1	2	2	2	3	3	1	4
Треонин	1	1	1	1	2	2	×	2	2	2	2	2	1	2	2	3	3	3	1	4
Аспарагиновая к-та	1	1	1	1	2	2	2	×	1	2	3	1	3	3	2	2	4	2	3	5
Лизин	2	1	2	1	1	2	1	1	×	1	2	1	3	2	2	1	3	1	2	4
Лейцин	1	1	1	2	1	1	2	1	1	×	2	1	2	1	1	2	3	2	1	4
Гистидин	1	2	2	2	1	2	2	2	2	2	×	3	1	1	3	1	1	3	3	2
Аспарагин	1	1	1	2	1	2	2	1	1	1	2	×	3	2	2	2	4	2	2	5
Изолейцин	1	2	2	2	2	1	1	2	2	1	1	2	×	1	3	2	2	4	1	3
Фенилаланин	1	1	1	2	2	2	2	2	2	1	1	2	1	×	2	2	2	3	2	3
Пролин	1	1	2	1	2	2	2	2	2	1	2	2	2	2	×	3	4	3	2	5
Глютамин	2	2	2	1	2	1	1	2	1	2	1	2	2	2	2	×	2	2	3	3
Тирозин	2	2	2	2	2	1	3	3	2	2	1	3	2	1	3	2	×	4	4	1
Триптофан	2	2	2	1	2	2	2	2	1	2	2	2	3	2	2	2	1	×	3	5
Метионин	2	2	2	2	2	1	1	2	2	1	2	2	1	2	2	2	2	3	×	5
Цистеин	2	2	2	2	3	2	1	3	3	3	2	3	2	2	3	2	1	2	2	×

Примечание. В левой нижней части — для всего материала, в правой верхней — с учетом только заведомо разрешенных замещений.

в табл. 8 и 12, указано для каждой пары аминокислот наименьшее число связывающих их переходов, причем в нижней левой части данные получены на основании всего материала, а в пра-

вой верхней учтены только заведомо разрешенные замещения (см. табл. 19). Поясним на примере способ составления таблицы. Аргинин дает замещения с серином, но не дает их с треонином и цистеином; следовательно, с серином он связан одним, а с треонином и цистеином более чем одним переходом. Для пары аргинин — треонин можно найти аминокислоту (таковой оказывается в данном случае серин)), которая дает замещения с ними обеими. А так как возможно превращение арг→сер → тре, то аргинин и треонин связаны двумя переходами. Для пары арг→цис такой связующей аминокислоты нет. Здесь возможен только тройной переход: арг→сер→тре →цис. Из таблицы видно, что число тройных переходов очень мало. Из 190 возможных пар аминокислот, при рассмотрении всех замещений, 75 связаны одним переходом, 103 — двумя и только 12 — тремя.

Бросается в глаза то обстоятельство, что тремя переходами связаны лишь такие пары, в которых, по крайней мере, одна из аминокислот относится к числу редких, давших наименьшее число замещений. Если ограничить рассмотрение шестнадцатью более распространенными аминокислотами, исключив из таблицы тирозин, триптофан, цистеин и метионин, то из 120 оставшихся пар 63 оказываются связаны одним, 57 — двумя переходами, а пар, связанных тремя или более переходами, не остается ни одной. Такое выборочное рассмотрение, вероятно, допустимо, так как естественно полагать, что аминокислоты, по которым обнаружено очень мало замещений, должны, при прочих равных условиях, дать и пониженное число двойных переходов. Впрочем, результат здесь получается близким к тому, который получился вначале.

В связи с тем, что во всем материале некоторые из замещений являются, по-видимому, запрещенными, мы провели еще один расчет, при котором принимали во внимание только те замещения, в отношении которых трудно сомневаться в том, что они действительно относятся к разрешенным, т. е. такие, которые встречались более одного раза и притом с частотой, близкой к ожидаемой, исходя из встречаемости соответствующих аминокислот (ср. табл. 19). При таком жестком отборе остается довольно мало замещений, поэтому можно рассматривать только данные, касающиеся наиболее распространенных аминокислот.

Если рассмотреть данные, касающиеся десяти наиболее распространенных аминокислот, то из 45 пар 19 окажутся связаны одним и 26 — двумя переходами, при отсутствии больших чисел переходов. В табл. 29 приведены результаты этих подсчетов. Во всех трех случаях они не сильно отличаются друг от друга. Основной результат, полученный при этом, — отсутствие или, в крайнем случае, очень низкое число тройных переходов—кажется парадоксальным, так как противоречит наиболее распространенному взгляду о том, что код является триплетным. И, конеч-

Таблица 29

Распределение пар аминокислот по числу переходов

Число переходов	Все известные замещения для всех аминокислот		Все замещения без участия четырех редких аминокислот		Заведомо разрешенные замещения для десяти наиболее распространенных аминокислот.	
	число	%	число	%	число	%
Один	75	39,4	63	52,5	19	42,2
Два	103	54,2	57	47,5	26	57,8
Три	12	6,3	0	—	0	—
Сумма . . .	190	100,0	120	100,0	45	100,0

но, если бы можно было ограничить рассмотрение невырожденными кодами, то такой результат вполне однозначно говорил бы о том, что код является диплетным. Но низкое число тройных переходов может быть объяснено и большой вырожденностью кода. Поэтому для того, чтобы можно было истолковать полученные результаты более определенно, мы вычислили число одиночных, двойных и тройных переходов, ожидаемое в случае разных предположений о характере аминокислотного кода.

Во всех случаях вычисление производилось, исходя из предположения, что кодовые слова для соответствующих аминокислот выбираются случайным образом, т. е. для кода с заданными свойствами определяли вероятность того, что случайно выбранная пара слов будет различаться на одну, две, и т. д. буквы, а в случае вырожденных кодов считали, что вырожденность по всем аминокислотам примерно одинакова и что слова для кодирования одной и той же аминокислоты выбираются случайным образом. Результаты расчетов для шести разных гипотез об общих свойствах кода приведены в табл. 30. Были рассмотрены: диплетный код, невырожденные триплетные коды — урацилсодержащий и код без ограничений для состава слов, а также вырожденные триплетные коды — максимально вырожденный урацилсодержащий (словарь из 37 слов) и два кода без ограничений в составе слов, с объемом словаря в 40 и 60 слов.

Так как в предыдущей главе было выяснено, что число разрешенных замещений должно лежать в пределах 60—80 (что соответствует 32—42% пар аминокислот), то вырожденный урацилсодержащий код и наиболее вырожденный триплетный код без ограничений в составе триплетов сразу придется отбросить, так как они дают слишком много пар аминокислот, связанных через один переход. Оба варианта невырожденных триплетных кодов должны быть отброшены, так как дают слишком много

Таблица 30

Вычисленное распределение (в процентах суммы) чисел переходов для разных вариантов аминокислотного кода

Число переходов	Тип аминокислотного кода					
	диплетный (16 слов)	триплетный не вырожденный (20 слов)		триплетный вырожденный		
		урацилсодержащий	без ограничения выбора	урацилсодержащий (37 слов)	без ограничения выбора	
					40 слов	60 слов
Один	40,0	30,5	14,3	54,5	36,0	75,0
Два	60,0	50,5	42,8	43,7	60,6	24,4
Три	0	19,0	42,8	1,8	3,4	0,6

тройных переходов. В результате остается только два варианта, дающих удовлетворительное соответствие полученным цифрам — диплетный код и триплетный код без ограничения выбора букв с двукратной вырожденностью. Но против обоих этих кодов говорят другие соображения. В отношении диплетного кода они ясны — диплетный код дает словарь только из 16 слов, что недостаточно для кодирования 20 аминокислот.

Что же касается триплетного кода со случайным выбором кодирующих слов и с двукратной вырожденностью, то он противоречит результатам, полученным в опытах с урацилсодержащими СПРН. В этих опытах [231, 290] были найдены кодирующие слова для всех аминокислот, кроме глютамина, а для последнего, на основании нитритного замещения глю-NH_2 → вал, также было предсказано существование урацилсодержащего триплета [290]. Кажется странным, что если словарь не ограничен словами, содержащими У, то у каждой аминокислоты, по крайней мере, одно из двух кодирующих ее слов содержит У. Проверим это более точно. Так как из 64 возможных триплетов содержат урацил только 37, то, при случайном выборе триплетов, вероятность урацилсодержащего триплета равна 37/64=0,58; а триплета, не содержащего урацил: 1—0,58=0,42. Если каждой аминокислоте соответствует два кодирующих слова, то вероятность того, что хотя бы одно из них будет содержать У, равна $1-0{,}42^2=0{,}82$. Вероятность же того, что среди урацилсодержащих триплетов будут найдены слова, соответствующие всем двадцати аминокислотам, равна $0{,}82^{20}=0{,}019$, что является малой величиной. Эта вероятность должна увеличиться при еще большей вырожденности, но большая вырожденность разрешает слишком много замещений.

Таким образом, ни один из рассмотренных кодов не соответствует экспериментальным данным. Полное или почти полное

отсутствие тройных переходов говорит в пользу диплетного кода, а обилие урацилсодержащих триплетов — в пользу очень высокой (более, чем двухкратной) вырожденности, так как урацилсодержащие коды исключаются другими соображениями, а в последнее время и прямыми опытами [156, 201, 277, 320]. Противоречивость этих двух выводов очевидна. Сейчас мы не будем искать возможности для их примирения, а вначале применим для той же цели — выяснения общих свойств кода — второй метод.

ОБОБЩЕНИЕ РЕШЕТКИ ЗЬЮБИ – КВАСТЛЕРА

Г. Зьюби и Г. Квастлер [352] предложили недавно остроумный метод анализа замещений в связи с проблемой кодирования, основанный на том, что все возможные триплеты можно расположить в виде трехмерной прямоугольной решетки таким образом, что все разрешенные замещения будут соответствовать только перемещениям по ребрам решетки, и обратно, любое перемещение по ребру соответствует разрешенному замещению. Обобщение этого метода может позволить сделать некоторые выводы об общих свойствах кода. Не будем связывать себя никакими предположениями об объеме алфавита и числе букв в словах.

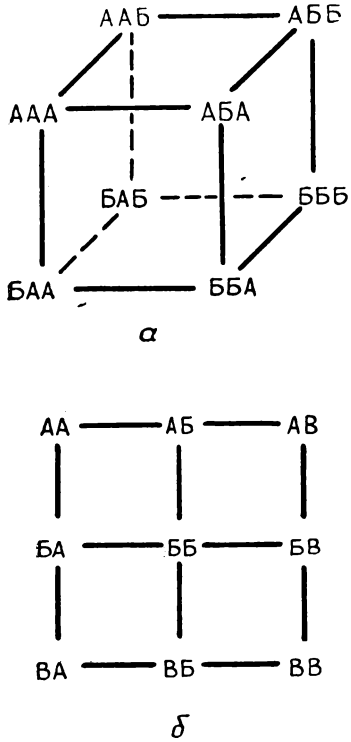

Рис. 12. Обобщенные решетки Зьюби—Квастлера:

а — для алфавита из двух букв и триплетного кода; *б* — для алфавита из трех букв и диплетного кода.

Расположим слова в n-мерной прямоугольной решетке таким образом, чтобы в каждой строке находились слова, в которых совпадают все буквы, кроме одной, и чтобы несовпадающие буквы стояли во всех словах строки на одном и том же месте. Очевидно, что число измерений этой решетки должно быть равно числу букв в словах, а число слов в каждой строке равно числу букв в алфавите. Для иллюстрации этого принципа на рис. 12 изображены в качестве примера две таких решетки: первая для триплетного кода при алфавите, состоящем всего из двух букв, а вторая для диплетного при трехбуквенном алфавите. Рассмотрение рисунка особенно ясно показывает, что аминокислоты, замещения между которыми разрешены, должны располагаться на одном ребре решетки. Поэто-

му можно попытаться расположить аминокислоты таким образом, чтобы те из них, которые дают разрешенные замещения, находились на одних и тех же ребрах, и по форме получившейся решетки судить об объеме алфавита и числе букв в слове. Если бы все известные замещения были заведомо разрешенными, а код не вырожденным, то задача решалась бы просто и однозначно. Но, к сожалению, ни одно из этих условий не выполняется. В проводимом ниже анализе мы будем считать те замещения, которые наблюдались более одного раза, заведомо разрешенными, а встречавшиеся однократно — сомнительными.

Нетрудно убедиться в том, что если три аминокислоты дают «треугольник», т. е. связаны всеми возможными переходами (например, для аланина, аспарагиновой кислоты и глицина: ала — асп, асп — глиц, глиц — ала), то значит слова, кодирующие эти аминокислоты, отличаются только по одной букве и должны находиться в одной строке решетки. Если учитывать только замещения, наблюдавшиеся неоднократно, то из материала, приведенного в табл. 1, можно составить 16 таких «треугольников»; все они приведены в табл. 31. Если в двух треугольниках совпа-

Таблица 31

Тройки аминокислот, дающие все взаимные замещения

ала,	асп,	глиц	арг,	глиц,	сер
ала,	асп,	глют	арг,	лиз,	сер
ала,	вал,	глиц	асп,	асп-NH_2,	лиз
ала,	вал,	глют	асп,	глиц,	глют
ала,	глиц,	глют	асп-NH_2,	лиз,	сер
ала,	глиц,	сер	вал,	глиц,	глют
ала,	про,	сер	вал,	глиц,	сер
ала,	тре,	сер	вал,	глют,	глю-NH_2

Примечание. При составлении таблицы учитывались только замещения, которые наблюдались более одного раза.

дают по две аминокислоты, (например, треугольники ала—асп—глиц и ала — асп — глют), то значит, что все 4 аминокислоты, входящие в эти треугольники, стоят в одной строке. Проведение такой процедуры приводит к странному результату. Оказывается, что все аминокислоты могут быть расположены в две строки, из которых одна состоит из шести, а другая из девяти аминокислот:

асп — асп-NH_2 — арг — глиц — лиз — сер,

ала — асп — вал — глиц — глю — глю-NH_2 — про — сер — тре.

Если не допускать возможности вырождения, то из такого результата следовало бы сделать вывод о том, что код является диплетным, а алфавит состоит из девяти букв. Однако, хорошо известно, что в состав РНК входит только 4 нуклеотида. Поэтому такой результат может быть объяснен только довольно значительной вырожденностью кода.

Возможен и другой способ анализа решетки. Так как нуклеиновокислотный алфавит состоит из четырех букв, можно расположить аминокислоты в соответствии с даваемыми ими замещениями таким образом, чтобы на каждом из ребер решетки находилось не более четырех аминокислот. Число измерений получившейся при этом решетки даст кодовое число. Нам нет надобности производить эту работу, так как она уже проделана другими авторами. Еще Г. Зьюби и Г. Квастлер [352] обратили внимание на то, что при проведении такой процедуры подавляющее большинство аминокислот располагается в одной плоскости. Плоскости соответствует диплетный код. Поэтому для заполнения решетки можно воспользоваться уже предлагавшимися диплетными кодами [261, 262]. Будем вначале полностью следовать диплетному коду. Так как он неоднозначен, то четыре слова окажутся изокодами, т. е. каждое из них будет кодировать по две аминокислоты. Из всех замещений мы исключили те, что наблюдались только один раз в естественных белках, так как они с большой вероятностью являются запрещенными и, расположив аминокислоты в плоской решетке 4×4, подсчитали процент замещений, не противоречащих коду, — для замещений разных сортов и для всех замещений.

То же самое было проделано для исключенных редких замещений, а также отдельно для нитритных замещений, считая для них разрешенными только те, которые происходят в результате транзиций Ц → У и А → Г. Полученные результаты приведены в табл. 32. Там же приведены результаты аналогичных расчетов для всех триплетных (т. е. соответствующих трехмерной решетке) кодов с определенным порядком букв в словах (ср. табл. 4). Рассмотрение таблицы приводит к неожиданному выводу — диплетные коды дали большее число совпадений, чем любой из триплетных кодов, что снова как будто говорит в пользу того, что код диплетный. С другой стороны, большое число замещений остается необъясненным, несмотря на то, что по крайней мере большинство наблюдавшихся запрещенных замещений было исключено из рассмотрения. Особенно велики расхождения для нитритных замещений — около половины их противоречат кодам. Отсюда ясно, что должны существовать слова сверх тех, которые принимались во внимание при анализе, причем для объяснения всех замещений требуется довольно много добавочных слов, следовательно, код должен быть довольно сильно вырожденным.

Таким образом, анализ замещений с помощью обобщенной решетки Зьюби — Квастлера, с одной стороны, совершенно ясно указывает на довольно значительную вырожденность кода, а с другой — делает правдоподобным предположение о том, что код является диплетным, что совпадает с выводами, полученными на основании анализа чисел переходов.

Таблица 32

Процент замещений, объясняемых разными кодами

Автор	Тип кода	Объем словаря	Нитритные замещения в белке ВТМ		Все замещения, кроме однократно встреченных в естественных белках		Замещения, однократно встреченные в естественных белках
			разные	всего	разные	всего	
И. Рыхлик и Ф. Шорм [262]	диплетный	16	55	61	74	79	18
Р. Робертс [261] . . .	»	16	50	55	76	75	21
Н. В. Лучник [38] .	триплетный	22	56	61	59	67	18
Т. Юкс [203]	»	25	56	61	63	66	21
Г. Зьюби и Г. Квастлер [352]	»	18	44	34	63	66	43
Р. Безе [3[illegible]]	»	20	50	55	59	64	18
Е. Смит [280] . . .	»	20	56	61	57	62	14
В. Г. Туманян и сотрудники [54] . . .	»	24	56	59	54	59	18

Примечание. Все рассмотренные коды приведены в табл. 4.

ОБЩИЕ СВОЙСТВА КОДА

Итак, анализ некоторых общих свойств кода, проведенный двумя методами, привел к одинаковому и довольно парадоксальному результату. С одной стороны, код обладает свойствами диплетного, а с другой — он должен быть сильно вырожденным. Для объяснения этих результатов могут быть предложены две гипотезы:

1. Код триплетный, сильно вырожденный, не имеет ограничений в составе слов, но имеет ограничения в выборе триплетов, кодирующих одни и те же аминокислоты. Для того чтобы код разрешал не слишком большое число замещений и при этом только очень ограниченно разрешал тройные переходы, он должен быть устроен примерно таким образом, как гипотетический код, рассмотренный в конце предыдущей главы: слова, кодирующие одну и ту же аминокислоту, должны, как правило, отличаться только одной буквой. Ясно, что такой код должен обладать некоторыми свойствами диплетного.

2. Код диплетный. Так как простой диплетный код недостаточен, то при этом должен существовать еще какой-то дополнительный источник информации.

Как будет показано ниже, эти две гипотезы не являются альтернативными. Но вначале рассмотрим более подробно вто-

рую гипотезу, с целью выяснения вопроса, что может служить источником дополнительной информации.

В синтезе белка непосредственно участвуют, как теперь достаточно хорошо известно: рибосомы, и-РНК, комплексы т-РНК с активированными аминокислотами и соответствующая ферментная система. Ясно, что ни т-РНК, ни ферментная система не могут быть искомым источником информации, так как они не связаны однозначным образом с конкретными местами присоединения аминокислот. Заманчивая на первый взгляд гипотеза, согласно которой диплет может приобретать разный смысл в зависимости от локальных свойств рибосомы, находится в резком противоречии с фактами, которые однозначно указывают на то, что рибосома не играет в синтезе белка специфической роли [101, 309]. Кроме того, можно попытаться принять «код последовательного ряда» [16], по которому аминокислота определяется диплетом и предыдущей аминокислотой. В пользу этого предположения могли бы говорить полученные в последнее время данные о том, что синтез полипептидных цепей происходит, действительно, последовательно, начиная с N-концевой аминокислоты и кончая С-концевой [87, 131а, 161а]. Однако код последовательного ряда требует существования запрещенных соседств в белках; между тем известно, что таких запрещенных соседств не существует. В результате остается единственный возможный носитель информации— и-РНК.

Казалось бы, что код может иметь словарь с непостоянным числом букв в словах [260]. Можно, например, предположить, что 16 аминокислот кодируются диплетами, а остальные — триплетами, или предложить какие-либо другие схемы. Однако при прочих общепринятых свойствах кода такое его устройство маловероятно. Действительно, если не существует специального механизма, препятствующего этому (например, запятых), то к триплету и-РНК может приосединиться не только т-РНК с соответствующим триплетом, но и такая т-РНК с диплетом, диплет которой комплементарен двум буквам рассматриваемого триплета. При этом становится невозможным однозначное чтение текста. Правда, можно придумать систему кодирования, при которой снимается неоднозначность, например, считать, что все триплетные слова начинаются с одной и той же буквы [260]. Но, помимо большой искусственности, против такой системы, как и против любого другого варианта диплетно-триплетного кода, говорят и другие факты: так, в случае диплетно-триплетного кода должно сильно ограничиваться число возможных замещений между аминокислотами, кодируемыми диплетами и триплетами — они возможны только за счет делеций и вставок, что сильно ограничивает их число. Кроме того, поскольку диплетных слов больше, чем триплетных, то известный результат по восстановлению нормального фенотипа у фага Т4 при совме-

щении трех профлавиновых мутаций одного типа с большим основанием можно было бы ожидать при совмещении двух, чего, однако, несмотря на довольно большое число опытов, не было ни разу обнаружено [128].

По-видимому, единственный способ преодолеть перечисленные затруднения заключается в принятии предположения, что в и-РНК каждый диплет включен с добавочной буквой, которая, однако, не всегда несет информационную нагрузку. В тех случаях, когда диплет кодирует одну аминокислоту, он будет действовать одинаково при любом третьем нуклеотиде, если же он кодирует две аминокислоты, то третий нуклеотид определяет, какая именно из двух аминокислот будет включена.

Тот вывод, к которому мы пришли при анализе условий, допускающих применимость диплетного кода, не отличается, по сути дела, от той схемы, которая была принята вначале в первой гипотезе. Действительно, триплеты, которые при кодировании одной и той же аминокислоты должны иметь две общих буквы, или диплеты, к которым присоединяется третья буква, не несущая существенной информации, — это два способа описания одной и той же модели.

Сейчас трудно сказать, какой из двух способов описания кода является более целесообразным. Это должно выясниться в ходе дальнейших исследований. Если с разными словами и-РНК, кодирующими одни и те же аминокислоты, взаимодействуют разные т-РНК, то тогда, конечно, следует говорить о триплетном коде. Если же т-РНК, взаимодействуя только с двумя буквами и-РНК, оказывается одинаковой для нескольких «триплетов» и-РНК, то тогда, возможно, правильнее будет говорить о том, что код диплетный.

Подробное обсуждение этого вопроса преждевременно, но, поскольку нам предстоит заниматься расшифровкой кода, причем все его свойства должны при этом учитываться, это новое свойство должно быть отражено в терминологии. Мы будем говорить в дальнейшем о триплетном коде, различая в триплетных словах **облигатные** и **факультативные** буквы. Так, если например, окажется, что пролин кодируется триплетами ЦЦ. Ц и ЦЦ. У, то условимся говорить, что слова, кодирующие пролин, содержат облигатную пару ЦЦ, которая имеет этот смысл при факультативных третьих буквах Ц и У.

Экспериментальные данные говорят в пользу только что высказанной гипотезы. Эти данные будут подробно анализироваться в двух последующих главах. Здесь мы приведем только несколько примеров. Недавно, рассматривая степень соответствия кодов, наблюдавшимся замещениям, мы имели возможность убедиться в том, что диплетные коды лучше соответствуют фактам, чем слабо вырожденные триплетные. Мысленно преобразуем диплетный код в код, предлагаемый нами. Для этого, сохранив

смысл всех диплетов, которые теперь становятся облигатными парами, присоединим к ним третьи буквы, скажем, У и Ц. В случае изокодов два триплета с данной парой приобретают разный смысл, в случае однозначных диплетов — одинаковый. В результате получится триплетный код со словарем из 32 букв, объясняющий около 80% замещений, т. е. дающий лучшее соответствие фактам, чем другие триплетные коды со словарем почти такого же объема.

В пользу предлагаемой гипотезы говорят также результаты опытов с СПРН. Так как обычно считают, что расположение нуклеотидов в СПРН случайное, то все слова одинакового состава должны встречаться одинаково часто. Поэтому степень включения аминокислот, кодируемых триплетами одного и того же состава, должна быть одинаковой.

Рассмотрим в качестве примера результаты опытов И. Маттеи и сотрудников [231] по влиянию поли-УЦ на включение аминокислот. Результаты этих опытов приведены в табл. 33. Бро-

Таблица 33

Стимуляция включения аминокислот в присутствии статистического полимера УЦ (39 : 61)

Аминокислоты	Диплет	Триплет	Относительное включение	
			вычисленное	наблюдаемое
Финилаланин . .	УУ	УУУ	100	100
Пролин	ЦЦ	1У, 2Ц	244	285
Серин	ЦУ	2У, 1Ц	157	160
Лейцин	УЦ	2У, 1Ц	157	79

сается в глаза резкое расхождение между вычисленным и наблюдаемым выходом для лейцина, которое особенно заметно рядом с очень хорошим соответствием для пролина и серина. До сих пор такой результат не получал объяснения. Что же касается предложенной гипотезы, то она для данного опыта предсказывает возможность выходов, составляющих 39% и 61% теоретических, для тех случаев, когда облигатная для данной аминокислоты пара имеет смысл только при одной из возможных в данном СПРН факультативных третьих букв. Для аминокислот, облигатная пара для которых имеет состав У, Ц, это дает, соответственно, выход 61% и 97% (при принятии выхода фенилаланина за 100%), что очень близко к цифре полученной для лейцина. Таким образом, полученные результаты могут быть объяснены тем, что слова, кодирующие фенилаланин, пролин и серин, имеют соответствующий смысл при любых факультативных буквах

(У и Ц), а лейцин только при одной из них. Следует заметить, что результат, полученный в этом опыте, нельзя объяснить случайностью. В аналогичных опытах И. Спейера и сотрудников [289] с поли-УЦ состава 5 : 1 и 3 : 1 лейцин давал выход примерно в полтора раза меньше, чем серин.

Против той интерпретации результатов, которая давалась до сих пор, говорит еще следующее соображение. Вернемся к данным, приведенным в табл. 33. Примененный полимер состоял на 61% из цитозина, следовательно, $0{,}61^3 \cdot 100 = 22{,}6\%$ всех триплетов должны иметь состав ЦЦЦ, которому по расшифровке авторов ничто не соответствует. Другими словами, почти что каждый четвертый триплет не должен был бы давать никакого включения, следовательно, и не могло бы образоваться сколько-нибудь длинных пептидов, которые могли бы быть обнаружены в опыте. По предложенной же нами гипотезе пролин кодируется диплетом ЦЦ с обеими факультативными буквами, поэтому триплет ЦЦЦ приобретает смысл. Как показали результаты работ, опубликованных уже после окончания этого анализа, пролин, действительно, включается триплетом ЦЦЦ [320].

В пользу предлагаемого кода можно привести и другие соображения и факты.

Предложенный код обладает преимуществом, по сравнению с большинством других, с точки зрения эффективности кодирования. Действительно, диплетный код, как это уже неоднократно говорилось, не обладает достаточным объемом информации. В триплетном коде, наоборот, она слишком велика, в связи с чем должно быть либо большое число нонсенсов, что повысит вероятность помех при передаче и хранении информации, либо излишне большое число разных молекул т-РНК, что также может уменьшить надежность кода.

Неоднократно уже указывалось на то, что некоторые твердо установленные экспериментальные факты находятся в противоречии с предлагавшимися до сих пор вариантами триплетного кода. Кратко рассмотрим их и попытаемся показать, что предложенный код, в отличие от других, не противоречит им.

1. Все предлагавшиеся до сих пор варианты триплетного кода сконструированы таким образом, что каждый триплет содержит, по крайней мере, один У. Если бы это было так, то и-РНК должна была бы почти наполовину состоять из У. На самом же деле все проведенные до сих пор анализы РНК показывают, что содержание У в ней варьирует в пределах от 13,3 до 26,4% [3, 5, 6]. Р. Робертс [260] подсчитал, что при принятии диплетного кода, полученного путем «сокращения» всех триплетов Ниренберга-Очоа на один У, содержание нуклеотидов в и-РНК оказывается таким, как этого требуют современные знания о составе РНК. Он же показал, что расчет состава матрицы на основании аминокислотного состава некоторых видов бактерий приводит к та-

кому составу РНК, который наблюдается именно у этих видов. Совершенно ясно, что если к диплетам прибавить в равной пропорции все 4 нуклеотида, то результаты этих расчетов не могут существенно измениться, даже если третий нуклеотид в небольшом проценте случаев сообщает облигатным парам разный смысл. Поэтому даже до конкретной расшифровки состава слов можно предсказать, что предложенный код должен требовать состава матрицы, соответствующего реальным РНК.

2. В работе Н. Суеока [293] было установлено, что у бактерий содержание определенных аминокислот коррелирует с содержанием Г + Ц в их ДНК, причем в одних случаях наблюдается прямая, в других обратная связь, тогда как в третьих содержание аминокислоты не зависит от содержания Г + Ц. Последнее в невырожденном коде возможно только при четном числе букв в словах. С другой стороны, в случае сильной вырожденности кода следовало бы ожидать отсутствия корреляции с содержанием в ДНК Г + Ц для всех или для большинства аминокислот. Однако было замечено [217, 348], что если триплеты «сократить» на один урацил, то получится идеальное соответствие между коэффициентами регрессии, полученными в опытах Н. Суеока, и составом диплетов. И в этом случае ничего не изменится, если к диплетам прибавить в примерно равной пропорции разные нуклеотиды, поэтому наш код должен давать такое же хорошее соответствие опыту, как и чистый диплетный.

3. В опытах Ф. Крика и сотрудников [128] было показано, что при вызывании мутаций профлавином, который оказывает эффект путем вставки или делеции одного основания, совмещение в одном цистроне трех однотипных мутаций ведет к восстановлению дикого фенотипа. Эти блестящие и широко известные опыты рассматриваются как доказательство того, что код является триплетным. Однако Р. Уолл [322] указал на то, что, строго говоря, эти данные однозначно говорят не о триплетности кода, а о том, что трем равно кодовое отношение, т. е. отношение числа букв в тексте и-РНК к числу аминокислот в кодируемом белке. Независимо от того, называть ли предложенный нами код триплетным или диплетным, но кодовое отношение в нем равно трем. Таким образом, он, в отличие от обычных диплетных кодов, не противоречит упомянутым данным.

4. В той же работе [128] было показано, что если довольно большой участок и-РНК заменен случайной последовательностью букв, то, тем не менее, белок все же строится. Это говорит о том, что число нонсенсов очень невелико, а следовательно код сильно вырожден. Между тем, результаты, полученные нами в этой, а особенно в предыдущей главе, показывают, что при случайной вырожденности степень ее не может быть слишком большой. Однако, если предположить, что вырожденность осуществляется преимущественно за счет варьирования факультативных

букв, то, даже при крайней степени вырожденности, число разрешенных замещений не выйдет за установленные выше пределы. Таким образом, предложенный код лучше соответствует совокупности имеющихся фактов, чем другие.

На основании результатов, полученных в двух последних главах, можно следующим образом сформулировать основные положения о свойствах аминокислотного кода.

1. Код триплетный, однако облигатными для кодирования аминокислоты являются в большинстве случаев лишь две буквы, третья часто оказывается факультативной.

2. Ограничений в выборе букв для составления кодирующих слов не существует.

3. Вырожденность кода большая. Возможно, что большинство триплетов имеет смысл. Однако вырожденность по облигатным парам не должна быть большой, она осуществляется в основном за счет различий в факультативных третьих буквах.

4. Слова, кодирующие одну и ту же аминокислоту, в большинстве случаев должны иметь одинаковые облигатные пары.

Интересно, что в результате проведенного анализа мы пришли к тем же выводам об общих свойствах кода, к которым приходит и Ф. Крик [128] на основании совершенно других данных. Это, конечно, делает выводы значительно более прочными. Помимо подтверждения уже известных свойств кода, удалось найти правило, по которому преимущественно должно осуществляться его вырождение, что нашло отражение во введении понятий облигатных и факультативных частей слов.

После этого можно со значительно большей уверенностью, чем раньше, приступить к расшифровке словаря. Этому будут посвящены две следующие главы.

ГЛАВА 7

ОПРЕДЕЛЕНИЕ ОБЛИГАТНЫХ БУКВ В КОДОВЫХ СЛОВАХ

Так как в настоящее время основным источником информации для расшифровки кода являются результаты опытов с СПРН, то в связи с очень большим числов вариантов, которые могут встретиться при сравнении выходов аминокислот в случае вырожденного кода и невысокой точностью получаемых цифр, непосредственная расшифровка смысла отдельных триплетов вряд ли возможна. Но новые представления об общих свойствах кода позволяют разделить эту работу на два этапа, благодаря чему она становится осуществимой. Если вначале определить состав облигатных пар в триплетах и порядок букв в них, то окончательная расшифровка сведется лишь к определению факультативных букв при уже известных облигатных парах, что, конечно, несравненно более просто. Между тем, можно предложить методы для достаточно надежного определения облигатных пар.

Возражения, касающиеся интерпретации результатов опытов с СПРН, относятся главным образом к ее количественному аспекту. Между тем, как будет ниже показано, для определения состава облигатных пар достаточно лишь качественного анализа этих результатов. Кроме того, можно предложить для той же цели другой метод, совершенно независимый от первого, дающий согласные результаты.

Поэтому в настоящей главе будет произведено определение облигатных пар двумя независимыми методами, а затем будет выяснен их относительный порядок, что даст достаточно твердую основу для последующей расшифровки триплетов.

КАЧЕСТВЕННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ОПЫТОВ С ИСКУССТВЕННЫМИ ПОЛИНУКЛЕОТИДАМИ

Основы метода. Будем считать для простоты, что каждой аминокислоте соответствует только одна облигатная пара букв и что третья буква триплета всегда является факультатив-

ной. Рассмотрим ожидаемые качественные результаты, которые должны при этом получаться в опытах с СПРН. Если некоторой аминокислоте соответствует облигатная пара, состоящая из двух одинаковых букв (например, АА), то эта аминокислота должна включаться всеми СПРН, в состав которых входит данная буква (т. е. поли-А, поли-АГ, поли-АЦ, поли-АУ, поли-АГУ, поли-АГЦ и поли-АУЦ). Если аминокислоте соответствует облигатная пара, состоящая из двух разных букв (например, АГ), то точно так же включение должно происходить в присутствии всех полимеров, в состав которых входит данная пара букв (т. е. поли-АГ, поли-АГУ и поли-АГЦ).

Трехкомпонентные СПРН для нашего рассмотрения удобнее обозначать не теми буквами, которые входят в их состав, а той буквой,которая в них отсутствует, например, поли-АГУ будем обозначать как поли (—Ц). В таком случае в опытах с трехкомпонентными СПРН включение будет стимулироваться всеми СПРН, кроме тех, в которых отсутствуют буквы, необходимые для включения данной аминокислоты. Так, в рассмотренном примере, где аминокислота в опытах с двухкомпонентными СПРН должна была включаться только в присутствии поли-АГ, в опытах с трехкомпонентными она будет включаться всеми СПРН, кроме поли (—А) и поли (—Г). Поэтому качественный анализ результатов опытов с двухкомпонентными и трехкомпонентными СПРН может дать независимо друг от друга состав облигатных пар.

В действительности ситуация несколько сложнее рассмотренной, так как некоторым аминокислотам может соответствовать более чем по одной облигатной паре, а также потому, что число аминокислот больше, чем число возможных пар букв, в связи с чем не все третьи буквы могут быть факультативными. Однако, если эти отклонения от рассмотренной модели не слишком велики, то этот метод все равно может быть использован. Этому помогает прежде всего то обстоятельство, что анализ двухкомпонентных СПРН дает *достаточные* буквы, а трехкомпонентных — *необходимые*. Поэтому сравнение результатов, полученных в двух группах опытов, должно помочь выяснению возможных недоразумений. Тому же в некоторой степени могут служить и результаты опытов с однокомпонентными СПРН. Наконец, в случаях, если аминокислота кодируется лишь одной облигатной *тройкой*, то это также должно быть сразу заметно при качественном рассмотрении результатов: включение должно стимулироваться только одним единственным трехкомпонентным СПРН.

Экспериментальный материал. В настоящее время проведены успешные опыты со всеми возможными СПРН, кроме поли-Г. При этом по большинству СПРН имеются независимые данные, полученные в разных лабораториях. В табл. 34 дана качественная сводка имеющихся к настоящему времени дан-

Таблица 34

Качественные данные о влиянии различных СПРН на включение аминокислот в белки

Аминокислота	Состав СПРН												
	А	У	Ц	АГ	АУ	АЦ	ГУ	ГЦ	УЦ	АГУ (—Ц)	АГЦ (—У)	АУЦ (—Г)	ГУЦ (—А)
Аланин	—	—	—	—	—	—	—	+	—	—	+	—	+
Аргинин	—	—	—	+	—	—	—	+	—	(+)	+	—	+
Аспарагиновая к-та	—	—	—	—	—	—	—	—	—	(+	+	—	—
Аспарагин	—	—	—	+	+	+	—	—	—	+	+	+	—
Валин	—	—	—	—	—	—	+	—	—	+	—	—	+
Гистидин	—	—	—	—	—	+	—	—	—	—	+	+	—
Глицин	—	—	—	+	(—)	—	+	+	—	+	(+)	—	+
Глютаминовая к-та	—	—	—	+	—	—	—	—	—	+	+	—	—
Глютамин	—	—	—	(—)	—	+	—	—	—	—	+	—	—
Изолейцин	—	—	—	—	+	—	—	—	—	+	—	+	+
Лейцин	—	(+)	—	—	+	—	+	—	+	+	—	+	+
Лизин	+	—	—	+	+	+	—	—	—	?	+	+	—
Метионин	—	—	—	—	—	—	—	—	—	+	—	—	—
Пролин	—	—	+	—	—	+	—	+	+	—	+	+	+
Серин	—	—	—	—	—	(—)	—	(—)	+	(—)	+	+	+
Тирозин	—	—	—	—	+	—	—	(—)	—	+	—	+	—
Треонин	—	—	—	—	—	+	—	(+)	—	—	+	+	—
Триптофан	—	—	—	—	—	—	+	(—)	—	+	—	—	+
Фенилаланин	—	+	—	—	+	—	+	(—)	+	+	—	+	+
Цистеин	—	—	—	—	(—)	—	+	—	—	+	—	—	+

ных. Цифры, на основании которых она составлена, мы здесь не приводим, так как они будут использоваться в главе 8 для количественного анализа, где и дается соответствующий фактический материал. В таблице знак «плюс» означает включение данной аминокислоты, знак «минус» — его отсутствие. Если наблюдались противоречия в работах разных авторов (их было не так много), то соответствующий знак заключали в скобки. Если в противоречивом случае включение было значительным, то в скобки заключался «плюс», если низким, то «минус».

Состав облигатных пар. Результаты анализа приведены в табл. 35. Как видно из этой таблицы, в большинстве случаев анализ двух- и трехкомпонентных СПРН дал вполне согласные результаты. Случаи расхождений почти всегда объяснимы. Рассмотрим их.

Аргинин на основании опытов с двухкомпонентными СПРН имеет две облигатных пары — А, Г и Г, Ц, а на основании трехкомпонентных — одну облигатную букву Г. Это может быть объяснено двояко: либо тем, что аргинину соответствуют две разных облигатных пары — А, Г и Г, Ц или тем, что ему соответствует одна пара — ГГ.

Таблица 35

Состав облигатных пар на основании качественного анализа данных опытов с СПРН

Аминокислота	С однокомпонентными СПРН	С двухкомпонентными СПРН	С трехкомпонентными СПРН	На основании всего материала
Аланин	—	Г,Ц	Г,Ц	Г,Ц
Аргинин	—	А,Г + Г,Ц	ГГ	А,Г + Г,Ц или ГГ
Аспарагиновая к-та	—	—	А,Г	А,Г
Аспарагин	—	А,У+А,Ц	АА	А,У+А,Ц или АА
Валин	—	Г,У	Г,У	Г,У
Гистидин	—	А,Ц	А,Ц	А,Ц
Глицин	—	ГГ	ГГ	ГГ
Глютаминовая к-та	—	А,Г	А,Г	А,Г
Глютамин	—	А,Ц	А,Г,Ц	А,Ц
Изолейцин	—	А,У	А,У	А,У
Лейцин	(УУ)	УУ	УУ	УУ
Лизин	АА	АА	АА	АА
Метионин	—	—	А,Г,У	А,Г,У
Пролин	ЦЦ	ЦЦ	ЦЦ	ЦЦ
Серин	—	У,Ц	Ц	У,Ц + ?
Тирозин	—	А,У	А,У	А,У
Треонин	—	А,Ц	А,Ц	А,Ц
Триптофан	—	Г,У	Г,У	Г,У
Фенилаланин	УУ	УУ	УУ	УУ
Цистеин	—	Г,У	Г,У	Г,У

Аспарагиновая кислота не включается ни одним из двухкомпонентных СПРН, а трехкомпонентные указывают на облигатную пару А, Г. Отрицательный результат с двухкомпонентными матрицами может быть объяснен тем, что в триплет для аспарагиновой кислоты в качестве факультативной буквы не могут входить ни А, ни Г.

Аспарагин, также как аргинин, получает на основании двухкомпонентных СПРН две облигатных пары — А, У и А, Ц, а на основании трехкомпонентных — одну (АА).

Глютамин, согласно анализу двухкомпонентных СПРН, получает облигатную пару А, Ц, тогда как на основании трехкомпонентных ему должна принадлежать облигатная тройка А, Г, Ц. Это явное противоречие, так как если поли-АЦ стимулирует включение, то оно должно стимулироваться также и поли-УАЦ. Однако включение глютамина с поли-УАЦ было проверено только в одном опыте с полимером, содержавшим низкий процент А и Ц. Поэтому для глютамина более вероятно существование облигатного диплета, а не триплета.

С серином получены не вполне ясные данные, которые скорее всего могут объясняться тем, что облигатная пара У, Ц — не единственная.

В последней колонке табл. 35 приведены выводы о составе облигатных групп на основании сравнения всех результатов. Только в одном случае (метионин) получена облигатная тройка, для всех остальных кодовых слов облигатными оказываются только две буквы.

СРАВНЕНИЕ НЕЗАВИСИМЫХ КЛАССИФИКАЦИЙ АМИНОКИСЛОТ И ОБЛИГАТНЫХ ПАР НУКЛЕОТИДОВ

Основа метода. Кроме примененного выше метода определения облигатных пар в кодовых словах, возможен еще один, совершенно независимый от первого. Сущность его сводится к следующему. Если есть возможность расклассифицировать аминокислоты и соответствующие им облигатные пары на несколько групп более чем одним способом, то сравнение этих классификаций может дать возможность и для индивидуальных соподчинений отдельных облигатных пар вполне определенным аминокислотам. Пусть, например, на основании одной классификации установлено, что аланину, аргинину, глицину и пролину соответствуют облигатные пары ГГ, ЦЦ, ГЦ и ЦГ, а на основании другой, что аспарагину, лизину, пролину и треонину соответствуют пары АА, ЦЦ, АЦ и ЦА. При сравнении результата этих двух классификаций бросается в глаза, что общими для обеих классификаций являются только одна аминокислота (пролин) и одна пара букв (ЦЦ), из чего следует, что ЦЦ является облигатной парой для пролина. Имеющиеся в настоящее время данные дают два независимых способа классификации, путем сравнения которых можно расшифровать значительную часть облигатных пар.

Экспериментальный материал. Материал для первой классификации дают результаты опытов Н. Суеока [293]. Этот автор определил процентное содержание различных аминокислот в белках ряда бактерий с сильно отличавшимся содержанием комплементарной пары Г + Ц в их ДНК. Оказалось, что содержание аминокислот в белках коррелировало с содержанием Г + Ц, но коррелировало по-разному. Если содержание одних аминокислот возрастало с содержанием Г + Ц, то содержание других падало, а в ряде случаев содержание аминокислоты не зависело заметным образом от содержания Г + Ц. Результаты этих опытов приведены в табл. 36, где аминокислоты сгруппированы в соответствии с теми коэффициентами регрессии, которые дает зависимость их содержания в белках от содержания в ДНК пары Г + Ц. Можно считать, что те аминокислоты, которые дали положительные коэффициенты, имеют облигатные пары состава ГГ, ЦЦ, или Г, Ц, те, которые дали отрицательные — АА, УУ или А, У. Остальным аминокислотам должны принадлежать облигатные пары У, Г; А, Г; Ц, У, или А, Ц.

Таблица 36

Зависимость содержания аминокислот в белках бактерий от содержания Г + Ц в их ДНК [по Суеока, 293]

Группа	Аминокислоты	Коэффициент регрессии	P
I. Достоверная прямая зависимость	ала	+0,164±0,055	<0,001
	арг	+0,089±0,043	<0,001
	глиц	+0,051±0,034	<0,02
	про	+0,024±0,020	~0,02
II. Достоверная обратная зависимость	илей	—0,098±0,032	<0,001
	лиз	—0,084±0,046	<0,01
	тир	—0,047±0,035	<0,02
	фала	—0,040±0,022	<0,001
III. Отсутствие достоверной связи	вал	+0,008±0,023	>0,30
	тре	0,000±0,028	>0,90
	лей	—0,006±0,020	>0,50
	гис	—0,010±0,015	>0,10
	сер	—0,017±0,033	>0,30
	мет	—0,024±0,062	>0,30
	глюх	—0,052±0,050	>0,05
	аспх	—0,053±0,054	>0,05

Для второй классификации, независимой от первой, можно использовать данные по замещениям, вызванным азотистой кислотой.

Как уже указывалось во второй главе, в зависимости от того, играет ли аминокислота только роль замещаемой, только замещающей или может быть и той и другой, можно все аминокисло-

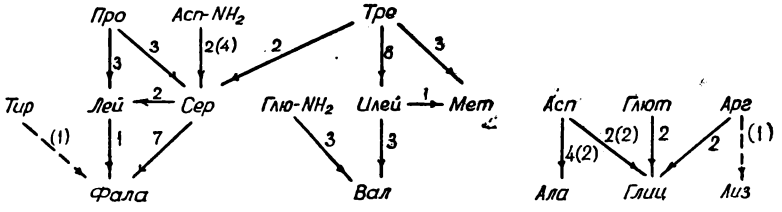

Рис. 13. Нитритные замещения в белке ВТМ.
В скобках указано число случаев, в достоверности которых могут быть сомнения.

ты разбить на три группы. Им, очевидно, должны соответствовать облигатные пары, в первом случае состоящие из А и Ц, во втором — из А и У, а в третьем — из букв обоих сортов. На рис. 13 изображены все известные до сих пор замещения в белке

ВТМ, вызванные действием азотистой кислоты на вирусную РНК. Эти данные несколько отличаются от приведенных в табл. 5, так как при составлении этого рисунка были учтены данные последней работы Г. Виттмана [342]. Замещения, в отношении происхождения которых имеются сомнения, показаны пунктиром.

Состав облигатных пар. В табл. 37 приведены две независимые классификации аминокислот и облигатных пар на

Таблица 37

Две независимые классификации аминокислот и диплетов

Группа	Облигатные пары	Аминокислоты
	А. По корреляции между составом белков и ДНК	
I	ГГ, ЦЦ, ГЦ, ЦГ	ала, арг, глиц, про
II	АА, УУ, АУ, УА	лиз, илей, тир, фала
III	ГУ, УГ, ГА, АГ, ЦУ, УЦ, ЦА, АЦ	вал, гис, лей, мет, сер, тре, (асп*x*), (глю*x*)
	Б. По замещениям в нитритных мутациях	
I	АА, ЦЦ, АЦ, ЦА	асп-NH_2, про, тре
II	УУ, ГГ, УГ, ГУ	ала, вал, глиц, фала (лиз?)
III	АГ, ГА, АУ, УА	арг, асп, глют, глю-NH_2
	ЦГ, ГЦ, ЦУ, УЦ	илей, лей, мет, сер, (тир)

Примечание. В первой классификации отсутствует материал для триптофана и цистеина, а данные по аспарагиновой и глютаминовой кислотам приведены в сумме с их амидами. Во второй классификации отсутствуют данные по гистидину, триптофану и цистеину, а данные по лизину и тирозину мало достоверны.

основании только что обсуждавшихся данных. Далее можно, сравнивая каждую группу первой классификации с каждой группой второй, по наблюдающимся совпадениям расшифровать ряд облигатных пар, что сделано в табл. 38. В опытах Н. Суеока для аспарагина (+ аспарагиновая кислота) и для глютамина (+ глютаминовая кислота) были получены отрицательные коэффициенты регрессии, довольно большие по абсолютной величине, которые, в связи со значительной ошибкой, оказались статистически недостоверными. Поэтому в табл. 38 данные, касающиеся этих аминокислот, расшифрованы в соответствии с обеими возможностями — при отнесении их ко второй и к третьей группе. Дальнейшее уточнение принадлежности облигатных пар определенным аминокислотам может быть сделано путем сравнения уже полученных результатов с нитритными замещениями.

Так, согласно данным табл. 38, лейцин и серин могут иметь облигатные пары либо состава А, Г, либо У, Ц. Однако, посколь-

Таблица 38

Частичная расшифровка облигатных пар путем сравнения двух классификаций

Сравниваемые группы		Общие облигатные пары	Общие аминокислоты	При отнесении асп*х* и глю*х* по классификации А во вторую группу
Из классификации А	Из классификации Б			
1	1	ЦЦ	про	
1	2	ГГ	ала, глиц	
1	3	ГЦ, ЦГ	арг	
2	1	АА	—	(Асп-NH_2)
2	2	УУ	фала, (лиз?)	
2	3	АУ, УА	илей, (тир)	(асп), (глют) (глю-NH_2)
3	1	АЦ, ЦА	тре, (асп-NH_2)	
3	2	УГ, ГУ	вал	
3	3	АГ, ГА, ЦУ, УЦ	лей, мет, сер, (асп), (глют), (глю-NH_2)	

ку обе эти аминокислоты замещают пролин (ЦЦ) и замещаются фенилаланином (УУ), из двух возможных пар каждой из них может соответствовать только пара У, Ц. Точно так же, на основании замещений глют → глиц, асп → глиц и асп → ала можно внести однозначность в вопрос об облигатных парах для аспарагиновой и глютаминовой кислот; для обеих такой парой должна быть А, Г. При дальнейшем сравнении обнаруживается, что некоторые расшифровки не находят себе подтверждения в нитритных замещениях (например, пара Г, Ц для аргинина или АА для аспарагина), с другой стороны, замещения указывают на существование облигатных пар, не обнаруженных при сравнении двух классификаций (например, А, Г для аргинина, А, У — для метионина).

В табл. 39 приведены окончательные результаты расшифровки, произведенной этим методом, причем менее достоверные (не подтверждаемые замещениями, или устанавливаемые только на их основании) заключены в скобки. В той же таблице приведены результаты расшифровки, произведенной первым методом — путем качественного анализа данных опытов с СПРН. Нетрудно заметить очень большое сходство в результатах обеих расшифровок. В последней колонке таблицы приведен состав облигатных пар на основании рассмотрения результатов обоих методов. В тех случаях, когда пара определена только одним из методов, она заключена в скобки. Таким образом и в большинстве случаев с полной достоверностью удается определить облигатные пары для всех аминокислот. Для некоторых из них оказывается весьма вероятным существование двух разных пар.

Таблица 39

Состав облигатных пар

Аминокислоты	Метод		Принимаемый состав
	Качественный анализ данных с СПРН	Сравнение независимых классификаций	
Аланин	Г, Ц	ГГ	Г, Ц (+ ГГ)
Аргинин	А,Г+Г,Ц или ГГ	(А, Г) + (Г, Ц)	А, Г+Г, Ц
Аспарагиновая к-та	А,Г	А, Г	А, Г
Аспарагин	А,У+А,Ц или АА	АА	АА
Валин	Г, У	Г, У	Г, У
Гистидин	А, Ц	(Ц, У; Г, У; А, Г или А, Ц)	А, Ц
Глицин	ГГ	ГГ	ГГ
Глутаминовая к-та	А, Г	А, Г	А, Г
Глютамин	А, Ц	?	(А, Ц)
Изолейцин	А, У	А, У	А, У
Лейцин	УУ	У, Ц	(УУ) + (У, Ц)
Лизин	АА	?	(АА)
Метионин	А, Г, У	?	(А, Г, У)
Пролин	ЦЦ	ЦЦ	ЦЦ
Серин	У, Ц + ?	У,Ц	У, Ц (+ ?)
Тирозин	А, У	А, У	А, У
Треонин	А, Ц	А, Ц	А, Ц
Триптофан	Г, У	?	(Г, У)
Фенилаланин	УУ	УУ	УУ
Цистеин	Г, У	?	(Г, У)

ПОРЯДОК БУКВ В ОБЛИГАТНЫХ ПАРАХ

Остается определить порядок букв в облигатных парах. Пока что имеется единственный источник информации об абсолютном порядке букв в кодовых словах. В работе А. Ваба и сотрудников [319] было показано, что смесь СПРН АУУУУ... и ААУУУ... стимулирует включение тирозина и не влияет на включение изолейцина, аспарагина и лизина, а смесь ГУУУУ... и ГГУУУ... стимулирует включение цистеина, не влияя на включение валина, глицина и триптофана. Из этого следует, что порядок букв в облигатной паре для тирозина должен быть АУ, а для цистеина — ГУ (считая, что справа находится свободная 3′-гидроксильная группа). Вероятным также кажется, что порядок букв для изолейцина — не АУ (следовательно, УА) и для валина — не ГУ (т. е. УГ); этот вывод менее достоверен, чем в отношении тирозина и цистеина, так как отрицательный результат опыта можно объяснить низкой чувствительностью системы. Если исходить из этих результатов, то нетрудно на основании данных о замещениях расшифровать порядок и в большинстве остальных облигатных пар.

В связи с вырожденностью по облигатным парам для некоторых аминокислот, надежнее начинать расшифровку с тех случаев, когда одни и те же аминокислотные локусы давали разные замещения. Таких случаев не так много: группа мутаций и рекомбинаций в белке А триптофансинтетазы [181, 182], объединяющая 6 аминокислот (данные были приведены в табл. 7), и

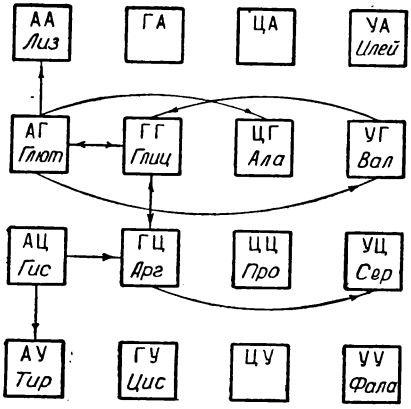

Рис. 14. Облигатные пары и множественные замещения в решетке Зьюби—Квастлера. Стрелки показывают направление замещения.

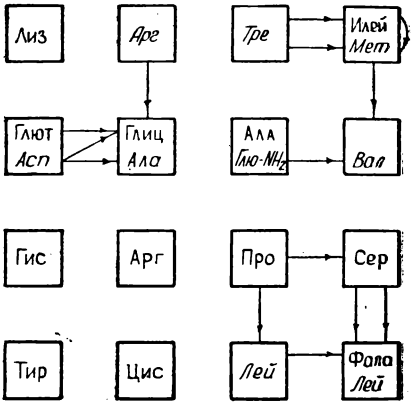

Рис. 15. Нитритные замещения в решетке Зьюби—Квастлера. Расположение облигатных пар то же, что и на рис. 14.

две группы среди мутаций гемоглобина, каждая из которых объединяет по три аминокислоты (ср. табл. 6).

Для расшифровки порядка букв очень удобно пользоваться матрицей Зьюби—Квастлера, которая уже была описана в предыдущей главе. Напомним, что слова, дающие друг с другом разрешенные замещения, должны стоять в ней в узлах одного и того же ребра. После того, как размещены в решетке валин, изолейцин, тирозин и цистеин, порядок букв для облигатных пар которых мы только что установили, а также глицин, лизин, пролин и фенилаланин, облигатные пары которых состоят из одинаковых букв, упомянутые выше замещения располагаются непротиворечивым и притом единственным образом так, как это изображено на рис. 14. Это дает расшифровку порядка букв еще в нескольких парах.

Порядок букв в дальнейших облигатных парах можно определить, анализируя нитритные замещения. Для этого на решетку наносятся аминокислоты, для которых порядок букв в облигатных парах уже был выяснен (эти аминокислоты обозначены на рис. 15 прямым шрифтом), после чего размещаются нитритные замещения. Некоторая неясность возникает при этом в отношении аминокислот, облигатная пара для которых имеет

состав А, Г. Для глютаминовой кислоты порядок букв (АГ) уже установлен. Так как среди замещений в естественных белках несколько раз наблюдалось замещение асп — глют, то очевидно, что аспарагиновой кислоте должна принадлежать та же облигатная пара. В таком случае эти замещения могут быть объяснены изменением факультативной буквы; если же облигатной паре для аспарагиновой кислоты приписать порядок ГА, то эти замещения объяснить нельзя. Что касается второй облигатной

Таблица 40

Порядок нуклеотидов в облигатных парах

Аминокислоты	Облигатные пары	Аминокислоты	Облигатные пары
ала	ЦГ + ГГ	лей	ЦУ + УУ
арг	ГЦ + ГА	лиз	АА
асп	АГ	мет	УА
асп-NH_2	АА	про	ЦЦ
вал	УГ	сер	УЦ + ГА + ?
гис	АЦ	тир	АУ
глиц	ГГ	тре	ЦА + ?
глют	АГ	трип	Г, У
глю-NH_2	АЦ + ЦГ	фен	УУ
илей	УА	цис	ГУ

пары для аргинина (А, Г), то для нее более вероятен порядок ГА, так как порядок АГ уже принадлежит облигатной паре для двух аминокислот. Остальные соподчинения вряд ли требуют каких-либо пояснений.

Два нитритных замещения не удается однозначо поместить в решетку, а именно асп-NH_2 → сер и тре → сер. В обоих замещениях участвует серин, для которого уже казалось вероятным существование второй облигатной пары. Однако нет такой пары, приписав облигатность которой серину, можно было бы объяснить оба эти замещения. Таким образом, необходимо приписать серину целых три облигатных пары (например, кроме УЦ еще ГА и ЦГ), либо, приписав две пары серину, дать вторую пару также аспарагину или треонину (например: серин — ГА, треонин — ЦЦ). Но для выбора между этими возможностями пока нет данных. Это станет возможным при количественном анализе данных с СПРН. Что же касается второй облигатной пары для серина, то ею должна быть ГА, так как в пользу этого, кроме нитритного замещения асп-NH_2 → сер, говорят также замещения ала → сер, глиц → сер и лиз → сер, наблюдавшиеся каждое по нескольку раз.

Из облигатных пар, приведенных в табл. 39, не выяснен еще порядок для пары, соответствующей триптофану (Г, У) и для второй облигатной пары глютамина (А, Ц). В последнем случае

порядок должен быть АЦ, так как имеются замещения гис — глю-NH_2 и глют — глю-NH_2. Что же касается триптофана, то для него пока неизвестно ни одного достоверного замещения, в связи с чем вопрос о порядке букв в облигатной для него паре остается открытым. Этот порядок может быть выяснен при количественном анализе данных с СПРН. В табл. 40 приведены окончательные результаты расшифровки состава облигатных пар и порядка букв в них. При этом остался неизвестным порядок букв в паре, облигатной для триптофана, и остается открытым вопрос о дополнительной паре, которая должна быть облигатной для серина (ЦГ или УА) или для треонина (АА или ЦЦ). После этого из достаточно достоверных замещений оказываются объясненными почти все. (Необъясненными остаются только 3 замещения из 46). Поэтому можно думать, что удалось расшифровать почти все облигатные пары, что дает достаточно прочную основу для расшифровки триплетов, которой будет посвящена следующая глава.

ГЛАВА 8

РАСШИФРОВКА ТРИПЛЕТНОГО КОДА

После определения состава облигатных пар и порядка букв в них задача расшифровки кода сильно облегчается. Хотя в принципе для окончательной расшифровки кода могли бы служить данные разного рода, причем не только теоретически мыслимые, но и вполне доступные современной технике эксперимента, однако только опыты по влиянию СПРН на включение аминокислот в белки дают относительно большой материал. Вероятно, более надежные результаты дал бы анализ аминокислотных «множественных аллей», т. е. многократные замещения вполне определенных аминокислотных локусов белка. Но таких данных пока что имеется очень мало. Опыты по изучению мутаций триптофансинтетазы дают одну группу из шести аминокислот, связанных значительным числом взаимных переходов [181, 182]. Среди мутаций гемоглобина человека имеется два случая, когда одна и та же аминокислота дала две разных мутации (6-я и 63-я аминокислоты β-цепи). Этим исчерпывается весь известный нам материал такого рода.

Довольно большое число «множественных аллелей» можно обнаружить при сравнении разных цепей гемоглобина человека и животных и миоглобина, но, поскольку генетические связи соответствующих цепей неизвестны и, несомненно, отсутствуют многие промежуточные звенья, использование этих данных для расшифровки могло бы только привести к недоразумениям; эти данные целесообразнее использовать для проверки правильности уже произведенной расшифровки. Другие же данные, вроде корреляций в составе естественных белков и нуклеиновых кислот [293], не только малочисленны, но и не более надежны, чем опыты с СПРН. Поэтому в основу расшифровки кода мы должны положить данные опытов с СПРН. Сразу скажем, что и этих данных оказывается далеко недостаточно, поэтому, кладя их в основу анализа, мы будем дополнять его и привлечением данных по замещениям. Но даже и в этом случае полной расшифровки дать не удастся. Смысл нескольких триплетов оста-

нется вообще неизвестным, расшифровка части остальных останется более или менее предположительной. Впрочем, нас будет не столько интересовать окончательная расшифровка всех сочетаний нуклеотидов, сколько описание и иллюстрация методов, которые могут дать такую расшифровку после получения соответствующих цифр, а также подсказать постановку конкретных опытов, необходимых для полной расшифровки.

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ОПЫТОВ С ИСКУССТВЕННЫМИ ПОЛИНУКЛЕОТИДАМИ

Если при определении состава облигатных пар мы могли удовольствоваться качественным анализом, то для расшифровки их смысла при разных факультативных буквах необходим количественный анализ результатов, полученных в опытах с СПРН. Однако, как неоднократно указывалось [125, 343, 352], эти результаты, по крайней мере при современной технике эксперимента, дают настолько большие ошибки, что возникают сомнения в допустимости их количественного анализа. Для того чтобы убедиться в этом, достаточно сравнить результаты, полученные с полимерами одинакового состава в разных лабораториях. Так, поли-АУ с соотношением нуклеотидов 1 : 5 дал в лабораториях Очоа [290] и Ниренберга [231], включение (в процентах фенилаланина) для изолейцина, соответственно, 19 и 12, для лейцина 14 и 4,9, для тирозина — 26 и 13, глицин и цистеин в первом случае не включались вообще, во втором их включение составляло около 5%. Такие расхождения встречаются на каждом шагу. Даже в параллельных опытах в одной и той же лаборатории с одним и тем же препаратом наблюдаются иногда значительные расхождения; например, в работе [290] поли-УАГ (6 : 1 : 1) стимулировал включение аспарагина на 6,0 и на 3,6%. Совершенно очевидно, что при наличии одной единственной цифры использование ее может привести к совершенно ложным выводам. Другое дело, если имеются несколько сходных опытов, давших согласные результаты.

Второе затруднение — методика анализа. Применявшиеся до сих пор методы анализа справедливы только для однозначного кода. В случае же вырожденного кода эта методика неприменима даже в случае, если вырожденность будет учитываться.

Рассмотрим, какие типы сравнений выходов могут встретиться при анализе результатов опытов с двухкомпонентным СПРН. Пусть используется полимер, состоящий из нуклеотидов А и Б, взятых в пропорции $a : б$ (причем $a + б = 1$) и распределенных вдоль цепочки случайно. В этом полимере должны встречаться следующие триплеты со следующей вероятностью:

АА. А — с вероятностью a^3
АА. Б — с вероятностью $a^2б$

АБ. А — с вероятностью $a^2б$
БА. А — с вероятностью $a^2б$
АБ. Б — с вероятностью $аб^2$
БА. Б — с вероятностью $аб^2$
ББ. А — с вероятностью $аб^2$
ББ. Б — с вероятностью $б^3$

Для простоты будем учитывать только вырожденность по факультативным буквам, т. е. предположим, что если в данном полимере имеется несколько триплетов, кодирующих данную аминокислоту, то они отличаются только факультативной буквой, имея одинаковую облигатную пару. Хотя, как было показано уже в предыдущей главе, вырожденность имеет место не только по факультативным буквам, но и по облигатным парам, однако там же указывалось на то, что эта последняя вырожденность невелика. Поэтому для небольшой выборки — 8 слов из 64 возможных этим можно пренебречь. Будем сравнивать выходы двух аминокислот, выражая выход одной из них в процентах от их суммы. Для того чтобы всегда иметь дело с прямой зависимостью, пусть этой аминокислотой всегда будет та из двух, относительный выход которой (если он варьирует) возрастает с возрастанием *а*.

В табл. 41 приведены те сравнения диплетов, которые могут встретиться при сравнении выхода двух аминокислот, включение которых стимулируется поли-АБ. Как видно из таблицы, всего может встретиться 22 разных сравнения, которым соответствует всего 10 разных относительных выходов. Уже одно это само по себе вносит неоднозначность в проводимый анализ. Кроме того, как видно из рис. 16, где изображены все эти кривые, многие из них идут сходно. Заметим, что в проводившихся до сих пор анализах выход аминокислоты, триплет которой подлежит расшифровке, сравнивался с выходом фенилаланина или лизина, которым приписывались триплеты УУУ и ААА. При этом наиболее употребительными были полимеры, содержавшие 83% У или, соответственно, А. Этому соответствуют начальные точки на особенно трудно различимых S-образных кривых. Из рисунка ясно, что разница между соответствующими кривыми на этом участке больше тех ошибок, которые дает эксперимент. К этому нужно добавить, что настоящее рассмотрение проводилось без учета возможной вырожденности по облигатным парам. Учет ее еще более осложняет дело. Что же касается трехкомпонентных СПРН, то количество сравнений, возможных в опытах с ними, так возрастает, что подобный анализ для них вообще немыслим.

Однако количественный анализ опытов с СПРН даже при современной точности эксперимента вполне возможен, если применять другие методы, чем использовавшиеся до сих пор. Для

Таблица 41

Зависимость относительных выходов включения аминокислот от состава двухкомпонентных СПРН

Триплеты, кодирующие сравниваемые аминокислоты		Зависимость относительного выхода первой кислоты от содержания А и Б в СПРН	Тип кривой
Первую	Вторую		
3А	2А, Б	$a/a+б$	I
	А, 2Б	$a^2/a^2+б^2$	V
	3Б	$a^3/a^3+б^3$	VI
	АБ (А + Б)	$a^2/a^2+аб+б^2$	VII
	ББ (А + Б)	$a^3/a^3+аб^2+б^3$	X
2А,Б	2А, Б	0,5	II
	А, 2Б	$a/a+б$	I
	3Б	$a^2/a^2+б^2$	V
	АБ (А+Б)	$a/2a+б$	III
	ББ (А+Б)	$a^2/a^2+аб+б^2$	VII
АА (А+Б)	2А, Б	$a+б/a+2б$	IV
	А, 2Б	$a^2+аб/a^2+аб+б^2$	VIII
	3Б	$a^3+a^2б/a^3+a^2б+б^3$	IX
	АБ (А+Б)	$a/a+б$	I
	ББ (А+Б)	$a^2/a^2+б^2$	V
АБ (А+Б)	А, 2Б	$a+б/a+2б$	IV
	3Б	$a^2+аб/a^2+аб+б^2$	VIII
	БА (А+Б)	0,5	II
	ББ (А+Б)	$a/a+б$	I
А, 2Б	А, 2Б	0,5	II
	3Б	$a/a+б$	I
	ББ (А+Б)	$a/2a+б$	III

этого необходимо соблюдение двух условий. Первое — анализировать относительные выходы не по одной точке, а по кривой зависимости относительного выхода от состава полинуклеотида и второе — сравнивать относительные выходы по возможности для аминокислот, которым соответствуют одинаковые по составу облигатные пары (что возможно, так как последние были достаточно точно определены в предыдущей главе). В опыте с поли-АБ возможны при этом следующие четыре случая: 1) обе аминокислоты включаются при одинаковых факультативных буквах; выход при этом не будет меняться с изменением состава СПРН, т. е. получится кривая II типа; 2) одна аминокислота включается с факультативной буквой А, другая — с Б; относительный выход будет линейно возрастать от 0 до 100% (кривая I типа); 3) одна аминокислота включается с обеими факультативными буквами, вторая только с А; выход с увеличе-

нием А будет возрастать от 0 до 50% (кривая III типа); 4) одна аминокислота включается с А и Б, другая только с Б; с увеличением содержания А относительный выход будет возрастать от 50 до 100% (кривая IV типа). Эти четыре типа кривых, ко-

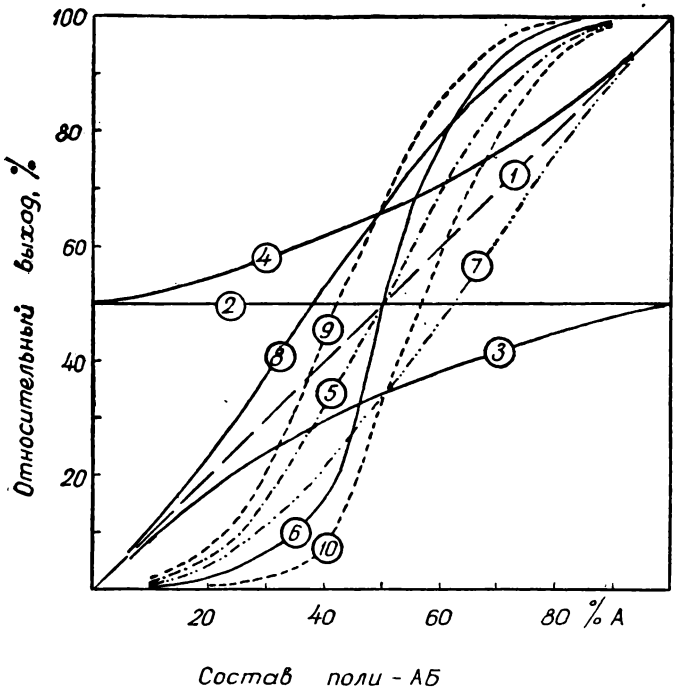

Рис. 16. Теоретические кривые для зависимости относительных выходов аминокислот от состава двухкомпонентных СПРН. Цифрами обозначены типы кривых в соответствии с их классификацией в табл. 41.

торые могут получиться при сравнении относительных выходов аминокислот с облигатными парами одинакового состава, настолько резко отличаются друг от друга, что сделать выбор между ними возможно даже в случае довольно большой ошибки опыта, если имеется несколько опытов с СПРН достаточно варьирующего состава.

Анализ результатов опытов с двухкомпонентными СПРН не может дать расшифровки всех триплетов, так как 24 триплета из возможных 64 состоят из букв трех разных сортов. Для их расшифровки необходимо использование результатов опытов с трехкомпонентными СПРН. Вероятно, наиболее простой, метод анализа таких опытов состоит в том, что после анализа опытов с двухкомпонентными СПРН и выяснения смысла большей части триплетов производится, исходя из состава СПРН,

вычисление ожидаемого выхода для известных триплетов, который сравнивается с наблюдаемым выходом. Из степени расхождения между вычисленным и ожидаемым выходом для отдельных аминокислот можно делать вывод о существовании еще неизвестных триплетов, состоящих из трех разных букв.

Если бы по всем СПРН имелись достаточно достоверные данные с достаточной вариацией соотношения нуклеотидов в полимерах, то только с помощью такого анализа можно было бы расшифровать смысл всех 64 триплетов. С теми же данными, которые имеются в настоящее время, придется привлекать и дополнительные материалы, полученные с помощью других методов.

АНАЛИЗ ДАННЫХ ОПЫТОВ С ДВУХКОМПОНЕНТНЫМИ ПОЛИМЕРАМИ

Опыты с поли-АУ. Результаты проведенных до сих пор разными авторами опытов с поли-АУ с разным отношением А : У приведены в табл. 42. Как видно из этой таблицы[1], по-

Таблица 42

Влияние поли-АУ на включение аминокислот

Аминокислоты	Облигатные пары	Процентное отношение А : У и литература					Триплеты
		83 : 17 [156]	20 : 80 [319]	17 : 83 [290]	17 : 83 [80]	13 : 87 [231]	
Аспарагин	АА	28	5,2	6,7	—	—	АА. У
Глицин	ГГ	0	—	0	—	4,9	—
Изолейцин	УА	20	16	20	19	12	УА. (А+У)
Лейцин	ЦУ+УУ	3,4	—	14	14	4,9	УУ. А
Лизин	АА	100	1,9	3,1	—	—	АА. А
Тирозин	АУ	3,2	17	25	26	13	АУ. У
Фенилаланин	УУ	0	100	100	100	100	УУ. У
Цистеин	ГУ	0	—	0	0	4,9	—
Метионин	УА	0	—	0	—	0	—

ли-АУ стимулирует включение всех аминокислот, имеющих облигатные пары состава АА, УУ и А, У, кроме метионина; очевидно, что метионин не включается соответствующим диплетом с А или У в качестве факультативной буквы. Кроме того, в одном из опытов было отмечено небольшое, но, по-видимому, реальное включение глицина и цистеина, которым по нашим данным соответствуют только облигатные пары, содержащие

[1] Здесь, так же как и дальше, используются только данные, полученные на бесклеточной системе из кишечной палочки. Появившиеся в последнее время данные на системах из клеток млекопитающих [75, 146, 232, 328] не принимаются во внимание, так как они не добавляют ничего нового и пока что малочисленны.

Г, и включение которых поэтому не должно было бы стимулироваться поли-АУ.

Будем сравнивать, как было только что сказано, относительные выходы аминокислот, кодовые слова которых имеют облигатные пары одинакового состава. Таковыми являются в данном случае пары аспарагин — лизин, фенилаланин — лейцин и тирозин — изолейцин. Зависимость относительного выхода аспарагина (в процентах суммы аспарагин + лизин) от процентного содержания У в поли-АУ изображена на рис. 17. На том же

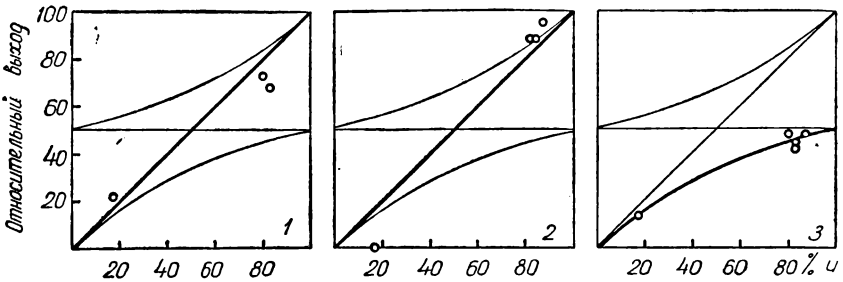

Рис. 17. Зависимость относительных выходов аминокислот от процентного содержания У в поли-АУ.

1 — выход аспарагина в процентах суммы асп-NH_2 + лиз; *2* — выход фенилаланина в процентах суммы фала + лей; *3* — выход тирозина в процентах суммы тир + илей.

рисунке изображены аналогичные зависимости для выхода фенилаланина (в процентах суммы фенилаланин + лейцин) и для тирозина (в процентах суммы тирозин + изолейцин). На этом рисунке проведены также все четыре теоретические кривые, которые вообще могут получиться при подобных сравнениях. Ход этих теоретических кривых настолько различен, что, даже несмотря на явно довольно значительные ошибки опыта, выбор кривых, лучше всего соответствующих экспериментальным точкам, не вызывает никаких сомнений. Для первых двух сравнений это — кривые I типа, для последнего — III типа (соответствующие кривые выделены на рисунке жирными линиями).

Кривые I типа соответствуют тому случаю, когда сравниваемые аминокислоты кодируются с разными факультативными буквами. Поскольку с увеличением содержания в полимере У возрастает относительное включение аспарагина и фенилаланина за счет лизина и лейцина, то значит, что аспарагин кодируется триплетом АА.У, лизин — АА.А, фенилаланин — УУ.У и лейцин — УУ.А. Кривая III типа получается в том случае, когда аминокислота, относительный выход которой возрастает, включается только с одной из факультативных букв, причем не с той, содержание которой возрастает в полимере; а вторая аминокислота включается с обеими буквами. Отсюда следует, что изолейцину соответствуют два триплета — УА. А и УА. У.

а тирозину один триплет — АУ. У. Незанятым остается один триплет — АУ.А. Поэтому аминокислоты, кроме уже рассмотренных, включение которых также наблюдалось в одном из опытов, должны были бы иметь этот триплет, если их включение не является артефактом. В таком случае их выход должен был бы быть наименьшим при наибольшем содержании в полимере У, т. е. в том единственном опыте, где они показали включение. Поэтому, если полимеры с большим содержанием А вообще не дали включения глицина и цистеина, можно считать, что их включение в одном из опытов связано с какими-то случайными причинами.

Анализ опытов с поли-АУ мы описали довольно подробно, чтобы проиллюстрировать применение предложенных методов. Общий ход рассуждений при анализе опытов с другими двухкомпонентными полимерами совершенно такой же.

Рассмотрим вопрос о порядке букв в триплетах. Поскольку порядок букв в облигатных парах уже известен, задача сводится к определению положения факультативной третьей по отношению к буквам облигатной пары. В табл. 43 выписаны три

Таблица 43

Ожидаемые при разном положении факультативной буквы и наблюдаемые включения аминокислот в присутствии АУУУУ ...+ААУУУ...

Аминокислоты	Порядок букв в триплетах для трех возможных гипотез и ожидаемое для них включение			Наблюдаемое включение
	I	II	III	
Аспарагин	уАА —	АуА —	ААу +	—
Изолейцин	аУА — уУА —	УаА — УуА —	УАа — УАу —	—
Лейцин	аУУ +	УаУ —	УУа —	Не исследовано
Лизин	аАА —	АаА —	ААа —	—
Тирозин	уАУ —	АуУ +	АУу +	+
Фенилаланин	уУУ +	УуУ +	УУу +	+

возможных варианта для порядков букв в триплетах аминокислот, включаемых поли-АУ, причем облигатная пара обозначена большими буквами, а факультативная третья — малой. Выбор между этими тремя вариантами может быть сделан на основании уже использовавшихся результатов опытов А. Ваба и сотрудников [319]. Напомним, что в этих опытах исследовали включение аминокислот в присутствии смеси полимеров состава АУУУУ... и ААУУУ... В таблице знаками плюс и минус показано, в каких случаях следует ожидать стимуляции включения тех или иных аминокислот. В последней графе таблицы приведены результаты опытов. Из сравнения ожидаемых и наблюда-

емых включений кажется наиболее вероятным, что факультативная буква стоит на втором месте. Однако, если ее положение на первом месте можно смело исключить, так как при этом не должно было бы наблюдаться включение тирозина, которое стимулировалось в опытах, то этого нельзя с полной уверенностью сделать для предположения о том, что факультативная буква является третьей. В последнем случае следовало бы ожидать стимуляции включения, наряду с тирозином, также аспарагина. Однако отсутствие включения аспарагина в опыте не исключает того, что он может включаться. Отрицательный результат мог бы объясняться в случае справедливости третьего варианта, например, тем, что обнаружение включения тирозина — предел чувствительности примененного метода, а если в смеси полимеров ААУУУ... было меньше, чем АУУУУ..., то выход аспарагина лежал бы ниже этого предела. Поэтому, хотя мы должны считать за наиболее правдоподобное положение факультативной буквы на втором месте, достоверным этот вывод считать нельзя. Помня об этом, мы, однако, будем и дальше продолжать ставить факультативную букву на третье место, отделяя ее точкой, так как это значительно удобнее.

Опыты с поли-УЦ. Результаты опытов с поли-УЦ приведены в табл. 44, а результаты их анализа на рис. 18. Опыты

Таблица 44

Влияние поли-УЦ на включение аминокислот

Белок	Облигатные пары	Процентное отношение У : Ц и литература			Кодирующие триплеты
		82 : 18 [290]	75 : 25 [289]	39 : 61 [231]	
Лейцин	ЦУ+УУ	20	25	79	ЦУ. У
Пролин	ЦЦ	7,7	—	285	ЦЦ. (У+Ц)
Серин	УЦ+ГА	25	40	160	УЦ. (У+Ц)
Фенилаланин	УУ	100	100	100	УУ. (У+Ц)
Треонин	ЦЦ+АЦ	0	—	—	—

с поли-УЦ, несмотря на небольшое их число, дали очень ясные результаты и потому нуждаются лишь в немногих замечаниях. Лейцину соответствуют две облигатные пары — ЦУ и УУ. Это могло бы осложнить анализ, так как в поли-УЦ присутствуют обе эти пары. Однако хорошее соответствие экспериментальных точек теоретической кривой IV типа при сравнении выходов серина и лейцина делает вероятным, что следующий отсюда для лейцина триплет ЦУ, У является единственным в этом полимере, а серин включается при обеих факультативных буквах. Отсутствие в данном полимере соответствующих лейцину слов с диплетом УУ подтверждается также сравнением выходов

серина и фенилаланина, давшим кривую I типа, из чего следует, что фенилаланин кодируется триплетами УУ.У и УУ.Ц. Точно так же, из сравнения выходов серина и пролина следует, что пролин кодируется триплетами ЦЦ.У и ЦЦ.Ц. В двух последних случаях нам пришлось отступить от правила сравнивать выходы

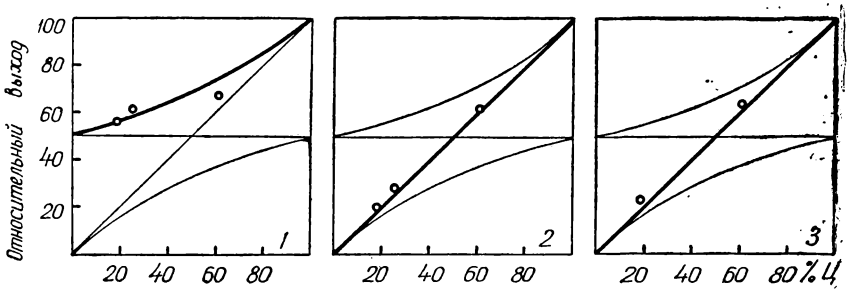

Рис. 18. Зависимость относительных выходов аминокислот от процентного содержания Ц в поли-УЦ.

1 — выход серина в процентах суммы сер + лей; *2* — выход серина в процентах суммы сер + фала; *3* — выход пролина в процентах суммы сер + про.

аминокислот, которым соответствуют факультативные пары одинакового состава. Однако в этих случаях у сравниваемых аминокислот тождественны факультативные буквы, и мы не выходим здесь за рамки сравнений четырех первых типов кривых. В связи с тем, что, кроме рассмотренных четырех аминокислот, больше ни одна не включается в белки в присутствии поли-УЦ, и так как наблюдаемые выходы хорошо соответствуют теоретическим, оставшийся незаполненным триплет ЦУ.Ц, по-видимому, является нонсенсом или точкой.

Опыты с поли-АЦ. Как видно из табл. 45, с поли-АЦ было поставлено больше всего опытов. Поэтому анализ их ре-

Таблица 45

Влияние поли-АЦ на включение аминокислот

Аминокислоты	Облигатные пары	Процентное отношение А : Ц и литература									Кодирующие триплеты
		3 : 97 [201]	6 : 94 [201]	12 : 88 [201]	15 : 85 [102]	30 : 70 [201]	37 : 63 [102]	67 : 33 [201]	80 : 20 [201]	83 : 17 [156]	
Аспарагин .	АА	0	0	1,4	0	4,6	0	17	12	30	АА. Ц
Гистидин . . .	АЦ	0	0	0	13	18	31	9,4	3,9	9,1	АЦ. Ц
Глютамин .	АЦ+ЦГ	0	1,7	3,7	4,3	4,8	6,8	15	12	44	АЦ. А
Лизин	АА	0	0	2,2	0	4,3	0	100	100	100	АА. А
Пролин	ЦЦ	100	100	100	100	100	100	15	4,5	5,4	ЦЦ. (А+Ц)
Серин	УЦ+ГА	1,8	1,7	2,8	0	1,5	0	1,0	1,8	0	—
Треонин . . .	ЦА+ЦЦ	4,8	5,7	12	36	15	63	19	24	23	ЦА. (А+Ц)

зультатов можно провести вполне однозначно. Данные этого анализа представлены на рис. 19. Сравнение выходов глютамина и гистидина, давшее кривую I типа, показывает, что глютамин в данном полимере кодируется триплетом АЦ.А, а гистидин — АЦ.Ц. Сравнения выходов пролина и гистидина (так как вероятность триплета АЦ.Ц равна вероятности триплета ЦЦ.А) дает

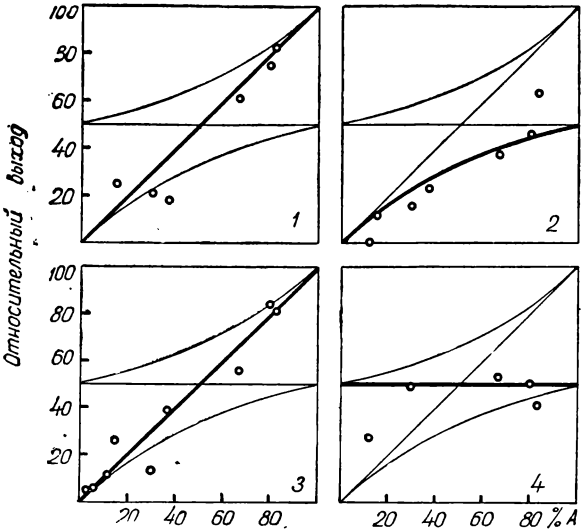

Рис. 19. Зависимость относительных выходов аминокислот от процентного содержания А в поли-АЦ:
1 — выход глютамина в процентах суммы глю-NH_2+гис; *2* — выход гистидина в процентах суммы гис + про; *3* — выход треонина в процентах суммы тре + про; *4* — выход аспарагина в процентах суммы асп-NH_2 + глю-NH_2.

кривую III типа, следовательно, пролин включается парой ЦЦ с обеими факультативными буквами. Из того, что сравнение треонин — пролин дает кривую I, а аспарагин — глютамин — II типа, в то время как триплет АА.А уже приписан лизину, следует расшифровка всех остальных триплетов, которая приведена в последней колонке табл. 45. Очень слабое включение серина в некоторых опытах с поли-АЦ следует признать артефактом, тем более, что все 8 триплетов, встречающиеся в этом полимере, оказываются заняты другими аминокислотами.

Опыты с поли-УГ. Хотя, как видно из табл. 46, с поли-УГ было поставлено довольно много опытов, большинство полимеров содержало низкие количества гуанина и далеко не всегда было исследовано включение всех аминокислот. Поэтому здесь анализ провести труднее и выводы получаются не столь уверенные, как раньше. То же относится и к другим СПРН, содержащим гуанин, которые рассматриваются ниже. Это связано с тем, что полимеры, содержащие высокий процент гуани-

Таблица 46

Влияние поли-УГ на включение аминокислот

Аминокислота	Облигатные пары	Процентное отношение У : Г и литература							Кодирующие триплеты
		87 : 13 [277]	84 : 16 [290]	83 : 17 [80]	77 : 23 [277]	76 : 24 [231]	62 : 38 [277]	37 : 63 [277]	
Валин	УГ	32	20	23	61	37	99	406	УГ(Г+У)
Глицин	ГГ	—	4,2	4,4	—	12	—	—	ГГ. У (+ГГ.Г?)
Лейцин	ЦУ+УУ	30	12	26	49	36	80	233	УУ. Г
Триптофан . . .	Г, У	4,2	5,0	4,4	12	14	26	67	ГУ. Г
Фенилаланин . .	УУ	100	100	100	100	100	100	100	УУ. У
Цистеин	ГУ	—	20	13	—	35	—	—	ГУ. У
Аланин	ЦГ+ГГ	—	0	—	—	0	—	—	—

на, образуют вторичную структуру, делая тем самым матрицу недоступной для контакта с т-РНК [157, 277]. Полимеры с высо-

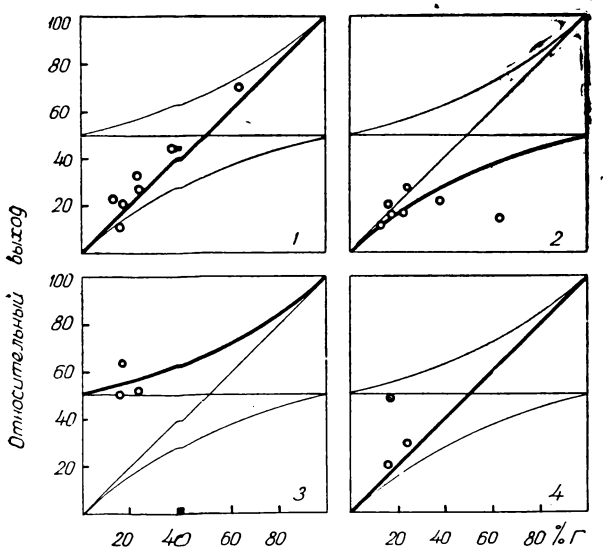

Рис. 20. Зависимость относительных выходов аминокислот от процентного содержания Г в поли-УГ:

1 — выход лейцина в процентах суммы лей + фала; *2* — выход триптофана в процентах суммы трип + вал; *3* — выход валина в процентах суммы вал + цис; *4* — выход триптофана в процентах суммы трип + цис.

ким содержанием Г дают всегда количественно низкие выходы и изменчивые результаты. Анализ данных из табл. 46 представлен на рис. 20.

Сравнение выходов лейцина и фенилаланина основано на достаточно большом числе точек, хорошо следующих кривой I типа, из чего следует, что фенилаланин кодируется триплетом УУ.У, а лейцин — УУ.Г. Что же касается сравнения выходов аминокислот, которым соответствуют облигатные пары состава Г, У, то они дают значительно менее ясные результаты. Наиболее правдоподобна интерпретация, согласно которой сравнение триптофан — валин дает кривую III типа, валин — цистеин — IV, а триптофан — цистеин — I. Но можно приписывать им также, соответственно, кривые I, II и II типов. Однако в пользу правильности первой интерпретации говорят два обстоятельства. Во-первых, для этой интерпретации сумма квадратов отклонений оказывается наименьшей. Во-вторых, это единственная интерпретация, приводящая к непротиворечивой расшифровке триплетов; если принять любую другую интерпретацию, то разные сравнения приводят для одних и тех же аминокислот к разным триплетам. Поэтому делаем заключение, что цистеин включается с факультативной буквой У, триптофан — с Г, а валин — с обеими. До сих пор порядок букв в облигатной паре, соответствующей триптофану, оставался неизвестным. Однако, поскольку валин кодируется обоими триплетами с облигатной парой УГ, то для триптофана остается только порядок ГУ — тот же, что и для цистеина. Для глицина дать однозначную расшифровку не удается, так как его включение исследовано только в трех случаях в опытах с полимерами сходного состава, а сравнение возможно только с триптофаном. Так как выходы у этих двух аминокислот довольно близки, то значит, что глицину соответствует или триплет ГГ.У, или два триплета — ГГ.У и ГГ.Г.

Опыты с поли-ГЦ. Удовлетворительные данные по поли-ГЦ имеются только для полимеров, содержащих не более 18% гуанина. Опыт с поли-ГЦ, содержавшим 91% Г, дал выходы, не отличающиеся достоверно от нуля. Однако среди полимеров более сложного состава имеются два, содержащие другие нуклеотиды, кроме Г и Ц, в очень низком проценте — поли-АЦГ (4 : 77 : 19) и поли-УГАЦ (3 : 45 : 9 : 43). Эти данные мы также включили в табл. 47, где приведены результаты всех опытов с поли-ГЦ. Результаты анализа даны на рис. 21. Сравнение выходов аминокислот, имеющих облигатные пары ЦЦ, — пролина и треонина — дает зависимость, наиболее близкую к кривой III типа. Однако такая кривая невозможна, так как в этом случае один из триплетов серина должен был бы быть идентичен триплету треонина, поэтому, если включение треонина не является артефактом, он должен кодироваться триплетом ЦЦ.Г, а пролин — ЦЦ.Ц. Несмотря на значительное отклонение третьей точки от кривой I типа, мы примем такую интерпретацию, так как это хорошо соответствует наблюдавшемуся не-

Таблица 47

Влияние поли-ГЦ на включение аминокислот

Аминокислота	Облигатные пары	Процентное отношение Ц:Г и литература							Кодирующие триплеты
		90:10 [201]	87:13 [201]	83:17 [320]	82:18 [201]	80:20* [201]	49:51* [201]	9:91** [201]	
Аланин	ЦГ+ГГ	21	9,9	24	18	30	50	0	ЦГ (Г+Ц)
Аргинин	ГЦ+ГА	27	7,9	19	24	22	54	1	ГЦ (Г+Ц)
Глицин	ГГ	3,5	0	5,2	2,2	0	20	0	ГГ. Ц или (и) ГГ. Г
Пролин	ЦЦ	100	100	100	100	100	100	2	ЦЦ. Ц
Серин	УЦ+ГА	13	0	0	1,7	16	145	0	—
Тирозин	АУ	1,4	3,0	0	5,6	0	8,8	0	—
Треонин	ЦА+ЦЦ	0	0	17	0,3	7,0	15	0	ЦЦ. Г
Триптофан . . .	Г, У	0	0,5	0,7	4,8	0	13	1	—
Фенилаланин . .	УУ	3,5	2,5	0	1,1	4,3	69	1	—
Глютамин	АЦ+ЦГ	0	0	0	0	0	0	0	—

* Из полимеров более сложного состава (см. в тексте)
** Приведен абсолютный выход в *ммкМ*.

сколько раз при действии азотистой кислоты замещению тре → сер, которое иначе необъяснимо.

Сравнение выходов аргинина и аланина дает кривую II, а аргинина и треонина — IV типа. Отсюда следует, что аланин

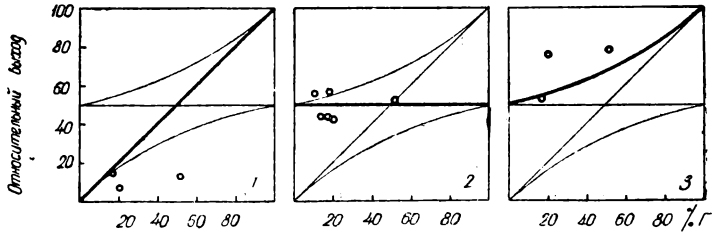

Рис. 21. Зависимость относительных выходов аминокислот от процентного содержания Г в поли-ГЦ:

1 — выход треонина в процентах суммы тре + про; *2* — выход аргинина в процентах суммы арг + ала; *3* — выход аргинина в процентах суммы арг + тре.

и аргинин кодируются с обеими факультативными буквами. С расшифровкой триплетов для глицина и здесь возникают трудности в связи с малым выходом, большой изменчивостью результатов и отсутствием подходящего партнера для однозначного сравнения. Имеющиеся данные не исключают ни одной из трех возможностей. Может быть, что глицин кодируется только ГГ.Г, или только ГГ.Ц, или обоими этими триплетами.

Опыты с поли-АГ. Результаты опытов с поли-АГ разного состава приведены в табл. 48. Всего было исследовано

Таблица 48

Влияние поли-АГ на включение аминокислот

Аминокислота	Облигатные пары	Процентное отношение А : Г и литература			Кодирующие триплеты
		83 : 17 [156]	73 : 27 [201]	48 : 52* [201]	
Аргинин	ГЦ+ГА	14	9,1	100	ГА. А
Глицин	ГГ	5,4	1,8	0	ГГ. А (+ ГГ. Г?)
Глютаминовая к-та	АГ	30	40	63	АГ (А+Г)
Глютамин	АЦ+ГЦ	5,9			—
Лизин	АА	100	100	100	АА (А+Г)
Асп*x*	АГ+АА	0	11	125	—
Аланин	ЦГ+ГГ	0	0	0	—
Серин	УЦ+ГА	0	0	0	—

* Абсолютные выходы в этом СПРН очень низкие.

только три полимера, причем в последнем абсолютные значения выходов были очень низкими, в связи с чем цифры, конечно, имеют довольно большую ошибку. На рис. 22 изображены ре-

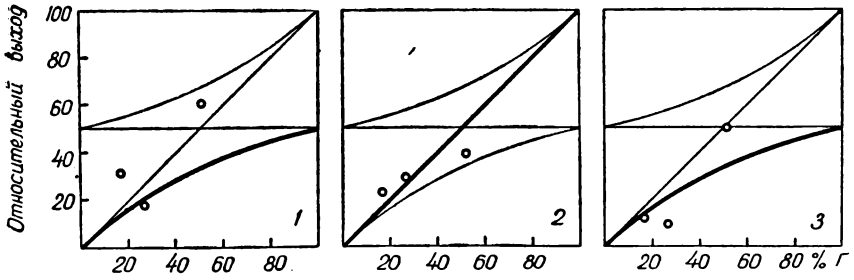

Рис. 22. Зависимость относительных выходов аминокислот от процентного содержания Г в поли-АГ.

1 — выход аргинина в процентах суммы арг + глют; *2* — выход глютаминовой кислоты в процентах суммы глют + лиз; *3* — выход аргинина в процентах суммы арг + лиз.

зультаты сравнений относительных выходов для пар аргинин — глютаминовая кислота, глютаминовая кислота — лизин и аргинин — лизин. Как видно из этого рисунка, экспериментальные точки дают большое рассеяние, и в каждом из трех случаев им можно примерно с одинаковым основанием соподчинить кривую I или III типа, тогда как кривые II и IV типа им явно не могут соответствовать.

Для расшифровки значения триплетов, входящих в состав поли-АГ, применим тот же прием, который уже был использован при анализе опытов с поли-УГ. Попытаемся произвести

расшифровку для всех восьми правдоподобных случаев (т. е. считая, что все кривые I типа, или, что первая — III, остальные две — I, или что вторая — III типа, остальные — I и т. д.). В семи случаях из восьми результаты расшифровки оказываются противоречивыми и только в одном полностью соответствуют друг другу, а именно при предположении, что в первом случае имеет место кривая III типа, во втором — I типа и в третьем — III типа. Из этого следует, что аргинин кодируется триплетом ГА.А, лизин — АА.А и АА.Г, глютаминовая кислота — АГ.А и АГ.Г. Триплеты, соответствующие глицину, и здесь оказывается труднее всего расшифровать. Анализ данных по поли-АГ состава 83 : 17 показывает, что относительный выход для глицина близок к ожидаемому для триплета ГГ.А, что, однако, не исключает возможности того, что наряду с этим глицин кодируется и триплетом ГГ.Г, так как ожидаемый выход для этого триплета очень мал.

Таким образом удалось определить кодовое значение почти всех триплетов, состоящих из букв двух сортов. Как видно из рисунков 17—22, соподчинение экспериментальных точек определенной кривой большей частью не встречает затруднений, а возникающие сомнения носят не принципиальный характер, будучи связаны с отсутствием достаточного материала. Поэтому можно считать, что большинство сделанных выводов правильно, а после получения дополнительных экспериментальных данных все триплеты, состоящие из букв двух сортов, можно будет с помощью примененного метода расшифровать вполне однозначно.

Хочется сказать пару слов о методах количественного анализа результатов опытов с СПРН, применявшихся раньше. Из рисунков ясно видна, по степени отклонения точек от теоретических кривых, величина ошибок, которую несут почти все произведенные до сих пор определения, что находится в согласии с утверждением, сделанным в начале этой главы о том, что количественный анализ на основании одной точки является крайне ненадежным.

Значение триплетов, состоящих из букв трех разных сортов, может быть определено с помощью анализа результатов опытов с трехкомпонентными СПРН. Однако, поскольку мы этот анализ будем производить путем сравнения наблюдаемых выходов с ожидаемыми для уже известных триплетов, важно, чтобы до этого было определено возможно большее число триплетов.

Три триплета, состоящие из букв трех разных сортов, уже можно определить на основании качественного анализа результатов опытов с СПРН (см. выше табл. 34 и 35) — УА.Г для метионина, а также АГ.Ц и АГ.У для аспарагиновой кислоты. Ими можно дополнить список триплетов, определенных в результате анализа двухкомпонентных СПРН.

Кроме того, ряд триплетов может быть определен путем анализа замещений аминокислот. Поэтому сейчас мы займемся анализом замещений, после чего, зная смысл бо́льшего числа триплетов, перейдем к рассмотрению результатов, полученных с трехкомпонентными СПРН.

АНАЛИЗ ДАННЫХ ПО ЗАМЕЩЕНИЯМ

На рис. 23 изображена трехмерная решетка Зьюби — Квастлера со всеми расшифрованными выше триплетами. Для

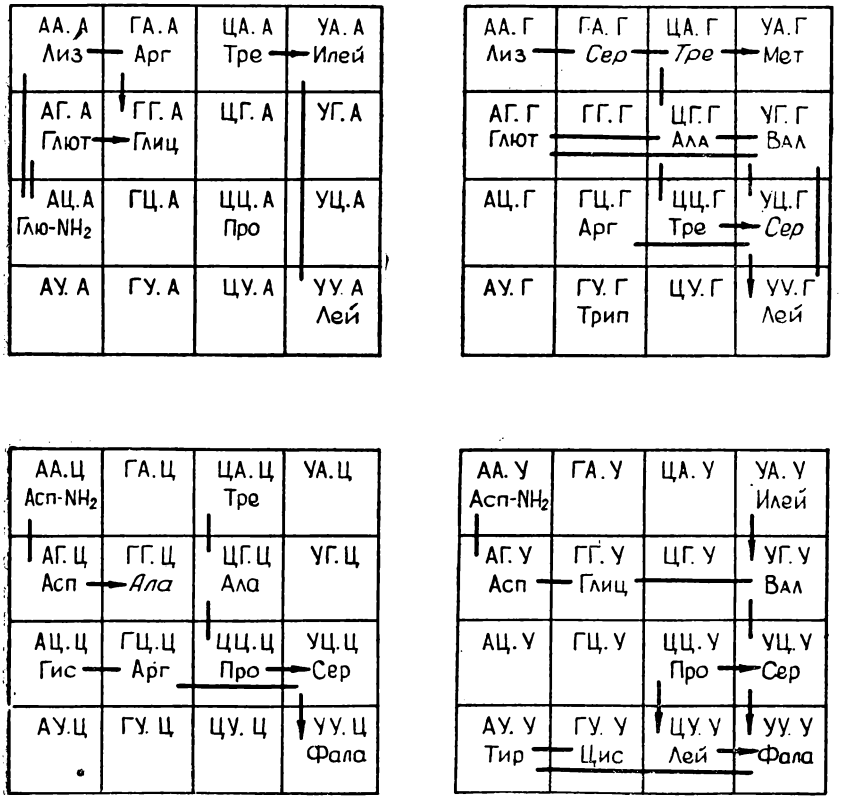

Рис. 23. Расшифрованные триплеты и объясняемые ими замещения (из рис. 13 и табл. 49). Стрелками показаны нитритные замещения в белке ВТМ. На рисунке отдельно представлены четыре плоскости, которые, при расположении их друг над другом, дают трехмерную решетку.

большей ясности она изображена не в виде куба, а разложена на отдельные плоскости. В каждой из плоскостей представлены все возможные облигатные пары, которые расположены во всех плоскостях одинаково. Плоскости отличаются друг от друга

факультативными третьими буквами. Для рассмотрения переходов в связи с изменением факультативных букв (что понадобится сравнительно редко) следует представить изображенные четыре квадрата лежащими друг над другом.

Рассмотрим вначале замещения, вызванные азотистой кислотой. Большинство их (10 из 16) объясняются уже известными триплетами. Остальные шесть могут служить для расшифровки еще неизвестных. Таковыми являются: асп → ала, асп-NH_2 → → сер, глю-NH_2 → вал, тре → мет, тре → сер и сер → лей. Первое из этих замещений объясняется только в том случае, если аланин кодируется триплетом ГГ.Ц, так как аспарагиновая кислота имеет только два триплета — АГ.Ц и АГ.У, которые могут под влиянием азотистой кислоты перейти, соответственно, в ГГ.Ц и ГГ.У, но последний уже занят глицином.

Следующие два замещения пока не могут быть однозначно расшифрованы. Для их объяснения серину можно приписать триплеты ГА.Ц или ГА.У, а глютамину — ЦГ.А или ЦГ.У. Поскольку метионин кодируется единственным триплетом УА.Г, то замещение тре → мет может быть объяснено только в том случае, если треонин кодируется триплетом ЦА.Г. Точно так же для замещения тре → сер остается, если не приписывать серину третьей облигатной пары, одна возможность считать, что серин кодируется триплетом УЦ.Г. После этого само собой объясняется и последнее замещение сер → лей. На рисунке изображены все нитритные замещения, кроме замещения илей → → мет, которое связано с заменой третьей буквы УА.А → УА.Г и соответствует переходу из одной плоскости в другую. Триплеты, расшифрованные на основании анализа нитритных мутаций, выделены курсивом.

Известные аминокислотные «множественные аллели» пока не могут ничего к этому добавить. Данные по мутациям и рекомбинациям в белке А могут быть помещены как в квадрат «А», так и в квадрат «Г». Что же касается одной из групп, обнаруживаемых среди мутаций гемоглобина, то она полностью объясняется уже известными триплетами, тогда как вторая (связывающая глютамин, тирозин и аргинин) может быть помещена в разные квадраты.

Обратимся к прочим замещениям (кроме вызванных азотистой кислотой). Все наблюдавшиеся замещения приведены в табл. 49. Они разбиты на 5 групп в связи с предполагаемой степенью их достоверности. В первую группу объединены замещения, наблюдавшиеся более одного раза в опытах по вызыванию мутаций, ко второй отнесены замещения, наблюдавшиеся один раз среди мутаций и, кроме того, наблюдавшиеся при сравнении естественных белков, третью составили однократно наблюдавшиеся мутации, не встреченные в другом материале, четвертую — замещения, наблюдавшиеся более одного раза при

Таблица 49

Замещения аминокислот, кроме вызванных действием азотистой кислоты

Группа	Замещение	Число наблюдавшихся случаев	Индекс
1	ала—глиц	2+7	0
	арг—сер	2+2	0,96
	асп—лиз	2	0,93
	асп-NH_2—лиз	2	0,23
	вал—глют	3+2	0
	гис—тир	2	0
	глют—глю-NH_2	2+2	0,04
	глю-NH_2—лиз	2	0,58
2	ала—глют	1+3	0,78
	арг—гис	1+2	0,50
	арг—лиз	1+5	0,01
	вал—глиц	1+2	1,26
	вал—сер	1+1	2,40
3	арг—асп-NH_2	1	2,46
	илей—лей	1	1,80
	тир—фала	1	0,43
	тир—цис	1	0,07
4	ала—вал	4	0,61
	ала—гис	2	1,80
	ала—про	3	1,07
	ала—сер	6	0,12
	ала—тре	7	0,25
	арг—лей	2	0,85
	асп—асп-NH_2	2	0,58
	асп—глют	4	0,17
	вал—лей	2	0,58
	гис—глю-NH_2	2	0,45
4	глиц—сер	4	1,16
	глют—про	2	1,10
	лиз—сер	3	0,80
5	ала—арг	1	4,85
	ала—илей	1	1,57
	ала—лей	1	2,54
	ала—фала	1	1,57
	арг—вал	1	3,60
	асп—илей	1	1,33
	асп—сер	1	3,80
	асп-NH_2—глиц	1	2,74
	асп-NH_2—лей	1	0,36
	вал—мет	1	1,16
	вал—тир	1	1,43
	гис—илей	1	0,90
	гис—фала	1	0,90
	глиц—лей	1	2,34
	глиц—тре	1	3,79
	глиц—фала	1	2,74
	глют—лиз	1	2,54
	глют—сер	1	4,17
	глют—тре	1	3,02
	глют—трип	1	1,42
	глю-NH_2—тре	1	2,25
	илей—фала	1	0,65
	лей—лиз	1	0,73
	лей—мет	1	0,03
	лиз—тре	1	2,06
	лиз—трип	1	0,58
	тре—трип	1	1,07
	тре—цис	1	1,07

Примечание: В двух первых группах первая цифра—замещения, наблюдавшиеся в мутационных опытах, вторая—при сравнении естественных белков; в третьей группе—только мутационные данные, в четвертой и пятой—только данные по естественным белкам.

сравнении естественных белков и, наконец, в пятую — встреченные при сравнении естественных белков однократно. Для каждого замещения приведено значение соответствующего ему индекса, являющегося мерой отклонения наблюдаемой частоты

от ожидаемой при предположении, что замещение является разрешенным (ср. выше, табл. 19).

При рассмотрении таблицы бросается в глаза большое, в среднем, значение индексов для последней пятой группы, по сравнению с остальными. Это особенно ясно видно на рис. 24, где изображены вариационные кривые для индексов первых четырех групп и пятой. Отсюда следует, что замещения пятой

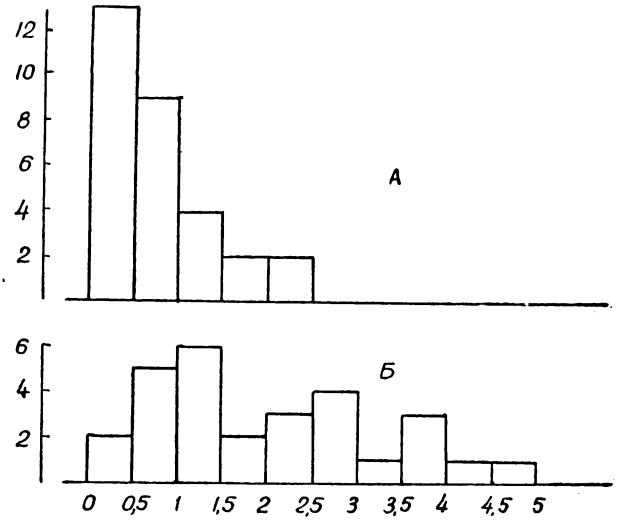

Рис. 24. Распределения нормированных квадратов отклонений для замещений, приведенных в табл. 49.
А — замещения первой — четвертой групп; Б — замещения пятой группы.

группы в значительной своей части являются запрещенными и не должны использоваться при анализе. Подавляющее большинство замещений из четырех первых групп объясняется уже известными триплетами. Не объясняются девять замещений. Из них четыре (ала — гис, арг — лей, асп — лиз и глют — про) противоречат даже облигатным парам, поэтому мы их будем считать запрещенными. Впрочем, если в двух последних из них идет речь не об аспарагиновой и глютаминовой кислотах, а об аспарагине и глютамине, то они станут разрешенными. Остальные замещения, не объясняющиеся известными триплетами, — гис — тир, арг — асп-NH_2, ала — сер, глиц — сер, лиз — сер. Из них только последнее позволяет найти новый триплет, так как только оно может быть объяснено единственным образом, а именно при принятии для серина триплета ГА.Г. Этот триплет также показан на рис. 23 курсивом.

АНАЛИЗ ДАННЫХ ОПЫТОВ С ТРЕХКОМПОНЕНТНЫМИ ПОЛИМЕРАМИ

После того как исчерпаны остальные возможности, перейдем к анализу результатов опытов с трехкомпонентными СПРН. При этом будем учитывать также данные по тем замещениям, которые до сих пор не были объяснены известными триплетами.

Опыты с поли-АЦГ. Из всех трехкомпонентных СПРН наибольший материал имеется по поли-АЦГ, поэтому с него и целесообразнее всего начать анализ. В табл. 50 приведены данные, полученные с поли-АЦГ разными авторами. Сравним наблюдаемые для отдельных аминокислот выходы с теоретиче-

Таблица 50

Влияние поли-АЦГ на включение аминокислот

Аминокислота	Процентное отношение А:Ц:Г и литература						Качественный результат
	75:13:13 [320]	67:17:17 [320]	46:32:22 [201]	2:89:9 [201]	4:77:19 [201]	16:72:12 [201]	
Аланин	2,3	5,9	15	14	30	207	+
Аргинин	12	27	16	9,4	22	180	+
Аспарагиновая к-та	2,0	3,8	20	0	0	59	+
Аспарагин	14	17					+
Валин	0	0	0	0	0	0	—
Гистидин	11	13	8,7	1,9	4,9	232	+
Глицин	1,0	3,4	0,6	0	0	0	(+)
Глютаминовая к-та	16	21	40	0	0	81	+
Глютамин	22	30					+
Изолейцин	0	0	0	0	0	0	—
Лейцин	0	0	0	0	0	0	—
Лизин	100	100	100	1,6	0	56	+
Метионин	0	0	0	0	0	0	—
Пролин	2,8	8,8	18	100	100	100	+
Серин	3,5	7,6	22	7,5	16	134	+
Тирозин	0	0	0	0	0	0	—
Треонин	20	24	31	3,4	7,0	27	+
Триптофан	0	0	0	0	0	0	—
Фенилаланин	0	0	0	0	0	0	—
Цистеин	0	0	—	—	—	—	—

ски ожидаемыми для уже известных триплетов. В табл. 51 произведено такое сравнение для двух опытов. Для бо́льшей точности здесь цифры приведены не в процентах от включения лизина, как в табл. 50, а в миллимикромолях аминокислоты на 1 *мг* рибосомного белка. Только для четырех аминокислот (аланин глютамин, гистидин и серин) наблюдаемый выход оказывается больше вычисленного в обоих опытах, причем для аланина

Таблица 51

Сравнение вычисленных и наблюдаемых выходов аминокислот в опытах с поли-АГЦ

Аминокислота	Известные триплеты	Опыт с поли-АГЦ (4:1:1)			Опыт с поли-АГЦ (6:1:1)		
		вычисленный выход	наблюдаемый выход	разница	вычисленный выход	наблюдаемый выход	разница
Аланин	ГГ. Ц+ГЦ. Г	81	121	—40	36	66	—30
Аргинин	ГЦ (Ц+Г)+ +АГ. А	464	550	—86	444	350	+94
Аспарагиновая к-та	АГ. Ц	102	78	+24	70	59	+11
Аспарагин	АА. Ц	410	354	+56	420	389	+31
Глютаминовая к-та	АГ (А+Г)	512	431	+181	490	473	+17
Глютамин	АЦ. А	410	620	—210	420	636	—216
Глицин	ГГ. А	102	69	+33	70	30	+40
Гистидин	АЦ. Ц	102	261	—159	70	320	—250
Лизин	АА (А+Г)	1920	2060	—140	2940	2920	+20
Пролин	ЦЦ (А+Ц)	129	182	—53	82	82	0
Серин	АГ. Г	102	156	—54	70	103	—33
Треонин	АЦ (А+Г+Ц)+ +ЦЦ. Г	641	494	+147	672	568	+104

Примечание. Вычисленные в соответствии с отношением А:Г:Ц выходы для триплетов, содержащих 3,2, 1 и 0 А, равны для первого опыта, соответственно, 1610, 410, 102 и 27, а для второго—2520, 420, 70 и 12.

отклонение оказывается наименьшим. Поли-АГЦ содержит 27 триплетов — все, не имеющие в своем составе У. В вычислении ожидаемых выходов фигурировал 21 триплет, следовательно, опыты с поли-АЦГ могут, теоретически рассуждая, дать сведения о пяти новых триплетах — ГГ.Г, ЦГ.А, ГЦ.А, АЦ.Г и ГА.Ц. Однако для первого из них ожидается слишком низкий выход (12 и 27), чтобы можно было при современной точности опытов идентифицировать соответствующую ему аминокислоту.

Что же касается остальных триплетов, то для трех из них первые буквы как раз соответствуют облигатным парам для тех аминокислот, для которых наблюдались постоянные расхождения между вычисленным и наблюдаемым выходом. Совершенно очевидно соподчинение триплета ГА.Ц серину. Что же касается остальных двух, то АЦ.Г мог бы принадлежать либо гистидину, либо аспарагину, а ЦГ.А глютамину или аланину. Однако, поскольку гистидину не может принадлежать другой триплет, кроме АЦ.Г, а для аланина расхождение между вычисленным и наблюдаемым отклонением наименьшее, а также учитывая, что до сих пор остается необъясненным нитритное замещение глю-NH_2 → вал, мы считаем наиболее правдоподобным соподчи-

нение гистидину АЦ.Г, а глютамину — ЦГ.А. Что касается триплета ГЦ.А, то для соподчинения его какой-либо аминокислоте нет достаточных оснований.

После сделанных соподчинений обратимся к еще не объясненным замещениям. Так как триплет ЦГ.А теперь занят глютамином, то данные по замещениям в белке А могут быть объяснены только в рамках квадрата, соответствующего факультативной третьей «Г», в результате чего глицин получает, наконец, триплет ГГ.Г, в отношении которого было много сомнений при анализе данных опытов с СПРН. Из двух остававшихся нитритных замещений асп-NH_2→ сер объясняется теперь само собой, тогда как для объяснения замещения глю-NH_2 → → вал нужно дать валину триплет УГ.А (не вполне исключенная возможность объяснения этого замещения в рамках У-квадрата мало вероятна, в связи с безуспешностью поисков для глютамина урацилсодержащих триплетов [290]). Из необъясненных ранее прочих замещений получают теперь естественное

Таблица 52

Влияние поли-УАЦ, поли-УЦГ и поли-УАГ на включение аминокислот

Аминокислота	Состав и процентное отношение, и литература								
	У : А : Ц			У : Ц : Г			У : А : Г		
	74 : 15 : 11 [290]	83 : 5 : 12 [231]	Качественный результат	79 : 8 : 13 [290]	34 : 50 : 15 [231]	Качественный результат	72 : 16 : 12 [290]	68 : 3 : 29 [231]	Качественный результат
Аланин	0	(1,0)	—	3,2	40	+	0	(0,9)	—
Аргинин	0	0	—	3,3	49	+	0	2,9	(+)
Аспарагиновая к-та	0	—	—	0	—	—	2,4	—	(+)
Аспарагин	6,7	—	+	0	—	—	5	—	+
Валин	0	(0,4)	—	20	30	+	25	75	+
Гистидин	3,4	—	(+)	0	—	—	0	—	—
Глицин	0	(0,5)	(—)	2,5	9,7	+	?	5	+
Глютаминовая к-та	0	(1,2)	—	0	(0,4)	—	1,6	6,2	+
Глютамин	0	—	—	0	—	—	0	—	—
Изолейцин	17	4,8	+	6	5,4	+	35	8,4	+
Лейцин	25	5,1	+	25	157	+	30	44	+
Лизин	2,2	—	+	0	—	—	?	—	?
Метионин	0	(0,6)	—	0	(1,5)	—	4,4	8,0	+
Пролин	3,4	0	+	6,5	188	+	0	0	—
Серин	25	3,6	+	—	170	+	—	2,3	(—)
Тирозин	25	8,0	+	—	1,0	—	20	8,6	+
Треонин	9,1	—	+	0	—	—	0	—	—
Триптофан	0	0	—	4,2	1,6	+	?	23	+
Фенилаланин	100	100	+	100	100	+	100	100	+
Цистеин	0	0	—	25	5,4	+	34	46	+

Сравнение вычисленных и наблюдаемых выходов аминокислот в опытах с поли-УАЦ, поли-УЦГ и поли-УАГ

Аминокислота	Поли-УАЦ						Поли-УЦГ						Поли-УАГ					
	74 : 15 : 11 [290]			83 : 5 : 12 [231]			79 : 8 : 13 [290]			34 : 50 : 15 [231]			72 : 16 : 12 [290]			68 : 3 : 29 [231]		
	Вычисленный выход	Наблюдаемый выход	Разница	Вычисленный выход	Наблюдаемый выход	Разница	Вычисленный выход	Наблюдаемый выход	Разница	Вычисленный выход	Наблюдаемый выход	Разница	Вычисленный выход	Наблюдаемый выход	Разница	Вычисленный выход	Наблюдаемый выход	Разница
Аланин	—	—	—	—	—	—	2	3	—1	67	40	+27	—	—	—	—	—	—
Аргинин	—	—	—	—	—	—	1	3	—2	54	49	+5	—	—	—	—	—	—
Валин	—	—	—	—	—	—	20	20	0	28	30	—2	—	—	—	—	—	—
Гистидин	1	3	—2	—	?	—	—	—	—	—	—	—	—	—	—	—	—	—
Глютамин	0	0	0	—	?	—	0	0	0	0	0	0	—	—	—	—	—	—
Изолейцин	20	17	+3	5	5	0	—	—	—	—	—	—	—	—	—	—	—	—
Лейцин	34	25	+9	17	5	+12	34	25	+9	79	157	—78	—	—	—	—	—	—
Метионин	0	0	0	0	1	—1	—	—	—	—	—	—	—	—	—	—	—	—
Серин	20	25	—5	14	4	+10	—	—	—	—	—	—	—	—	—	—	—	—
Тирозин	17	25	—8	5	8	—3	—	—	—	—	—	—	18	20	—2	5	9	—4
Треонин	1	9	—8	—	?	—	—	—	—	—	—	—	—	—	—	—	—	—
Триптофан	—	—	—	—	—	—	3	4	—1	9	2	+7	—	?	—	19	23	—4
Цистеин	—	—	—	—	—	—	17	25	—8	19	5	+14	17	34	—17	43	46	—3

объяснение замещения ала — сер и глиц — сер. Для объяснения же замещения арг — асп-NH_2 остается единственная возможность — приписать аргинину триплет ГА.У. Из непротиворечащих облигатным парам замещений остается необъясненным только одно гис — тир.

Опыты с урацилсодержащими СПРН. Результаты всех проведенных до сих пор опытов с поли-УАЦ, поли-УЦГ и поли-УАГ приведены в табл. 52. К сожалению, по этим полинуклеотидам материал мал. Все эти опыты принадлежат к числу самых ранних с применением менее совершенной методики, причем не во всех опытах исследовалось включение всех аминокислот. К тому же количественный состав СПРН иногда был такой, что их результаты не годятся для нашего анализа. Так, в одном из опытов с поли-УАЦ ожидаемый выход для триплетов состава У, А, Ц составляет по отношению к фенилаланину меньше одного процента. В табл. 53 приведены результаты сравнения наблюдаемых выходов с вычисленными для тех аминокислот, для которых можно было в этих опытах ожидать обнаружения новых триплетов. Непротиворечивые данные получаются только для цистеина (ГУ.А), валина (УГ.Ц) и тирозина (АУ.Ц). С большим сомнением можно приписать новые триплеты тирозину (АУ.Г), треонину (ЦА.У) и гистидину (АЦ.У). На основании же того, что в трех случаях из четырех вычисленные выходы для лейцина оказались выше наблюденных, можно сделать вывод, что все триплеты лейцина уже известны, и что ЦУ.А и ЦУ.Г являются нонсенсами.

Этим исчерпываются имеющиеся в настоящее время возможности для расшифровки аминокислотного кода. Можно было бы попытаться использовать данные опытов с четырехкомпонентными СПРН, но их нельзя количественно анализировать.

В табл. 54 приведены результаты произведенной частичной расшифровки кода. Как видно из этой таблицы, удалось раскрыть смысл 59 триплетов из 64. Из этих 59 триплетов шесть расшифрованы с большими или меньшими сомнениями (возле этих триплетов в таблице стоит вопросительный знак). Порядок написания букв в триплетах в этой таблице, в отличие от всего предыдущего, соответствует наиболее вероятной интерпретации опытов А. Ваба и сотрудников [319], т. е. факультативные буквы стоят в середине триплета. Конечно, не все расшифровки, в том числе показанные и без вопросительного знака, должны быть безусловно правильными. Они соответствуют непротиворечивой и наиболее однозначной интерпретации имеющихся до сих пор экспериментальных данных, не все из которых можно считать вполне достоверными. Однако мы считаем, что при накоплении большого количества данных того же рода, что уже имеются и были рассмотрены в настоящей работе, применение предложенных методов должно дать полную и однозначную

Таблица 54

Аминокислотный код

Аминокислота	Кодирующие триплеты
Аланин	ЦгГ, ГцГ, ЦцГ
Аргинин	ГаА, ГгЦ, ГцЦ, ГуА
Аспарагиновая к-та	АцГ, АуГ
Аспарагин	АцА, АуА
Валин	УаГ, УгГ, УцГ, УуГ
Гистидин	АгЦ, АцЦ, АуЦ?
Глицин	ГаГ, ГгГ, ГуГ
Глютаминовая к-та	АаГ, АгГ
Глютамин	ЦаГ, АаЦ
Изолейцин	УаА, УуА
Лейцин	УаУ, УгУ, ЦуУ
Лизин	АаА, АгА
Метионин	УгА
Пролин	ЦаЦ, ЦцЦ, ЦуЦ
Серин	ГгА, УгЦ, ГцА, УцЦ, УуЦ
Тирозин	АгУ?, АцУ, АуУ
Треонин	ЦаА, ЦгА, ЦгЦ, ЦцА, ЦуА?
Триптофан	ГгУ
Фенилаланин	УцУ, УуУ
Цистеин	ГаУ, ГуУ
Нонсенсы или точки	АаУ, ЦцУ, ГаЦ?, ЦаУ?, ЦгУ?
Не расшифрованные триплеты	УаЦ, УцА, ГцУ, ЦуГ, ГуЦ

расшифровку аминокислотного кода. Некоторые методы проверки полученных результатов будут приведены в следующей, заключительной, главе.

ГЛАВА 9

ИТОГИ И ОБСУЖДЕНИЕ

ИТОГИ

Прежде, чем перейти к обсуждению полученных выводов, мы хотим в общих словах повторить основной ход рассуждений.

Подробное рассмотрение современного состояния проблемы показало, что большинство положений об общих свойствах кода обосновано недостаточно твердо. Наименее неисследован был вопрос о степени вырожденности кода, хотя не подлежало сомнению, что какая-то (и, вероятно, значительная) вырожденность имеет место. Что же касается всех попыток расшифровки кода, то они базировались на представлении о невырожденном (или очень слабо вырожденном) коде.

Мы решили использовать для анализа проблемы кода статистические методы, надеясь получить при этом независимую информацию об общих свойствах кода и найти методы для более однозначной расшифровки кодовых слов.

Первый вопрос, который мы попытались решить, касался степени вырожденности кода. В настоящее время уже накоплен большой материал по замещениям аминокислот в белках, причем наиболее частой причиной замещений является замена отдельных букв в кодовых словах. Замещения можно сравнить с опечатками, в результате которых получаются слова, имеющие смысл (примером подобных опечаток в русском языке будут для слова «КОД» следующие: КОТ, КОЛ, РОД, ГОД и т. д.). Если известны объем алфавита и словаря и число букв в словах (которое полагаем постоянным), то нетрудно вычислить число таких опечаток, в результате которых одно слово превращается в другое. Можно поставить и обратную задачу: если определить каким-то образом максимальное число «осмысленных опечаток», которое допускает данный словарь, то тем самым будет получена информация об объеме последнего, т. е. о степени вырожденности кода. Поэтому первой задачей, которую мы решили, была количественная оценка максимально

возможного числа таких «опечаток», т. е. замещений аминокислот.

Вначале был использован метод, позволяющий по средней частоте наблюдающихся замещений определить их общее число. Этот метод основан на том, что при увеличении объема выборки возрастает средняя частота замещений и обнаруживается больший процент от их общего числа. Этот метод давал бы простое и однозначное решение задачи, если бы все наблюдаемые замещения были действительно результатом замены одной буквы в слове (т. е. являлись бы «разрешенными») и обнаруживались все с одинаковой вероятностью. Поскольку на самом деле это не так, пришлось провести ряд дополнительных расчетов, чтобы учесть влияние различных факторов. Некоторые из таких побочных расчетов имеют самостоятельное значение. Так, важные результаты дало сравнение наблюдаемой частоты отдельных замещений с ожидаемой (ожидание соответствовало предположению о том, что частота замещения пропорциональна встречаемости участвующих в нем аминокислот). В результате этого сравнения выяснилось, что все возможные пары аминокислот довольно отчетливо распадаются на две группы, первая из которых дает небольшие, а вторая — довольно значительные расхождения между вычисленной и наблюдаемой цифрами. Логично считать (в пользу чего говорит ряд соображений), что эти группы соответствуют разрешенным и запрещенным замещениям. Эти результаты оказывают службу и в дальнейшем, при расшифровке кода, позволяя отобрать действительно разрешенные замещения.

С той же целью были, кроме того, применены четыре других метода. Они были развиты не столь подробно, как первый, так как служили только для контроля правильности результатов, полученных первым методом. Во втором методе число замещений оценивалось по числу совпадений в двух независимых выборках, в третьем — по распределению отклонений между вычисленной и наблюдаемой частотой замещений, в четвертом — по частоте замещений, в которых участвуют распространенные и редкие аминокислоты, и в пятом — по числу замещений, которое дают отдельные аминокислоты.

Все методы дали согласные результаты, после чего с достаточной надежностью можно говорить о том, что число разрешенных замещений должно быть около 70. Если бы были известны все остальные свойства кода, то с помощью этой цифры можно было бы количественно оценить и степень вырожденности. Но, так как и остальные свойства кода были известны недостаточно, то этого сделать было нельзя; однако знание числа замещений облегчило дальнейшую работу, потому что после этого сильно уменьшилось число приемлемых гипотез.

Вторая задача была — выяснить общие свойства кода, зна-

ние которых необходимо для его расшифровки. Для этого было использовано два метода. Первый из них основан на следующих соображениях. Если две аминокислоты дают разрешенное замещение, значит кодирующие их слова отличаются лишь одной буквой, если две аминокислоты непосредственно не дают замещения, но дают его с какой-либо третьей аминокислотой, то кодирующие их слова отличаются двумя буквами и т. д. Отсюда ясно, что, проанализировав число переходов, которыми связаны все возможные пары аминокислот, можно получить информацию о числе букв в словах. Второй метод основан на том, что если по определенным правилам расположить аминокислоты, в соответствии с даваемыми ими замещениями, в узлах прямоугольной решетки, то число измерений этой решетки должно быть равно числу букв в словах, а число узлов на каждом ребре — числу букв в алфавите.

Применение обоих методов привело к одним и тем же, на первый взгляд парадоксальным, выводам. Оказалось, что, с одной стороны, код должен быть сильно вырожденным, с другой — что он обладает рядом свойств диплетного кода. Эти данные могли быть согласованы друг с другом и с числом разрешенных замещений в случае следующей гипотезы. Код триплетный и сильно вырожденный, но в большинстве случаев для кодирования аминокислоты необходимы только две («облигатные») буквы, характер третьей («факультативной») довольно безразличен. Такой код, действительно, должен обладать рядом свойств диплетного и при очень большой степени вырожденности (осуществляемой преимущественно за счет вариации в факультативных буквах) должен разрешать довольно умеренное число замещений аминокислот. Таким образом были подтверждены современные представления об общих свойствах кода, развитые главным образом Ф. Криком, а также установлена неравноценность букв внутри слова. После этого стало возможно приступить к расшифровке кода.

Для расшифровки кода могут быть в первую очередь использованы данные о стимуляции включения аминокислот в пептиды в бесклеточной системе под влиянием СПРН разного состава, а также данные о замещениях аминокислот. Пути использования этих данных настолько ясны, что попытки расшифровки кода предпринимались независимо друг от друга целым рядом авторов. Во всех случаях предполагали, что код не вырожден или очень слабо вырожден, что сильно облегчало задачу, но не приносило правильного решения. В условиях сильной вырожденности использование тех же методов не представляется возможным, так как из-за большого числа синонимов в словаре одни и те же экспериментальные данные допускают разные объяснения. Однако решение становится возможным, если задачу разделить на две части: вначале определить облигатные

пары букв и только на их основе приступать к окончательной расшифровке триплетного кода.

Состав облигатных пар был определен двумя методами. В первом методе применялся качественный анализ данных с СПРН, который дает возможность определить буквы, необходимые и достаточные для кодирования каждой аминокислоты. Во втором методе было использовано сравнение двух независимых классификаций аминокислот и облигатных пар букв. Для первой классификации служила корреляция содержания аминокислот в белках с содержанием пары Г + Ц в ДНК бактерий, для второй — результаты опытов по вызыванию мутаций азотистой кислотой, которая вызывает в и-РНК только транзиции Ц → У и А ← Г. Каждая из классификаций дает по три группы. Произведя все возможные попарные сравнения этих групп по совпадениям в них аминокислот и пар букв, можно расшифровать большинство облигатных пар. Результаты обоих методов хорошо согласуются друг с другом. Относительный порядок букв в облигатных парах определялся на основании анализа замещений, а так как для двух аминокислот можно установить на основании прямых опытов абсолютный порядок, то его легко было распространить и на все остальные облигатные пары аминокислот.

На основе выясненных облигатных пар производится окончательная расшифровка триплетного кода. Благодаря тому, что облигатные пары уже известны, можно сравнивать выходы, получаемые в опытах с СПРН только у тех аминокислот, слова которых содержат облигатные пары одинакового состава, и задача сводится, таким образом, лишь к выяснению того, при каких третьих буквах облигатная пара кодирует данную аминокислоту. Основной анализ проводился на данных опытов с двухкомпонентными СПРН. В связи с неточностью экспериментальных цифр анализировались не отдельные выходы, а зависимость относительного выхода от состава СПРН. При введенном ограничении — сравнивать только выходы аминокислот с облигатными парами одинакового состава — теоретически возможны лишь 4 типа кривых, ясно отличающиеся друг от друга. Если же аминокислоты для сравнения брать произвольно, не считаясь с составом облигатных пар, то задача однозначно не решается. Результаты, полученные с двухкомпонентными СПРН были сопоставлены с заведомо разрешенными замещениями аминокислот, что дало возможность расшифровать еще некоторое количество слов, после чего были проанализированы результаты опытов с трехкомпонентными СПРН. При этом наблюдаемые выходы сравнивались с ожидаемыми для уже известных триплетов и по расхождениям делали заключения о существовании еще неизвестных слов. В результате удалось расшифровать смысл большинства слов. Оставшиеся неясности связаны

с отсутствием достаточно достоверных экспериментальных данных по некоторым СПРН.

ОБСУЖДЕНИЕ

После завершения работы напрашивается рассмотрение двух групп вопросов. Первая — насколько можно быть уверенным в том, что гипотеза о свойствах кода и результат его расшифровки верны. Вторая — какие выводы для молекулярной и общей биологии могут следовать из полученных результатов. Несмотря на всю заманчивость спекуляций в области второй группы вопросов, мы пока воздержимся от них; до подтверждения правильности полученных выводов это было бы преждевременно. Сейчас мы остановимся только на возможных методах проверки справедливости этих результатов.

Здесь возникает два вопроса — правильны ли представления об общих свойствах кода и правильна ли его расшифровка. Что касается общих свойств кода, то мы за одним, хотя и довольно существенным исключением, касающемся неравноценности букв в словах, пришли к тем же выводам, что и Ф. Крик [35, 128]. Поскольку наши методы совершенно не связаны с тем материалом, который анализировал Ф. Крик, такое совпадение выводов, конечно, говорит в пользу их правдоподобности. Новый вывод состоит в представлении о факультативности третьих букв триплетов, и, поскольку такой вывод раньше не делался, больше всего сомнений может быть именно в его справедливости. По этому поводу нам хочется сказать следующее. К этому выводу мы окончательно пришли в конце 1962 г., когда еще не были опубликованы работы по определению триплетов, не содержащих урацила [156, 201, 277, 320]. Тогда же была произведена первая расшифровка кода на новой основе. После появления упомянутых работ оказалось, что сделанные предсказания о составе кодирующих слов получили экспериментальное подтверждение. Никаких существенных противоречий теории эти новые данные не принесли.

Что касается расшифровки кода, то она должна быть, безусловно, правильной, если правильны положенные в ее основу предпосылки и достоверны использованные экспериментальные факты. Приходится пожалеть, что нет достаточно прямых данных, которые можно было бы использовать для проверки сделанных выводов. Впрочем, это вполне естественно, так как все подобные данные были уже использованы для самой расшифровки. Однако есть несколько возможностей для более косвенной проверки. Предложенный нами код позволяет сделать некоторые предсказания, материал для проверки которых имеется. В значительной части это именно тот материал, на основании которого высказывались сомнения в правильности кодов,

предлагавшихся раньше. Совпадение экспериментальных данных с предсказанием должно в большей или меньшей степени говорить в пользу правильности кода. Рассмотрим некоторые из таких данных.

Соотношение букв в матрице. Можно было бы попытаться вычислить соотношение нуклеотидов в и-РНК для

Таблица 55

Сравнение состава матричной РНК для белка E. coli

Тип РНК	Содержание, %			
	А	Г	Ц	У
Рибосомная	25,4	33,5	21,5	19,6
Свежесинтезированная	25,0	29,5	22,9	22,6
Вычисленный состав	28,2	29,0	20,4	22,4

какого-либо белка известного состава. Однако до сих пор нет данных о составе определенных и-РНК; поэтому мы должны удовольствоваться сравнением состава суммарных белка и РНК. Для этого мы избрали данные по E. Coli [293]. Состав матрицы вычислялся по данным табл. 54 в предположении, что все синонимы для данной аминокислоты встречаются одинаково часто.

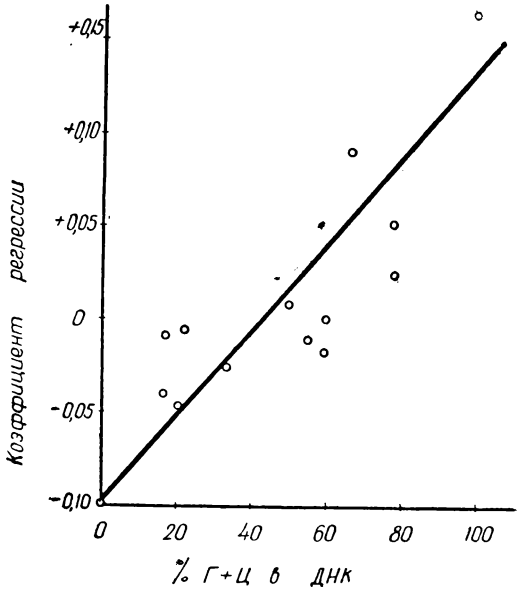

Рис. 25. Зависимость коэффициентов регрессии для отдельных аминокислот от вычисленного содержания Г + Ц в кодирующих их словах.

Результаты расчета приведены в табл. 55. Там же приведены данные о составе рибосомной РНК и свежесинтезированной РНК того же организма [354а]; есть основания считать, что последняя в значительной части состоит из и-РНК [260]. Соответствие между вычисленным и наблюдаемым составом нельзя не признать хорошим. Интересно отметить, что вычисленный состав предсказывает для ДНК содержание пары Г + Ц 49,4%, тогда как в действительности наблюдается 50%. Отношение числа 6-амино-

и 6-кетогрупп близко в вычисленной матрице к единице, что и наблюдается в естественных РНК.

Содержание пары Г + Ц. Нам уже неоднократно приходилось упоминать о работе Н. Суеока [293], в которой исследовалась корреляция между содержанием пары Г + Ц в ДНК разных видов бактерий и содержанием аминокислот в их белках. Полученные в этой работе цифры были приведены в табл. 36. Можно попытаться вычислить среднее содержание пары Г + Ц в словах, кодирующих отдельные аминокислоты, и сопоставить их с коэффициентами регрессии, полученными Н. Суеока. Наличие корреляции между ними будет говорить в пользу правдоподобности кода. Соответствующее сравнение производится на рис. 25. Как видно из рисунка, зависимость коэффициентов регрессии от вычисленного содержания Г + Ц примерно соответствует ожидаемой прямолинейной зависимости.

Встречаемость аминокислот. Не лишено известных оснований предположение о том, что число синонимов может быть связано с распространенностью соответствующей аминокислоты в белках. В пользу этого говорят данные об изменчивости состава ДНК у разных видов при значительной гомогенности ее внутри организма и незначительных вариациях в составе белков [278, 294, 355], из чего можно сделать вывод о том, что порядок оснований в ДНК не должен сильно отличаться от случайного. Мы сопоставили данные о числе синонимов (см. табл. 54) с распространенностью аминокислот. Результаты приведены в табл. 56, из которой видно, что такая связь действительно имеет место.

Таблица 56

Число триплетов и распространенность аминокислот

Число триплетов, кодирующих одну аминокислоту	Средняя встречаемость аминокислот в белках, %
1	0,8
2	4,5
3	5,7
4	6,6
5	7,2

Замещения. Хороший материал для проверки правильности кода давали бы замещения, если бы они сами не использовались для расшифровки. Правда, выше они использовались очень ограниченно, так как к тому времени, когда это понадобилось, значительная часть триплетов была уже расшифрована и объясняла большую часть замещений. В конечном счете, код, приведенный в табл. 54, объясняет подавляющее большинство

замещений. Из всех замещений, кроме встреченных однократно при сравнении искусственных белков (большинство которых являются запрещенными), противоречат коду только 4 сорта замещений, каждое из которых было обнаружено дважды: асп — лиз, глют — про, ала — гис, и арг — лей. Два первых из них могут быть объяснены; если допустить, что в действительности речь шла не об аспарагиновой и глютаминовой кислотах, а об их амидах. Замещения, встреченные однократно в естественных белках (5-я группа в табл. 49), в подавляющем большинстве противоречат коду: он объясняет только 6 замещений из 28.

Более уверенно можно привлечь для проверки кода запрещенные замещения, так как факт «запрещенности» при расшифровке не учитывался. В связи с тем, что и разрешенные замещения были использованы довольно ограниченно, можно думать, что «запреты» не использовались и в неявной форме. Если сопоставить 20 пар аминокислот, дающих наибольшие индексы при сравнении наблюденных и ожидаемых частот, которые были приведены в табл. 26 с кодом, то окажется, что 7 из них кодом разрешены. Это следующие пары: илей — сер, асп — вал, гис — сер, ала — арг, арг — про, арг — тре, ала — глю-NH_2.

Хотя 7 пар из 20 составляют для столь вырожденного кода небольшой процент, все же с ним нельзя не считаться. Поэтому рассмотрим эти пары аминокислот более подробно. При этом можно заметить две особенности. Во-первых, в большинстве случаев замещение в этих парах возможно только при одной определенной паре триплетов. Например, изолейцин кодируется двумя триплетами, серин — пятью, но замещение илей — сер возможно только для одной пары триплетов: УуА и УуЦ; если бы это обстоятельство было учтено при вычислении индекса, то он должен был бы оказаться значительно меньше. Во-вторых, все эти семь пар могут дать замещения только в случае трансверзий. Можно думать, что это процесс менее вероятный, чем транзиции.

В табл. 57 приведены данные о частоте замещений, объяс-

Таблица 57

Число транзиций и трансверзий среди разрешенных замещений

Механизм замещений	Разные типы замещений		Общая частота замещений	
	число	%	число	%
Транзиции	30	71,5	162	83,0
Трансверзии	12	28,5	33	17,0

няемых и не объясняемых транзициями. Если учесть, что в случае равной вероятности обеих реакций трансверзии должны были бы происходить вдвое чаще транзиций, то из этих результатов следует, что в действительности трансверзии происходят значительно более редко.

Таким образом, данные о замещениях находятся в хорошем соответствии с предложенным кодом.

Конечно, ни одно из этих сопоставлений не может служить подтверждением правильности расшифровки кода, для этого нужно ждать появления новых экспериментальных данных. Однако соответствие нашего кода фактам, которые противоречили ранее предлагавшимся кодам говорит в пользу его правдоподобности.

ВЫВОДЫ

В заключение кратко подытожим современные данные об общих свойствах аминокислотного кода, насколько о них можно судить на основании нашей и других работ.

1. Кодовое число равно трем, т. е. код является триплетным, причем никаких ограничений в выборе букв для кодовых слов не существует.

2. Кодовое отношение равно трем.

3. Код не имеет запятых. В пользу этого положения можно привести дополнительный довод. Если бы запятые являлись необходимой частью любого текста, то они должны были бы присутствовать и в СПРН, стимулирующих синтез пептидов в бесклеточных системах. Однако при анализе опытов с некоторыми СПРН удается расшифровать кодирующий смысл всех входящих в их состав триплетов.

4. Код неперекрывающийся. Новые доводы в пользу того, что код триплетный и не имеет запятых при кодовом отношении, равном трем, вряд ли оставляют в этом сомнение.

5. Считывание текста при синтезе белков происходит последовательно, начиная с некоторой фиксированной точки.

6. Код универсален.

7. Вырожденность кода очень велика. Однако имеются триплеты, не кодирующие аминокислот (точки или нонсенсы).

8. В большинстве случаев в триплете облигатными являются лишь две буквы. Третья почти всегда факультативна. Вырожденность осуществляется преимущественно за счет вариации в факультативных буквах.

9. В связи с неслучайной вырожденностью кода разрешены замещения далеко не между всеми парами аминокислот.

10. Большинство спонтанных замещений является разрешенными, причем наиболее частый их механизм — транзиции.

ЛИТЕРАТУРА

1. Анфинсен К. Молекулярные основы эволюции. Изд-во Иностр. Лит., М., 1962.
2. Баев А. А. «Ж. Всесоюз. хим. о-ва» т. 8, стр., 2, 1963.
3. Белозерский А. Н. В кн.: «Возникновение жизни на земле». Изд-во АН СССР, М., стр. 198, 1957.
4. Белозерский А. Н. Нуклеиновые кислоты и их биологическое значение. М., Изд-во «Знание», 1961.
5. Белозерский А. Н. Тр. 5-го Междунар. биохим. конгресса, т. 3, стр. 202, М., 1962.
6. Белозерский А. Н., Спирин А. С. В кн.: «Нуклеиновые кислоты» М., Изд-во иностр. лит., стр. 123, 1962.
7. Бензер С. В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 56, 1960.
8. Бензер С. В кн.: «Молекулярная генетика». М., Изд-во иностр. лит., стр. 11, 1963.

8 а. Бертон К. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 77, М., 1962.

9. Браунитцер Г., Геринг-Мюллер Р., Хильшман Н., Хильзе К., Хобом Г., Рудлов В., Виттман-Либольд Б., Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 16, М., 1962.
10. Браше Ж. Биохимическая цитология. М., Изд-во иностр. лит., 1959.
11. Вагнер Р., Митчелл Г. Генетика и обмен веществ. М., Изд-во иностр. лит., 1958.
12. Винер Н. Кибернетика, или управление и связь в животном и машине. М., Изд-во «Сов. радио», 1958.
13. Виттман Г. Г., Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 274, М., 1962.
14. Виттман Х., В кн.: «Молекулярная генетика». М., Изд-во иностр. лит., стр. 51. 1963.
15. Гамов Г., Ичас М., В кн.: «Теория информации в биологии». М., Изд-во иностр. лит., стр. 66, 1960.
16. Гамов Г., Рич А., Икас М., В кн.: «Вопросы биофизики». М., Изд-во иностр. лит., стр. 205, 1957.
17. Гирер А. Тр. 5-го Междунар. биохим. конгресса, т. 3, стр. 191, М., 1962.
18. Дельбрюк М., Стент Г., В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 562, 1960.
19. Дубинин Н. П. «Биофизика», т. 1, стр. 677, 1956.
20. Дэвидсон Д. Биохимия нуклеиновых кислот. М., Изд-во иностр. лит., 1952.
21. Жакоб Ф., Вольман Э. Пол и генетика бактерий. М., Изд-во иностр. лит., 1962.
22. Жакоб Ф., Моно Ж., Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 157, М., 1962.

23. Збарский Б. И., Иванов И. И., Мардашев С. Р. Биологическая химия. М., Медгиз, 1954.
24. Ингрэм В. М. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 21, М., 1962.
25. Ичас М. В кн.: «Теория информации в биологии». М., Изд-во иностр. лит., стр. 72, 1960.
26. Кастлер Г. В кн.: «Теория информации в биологии». М., Изд-во иностр. лит., стр. 9, 1960.
27. Киселев Л. Л. «Ж. Всесоюз. хим. о-ва» т. 8, стр. 11, 1963.
28. Кольцов Н. К. Физико-химические основы морфологии. М., ГИЗ, 1929.
29. Кольцов Н. К. Организация клетки М., Биомедгиз, 1936.
30. Конигсберг В., Хилл Р. Д. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 14, М., 1962.
31. Корнберг А., В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 468, 1960.
32. Крик Ф., В кн.: «Биологическое воспроизводство макромолекул» М., Изд-во иностр. лит., стр. 209, 1960.
33. Крик Ф. В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 428, 1960.
34. Крик Ф. В кн.: «Живая клетка». М., Изд-во иностр. лит., стр. 203, 1962.
35. Крик Ф., Барнетт Л., Бреннер С., Уоттс-Тобин Р. В кн.: «Молекулярная генетика». М., Изд-во иностр. лит., стр. 33, 1963.
36. Липманн Ф. Т. 5-го Междунар. биохим. конгресса, т. 1, стр. 147, 1962.
37. Лучник Н. В. Докл. АН СССР, т. 129, стр. 1168, 1959.
38. Лучник Н. В. «Биохимия», т. 27, стр. 1032, 1962.
39. Лучник Н. В. В сб.: «Применение математических методов в биологии». Вып. 2 стр. 177, Л., Изд-во ЛГУ, 1963.
40. Лучник Н. В., Царапкин Л. С. Докл. АН СССР, т. 124, стр. 213, 1959.
41. Лучник Н. В., Царапкин Л. С. «Цитология» т. 1, стр. 73, 1959.
42. Мацуда Г., Браунитцер Г. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 28, М., 1962.
43. Медведев Ж. А. «Ж. Всесоюз. хим. о-ва», т. 6. стр. 268, 1961.
44. «Молекулярная генетика». М., Изд-во иностр. лит., 1963.
45. Нейрат Г. В кн.: «Современные проблемы биофизики». М., Изд-во иностр. лит., т. 1, стр. 238, 1961.
46. Ниренберг М. В., Маттеи Г. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 211, М., 1962.
47. Ниренберг М., Маттеи Г. В кн.: «Молекулярная генетика». М., Изд-во иностр. лит., стр. 69, 1963.
48. Очоа С., Хеппел Л. В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 500, 1960.
49. Полетаев И. А. Сигнал. Изд-во «Сов. радио», М., 1957.
50. Смит Д. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 38, М., 1962.
51. Спейер Д., Ленгел Р., Базилио К., Очоа С. В кн.: «Молекулярная генетика». Изд-во иностр. лит., М., стр. 93, 1963.
52. Спирин А. С. «Ж. Всесоюз. хим. о-ва», т. 6, стр. 260, 1961.
53. Спирин А. С., Смирнов В. Н. «Биофизика», т. 7, стр. 501, 1962.
54. Туманян В. Г., Есипова Н. Г., Андреева Н. С. «Биофизика», т. 8, стр. 124, 1963.
55. Туппи Г. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 34, М., 1962.
56. Фолькин Э., Астрахан Л. В кн.: «Химические основы наследственности». М., Изд-во иностр. лит., стр. 556, 1960.
57. Фриз Э. Тр. 5-го Междунар. биохим. конгресса, т. 1, стр. 231, М., 1962.

58. Хесин Р. Б. «Ж. Всесоюз. хим. о-ва», т. 6, стр. 254, 1961.
59. Чавчанидзе В. В. «Биофизика», т. 3, стр. 391, 1958.
60. Чаргафф Э. В кн.: «Возникновение жизни на земле». М., Изд-во АН СССР, стр. 192, 1957.
61. Чаргафф Э., Девидсон Д. Нуклеиновые кислоты, том 1 и 2, М., Изд-во иностр. лит., 1957 и 1962.
62. Шорм Ф. В кн.: «Возникновение жизни на земле». М., Изд-во АН СССР, стр. 166, 1957.
63. Шустер Г., Гирер А., Мундри К. В. «Ж. Всесоюз. хим. о-ва» т. 6, стр. 293, 1961.
64. Эйбельсон Ф. Г. Тр. 5-го Междунар. биохим. конгресса, т. 1, М., 1962.
65. Эфроимсон В. П. «Ж. Всесоюз. хим. о-ва», т. 6, стр. 305, 1961.
66. Эшби У. Р. Введение в кибернетику. М., Изд-во иностр. лит. 1959.
67. Acher R., Chauvet J. Biochim. Biophys. Acta, v. **14**, p. 421, 1954.
68. Acher R., Chauvet J., Lenci M.—T. Biochim. Biophys. Acta, v. **31**, p. 545, 1959.
69. Acher R., Chauvet J., Lenci M.-T., Morel F., Maetz J., Biochim. Biophys. Acta, v. **42**, p. 379, 1960.
70. Allfrey V. G., Mirsky A. E., Proc. Natl. Acad. Sci., v. **48**, p. 1590, 1962.
71. Anderer F. A., Handschuch D., Z. Naturforsch., v. **17b**, p. 536, 1962.
72. Ando T., Tobita T., Yamasaki M., J. Biochem. (Tokyo), v. **45**, p. 285, 1958.
73. Aposhian H. V., Kornberg A. Fed. Proc., v. **20**, p. 361, 1961.
74. Aposhian H. V., Kornberg A. J. Biol. Chem., v. **237**, p. 519, 1962.
75. Arnstein H. V. R., Cox R. A., Hunt J. A. «Nature», v. **194**, p. 1042, 1962.
76. Avery O. T., MacLeod C. M., McCarty M. J. Exptl. Med., v. **79**, p. 137, 1944.
77. Baglioni C. J. Biol. Chem., v. **237**, p. 69, 1962.
78. Baglioni C., Ingram V. M. Biochim. Biophys. Acta., v. **48**, p. 253, 1961.
79. Barnabas J., Muller C. J. «Nature», v. **194**, p. 931, 1962.
80. Basilio C., Wahba A. J., Lengyel P., Speyer J. F., Ochoa S., Proc. Natl. Acad. Sci., v. **48**, p. 613, 1962.
81. Beer M., Moudrianakis E. N. Proc. Natl. Acad. Sci., v. **48**, p. 409, 1962.
82. Bell P. H. J. Amer. Chem. Soc., v. **76**, 5565, 1954.
83. Bell P. H., Howard K. S., Shepherd R. G., Finn B. M., Meisenhelder J. H. J. Amer. Chem. Soc., v. **78**, p. 5059, 1956.
84. Benzer S. Proc. Natl. Acad. Sci., v. **41**, p. 344, 1955.
85. Benzer S., Weisblum B. Proc. Natl. Acad. Sci., v. **47**, p. 1149, 1961.
86. Benzer S., Champe S. P. Proc. Natl. Acad. Sci., v. **48**, p. 1114, 1962.
87. Bishop J. O., Schweet R. Biochim. Biophys. Acta, v. **54**, p. 617, 1961.
88. Bloch D. P. Proc. Natl. Acad. Sci., v. **41**, p. 1058, 1955.
89. Blombäck B., Sjöquist J. Acta Chem. Scand., v. **14**, p. 493, 1960.
90. Blombäck B., Blombäck M., Edman P., Hessel B., «Nature», v. **193**, p. 883, 1962.
91. Bollum F. J. J. Biol. Chem. v. **235**, p. 2399, 1960.
92. Brachet J. Enzymologia, v. **10**, p. 87, 1941.
93. Brachet J. Arch. Biol., v. **53**, p. 207, 1942.

94. Brachet J. Embryologie Chimique. Desoer, Liège, 1944.
95. Brachet J. Biochemical Cytology. Acad. Press, N.-Y., 1958.
96. Braunitzer G., Rudloff V., Hilse K., Liebold B., Müller R., Z. Physiol. Chem., v. 320, p. 283, 1960.
97. Braunitzer G., Hilschmann N., Hilse K., Liebold B., Müller R. Z. Physiol. Chem., v. **322**, p. 96, 1960.
98. Braunitzer G., Matsuda G. Z. Physiol. Chem., v. **324**, p. 91, 1961,
99. Brenner S. Proc. Natl. Acad. Sci., v. **43**, p. 687, 1957.
100. Brenner S., Barnett L., Crick F. H. C., Orgel A. J. Molec. Biol., v. **3**, p. 121, 1961.
101. Brenner S., Jacob F., Meselson M. «Nature», **190**, 576, 1961.
102. Bretscher M. S., Grunberg-Manago M. «Nature», v. **195**, p. 283, 1962.
103. Bromer W. W., Sinn L. G., Staub A., Behrens O. K. J. Amer. Chem. Soc., v. **78**, p. 3858, 1956.
104. Brown G. L., Zubay G. J. Molec. Biol., v. **2**, p. 287, 1960.
105. Brown H., Sanger F., Kitai R. Biochem. J., v. **60**, p. 556, 1955.
106. Burdon R. H., Shellie R. M. S. Biochim. Biophys. Acta, v. **51**, p. 153, 1961.
107. Burma D. P., Krõger H., Ochoa S., Warner R. C., Weill J. D. Proc. Natl. Acad. Sci., v. **47**, p. 749, 1961.
108. Caldwell P. C., Hinshelwood C. N. J. Chem. Soc., N. **1950**, p. 3156, 1950.
109. Canellakis E. S. Biochim. Biophys. Acta, v. **23**, p. 217, 1957.
110. Carlton B. C., Janofsky C. J. Biol. Chem., v. **273**, p. 1531, 1962.
111. Caspersson T. Scand. Arch. f. Physiol., v. **73**, Suppl. 8, 1936.
112. Caspersson T. Naturwissenschaften, v. **28**, p. 33, 1941.
113. Champe S. P., Benzer S. Proc. Natl. Acad. Sci., v. **48**, p. 532, 1962.
114. Chapeville F., Lipmann F., Ehrenstein G., Weisblum B., Ray W. J., Benzer S. Proc. Natl. Acad. Sci., v. **48**, p. 1086, 1962.
115. Chargaff E. Experientia, v. **6**, p. 201, 1950.
116. Chargaff E. II-e Congrès International de Biochimie, Paris, p. 41, 1952.
117. Chargaff E. In: «The Nucleic Acids» Acad. Press, N.-Y., v. **1**, p. 307, 1955.
118. Chargaff E. «Nature», v. **194**, p. 86, 1962.
119. Cheng P. Y. Biochim. Biophys. Acta, v. **53**, p. 235, 1961.
120. Chung C. W., Mahler H. R., Enrione M., J. Biol. Chem., v. **235**, p. 1448, 1960.
121. Cochran G. W., Dhaliwal A. S., Welkie G. W., Chidester J. L., Lee M. H., Chandrasekhar B. K. Science, v. **138**, p. 46, 1962.
122. Cox R. A., Littauer U. Z. J. Molec. Biol., **2**, p. 166, 1960.
123. Crick F. H. C. In: «The Structure of Nucleic Acids and their Role in Protein Synthesis». Cambridge, p. 25, 1957.
124. Crick F. H. C. Biochem. J., v. **84**, p. 59, 1962.
125. Crick F. H. C. In: «Progress in Nucleic Acid Research». Acad. Press. (In Press), 1963.
126. Crick F. H. C., Watson J. D. Proc. Roy. Soc. (Lond.), v. **223A**, p. 80, 1954.
127. Crick F. H. C., Griffith J. S., Orgel J. F. Proc. Natl. Acad. Sci., v. **43**, p. 416, 1957.
128. Crick F. H. C., Barnett L., Brenner S., Watts-Tobin R. J., «Nature», v. **192**, p. 1227, 1961.

128a. Cullis A. F., Muirhead H., Perutz M. F., Rossman M. G., North A. C. T. Proc. Roy. Soc. (Lond.), v. **265A**, p. 161, 1962.
129. Darlington C. D. Recent Advances in Cytology. London, 1937.
130. Davis F. F., Allen F. W. J. Biol. Chem., v. **227**, p. 907, 1957.
131. Decken A. v. d., Hultin T. Biochim. Biophys. Acta, v. **45**, p. 137, 1960.
131a. Dintzis H. M., Proc. Natl. Acad. Sci., **47**, 247, 1961.
132. Dounce A. L. Enzymologia, v. **15**, p. 251, 1952.
133. Dounce A. L. «Nature», v. **172**, p. 541, 1953.
134. Dunn D. B., Smith J. D., Spahr P. F. J. Molec. Biol., v. **2**, p. 113, 1960.
135. Eck R. V. «Nature», v. **191**, p. 1284, 1961.
136. Eck R. V. J. Theoret. Biol., v. **2**, p. 139, 1962.
137. Edmonds M., Abrams R. J. Biol. Chem., v. **235**, p. 1142, 1960.
138. Edmundson A. B., Hirs C. H. W. «Nature», v. **190**, p. 663, 1961.
139. Ehrenberg A., Theorell H. Acta Chem. Scand., v. **9**, p. 1193, 1955.
140. Ehrenstein G. v., Lipmann F. Proc. Natl. Acad. Sci., v. **47**, p. 941, 1961.
141. Elliott D. F., Peart W. S. «Nature», v. **177**, p. 527, 1956.
142. Elliott D. F., Peart W. S. Biochem. J., v. **65**, p. 246, 1956.
143. Felix K., Hirohata R., Dirr K. Z. Physiol. Chem., v. **291**, p. 228, 1933.
144. Felix K. Experientia, v. **8**, p. 312, 1952.
145. Felix K., Goppold-Krekels A., Hübner L., Meissner P., Bull. Soc. Chim. Biol., v. **40**, p. 1973, 1958.
146. Fessenden J. M., Cairncross J., Moldave K., Proc. Natl. Acad. Sci., v. **49**, p. 82, 1963.
147. Feymann R. P., Benzer S., Freese E., Цитир. по [128].
148. Folk J. E., Gladner J. A. Biochim. Biophys. Acta, v. **44**, p. 383, 1960.
149. Folk J. E., Gladner J. A., Levin Y. J. Biol. Chem., v. **234**, p. 2317, 1959.
150. Fox A. S., Yoon S. B., Mead C. G. Proc. Natl. Acad. Sci., v. **48**, p. 546, 1962.
151. Franklin R. E., Gosling R. G. Acta Cristallogr., v. **6**, p. 673, 1953.
152. Freese E. Proc. Natl. Acad. Sci., v. **45**, p. 622, 1959.
153. Gamow G. «Nature», v. **173**, p. 318, 1954.
154. Gamow G. Dan. Biol Medd., v. **22**. p. 3, 1954.
155. Gamow G., Ycas M. Proc. Natl. Acad. Sci., v. **41**, p. 1011, 1955.
156. Gardner R. S., Wahba A. J., Basilio C., Miller R. S., Lengyel P., Speyer J. F. Proc. Natl. Acad. Sci., v. **48**, p. 2087, 1962.
157. Gellert M., Lipsett M. N., Davies D. R. Proc. Natl. Acad. Sci., v. **48**, p. 2013, 1962.
158. Gerald P. S., Efron M. L. Proc. Natl. Acad. Sci., v. **47**, p. 1758, 1961.
159. Gerald P. S., Ingram V. M. J. Biol. Chem., v. **236**, p. 2155, 1961.
160. Geschwind I. I., Li C. H., Barnafi L. J. Amer. Chem. Soc., v. **79**, p. 6394, 1957.
161. Gierer A., Schramm G. Z. Naturforsch., v. **11b**, p. 138, 1956.
161a. Goldstein A., Biochim. Biophys. Acta, v. **53**, p. 468.
162. Golomb S. W., Gordon B., Welch L. R. Canad. J. Mathem., v. **10**, p. 202, 1958.

163. Golomb S. W., Welch L. R., Delbrück M., Dan. Biol. Medd., v. **29**, p. 9, 1958.
164. Goodman H. M., Rich A. Proc. Natl. Acad. Sci., v. **48**, p. 2101. 1962.
165. Gordon W. G., Basch J. J., Kalan E. B. J. Biol. Chem., v. **236**, p. 2912, 1961.
166. Gray E. D., Weissmann S. M., Richards J., Bell D., Keir H. M., Shellie R. M. S., Davidson J. N., Biochim. Biophys. Acta, v. **45**, p. 111, 1960.
167. Griffin A. C., O' Neal M. A. Biochim. Biophys. Acta, v. **61**, p. 496, 1962.
168. Gros F. W., Gilbert W., Hiatt H. H., Kurland C. G., Risebrough R. W., Watson J. D. «Nature», v. **190**, p. 581, 1960.
169. Grossi L. G., Moldave K. J. Biol. Chem., v. **235**, p. 2370. 1960.
170. Grunberg-Manago M., Ochoa S., J. Amer. Chem. Soc., v. **77**, p. 3165, 1955.
171. Gulland J. M. Cold Spring Harbor Symp., v. **12**, p. 95, 1947.
171a. Hall B. D., Spiegelman S., Proc. Natl. Acad. Sci., v. **47**, p. 1564, 1961.
172. Hämmerling J., Pflügers Arch. ges. Physiol., v. **131**, p. 1, 1934.
173. Harris J. I. Biochem. J., v. **71**, p. 451, 1959.
174. Harris J. I., Roos P., Biochem, J., v. **71**, p. 445, 1959.
175. Harris J. I., Sanger F., Naughton M. A. Arch. Biochem. Biophys., v. **65**, p. 427, 1956.
176. Hartmann C., Lipmann F. Цитир. по [36].
177. Haselkorn R., Fried V. A., Dahlberg J. E. Proc. Natl. Acad. Sci., v. **49**, p. 511, 1963.
178. Hayashi M., Spiegelman S. Proc. Natl. Acad. Sci., v. **47**, p. 1564, 1961.
179. Helinski D. R., Yanofsky C. Proc. Natl. Acad. Sci., v. **48**, p. 173, 1962.
180. Hendler R. W. Proc. Natl. Acad. Sci., v. **48**, p. 1402, 1962.
181. Henning U., Yanofsky C. Proc. Natl. Acad. Sci., v. **48**, p. 183, 1962.
182. Henning U., Yanofsky C. Proc. Natl. Acad. Sci., v. **48**, p. 1497, 1962.
183. Hershey A. D., Chase M., J. Gen. Physiol., **36**, 39, 1952.
184. Hiatt H. H., J. Molec. Biol., **5**, 217, 1962.
185. Hill R. L., Schwartz H. C. «Nature», v. **184**, p. 641, 1959.
186. Hill R. L., Swenson R. T., Schwartz H. C., J. Biol. Chem., v. **235**, p. 3182, 1960.
187. Hirs C. H. W., Moore S., Stein W. H. J. Biol. Chem., v. **235**, p. 633, 1960.
188. Hoagland M. B., Askonas B. A., Proc. Natl. Acad. Sci., v. **49**, p. 130, 1963.
189. Hoffmann-Berling H. Цитир. по [343]
190. Huang R. S., Maheshwary N., Bonner J. Biochem. Biophys. Res. Comm., v. **3**, p. 689, 1960.
191. Hunt J. A., Ingram V. M. «Nature», v. **181**, p. 1062, 1958.
192. Hunt J. A., Ingram V. M. «Nature», v. **184**, p. 640, 1959.
193. Hunt J. A., Ingram V. M. «Nature , v. **184**, p. 870, 1959.
194. Hurwitz J., Furth J. J., Anders M., Ortiz P. J., August J. T. Cold Spring Harbor Symp., v. **26**, p. 91, 1961.
195. Ingram V. M. «Nature» , v. **180**, p. 326, 1957.
196. Ingram V. M. Biochim. Biophys. Acta, v. **36**, p. 402, 1959.
197. Ingram V. M., Stretton A. O. W. «Nature», v. **190**, p. 1079, 1961.

198. Isawa M., Allfrey V. G., Mirsky A. E. Proc. Natl. Acad. Sci., v. **49**, p. 544, 1963.
199. Jacob F., Wollman E. L. Sexuality and the Genetics of Bacteria. Acad. Press, N.-Y., 1961.
200. Jinks J. L. Heredity, v. **16**, p. 153, 241, 1961.
201. Jones O. W., Nirenberg M. W. Proc. Natl. Acad. Sci., v. **48**, p. 2115, 1962.
201a. Josse J., Kornberg A. Federat. Proc., v. **19**, p. 305, 1960.
202. Josse J., Kaiser A. D., Kornberg A. J. Biol. Chem., v. **236**, p. 864, 1961.
203. Jukes T. H. Proc. Natl. Acad. Sci,. v. **48**, p. 1809, 1962.
204. Kalan E. B., Gordon W. G., Basch J. J., Townend R. Arch. Biochem. Biophys., v. **96**, p. 376, 1962.
205. Karasek M., Schramm G. Biochem. Biophys. Res. Comm., v. **9**, p. 63, 1962.
206. Keil B., Sorm F., Holeysovsky V., Kostka V., Meloun B., Mikes O., Tomasek V., Vanecek J. Coll. Czech. Chem. Comm., v. **24**, p. 3491, 1959.
207. Keller E. B., Zamecnik P. C. J. Biol. Chem., v. **221**, p. 45, 1956.
208. Kendrew J. C., Watson H. C., Strendberg B. E., Dickerson R. E., Philips D. C., Share V. C. «Nature», v. **190**, p. 666, 1961.
209. Kim Y. T., Wildman S. G. Biochem. Biophys. Res. Comm., v. **8**, p. 294, 1962.
210. Kjeldgaard N. O. Biochim. Biophys. Acta, v. **49**, p. 64, 1961.
211. Klemperer H. G,. Kammen H. O. Biochem. Biophys. Res. Comm., v. **6**, p. 344, 1961.
212. Konigsberg W., Hill R. J. J. Biol. Chem., v. **237**, p. 3151, 1962.
213. Kornberg A., Lehman I. R., Bessman M. J., Simms E. S. Biochim. Biophys. Acta, v. **21**, p. 197, 1956.
214. Krakow J. S., Ochoa S. Proc. Natl. Acad. Sci., v. **49**, p. 88, 1963.
215. Lagerkvist U., Berg P. J. Molec. Biol., v. **5**, p. 139, 1962.
216. Lamfrom H. J. Molec. Biol., v. **3**, p. 241, 1961.
217. Lanni F. Proc. Natl. Acad. Sci., v. **48**, p. 1623, 1962.
218. Lehman I. R., Ann. N.-Y. Acad. Sci., **21**, 745, 1959.
219. Lengyel P., Speyer J. F., Ochoa S., 1961. Proc. Nat Acad. Sci., **47**, 1936.
219a. Lengyel P., Speyer J. F., Basilio C., Ochoa S., Proc. Natl. Acad. Sci., **48**, 282, 1962.
220. Lerman L. S., J. Molec. Biol., **3**, 18, 1961.
221. Lerman L. S., Proc. Natl. Acad. Sci., **49** 94, 1963.
222. Levinthal C., Keynan A., Higa A., Proc. Natl. Acad. Sci., **48**, 1631, 1962.
223. Li, C. H., Cole R. D., Chung D., Lèonis J., J. Biol. Chem., **227**, 207, 1957.
224. Li C. H., Dixon J. S., Chung D., J. Amer. Chem. Soc., **80**, 2587. 1958.
225. Li C. H., Parcells A. J., Papkoff H., J. Biol. Chem., **233**, 1143, 1958.
226. Li C. H., Geschwind J. J., Cole R. D., Raacke J. D., Harris J. J., Dixon J. S., «Nature», **176**, 687. 1955.
226a. Lindenström-Lang K. V., Lane Medical Lectures v. 6, Stanford Univ. Press, Stanford, 1952.
227. Lipshitz R., Chargaff E., Biochim. Biophys. Acta, **42**, 544. 1960.
228. Loeb T., Zinder N. D., Proc. Natl. Acad. Sci., **47**, 282, 1961.

229. Luborsky S. W., Cantoni G. L., Biochim. Biophys. Acta, **61**, 481, 1962.
230. Marks P. A., Burka E. R., Schlessinger D., Proc. Natl. Acad. Sci., **48**, 2163, 1962.
231. Matthaei J. H., Jones O. W., Martin R. G., Nirenberg M. W., Proc. Natl. Acad. Sci., **48**, 666, 1962.
232. Maxwell E. S., Proc. Natl. Acad. Sci., **48**, 1639, 1962.
233. Medvedev Z. A. «Nature», v. **195**, p. 38, 1962.
234. Melchers G. Naturwissenschaften, Bd. **30**, p. 48, 1942.
235. Meselson M., Stahl F. W., Proc. Natl. Acad. Sci., v. **44**, p. 671, 1958.
236. Meselson M., Weigle J. J. Proc. Natl. Acad. Sci., v. **47**, p. 857, 1961.
237. Meselson M. S., Stahl F. W., Vinograd J., Proc. Natl. Acad. Sci., v. **43**, p. 581, 1958.
238. Miescher F. Die histochemischen und physiologischen Arbeiten. Leipzig, 1898.
239. Moldave K. Biochim. Biophys. Acta, v. **43**, p. 188, 1960.
240. Monier R., Stephenson M. L., Zamecnik P. Biochim. Biophys. Acta, v. **43**, p. 1, 1960.
241. Monier R., Naono S., Hayes D., Hayes F. Gros F., J. Molec. Biol., v. **5**, p. 311, 1962.
242. Muller C. J., Kingma S. Biochim. Biophys. Acta, v. **50**, p. 595, 1961.
243. Mundry K. W., Gierer A. Z. Vererbungslehre, v. **89**, p. 614, 1958.
244. Murayama M. Fed. Proc., v. **19**, p. 78, 1960.
245. Natans D., Lipmann F. Proc. Natl. Acad. Sci. v. **47**, p. 497, 1961.
246. Nathans D., Notani G., Schwartz J. H., Zinder N. D. Proc. Natl. Acad. Sci., v. **48**, p. 1424, 1962.
247. Narita K. Biochim. Biophys. Acta, v. **31**, p. 372, 1959.
248. Naughton M. A., Dintzis H. M., Proc. Natl. Acad. Sci., Proc. Natl. Acad. Sci., **48**, 1822. 1962.
249. Neurath H. Adv. in Protein Chem., v. **12**, p. 320, 1957.
250. Nirenberg M. W., Matthaei J. H. Proc. Natl. Acad. Sci., v. **47**, p. 1588, 1961.
251. Nirenberg M. W., Matthaei J. H., Jones O. W. Proc. Natl. Acad. Sci., v. **48**, p. 104, 1962.
252. Nolan C., Smith E. L. J. Biol. Chem.,v **237**, p. 446, 453, 1962.
253. Osawa S. Biochim. Biophys. Acta, v. **42**, p. 244, 1960.
254. Otaka E., Mitsui H., Osawa S. Proc. Natl. Acad. Sci., v. **48**, p. 425, 1962.
255. Pauling L., Corey R. B. Proc. Natl. Acad. Sci., v. **37**, p. 235, 1951.
256. Pauling L., Corey R. B. Proc. Intern. Wool Textile Research Conf., v. **B**, p. 249, 1955.
257. Piez K. A., Davie E. W., Folk J. E., Gladner J. A. J. Biol. Chem., v. **236**, p. 2912, 1961.
258. Rich A. Proc. Natl. Acad. Sci., v. **46**, p. 1044, 1960.
259. Risebrough R. W., Tissiers A., Watson J. D. Proc. Natl. Acad. Sci., v. **48**, p. 430, 1962.
260. Roberts R. B. Proc. Natl. Acad.Sci., v. **48**, p. 897, 1962.
261. Roberts R. B. Proc. Natl. Acad. Sci., v. **48**, p. 1245, 1962.
262. Rychlik I., Sorm F. Coll. Czechosl. Chem. Comm., v. **27**, p. 2686, 1962.
263. Ryle A. P., Sanger F., Smith L. F., Kitai R. Biochem. J. v. **60**, p. 541, 1955.
264. Sanger F. Bull. Soc. Chim. Biol., v. **37**, p. 23. 1955.

265. Sanger F., Thompson E. O. P. Biochem. J., v. 53, 366, 1953.
266. Schachman H. K, Adler J., Rodding C. M., Lehman I. R., Kornberg A. J. Biol. Chem., v. **235**, p. 3242, 1960.
267. Scherrer K., Darnell J. E. Biochem. Biophys. Res. Comm., v. **7**, p. 486, 1962.
268. Scholtissek C., Potter V. R. Z. Naturforsch., v. **15b**, p. **453**, 1960.
269. Schramm G. G., Grötsch H., Pollmann N. Angew. Chem., v. **74**, p. 53, 1962.
270. Schroeder W. A., Shelton J. R., Shelton J. B., Cormik J. Proc. Natl. Acad. Sci., v. **48**, p. 284, 1962.
271. Schulman H. M., Bonner D. M., Proc. Natl. Acad. Sci., v. **48**, p. 53, 1962.
272. Sibatani A., Kloet S. R. de, Allfrey V. G., Mirsky A. E. Proc. Natl. Acad. Sci., v. **48**, p. 471, 1962.
273. Simon E. H. Virology, v. **13**, p. 105, 1961.
274. Sinex F. M. Цитир. по [25].
275. Sinex F. M., van Slyke D. D. J. Biol. Chem., v. **216**, p. 245, 1955.
276. Singer M. F., Cantoni G. L., Biochim. Biophys. Acta, v. **39**, p. 182, 1960.
277. Singer M. F., Jones O. W., Nirenberg M. W. Proc. Natl. Acad. Sci., v. **49**, p. 392, 1963.
278. Sinsheimer R. L. J. Molec. Biol., v. **1**, p. 218, 1959.
279. Smith E. L. Proc. Natl. Acad. Sci., v. **48**, p. 677, 1962.
280. Smith E. L. Proc. Natl. Acad Sci., v. **48**, p. 859, 1962.
281. Sorm F. Coll. Czechosl. Chem. Comm., v. **19**, p. 1003, 1954.
282. Sorm F. Coll. Czechosl. Chem. Comm., v. **24**, p. 3169, 1959.
283. Sorm F. Coll. Czechosl. Chem. Comm., v. **27**, p. 303, 1962.
284. Sorm F., Keil B. Coll. Czechosl. Chem. Comm., v. **23**, p. 1575, 1958.
285. Sorm F., Keil B., Holeysovsky V., Knesslova V., Kostka V., Mäsiar P., Meloun B., Mikes O., Tomasek V., Vanecek J., Coll. Czechosl. Chem. Comm., v. **22**- p. 1310, 1957.
286. Sorm F., Keil B., Holeysovsky V., Meloun B., Mikes O., Vanecek J. Coll. Czechosl. Chem. Comm., v. **23**, p. 985, 1958.
287. Sorm F., Keil B., Vanecek J., Tomasek V., Mikes O., Meloun B., Kostka V., Holeysovsky V. Coll. Czechosl. Chem. Comm., v. **26**, p. 531, 1961.
288. Spencer M., Fuller W., Wilkins M. H. F., Brown G. L. «Nature», v. **194**, p. 1014, 1962.
289. Speyer J. F., Lengyel P., Basilio C., Ochoa S. Proc. Natl. Acad. Sci., v. **48**, p. 63, 1962.
290. Speyer J. F., Lengyel P., Basilio C., Ochoa S. Proc. Natl. Acad. Sci., v. **48**, p. 441, 1962.
291. Streisinger G. Цитир. по [128].
292. Stretton A. O. W., Ingram V. M. «Nature», v. **190**, p. 1079, 1960.
293. Sueoka N. Proc. Natl. Acad. Sci., v. **47**, p. 1141, 1961.
294. Sueoka N., Marmur J., Doty P. «Nature», v. **183**, p. 1429, 1959.
295. Sueoka N., Yamane T., Proc. Natl. Acad. Sci., v. **48**, p. 1454, 1962.
296. Sutherland G. B. B. M., Tsuboi M. Proc. Roy. Soc., v. **239A**, p. 446, 1957.
297. Swartz M. N., Trautner T. A., Kornberg A., J. Biol. Chem., v. **237**, p. 1961, 1962.
298. Symons R. H., Rees M. W., Markham R. Biochem. J., v. **84**, p. 37, 1962.

299. Thompson E. O. P. J. Biol. Chem., v. **208**, p. 565, 1954.
300. Tissiers A., Schlessinger D., Gros F. Proc. Natl. Acad. Sci., v. **46**, p. 1450, 1960.
301. Tissiers A., Watson J. D. Proc. Natl. Acad. Sci., v. **48**, p. 1061, 1962.
302. Tissiers A., Watson J. D., Schlessinger D., Hollingworth B. R. J. Molec. Biol., v. **1**, p. 221, 1959.
303. Tomasek V. Coll. Czechosl. Chem. Commun., v. **22**, p. 1072, 1957.
304. Tso P. O. P., Bonner J., Dintzis H. Arch. Biochem. Biophys., v, **76**, p. 225, 1958.
305. Tsugita A. J Molec. Biol., v. **5**, p. 284, 1962.
306. Tsugita A. J. Molec. Biol., v. **5**, p. 293, 1962.
307. Tsugita A., Fraenkel-Conrat H. Proc. Natl. Acad. Sci., v. **46**, p. 636, 1960.
308. Tsugita A., Fraenkel-Conrat H. J. Molec. Biol., v.**4**, p. 73, 1962.
309. Tsugita A., Fraenkel-Conrat H., Nirenberg M. W., Matthaei H. J. Proc. Natl. Acad. Sci., v. **48**, p. 846, 1962.
310. Tuppy H., Palèus S. Acta Chem. Scand., v. **9**, p. 353. 1955.
311. Tuppy H. Z. Naturforsch., v. **12B**, p. 784, 1957.
312. Tuppy H. Naturwissenschaften, v. **46**, p. 35, 1959.
313. Tuppy H., Bodo G. Monatsh. f. Chem., v. **85**, p. 1182, 1954.
314. Tuppy H., Dus K. Monatsh. f. Chem., v. **89**, p. 407, 1958.
315. Vigneaud V. du, Lawler H. C., Popenoe E. A. J. Amer. Chem. Soc., v. **75**, p. 4880, 1953.
316. Vigneaud V. du, Ressler C., Trippet S. J. Biol. Chem., v. **205**, p. 949, 1953.
317. Vischer E., Zamenhof S., Chargaff E. J. Biol. Chem., v. **177**, p. 429, 1949.
318. Volkin E., Astrachan L., Virology, v. **2**, p. 149, 1956.
319. Wahba A. J., Basilio C., Speyer J. F., Lengyel P., Miller R. S., Ochoa S. Proc. Natl. Acad. Sci. v. **48**, p. 1683, 1962.
320. Wahba A. J., Gardner R. S., Basilio C., Miller R. S. Speyer J. F., Lengyel P., Proc. Natl. Acad. Sci., **49**, 116, 1963.
321. Waldschmidt-Leitz E., Gauss K., Z. Physiol. Chem., **293**, 10, 1953.
322. Wall R., Nature, **193**, 1269, 1962.
323. Warner J. R., Knopf P. M., Rich A., Proc. Natl. Acad. Sci., **49**, 122. 1963.
324. Warner R. C., Samuels H. H., Abbott M. T., Krakow J. S., Proc. Natl. Acad. Sci., **49**, 533, 1963.
325. Watson H. C., Kendrew J. C. «Nature», v. **190**, p. 670, 1961.
326. Watson J. D., Crick F. H. C. «Nature», v. **171**, p. 737, 1953.
327. Webster G. Fed. Proc., v. **20**, p. 387, 1961.
328. Weinstein B., Schechter A. N. Proc. Natl. Acad. Sci., v. **48**, p. 1686, 1962.
329. Webster G., Lingren J. B. In: Protein Sinthesis, Acad. Press, N. Y., 1961.
330. Weisblum B., Benzer S., Holley R. W., Proc. Natl. Acad. Sci., **48**, 1449, 1962.
331. Weiss S. B., Gladstone L., J. Amer. Chem. Soc., **81**, 4118, 1959.
332. Weiss S. B., Nakamoto T., Proc. Natl. Acad. Sci., **47**, 1400, 1960.
333. Weiss S. B., Nakamoto T., Proc. Natl. Acad. Sci., **47**, 694. 1961.
334. White W. F., Landmann W. A., J. Amer. Chem. Soc., **77**, 1711, 1955
335. White W. F., Shields J., Robbins K. C. J. Amer. Chem. Soc., **77** 1267, 1955.

336. Wilkins M. F. H., Stokes A. R., Wilson N. R. «Nature», v. **171**, p. 738, 1953.
337. Williams J., Clegg J. B., Mutch M. O. J. Molec. Biol., v. **3**, p. 532, 1961.
338. Wittmann H. G. Z. Vererb.-Lehre, v. **90**, p. 463, 1959.
339. Wittmann H. G. Virology, v. **12**, p. 609, 1960.
340. Wittmann H. G. Ber. Dtsch. Bot Ges., v. **73**, p. 15, 1960.
341. Wittmann H. G. Naturwissenschaften, v. **48**, p. 729, 1961.
342. Wittmann H. G. Z. Vererb.—Lehre, v. **93**, p. 491, 1962.
343. Wittmann H. G. Naturwissenschaften, v. **50**, p. 76, 1963.
344. Woese C. R. Biochem. Biophys. Res. Comm., v. **5**, p. 88, 1961.
345. Woese C. R. «Nature», v. **194**, p. 1114, 1962.
346. Yankofsky S. A., Spiegelman S. Proc. Natl. Acad. Sci., **48**, 1069, 1962.
347. Yankofsky S. A., Spiegelman S. Proc. Natl. Acad. Sci., v. **49**, p. 538, 1963.
348. Ycas M. Int. Rev. Cytol., v. **13**, p. 1, 1962.
349. Ycas M., Vincent W. S. Proc. Natl. Acad. Sci., v. **46**, p. 804, 1960.
350. Ycas M. In: Information Theory in Biology, p. 70, 1956.
351. Zillig W., Schachtschnabel D., Krone W. Z. Physiol. Chem., v. **318**, p. 100, 1960.
352. Zubay G., Quastler H. Proc. Natl. Acad. Sci., v. **48**, p. 461, 1962.
353. Bishop J., Leaky J., Schweet R. Proc. Natl. Acad. Sci., v. **46**, p. 1030, 1960.
354. Swenson R. T., Hill R. L., Lehman H., Jim R. T. S. J. Biol. Chem. Цитир. по [279].
354a. Midgley J. E. Цитир. по [260].
355. Sueoka N. Proc. Natl. Acad. Sci., v. **48**, p. 582, 1962.

SUMMARY

STATISTICAL ANALYSIS OF AMINO ACID CODE PROBLEM

N. V. Luchnik

A critical discussion of available data related with the problem of amino acid code, as well as an attempt to reveal the general features of the code and to decipher it by statistical means are made in the summarized book. In the abstract to follow, the book is recapitulated by chapters, and particular attention is paid to new results and conclusions attained.

Chapter 1: Contemporary views on the mechanism of protein biosynthesis.

Chapter 2: Contemporary state of the coding prvblem.

Both these chapters contain no original data.

Chapter 3: Basic concepts and formulation of the problem. In this chapter the terms being related with the coding proble m are defined; among these the most important and new ones are the following: *obligatory letters* are such letters of a codeword which cannot be changed without altering the meaning of this word, *facultative letters* are the letters which can be changed for any other (or some definite one) without altering the meaning of the word, *permitted replacements* of amino acids are such replacements which originate as a result of a single letter exchange in a codeword (i. e. transitions and transversions), all other replacements despite of their occurence in Nature we call *forbidden* ones. In the same chapter a plausibility of modern viewpoints on the general features of the code is critically considered. Only statements about consecutive reading of the text and universality of the code seem to be firmly grounded. All the rest require further confirmations. The more obscure feature of the code is the degree of its degeneracy. On the basis of these considerations the problem under investigation is formulated.

Chapter 4: Statistics of amino acids replacements. In this chapter a compilation of modern data about amino acid replace-

ments (i) arising as a result of mutations (table 5: TMV, table 6: human haemoglobins, table 7: triptophansynthetase, table 8: general survey) and (ii) observed in comparison of natural proteins (table 9: chains of human haemoglobine, table 10: other similar polypeptides, table 11: analogous proteins in various species, table 12: general survey) is given. Statistical analysis shows the heterogeneity of these data (fig. 8). It prooves that the main cause for this heterogeneity is various frequency of natural occurence of amino acids in proteins (table 14 and fig. 9).

Chapter 5: On the degree of degeneracy of the code. If all observable replacements are permitted and all of them are found with the same probability, than

$$R = \frac{-\ln (1-P)}{P},$$

where R is mean frequency of observed replacements and P is a fraction of observed types of replacements among all permitted ones. It follows from this equation and the available data on the statistics of the replacements that the number of permitted replacements (N) is 78. This estimation is rather crude because the real situation may be far from the above simplification. If one takes into consideration the frequency of occurence of amino acids, assuming that all the observed replacements are permitted, than N is about 90. The number of forbidden replacements can be estimated by comparing the expected and observed frequency of occurence of individual replacements (table 19). Assumption that all replacements with reduced frequencies are forbidden gives value for N about 50. Two last figures (50 and 90) are under- and overestimations of the real value, which should lie between them, i. e. about 70.

As the first method is indirect, four other methods are applied for the same purpose. In the second method (table 21) the number of replacements is estimated on the basis of frequency of coincidences of the same replacements in two independent samples. In the third method the analysis of distribution of deviations between observed and expected frequencies of individual replacements is used (fig. 10). In the fourth method (table 23) the same estimation is made on the ground of comparison of mean frequencies of replacements with participation of common amino acids, rare amino acids and both of them. In the fifth method, at last, investigation of distribution of replacement frequencies for individual amino acids is employed (fig. 11).

All these methods give similar results and N can be estimated confidently enough to be lying near 70 (table 24). If all other general features of the code were known with certainty, it would be possible to estimate the degree of its degeneracy from the

above figures; but unfortunately it is impossible to realize for the questionability of some statements on the general features of the code. However the estimation of N can reduce considerably the number of acceptable hypotheses about the code, which are considered later.

Chapter 6: On the general features of the code. In this chapter two methods are applied for determining the general features of the code.

First method. If two amino acids give permitted replacement it follows that the words which code them differ in a single letter. If two amino acids do not give replacements with each other but both give replacements with some third one the words which code them should differ in two letters and so forth. We have calculated the number of replacements which bound each possible amino acids pair (table 28). The comparison of distribution of observed frequencies of different types of amino acids pairs (table 29) with expected ones for various hypotheses about the general features of the code (table 30) gives possibility to obtain some information as for the length of codewords and the degree of code degeneracy.

Second method. A similar information can be recieved from a generalization of Zubay-Quastler lattice (ref. 352). If one arranges the amino acids in such a way that amino acids which give permitted replacements are placed in the knots of the same lattice edges, than the number of dimensions of the lattice will correspond to number of letters in the word; the length of edges will be equal to number of letters in alphabet (fig. 12).

Both these methods as well as the data of preceeding chapter give the same paradoxal result: the degree of degeneracy of the code is high, but nevertheless it has some properties of doublet code. In order to reconcile that, we put forward the following hypothesis. The code is highly degenerate and the code number is three, but as a rule only two letters in triplet are «obligatory» for coding. The third («facultative») letter is rather indifferent. This hypothesis is in agreement with all data so far available and gives basis for decoding the meaning of triplets.

Chapter 7: Determining of obligatory letters in codewords. The problem of deciphering of the code is divided into two parts. First, the obligatory letters are determined, and only then the meaning of tripets is decoded. The obligatory letters are determined in two independent ways outlined below.

Qualitative analysis of the experiments with artificial polyribonucleotides. Comparison of two independent classifications of amino acids and pairs of nucleotides.

This analysis makes it possible to determine the letters, the presence of which is necessary and sufficient for incorporation of definite amino acids into proteins (tables 34—35).

For instance, if we know that pairs of obligatory etters GG, CC, GC, CG correspond to ala, arg, gly, pro, and the obligatory pairs AA, CC, AC, CA correspond to asp-NH_2 lys, pro, tre, then by comparing these groups it becomes clear that the obligatory pair for proline is CC, as only pro and CC are common for both groups. The available data give possibility to classify the amino acids and corresponding pairs of letters in two independent ways. The results of well-known experiments of Sueoka (ref. 293) revealing correlations between G + C and amino acids content in various bacteria (table 36) give a basis for one classification. Another classification can be made on a basis of data on experimental induction of mutations with HNO_2 (fig. 13). Each classification gives three groups of amino acids and corresponding pairs of nucleotides. By making all possible comparisons (tables 37—39) the obligatory pairs for the most amino acids are found.

The results of both methods agree well with one another. A relative order of letters in these pairs may be obtained from the analysis of data on replacements (fig. 14—15). Choise between two possible absolute order (table 40) is made on a ground of the experiments of Wahba *et al.* (ref. 319).

Chapter 8: Deciphering of the triplet code. This chapter is started by a discussion of the problem of deciphering of the degenerate code. Let us consider the results of experiments with two-component artificial polynucleotide, say poly-AB. This polynucleotide contains 8 various triplets, relative number of which depends on a quantitative composition of poly-AB. If we suppose that the degeneracy takes place only at the cost of the facultative letters, then we can obtain, theoretically, 22 possible comparisons of the yields of two amino acids which are enumerated in table 41. One can calculate the expected dependence of relative yield from A content in poly-AB. It appears that only 10 various dependencies correspond to 22 various comparisons (fig. 16). In spite of this ambiguity it seems to be impossible to analyse such data with the common methods. The problem can be solved, however, when we, first, compare yields only of such amino acids which have the obligatory pairs of the same composition and, second, when we analyse the whole curve indicating the dependence of relative yield on the composition of poly-AB, but not the individual yields. With such selection being made, only curves of I—IV types may be met which are easily distinguishable one from another despite of considerable experimental errors. Such procedure applied to available data enables to decipher a number of triplets (fig. 17—22 and tables 42—48). It is most probable that every facultative third letter occupies the second place in triplet. Additional triplets are get from analysis of data on replacements (fig. 23—24 and table 49), and from experiments with three-com-

ponent polymers (tables 50—53). The result of deciphering is represented in table 54. The meaning of 5 triplets is still unknown, some of them are uncertain. A small number of triplets code no amino acids. The incompleteness of deciphering is connected with shortage in experimental data but not with the capability of the proposed methods.

Chapter 9: Summary and discussion After brief recapitulation of the course of the analysis, the results which were obtained are compared with some experimental facts and a good agreement is found.

In the course of this work we have obtained some information about the general properties of the code. These conclusions obtained by statistical means agree well with those achieved by Crick *et al.* (ref. 128) with the aid of quite different and independent methods. Besides, the new property was found, namely the unequal role of letters in a codeword. This property permits to make some definite conclusions concerning the degree and the mode of the code degeneracy and give more effective methods for its deciphering.

ОГЛАВЛЕНИЕ

Предисловие 3

Глава 1. **Современные представления о механизме биосинтеза белка** 5

Проблема биосинтеза белка . . . . . —
Механизм биосинтеза белка 16

Глава 2. **Современное состояние проблемы аминокислотного кода** . . 24

Первые подходы к решению проблемы . . . . . —
Экспериментальные подходы к расшифровке кода 32

Глава 3. **Основные понятия и постановка задачи** 45

Термины и определения . . . . . —
Сокращенные обозначения 47
Обоснованность представлений об общих свойствах кода 48
Постановка задачи 51

Глава 4. **Статистика аминокислотных замещений** 55

Глава 5. **О степени вырожденности кода** 70

Число разрешенных замещений 70
О степени вырожденности кода 86

Глава 6. **Об общих свойствах кода** 91

Число переходов и величина слов . . . . . —
Обобщение решетки Зьюби-Квастлера 96
Общие свойства кода 99

Глава 7. **Определение облигатных букв в кодовых словах** 106

Качественный анализ результатов опытов с искусственными полинуклеотидами . . . . . —
Сравнение независимых классификаций аминокислот и облигатных пар нуклеотидов 110
Порядок букв в облигатных парах 114

Глава 8. **Расшифровка триплетного кода** 118

Количественный анализ результатов опытов с искусственными нуклеотидами 119
Анализ данных опытов с двухкомпонентными полимерами 123
Анализ данных по замещениям 134
Анализ данных опытов с трехкомпенентными полимерами 138

Глава 9. **Итоги и обсуждение** 144

Итоги . . . . . —
Обсуждение 148
Выводы 152

Литература 153
Summary 164

Н. В. Лучник

СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРОБЛЕМЫ АМИНОКИСЛОТНОГО КОДА

Редактор изд-ва *Л. П. Ардасенова* Техн. редактор *Н. Ф. Тамкова*
Корректоры: *Н. В. Семенова, П. В. Винокурова*

Сдано в производство 24/VIII 1963 г. Подписано к печати 13/II 1964 г.
Бумага 60 × 90 1/16 Объем п. л. 10,75
НС 23023 Тираж 1200 Цена 88 коп. Заказ 2325

Свердловская типография № 1 Облполиграфиздата. г. Свердловск, Дом промышленности