

### § 9.1. ПРОБЛЕМА ГЕНЕТИЧЕСКОГО КОДА

Ген (цистрон, см. [1, 2]) есть часть макромолекулы ДНК, ответственная за синтез белковой цепи. Является ли роль ДНК инструктивной или разрешающей? Многочисленные данные свидетельствуют в пользу первой возможности. Синтез белка программируется молекулой ДНК, содержащей информацию о первичной структуре белковых цепей, иначе говоря, генетическую информацию. Эта информация записана в первичной структуре ДНК, т. е. в последовательности нуклеотидов. В частицах вирусов РНК-типа генетическая программа заложена в молекулах РНК.

Каково соответствие между последовательностью нуклеотидов в ДНК (и вирусной РНК) и последовательностью аминокислотных остатков в белковой цепи? В этом и состоит проблема генетического кода.

Это физическая проблема. Во-первых, ставится вопрос о соответствии информационного содержания ДНК и белка, во-вторых — о количественных взаимоотношениях между нуклеотидами и аминокислотами, определяемых в конечном счете взаимодействиями молекул в матричном синтезе белка. В-третьих, возникает вопрос о физическом смысле генетического кода, т. е. о степени неслучайности этого соответствия. Характерно, что проблема генетического кода была впервые сформулирована физиком Гамовым [3], и ряд физиков принял участие в ее решении. Однако код был расшифрован биологическими и химическими методами.

Постановка вопроса о коде, определяющем атомно-молекулярное строение и свойства вещества, типична для физики. Химические свойства атомов (т. е. структура периодической системы Менделеева) закодированы числом электронов в атоме в соответствии с принципами квантовой механики. То же относится к атомным спектрам. Зная последовательность квантовых уровней, мы получаем кодовые условия для химии и оптики. Число и природа нуклонов в атомном ядре кодируют свойства ядра. Систематика элементарных частиц и их превращений — одна из актуальных проблем современной физики — также

означает нахождение некоторого кода. Для кодовых проблем характерна обязательность реализуемых количественных взаимоотношений.

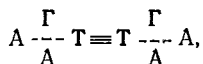
Первая попытка нахождения генетического кода, принадлежавшая Гамову [3,4], была предпринята сразу же после открытия двуспиральной структуры ДНК. Гипотеза Гамова оказалась несостоятельной. Тем не менее она заслуживает внимания, так как ею был поставлен ряд важных проблем. Работа Гамова сыграла большую роль и в привлечении внимания физиков к молекулярной биологии.

Белковый текст написан 20-буквенным алфавитом, текст ДНК (или РНК) — 4-буквенным. Из элементарных соображений следует, что кодовое отношение, т. е. число нуклеотидов, кодирующих одну аминокислоту, не может быть меньше трех. Количество информации, приходящейся на одну аминокислоту, равно  $i_a = \log_2 20 = 4,322$  бит, а на один нуклеотид  $i_n = \log_2 4 = 2,000$  бит. Следовательно, наименьшее число нуклеотидов, приходящееся на одну аминокислоту, равно  $i_a/i_n = 2,161$ .

Аминокислота не может кодироваться нецелым числом нуклеотидов. Значит, кодовое отношение должно быть не меньше трех. В самом деле, число размещений с повторениями из четырех по два равно 16, что меньше 20 — числа аминокислот. Число размещений с повторениями из четырех по три равно 64. Трех нуклеотидов в кодоне, т. е. в совокупности нуклеотидов, кодирующих одну аминокислоту, более чем достаточно.

Гамов предполагал, что белковая цепь синтезируется непосредственно на двойной спирали ДНК, причем каждая аминокислота располагается в выемке между четырьмя нуклеотидами. Эта выемка имеет примерно ромбическую форму. Два нуклеотида принадлежат одной цепи, два — другой. Один из нуклеотидов первой цепи образует уотсон-криковскую пару с одним нуклеотидом второй цепи.

«Бубновый код» Гамова обеспечивает именно 20 «букв». Каждая «буква», т. е. ромб, состоит из четырех нуклеотидов. Число сочетаний из 4 по 4 равно 256. Но здесь наложены ограничения — малая диагональ ромба соединяет А с Т или Г с Ц. Если считать тождественными правые и левые формы несимметричных ромбов, например



то получится 20 различных ромбов.

«Бубновый код» перекрывающийся. Так как в каждый ромб входят нуклеотиды из трех последовательных пар, два нуклеотида, находящиеся на одной стороне ромба, являются общими для двух соседних ромбов. Тем самым, возникает корреляция

между соседними аминокислотами. За данным остатком может следовать не любой из 20, но лишь некоторые (см. [1]).

Гамов попытался проверить правильность своего кода, сопоставив возможность сочетания ромбов с известной первичной структурой инсулина и адренкортикотропина. При этом возникли неразрешимые противоречия. Дальнейшие исследования показали, что никакие перекрывающиеся коды нельзя согласовать с опытом. Наличие перекрытий в кодонах может выражаться в корреляциях между соседними аминокислотными остатками. Иными словами, некоторые парные сочетания остатков должны быть запрещены. Анализ первичных структур белков показал, что таких корреляций нет — любой остаток может следовать за любым, хотя разные остатки встречаются с различными частотами [4, 5]. Можно, однако, представить себе перекрывающиеся нуклеотидные коды, допускающие любую последовательность аминокислот [6].

Допустим, что кодон содержит четыре нуклеотида, но последний нуклеотид данного кодона одновременно является первым в следующем кодоне. Кодовое отношение равно трем. Если все кодирующие квартеты начинаются, скажем, с Г или У, то они должны заканчиваться также Г и У. Общее число таких кодонов равно 64 ( $2 \cdot 16 \cdot 2$ ). Эти кодоны распределяются по 20 аминокислотам так, что каждая из них имеет по крайней мере один квартет, начинающийся с Г, и один, начинающийся с У. Так как каждый квартет должен заканчиваться также Г или У, возможны любые последовательности аминокислот.

Отсутствие корреляций в полипептидных цепях, таким образом, нельзя считать решающим аргументом против перекрывающихся кодов. Но отсутствие корреляций делает невозможной умозрительную расшифровку кода.

Более существенный аргумент состоит в том, что не удается наблюдать мутационные замещения двух соседних аминокислотных остатков в белке. Между тем мутация нуклеотида, общего для двух соседних аминокислот, должна приводить к такому двойному замещению. Уолл привел соображения, снижающие силу этого аргумента [6]. Однако, поскольку не было найдено ни одного экспериментального доказательства перекрывания, есть веские основания считать генетический код неперекрывающимся. Полная расшифровка кода это подтвердила.

Изучение мутаций показало, что код коллинеарен, т. е. кодоны в нуклеиновой кислоте и соответствующие аминокислотные остатки в белке расположены в той же линейной последовательности. Ряд доказательств этого положения приведен в монографии Ичаса [5]. Там же описаны другие попытки умозрительного нахождения кода, представляющие сегодня исторический интерес (см. также [1, 7]).

Кодовое отношение может быть найдено лишь экспериментально. Это было сделано в результате генетического исследования, проведенного Криком с сотрудниками [8], изучавшими цистрон *B* в области *r* II бактериофага T4, паразитирующего на *E. coli*. Ранее Бензер детально проанализировал генетические свойства системы фаг T4 — *E. coli* [9]. Среди точечных мутаций *r* II есть мутации, состоящие в выпадении нуклеотидов (делеции) и в их добавлении к цепи ДНК. Такие мутации могут быть вызваны акридиновыми красителями.

Дикий тип фага  $\psi$  размножается на штаммах *B* и *K* 12 ( $\lambda$ ) *E. coli*. Мутантные фаги *r* размножаются только на *B*-штаммах, образуя резко ограниченные бляшки. Мутанты *FCO*, индуцируемые профлавином, относятся к типу *r*. Они обладают способностью спонтанно ревертировать, возвращаться к дикому типу  $\psi$ . Генетический анализ показал, что такие ревертанты возникают не в результате обратной мутации  $r \rightarrow \psi$ , но вследствие появления второй супрессорной мутации вблизи первой мутации  $\psi \rightarrow r$ . Супрессоры относятся к тому же фенотипу *r*, что и супрессируемые ими мутации. Каждая из двух мутаций порознь приводит к утрате способности синтезировать соответствующий белок, но сочетание двух мутаций в одном цистроне эту способность восстанавливает. Всего было изучено около 80 *r*-мутантов, в том числе двойные и тройные их комбинации — супрессоры супрессоров и супрессоры супрессоров супрессоров. Все супрессоры оказались относящимися к двум классам: + (добавление нуклеотида) и — (делеция). Если исходная мутация *r* есть +, то ее супрессор — и наоборот. Дикий фенотип дает комбинации + —, — +, + + +, — — —, но не + +, — —, + + + +, — — — — и т. д.

Эти факты нетрудно объяснить, если допустить, что код триплетный, неперекрывающийся и «читаемый», начиная с некоторого фиксированного нуклеотида. Представим цепь ДНК последовательностью букв ABC ABC ... (рис. 9.1). Чтение кода, начиная с определенной буквы, эквивалентно наложению на эту последовательность рамки с прорезью. Если одна из букв выпала (—) или, напротив, добавилась новая (+), то вся последовательность, начиная с места мутации, искажена, т. е. нормальный белок дикого типа не может синтезироваться (рис. 9.1, б). Если появилась супрессорная мутация (— + или + —), то последовательность искажена лишь в области между двумя мутациями (рис. 9.1, г и д). Если искаженная область не слишком длинна, то синтезируемый белок может сохранить свою функцию и будет наблюдаться реверсия. Рис. 9.1, е, ж соответствуют реверсии при наличии трех однотипных мутаций — — — или + + +. Понятно также, почему двойные, четверные и пятерные однотипные мутации не дают реверсий.

Из этих результатов следует кодовое отношение, кратное трем. Естественно предположить, что оно равно именно трем. Впрочем, Уолл показал, что данные Крика и сотрудников согласуются и с предложенным им квартетным перекрывающимся кодом [5, 6]. Однако код Уолла непригоден по другим причинам.

Ряд деталей, относящихся к описанным «мутациям сдвига рамки» (frame shift mutations), содержится в оригинальной работе [8] и в монографиях [1, 5]. Вначале отсутствовали прямые

С	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC	А	а)
С	ABC	A-CA	B-CA	B-CA	B-CA	B-CA	B-CA	B-CA	B-CA	Б	б)
С	ABC	ABC	ABC	ABC	A+AB	СAB	СAB	СAB	СAB	С	в)
С	ABC	A-CA	B-CA	B-CA	+ABC	ABC	ABC	ABC	ABC	А	г)
С	ABC	ABC	A+AB	СAB	СAB	С-BC	ABC	ABC	ABC	А	д)
С	ABC	A-CA	B-CA	B-AB	СAB	С-BC	ABC	ABC	ABC	А	е)
С	ABC	AB+B	СAB	CA+A	B-CA	BC+C	ABC	ABC	ABC	А	ж)

Рис. 9.1. Тексты ДНК.

а) Дикий тип; б) мутант (делеция); в) мутант с добавлением; г), д) двойные мутанты типа + -; е) тройной мутант - - -; ж) тройной мутант + + +. Заштрихованы искаженные участки текста.

доказательства белок-синтезирующей роли области  $rII$  цистрона. В дальнейшем Терзаги и сотрудники [10] изучили аналогичные мутации цистрона фага T4, продуцирующего лизоцим. Оказалось, что «сдвиги рамки» действительно искажают белковый текст. Двойной ревертированный мутант лизоцима отличается от дикого типа последовательностью шести остатков

Дикий  $\omega$ -тип ... Тре—Лиз—Сер—Про—Сер—Лей—Аси—Ала ...

Ревертант

(псевдо  $\omega$ -тип) ... Тре—Лиз—Вал—Гис—Гис—Лей—Мет—Ала ...

Эти биологические исследования привели к решению ряда физических вопросов. Установлено кодовое отношение, еще раз доказана коллинеарность кода, доказано, что код читается, начиная с определенного нуклеотида и не имеет запятых, т. е. материальных единиц, разделяющих кодоны. Из описанных работ следовало, что кодоном является линейная последовательность трех нуклеотидов. Общее число разных кодонов равно 64.

В связи с этим встал вопрос о вырождении кода. Так как общее число кодонов (64) больше числа аминокислот (20), то

нужно выяснить, сколькими различными кодонами кодируется каждая аминокислота. Следует обратить внимание на типично физическую постановку вопроса и физический термин «вырождение» в отличие от биологического смысла этого слова.

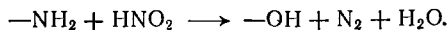
То же исследование мутаций сдвига рамки свидетельствует в пользу вырожденного кода. Это следует с особенной ясностью из работы [10]. Если код не вырожден, то оба кодона для Гис в ревертанте лизоцима должны быть одинаковыми, скажем АВС. Если сдвиг вызывается добавлением нуклеотида слева и кодон для Про есть ВСА, то соотношение между диким типом и ревертантом изобразится так:

... Про—Сер ...  
 А ВСА ВСА  
 АВС АВС  
 ... Гис—Гис ...

Но отсюда следует, что кодон для Сер есть также ВСА. Противоречие можно устранить, только допустив наличие различных кодонов для Гис.

Вырождение кода доказывается и тем, что при значительных вариациях в составе ДНК различных организмов доля Г + Ц варьирует от 25 до 75% (см. стр. 90), средний состав белков меняется мало. Вместе с тем ряд фактов указывает на универсальность кода. Предположение о том, что значительная часть нуклеотидов ДНК не функциональна, трудно согласовать с однородностью состава ДНК у данного биологического вида. Напротив, эти факты находят естественное объяснение, если код сильно вырожден.

Другие косвенные данные, указывающие на вырождение кода, получены при изучении мутаций, как индуцированных химическими агентами, так и спонтанных. Виттман исследовал мутации в белке вируса табачной мозаики (ВТМ), вызванные нитритом [11]. Под действием азотистой кислоты аминная группа заменяется на гидроксильную



Тем самым нитрит превращает Ц в У, Г в ксантин и А в гипоксантин. После репликации ксантин и гипоксантин замещаются на Г. Следовательно, происходят замены Ц→У и А→Г: 64 триплетных кодона можно представить восемью октетами. На рис. 9.2 изображены эти октеты и стрелками показаны нитритные точечные мутации. Наблюдаемые замещения в белке ВТМ соответственно располагаются по тем же октетам. Виттман установил, что по крайней мере Сер и Иле должны кодироваться несколькими триплетами (рис. 9.3), т. е. код вы-

рожден. Эти данные одновременно показывают, что большая часть замещений в белке происходит в результате замены одного нуклеотида в кодоне.

Всего может быть 190 (20·19:2) замещений одной аминокислоты на другую. Часть из них является результатом замены

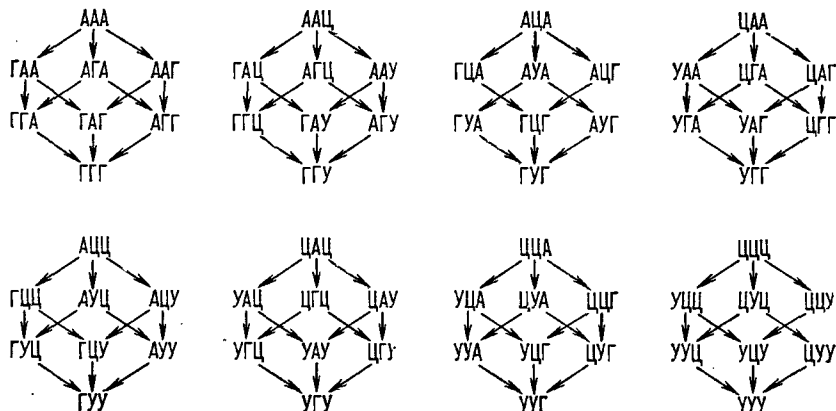


Рис. 9.2. Классификация кодонов по октетам.

одного нуклеотида в триплетном кодоне («разрешенные» замены), остальные замещения требуют изменения двух или даже трех нуклеотидов. Экк проанализировал большое число известных мутаций в белках, считая, что «разрешенные» замены должны быть распределены равномерно [12]. По его оценке

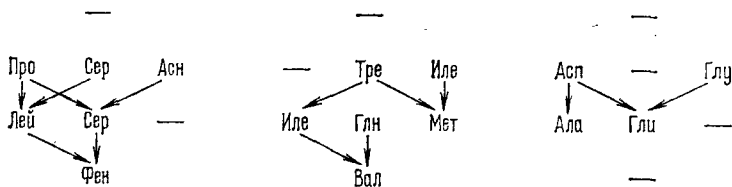


Рис. 9.3. Отнесение аминокислотных остатков к октетам на основании нитритных мутаций ВТМ.

число разрешенных замещений, практически реализующихся в белках,  $75 \pm 20$ . Более подробный статистический анализ также дал значение, близкое к 70 [13]. Само по себе это число свидетельствует о вырождении кода, но не указывает степень вырождения — число кодонов, приходящихся на одну аминокислоту. Однако статистический анализ показывает, что наряду с 75 разрешенными заменами имеется примерно 103 замены, определяемые изменениями двух нуклеотидов в кодоне и всего

лишь 12, определяемых изменениями трех нуклеотидов. Такое распределение согласуется как с двойным (40 кодонов), так и с тройным (60 кодонов) равномерным вырождением [13].

Итак, генетический код является *триплетным, непрерывающимся, лишенным запятых и вырожденным*. Полная расшифровка кода потребовала прямых биохимических опытов, ставших возможными лишь после раскрытия механизма биосинтеза белка.

## § 9.2. БИОСИНТЕЗ БЕЛКА

Современная молекулярная биология белкового синтеза изложена в монографии [14] (см. также [2]). Здесь мы ограничимся краткими сведениями, уделив в последующем изложении главное внимание физическим вопросам.

Информация, содержащаяся в ДНК, передается мРНК в процессе *транскрипции*. Белок синтезируется на мРНК и, таким образом, в этом процессе реализуется информационная цепь ДНК → РНК → Белок.

Первые экспериментальные доказательства существования мРНК получили Белозерский, Спирин и их сотрудники [15, 16]. Они показали, что состав тотальной РНК бактерий коррелирует с составом их ДНК и пришли к заключению о наличии по крайней мере двух типов РНК, один из которых (большая фракция) имеет состав, не отражающий состава ДНК, а второй воспроизводит состав ДНК. В дальнейшем выяснилось, что первая фракция это рибосомальная РНК (рРНК), а вторая — мРНК. Матричная функциональность РНК следовала из опытов с растительными вирусами. Было установлено, что РНК, выделенная из ВТМ, обладает инфицирующим действием. Это значит, что молекулы вирусной РНК, попадая в клетки растения, организуют синтез новых частиц ВТМ, т. е. синтез их белков [17, 18]. Френкель-Конрат восстановил инфицирующие свойства частиц ВТМ, смешивая РНК ВТМ с белком, и получил гибридные вирусы, комбинируя РНК одного штамма ВТМ с белком другого штамма [19].

Волкин и Астрахан изучали события, происходящие в клетке *E. coli* после заражения фагом Т2. При размножении фага синтезируется некоторое количество РНК, обладающей повышенной метаболической активностью: она быстро включает Р<sup>32</sup>. Состав синтезируемой РНК сходен с составом фаговой ДНК [20, 21].

В дальнейшем мРНК была обнаружена и в незараженных бактериальных клетках [22]. Ее состав аналогичен составу бактериальной ДНК. Было показано, что мРНК непосредственно синтезируется на ДНК как на матрице. Это доказывается, во-первых, наличием специального фермента — РНК-полимеразы