

ФИЗИКА БИОСИНТЕЗА БЕЛКА

§ 8.1. Проблема генетического кода

Ген есть часть макромолекулы ДНК, ответственная за синтез одной белковой цепи. ДНК содержит информацию о первичной структуре белка — генетическую информацию. Она записана в первичной структуре ДНК, т. е. в последовательности нуклеотидов. В вирусах РНК-типа генетическая программа заложена в молекулах РНК.

Каково соответствие между последовательностью нуклеотидов в ДНК и 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 — молекулярным узнаванием. В-третьих, возникает вопрос о физическом смысле генетического кода, в-четвертых, о его эволюционном происхождении.

Проблема генетического кода была впервые сформулирована физиком (Гамов, 1953) и ряд физиков принял участие в ее решении. Однако код был расшифрован не физическими, но биологическими и химическими методами.

Белковый текст написан алфавитом из двадцати букв, текст ДНК (или РНК) — из четырех. Из элементарных соображений следует, что кодовое отношение, т. е. число нуклеотидов, кодирующих один аминокислотный остаток, не может быть меньше трех. В самом деле, число сочетаний из 4 по 2 равно $4^2 = 16 < 20$, число сочетаний из 4 по 3 равно $4^3 = 64 > 20$.

Гамов предположил, что белковая цепь собирается 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 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 $4^4 = 256$. Но имеются ограничения, так как малая диагональ ромба обязательно соединяет А с Т

или Г с Ц. Если считать тождественными правые и левые формы ромбов, например,

$$\begin{array}{ccc} & \Gamma & \\ \text{A} \cdots \text{T} & \equiv & \text{T} \cdots \text{A}, \\ & \text{A} & \end{array}$$

то получится лишь 20 различных ромбов.

«Бубновый код» перекрывающийся. Так как в каждый ромб входят нуклеотиды из трех последовательных пар, два нуклеотида, находящихся на одной стороне ромба, являются общими для двух соседних ромбов. Тем самым возникает корреляция между двумя соседними аминокислотными остатками. За данным остатком может следовать не любой из 20, но лишь некоторые. Однако дальнейшие исследования показали, что перекрывающийся код нельзя согласовать с опытом, так как любой аминокислотный остаток может следовать за любым. Отсутствие корреляции делает невозможной теоретическую, умозрачительную расшифровку кода.

Невозможность перекрывающегося кода следует также из отсутствия мутационных замещений двух соседних аминокислотных остатков в белке. Мутация нуклеотида, общего для двух соседних остатков, должна приводить к такому двойному замещению.

Кодовое отношение было найдено экспериментально в результате генетического исследования, проведенного Криком с сотрудниками (1961), изучавшими область rII генома фага T4, размножающегося в культурах *E. coli*. Было установлено, что мутации в этой области, вызываемые акридиновыми красителями, состоят в выпадении, делеции, нуклеотидов и в их добавлении. Дикий тип w размножается на штаммах *B* и *K12 E. coli*. Мутанты r размножаются только на *B*-штаммах, образуя резко очерченные бляшки. Некоторые из мутантов этого типа способны спонтанно возвращаться к дикому типу w . Генетический анализ показал, что такие ревертанты возникают не в результате обратной мутации $r \rightarrow w$, но вследствие появления второй супрессорной мутации $w \rightarrow r$ вблизи первой. Каждая из двух мутаций порознь приводит к утрате способности синтезировать соответствующий белок, но сочетание двух мутаций в одном гене 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 дают комбинации +-, -+, +++, ---, но не ++, --, +++, ----.

Эти факты объясняются, если допустить, что код триплетный, неперекрывающийся и «читаемый» подряд, начиная с некоторого фиксированного нуклеотида. Представим цепь ДНК последовательностью букв ABCAVS... (рис. 8.1, а). На месте каждой буквы может находиться любой из четырех нуклеотидов А, Т, Г, Ц. Чтение кода, начиная с определенной буквы, эквивалентно нало-

жению 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. 8.1, б, в). Если появилась супрессорная мутация (—+ или

С	АВС	АВС	АВС	АВС	АВС	АВС	АВС	АВС	А	<i>а</i>
С	АВС	А-СА	ВСА	ВСА	ВСА	ВСА	ВСА	ВСА	В	<i>б</i>
С	АВС	АВС	АВС	АВС	А+АВ	САВ	САВ	САВ	С	<i>в</i>
С	АВС	А-СА	ВСА	ВСА	+АВС	АВС	АВС	АВС	А	<i>г</i>
С	АВС	АВС	А+АВ	САВ	САВ	С-ВС	АВС	АВС	А	<i>д</i>
С	АВС	А-СА	ВСА	В-АВ	САВ	С-ВС	АВС	АВС	А	<i>е</i>
С	АВС	АВ+В	САВ	СА+А	ВСА	ВС+С	АВС	АВС	А	<i>ж</i>

Рис. 8.1. Тексты ДНК: *а* — дикий тип, *б* — мутант (делеция), *в* — мутант (добавление), *г, д* — двойные мутанты типа + —, *е* — тройной мутант — — —, *ж* — тройной мутант + + +. Заштрихованы искаженные участки текста

+—), то последовательность искажена лишь в области между двумя мутациями (рис. 8.1, г и д). Если искаженная область не слишком длинна, то синтезируемый белок может сохранить свою функцию и будет наблюдаться *реверсия*. Рис. 8.1, е, ж поясняют реверсии при трех однотипных мутациях: — — — и + + +. 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. В дальнейшем это было подтверждено прямым изучением первичной структуры синтезированного белка. «Мутации сдвига рамки» (frame shift mutations) наблюдаются и для гена фага Т₄, продуцирующего лизоцим. Оказалось, что «сдвиги рамки» действительно искажают белковый текст. Двойной ревертированный мутант лизоцима отличается от дикого типа последовательностью пяти остатков:

Дикий *w*-тип ... Тре Лиз Сер Про Сер Лей Асп Ала ...

Ревертант

(псевдо-*w*-тип) ... Тре Лиз Вал Гис Гис Лей Мет Ала ...

Эти генетические исследования привели к решению ряда физических вопросов. Установлено кодовое отношение, доказана коллинеарность кода, доказано, что код читается, начиная с определенного нуклеотида, и не имеет запяты, т. е. материальных элементов, разделяющих *кодоны*. Кодонами называются тройки следующих друг за другом нуклеотидов, кодирующие аминокислотные остатки. Общее число разных кодонов равно $4^3 = 64$.

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 каждый остаток. Слово «вырождение» употребляется здесь в физическом, а не в биологическом смысле.

Исследование мутаций сдвига рамки непосредственно свидетельствует о вырождении кода. Вернемся к только что изображенным фрагментам лизоцима. Если код не вырожден, то оба кодона для Гис в ревертанте должны быть одинаковыми. Обозначим их ABCABC. 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 ...
 ... ABC ABC ...
 ... Гис Гис ...

Но отсюда следует, что кодон для Сер есть также ВСА. Противоречие снимается, лишь если имеются различные кодоны для Гис.

Вырождение кода доказывается и тем, что при значительных вариациях в составе ДНК различных организмов (доля Г + Ц

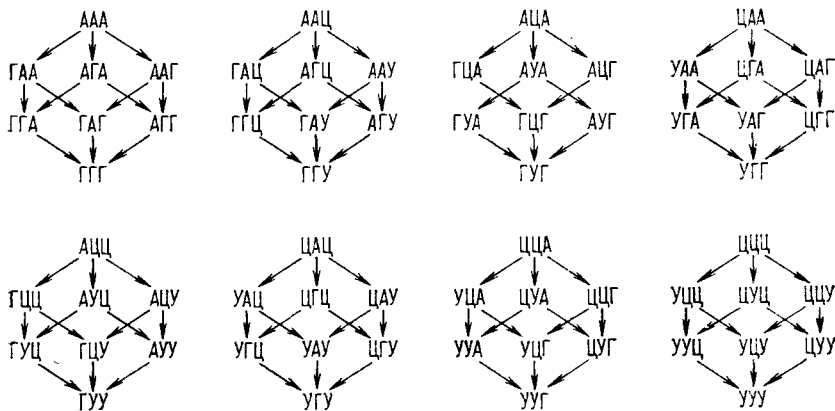
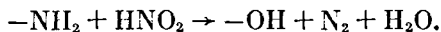


Рис. 8.2. Октеты кодонов

варьирует от 25 до 75%) средний состав белков меняется мало. Вместе с тем ряд факторов указывает на универсальность кода.

О вырожденности кода свидетельствует и исследование мутаций, как индуцированных химическими агентами, так и спонтанных. Азотистая кислота вызывает замещение аминной группы 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 восьмью октетами (Виттман). На рис. 8.2 изображены эти октеты, стрелками показаны нитритные точечные мутации. Наблюдаемые замещения в белке вируса

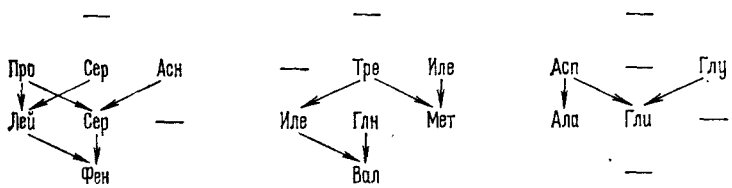


Рис. 8.3. Отнесение аминокислотных остатков к октетам на основании нитритных мутаций ВТМ

табачной мозаики (ВТМ) располагаются по тем же октетам. Виттман установил, что по крайней мере Сер или Иле (рис. 8.3) должны кодироваться несколькими триплетами, т. е. код вырожден. Эти данные одновременно показывают, что большая часть замещений в белке происходит в результате замены одного нуклеотида в кодоне.

В дальнейшем эти положения были многократно подтверждены. Генетический код является триплетным, неперекрывающимся, лишенным запятых и вырожденным. Код универсален для всех организмов, начиная с вирусов и кончая человеком. (Впрочем, недавно установлено, что при синтезе цитохромоксидазы в митохондриях код несколько отличен от универсального — см. с. 279.) Полная расшифровка кода потребовала прямых биохимических опытов, ставших возможными лишь в результате раскрытия механизма биосинтеза белка.

§ 8.2. Биосинтез белка

Само существование фиксированной первичной структуры у белковой цепи доказывает, что в клетке должна быть заложена программа построения этой структуры. «Текст» не может возникнуть в результате случайных встреч аминокислот — подобно типографскому тексту он должен «набираться» на некоторой матрице. Это понимал уже Кольцов задолго до открытия роли нуклеиновых кислот. Он считал, что роль матрицы, ответственной за синтез белка, играет также белок. Сейчас мы знаем, что матрицами служат молекулы ДНК и РНК. Для «набора текста» необходим генетический код. Матричный принцип биосинтеза белка является основным для молекулярной биологии и молекулярной биофизики.

Информация, содержащаяся в генах, т. е. в ДНК, передается мРНК в процессе *транскрипции*. Синтез белка происходит