

УСПЕХИ ФИЗИЧЕСКИХ НАУКФИЗИКА НАШИХ ДНЕЙ

575.1+547.963.3

ГЕНЕТИЧЕСКИЙ КОД**Ф. Крик, М. Ниренберг*

I

За прошедший год были достигнуты значительные успехи в решении «проблемы генетического кода». С биологической точки зрения она состоит в выяснении того, каким образом информация, заключенная в генах организма, определяет структуру белков.

Белки состоят из двадцати различных типов небольших молекул аминокислот, соединенных в длинные полипептидные цепи. Часто белки содержат несколько сотен связанных аминокислотных единиц, причем в каждом белке связи обладают специфичным, генетически определенным порядком. Белок, таким образом, является как бы длинным предложением, записанным с помощью двадцати букв.

Гены же состоят из других длинных молекул, а именно из молекул ДНК (дезоксирибонуклеиновая кислота) и в некоторых видах из близких к ДНК молекул РНК (рибонуклеиновая кислота). Недавно было обнаружено, что существует особая фракция РНК, названная РНК-посредником, которая переносит информацию от гена, расположенного в ядре клетки, в окружающую цитоплазму, где происходит основной синтез белков.

Нуклеиновые кислоты представляют собой длинные полинуклеотидные цепи, состоящие из соединенных между собой четырех типов нуклеотидов. Цепь состоит из остава, к которому на равных расстояниях присоединены боковые группы атомов четырех типов, известные под названием оснований. Однако последовательность расположения оснований не регулярна, и считается, что именно в этой последовательности зашифрована генетическая информация. Таким образом, проблема кода может быть сформулирована более точно: каким образом последовательность расположения четырех оснований в нуклеиновой кислоте определяет последовательность расположения двадцати аминокислот в белке?

*) F. H. C r i c k, Marshall W. N i e g e n b e r g, The Genetic Code, I—II, *Scientific American* 207 (4), 66 (1962) и 208 (3), 80 (1963). Часть I написана Ф. Криком, часть II — М. Ниренбергом. В оригинале — резюме: I. «Каким образом последовательность расположения оснований в нуклеиновой кислоте определяет последовательность аминокислот в белке? По-видимому, каждая аминокислота закодирована тройкой оснований, причем эти тройкичитываются в простой последовательности»; II. «Эта статья является продолжением статьи Ф. Крика, в которой обсуждается вопрос о том, каким образом наследственное вещество заключает в себе код, шифрующий синтез белка. В настоящей статье приводится дальнейшее описание природы этого кода». Перевод С. Н. Бреуса.

Эта проблема имеет две основные стороны: общую и более конкретную. Конкретная сторона этой проблемы состоит в том, чтобы выяснить, какой именно последовательности расположения оснований соответствует каждая аминокислота.

Более общая сторона проблемы кода, о которой я и буду рассказывать, состоит в выяснении следующих вопросов: какова длина генетических кодирующих единиц, каким образом они расположены в молекуле ДНК и каким образом считывается информация, посылаемая из ядра клетки. Эксперименты, о которых я буду говорить, проводились в Кембридже в Лаборатории молекулярной биологии Медицинского исследовательского совета.

В нашей работе мы использовали бактериофаг T4 — вирус, который инфицирует кишечную палочку и разрушает биохимический механизм бактерии, ответственный за ее размножение. Инфицирующий процесс начинается с того момента, когда T4 вводит свое генетическое ядро, состоящее из длинной нити молекулы ДНК, в бактериальную клетку. Менее чем за двадцать минут вирусная ДНК вызывает образование около ста вирусных частиц, состоящих из ядер и оболочек, которые содержат по крайней мере шесть различных белковых компонентов. В результате бактерия погибает, а вирусные частицы выходят наружу. Использование вирусов T4 в генетических экспериментах неоценимо, так как за очень короткий промежуток времени мы можем получить несколько их поколений и миллиарды вирусных частиц. Колонии, содержащие мутантные вирусы, могут быть выделены по появлению небольших круглых «блашек» на пластинах культуры. Более того, используя подходящие культуры, из миллиарда вирусных частиц можно выделить одну интересующую нас.

Используя такую же общую методику, С. Бензер из Университета в Пэрдью смог изучить тонкую структуру генов *A* и *B* (или цистронов, как он их называет), расположенных в области *rII* молекулы ДНК вируса T4.

Бензер показал, что гены *A* и *B*, находящиеся рядом на вирусной хромосоме, сами состоят из нескольких сотен различных областей, расположенных в линейном порядке. Это как раз то, что и следует ожидать, если предположить, что ген является частью, длиной, скажем, в 500 или 1000 оснований, очень длинной молекулы ДНК, которая образует вирусную хромосому (рис. 1). Полная молекула ДНК вируса T4 содержит около 200 000 пар оснований.

ЦЕННОСТЬ МУТАЦИЙ

Из работы Бензера и других мы знаем, что определенные мутации в области *A* и *B* приводят к полному нарушению работы обоих генов, другие же — лишь к частичному.

Наблюдались также мутации, способные подавить воздействие вредных мутаций и восстанавливающие работу либо одного, либо обоих генов. Мы подозревали, что различные, часто неожиданные, следствия разнообразных мутаций могут нам дать ключи к пониманию генетического кода.

Поэтому мы начали с опытов по скрещиванию вирусов T4, имеющих нарушения в различных частях гена. Выращивая в одной культуре две различные вирусные частицы, можно получить «рекомбинанты», обладающие свойствами обоих «родителей». Так, при скрещивании вируса, в определенной точке гена которого изменено основание, с вирусом, имеющим нарушение в другой точке, получаются фаги, обладающие дефек-

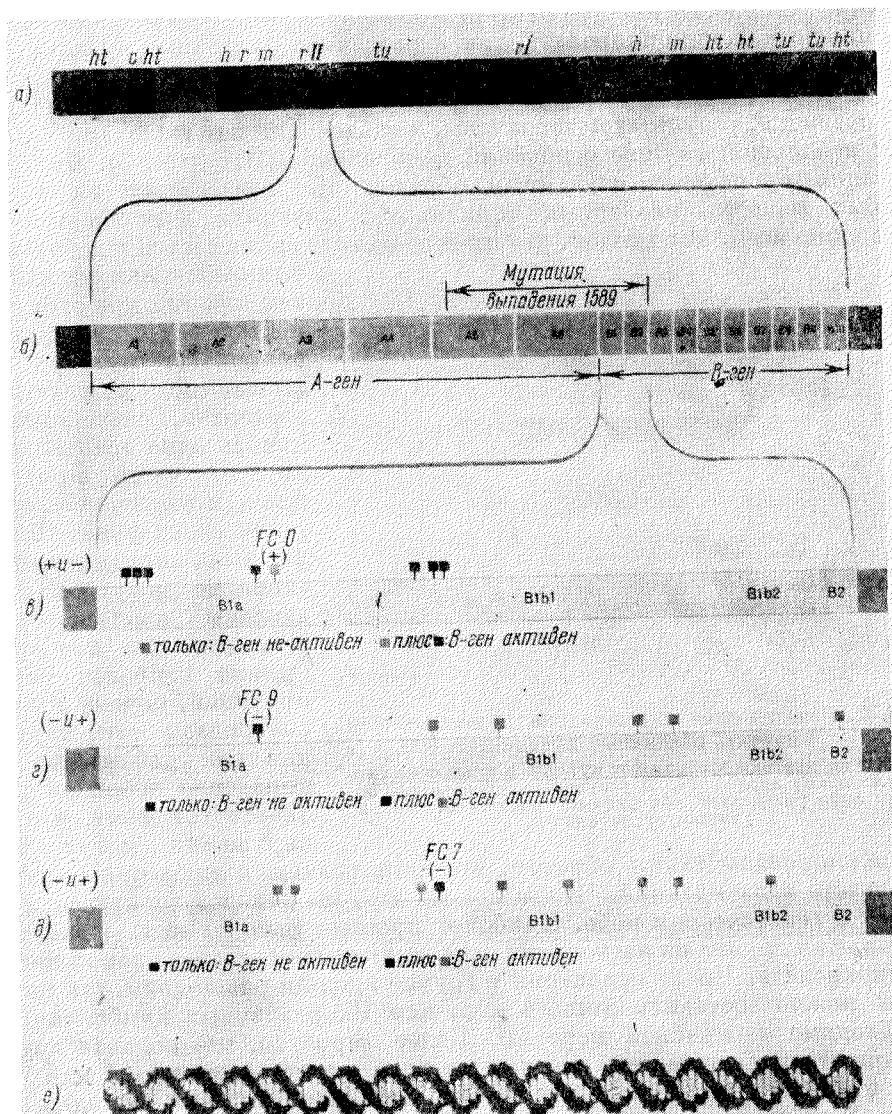


Рис. 1. Область rII вируса T4 составляет всего несколько процентов молекулы ДНК (дезоксирибонуклеиновой кислоты), содержащей полную информацию, необходимую для создания вируса.

Область состоит из двух генов, обозначенных здесь A и B. A-ген разбит на шесть основных участков, B-ген — на десять (б). В экспериментах, описанных в статье, использовались мутации в первом и втором участках B-гена. Активность B-гена нарушается при любой мутации, добавляющей (серый квадратик) или выбивающей (черный квадратик) молекулярную единицу, называемую основанием. Однако активность гена восстанавливается соответственно при выпадении или добавлении основания (в — д). Объяснение восстановления активности гена проиллюстрировано на рис. 5. Картина молекулярной структуры ДНК (е) находится примерно в масштабе с рисунками участков B1 и B2 B-гена. Эти два участка содержат около ста пар оснований.

тами в обоих точках (рис. 2). И наоборот: если фаг имеет несколько нарушений, то их можно разделить, скрещивая этот вирус с «диким», который, по определению, не имеет дефектов. Короче говоря, генетическими методами можно как комбинировать, так и разделять различные мутации, при условии, что они не совпадают.

Большинство нарушений, которые мы будем рассматривать, являются, очевидно, результатом не замены, а добавления или выпадания одного или небольшой группы оснований в молекуле ДНК (рис. 3). Подобные мутации происходят случайным образом при воздействии на вирусы особых веществ, называемых акридинами, механизм влияния которых не вполне ясен. Мы думаем, что их воздействие заключается в добавлении или выбивании оснований, так как ген при этом полностью перестает функционировать; мутации же, вызванные воздействием реагентов, способных заменить одно основание на другое, часто приводят лишь к частичной потере геном своих функций. Более того, акридиновые мутации не могут быть подавлены действием этих реагентов (и наоборот). Однако наиболее веской причиной, почему мы рассматриваем эти нарушения как добавление или выпадение оснований, является тот факт, что эти нарушения могут быть

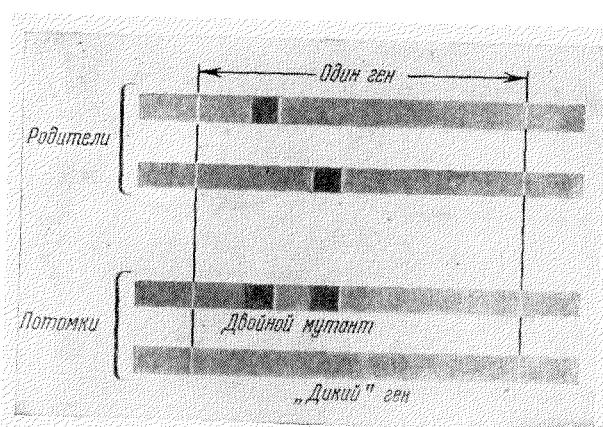


Рис. 2. Генетическое скрещивание позволяет нам изучать различные мутации.

Темные квадратики изображают мутации в хромосоме (молекуле ДНК) вируса T4. В результате скрещивания потомок либо унаследует оба дефекта родителей, либо не унаследует ни одного.

скомбинированы таким образом, который возможен лишь при подобном характере этих мутаций. Чтобы понять это, мы должны вернуться к вопросу о генетическом коде. Наиболее простым был бы код, в котором небольшая группа из нескольких оснований соответствовала какой-нибудь аминокислоте. Число оснований в группе вряд ли равно двум, так как из двух можно составить лишь 4×4 , или 16, различных комбинаций, а необходимо по крайней мере 20. Более вероятно, что кодовая группа содержит три основания, так как из трех можно составить $4 \times 4 \times 4$, или 64, сочетания. Небольшая группа оснований, шифрующая одну аминокислоту, получила недавно название кодона.

Первая определенная схема кодирования была предложена восемь лет назад физиком Георгом Гамовым. В этой схеме, как показано на рис. 4, соседние кодоны перекрываются. Из этого кода следует, во-первых, что за каждой аминокислотой могут идти только определенные аминокислоты. Во-вторых, изменение одного основания приводит к изменению трех соседних аминокислот. Опытные данные, собранные со временем выдвижения Гамовым своей схемы, говорят о том, что код не является перекрывающимся. Во-первых, в исследованных белках не было замечено каких-либо ограничений на порядок следования аминокислот. Во-вторых, было показано, что типичные мутации приводят к изменению только одной аминокислоты в полипептидной цепи молекулы белка. Хотя теоретически возможно, что генетический код может быть частично перекрывающимся, более вероятно, что кодоны не перекрываются совсем.

Так как остав молекулы ДНК совершенно регулярен, неясно, как отмечаются границы групп из трех или более оснований. Было предложено много хитроумных решений этой проблемы. Так, например, считалось, что код может быть устроен таким образом, что при считывании неправильной последовательности троек оснований переданная информация будет полным абсурдом, и белок синтезироваться не будет. Однако сейчас нам кажется, что правильным решением будет наиболее очевидное. А именно,

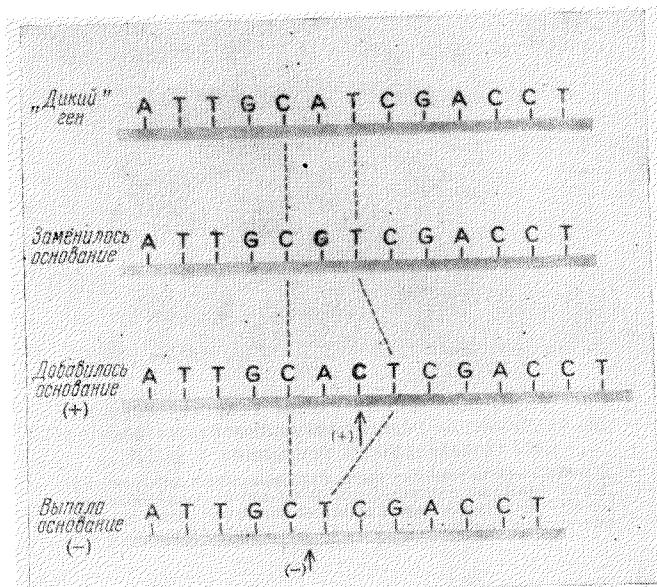


Рис. 3. Существуют два типа мутаций, вызванных нарушением последовательности расположения оснований (A, T, G, C), прикрепленных к оставу молекулы ДНК. Мутации первого типа просто заменяют одно основание на другое, например A на G. Мутации же второго типа либо добавляют, либо выбывают одно основание. Этими основаниями являются: аденин (A), гуанин (G), тимин (T) и цитозин (C).

«текст» начинается с фиксированной начальной точки, возможно, совпадающей с одним из концов гена, и далее просто считывается тройками оснований. Заметим, что, если бы считывание началось с другой начальной точки, текст разбился бы на неправильную последовательность троек и информация оказалась бы абсолютно неверной. Легко видеть, что для троичного кода существует один правильный и два неправильных способа считывания информации.

Если бы эта идея оказалась верной, то мы сразу бы смогли объяснить, почему выпадение или добавление одного основания в гене приводят к полной потере его функций, так как считывание генетической информации, начиная с места этого нарушения, оказалось бы полностью неправильным. Далее, мы обнаружили, что, хотя каждая из этих мутаций полностью нарушает работу гена, некоторые их комбинации из двух к этому не приводят. (Фактически мы обнаружили большинство парных мутаций, не нарушающих работу гена, наблюдая нефункционирующие гены и выбирая в дальнейшем те из них, в которых произошла редкая вторая мутация и восстановила работу гена. Однако это не меняет наше рассуждение.) Это дало нам возможность разбить все нарушения на два класса: плюс-мутации и минус-мутации. Мы обнаружили, что, пользуясь

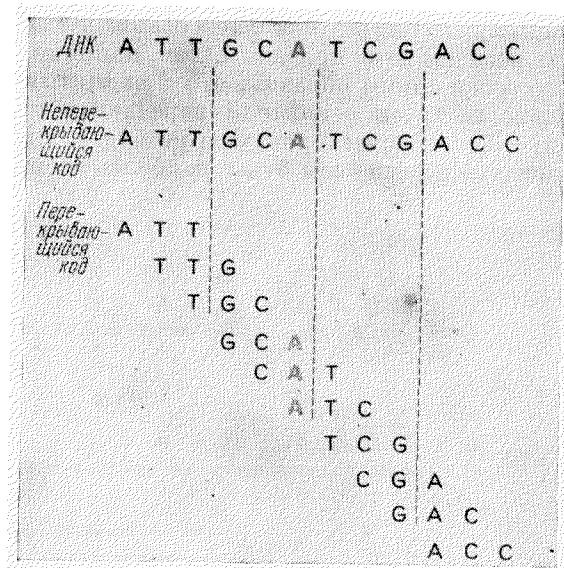


Рис. 4. Предложенные схемы генетического кода, поясняющие разные возможности считывания последовательности оснований в ДНК.

В неперекрывающемся коде, предложенном автором, кодовые единицычитываются в простой последовательности. В случае же перекрывающегося кода каждое основание содержится в трех последовательных кодовых группах.

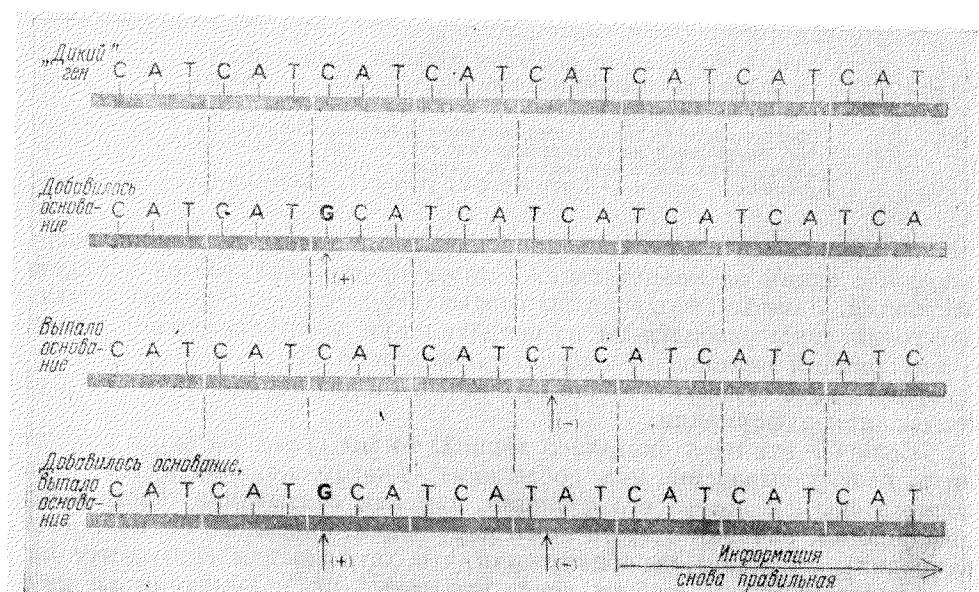


Рис. 5. Предположим, что считывание начинается в левом конце гена, мы видим, что в результате мутации, добавляющей или выбывающей одно основание, это считывание информации сдвигается.

Предполагаемая информация в диком гене записана в виде CAT, CAT, ... При добавлении основания получается последовательность TCA, TCA, ..., при выпадении — ATC, ATC, ... Соответственно выпадение или добавление еще одного основания восстанавливает первоначальную запись.

следующими правилами, можно всегда предсказать результат воздействия любой пары мутаций, находящейся в одном гене. Во-первых, комбинация плюс-мутации с плюс-мутацией нарушает работу гена. Во-вторых, минус-мутация с минус-мутацией также приводит к нарушению работы гена. И, в-третьих, плюс-мутация с минус-мутацией нарушает функции гена, если эти мутации находятся на большом расстоянии друг от друга, и не нарушает, если они расположены близко.

Последний случай представляет интерес. Нам удалось получить частично функционировавший ген при наличии плюс- и минус-мутаций, расположенных на не очень большом расстоянии друг от друга.

Для упрощения изложения предположим, что плюс-мутации вызваны добавлением одного основания, а минус-мутации — выпадением (доказать это весьма трудно). Тогда видно, что информация, считываемая с одного конца гена, будет правильной вплоть до добавившегося основания, затем до выпавшего основания она будет неверной, после чего она снова станет правильной. Таким образом, генетическая информация будет верной почти на всей длине гена, за исключением небольшого участка между плюс- и минус-мутациями. Таким же образом можно показать, что при троичном коде комбинации плюса с плюсом или минуса с минусом приведут к нарушению функций гена (рис. 5).

Нам удалось провести большую часть экспериментальной работы с мутациями в левом конце *B*-гена, расположенного в области *rII*. Оказалось, что функции этой части гена не очень важны, так как наличие в ней небольшого участка с нарушенной информацией может практически не влиять на работу гена. Однако если плюс- и минус-мутации расположены слишком далеко друг от друга, ген перестает функционировать.

БЕССМЫСЛЕННЫЕ ТРОЙКИ

Чтобы понять дальнейшее, необходимо снова вернуться к проблеме кода. Возможных сочетаний из четырех по три — 64, закодировать же надо 20 аминокислот. Возможно, что каждой аминокислоте соответствует несколько комбинаций из трех оснований. С другой стороны, логично предположить, что хотя бы одна или две тройки не кодируют аминокислоты, а имеют какой-нибудь другой смысл, например «начинай здесь» или «кончай здесь». Хотя такие гипотетические тройки могут иметь и другое значение, они были названы бессмысленными тройками. Мы предположили, что иногда в области, расположенной между плюс- и минус-мутациями, могут возникнуть бессмысленные тройки, в результате чего ген перестанет функционировать.

Мы исследовали целый ряд генов, имевших плюс- и минус-нарушения, расстояния между которыми были относительно небольшими, и обнаружили, что в некоторых случаях гены теряли свою активность, хотя можно было ожидать, что они будут работать. По-видимому, в этих случаях появлялись бессмысленные тройки оснований. Мы также обнаружили случаи, когда плюс-мутация с последующей минус-мутацией не приводила к нарушению работы гена, а минус-мутация с последующей плюс-мутацией приводила, даже если нарушения происходили в одних и тех же местах, но в обратном порядке. Как я уже говорил, генетическая информация будет неверной в двух случаях: если плюс-мутация расположена слева от минус-мутации или если плюс-мутация расположена справа от минус-мутации. В случае же, если плюс с минусом не нарушают работу гена, а минус с плюсом нарушают, даже если нарушения находятся в одних и тех же местах, мы заключаем, что во втором случае образовалась бессмысленная тройка в области между двумя мутациями. В подтверждение

этой гипотезы нам удалось с помощью мутагенных веществ, заменяющих одно основание на другое, уничтожить бессмысленные тройки, после чего ген восстанавливал свою активность. Одновременно мы зафиксировали положение бессмысленных троек.

Недавно мы проделали еще один удивительный эксперимент. Если в левом конце *B*-гена заменить одно основание, то можно ожидать, что ген будет функционировать, так как, во-первых, информация, заключенная в этой части гена, не очень существенна и, во-вторых, вся остальная информация останется прежней. Фактически, если бы ген продолжал функционировать, мы бы не узнали о том, что одно основание заменилось на другое. Однако в некоторых случаях нам удалось нарушить работу *B*-гена, заменив одно основание в его левом конце. Вероятно, в результате этой замены образуется бессмысленная тройка. Мы решили, что если бы нам удалось сместить считываение так, чтобы информация оказалась записанной другой последовательностью троек, бессмысленная тройка могла бы исчезнуть. Поэтому мы исследовали те случаи, когда плюс- и минус-нарушения в *B*-гене происходили по обеим сторонам от предполагаемой бессмысленной тройки. В результате этих трех мутаций ген восстанавливал свою активность (рис. 6).

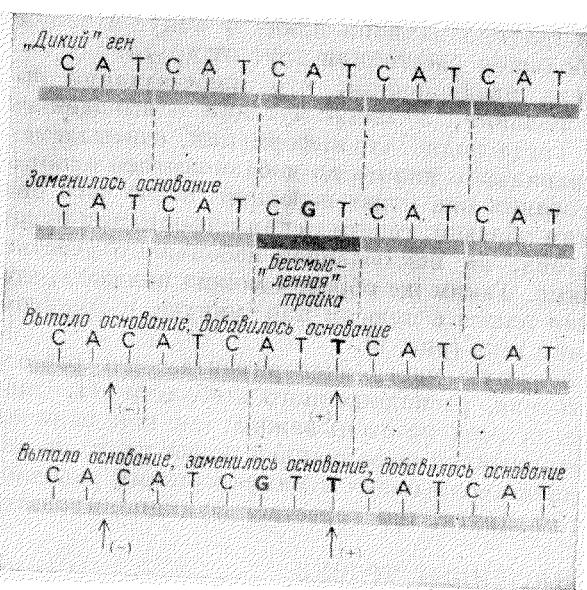


Рис. 6. «Бессмысленная» мутация состоит в том, что образуется кодовая группа, которая, очевидно, не соответствует ни одной из двадцати аминокислот, обнаруженных в белках.

В результате этого ген перестает функционировать. В представленном гипотетическом случае «бессмысленная» тройка CGT появилась при замене основания A в диком гене на G. «Бессмысленная» тройка может быть уничтожена, если сместить считываение, в результате чего G окажется в другой тройке. Это производится скрещиванием нефункционирующего гена с геном, содержащим плюс- и минус-мутации. Несмотря на наличие трех мутаций, получающийся ген активен.

нность (рис. 6). Другими словами, мы уничтожили тройку, сместив считываение. Все это говорит о том, что информация считывается с определенной фиксированной точки, возможно с какого-либо конца гена. Здесь возникает вопрос о том, как фиксируется конец одного гена и начало другого, поскольку мы ничего не знаем на оставе длинной молекулы ДНК, что могло бы их разделить. В то же время гены *A* и *B* вполне различны. Мы можем раздельно исследовать их свойства, и Бензер показал, что независимо от того, какая мутация происходит в *A*-гене, *B*-ген продолжает функционировать при условии, конечно, что мутация целиком лежит в *A*-гене. Таким же образом изменения в *B*-гене не влияют на активность *A*-гена.

ПРОСТРАНСТВО МЕЖДУ ГЕНАМИ

Поэтому разумно допустить, что что-то в молекуле ДНК между двумя генами разделяет их. Эта идея может быть проверена с помощью экспериментов на мутантном вирусе T4, в ядре которого выпала часть

области *rII*. У этого мутанта, известного под названием T41589, отсутствует значительная часть правого конца *A*-гена и небольшая часть левого конца *B*-гена. Удивительно, что *B*-ген продолжает частично функционировать; фактически мы поэтому и считаем, что эта часть *B*-гена не имеет существенного значения.

Хотя мы называем эту мутацию выпадением, так как генетическая карта показывает, что значительная часть генетической информации в этой области отсутствует, это не означает, что в молекуле ДНК действительно имеется разрыв. Более вероятно, что молекула ДНК составляет одно целое, из которого выпала часть генетического материала. Только сравнивая с целым вирусом («диким»), можно заметить, что часть информации отсутствует.

Мы предполагали, что между генами должна существовать небольшая разделяющая их область. Тогда можно ожидать, что если эта часть молекулы ДНК выпадет, два гена соединятся. Оказывается, что это предсказание легко проверить, так как с помощью генетических методов можно получить двойные мутанты. Поэтому мы скрещивали вирусы, которые обладали акридиновыми мутациями, расположенными в начале *A*-гена, с мутантами 1589. При наличии только акридиновой мутации *B*-ген продолжал полностью функционировать, что говорит о том, что гены были безусловно разделены. Однако при добавлении мутации 1589 активность *B*-гена нарушалась (рис. 7). Когда гены были соединены, нарушение в дальнем конце *A*-гена приводило к полной потере *B*-геном своих функций. Это говорит о том, что считывание начинается с одного конца.

Мы испытывали другие мутации в *A*-гене совместно с 1589. Все исследованные акридиновые мутации нарушали активность *B*-гена независимо от того, были они плюсовыми или минусовыми, однако комбинации из двух таких мутаций (плюс и минус) сохраняли его активность. С другой стороны, при использовании других мутаций (которые, как мы считаем, состоят не в добавлении или выпадении какого-нибудь основания, а в замене его на другое) в половине случаев *B*-ген функционировал, а в половине — нет. Мы считаем, что *B*-ген переставал работать из-за появления бессмысленных троек, и недавно Бензер использовал этот метод для определения бессмысленных комбинаций.

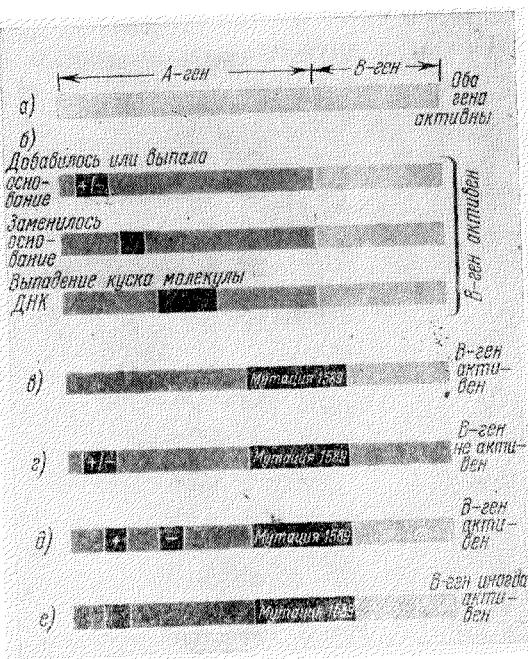


Рис. 7. Выпадение куска молекулы ДНК, соединяющее два гена, приводит к тому, что *B*-ген становится чувствительным к мутациям в *A*-гене.

Информации, заключенные в двух диких генах (*a*), считаются независимо, начиная с левого конца каждого из них. Независимо от вида мутации в гене *A*, *B*-ген сохраняет свою активность (*b*). Выпадение куска молекулы ДНК, известное под названием мутации 1589, нарушает активность *A*-гена, но не нарушает активности *B*-гена (*c*). Однако теперь нарушение в *A*-гене часто приводит к тому, что *B*-ген перестает функционировать. Это говорит о том, что в результате мутации 1589 оба гена каким-то образом соединились и считывание информации происходит таким образом, как если бы они составляли один ген (*d-e*).

Мы испытывали другие мутации в *A*-гене совместно с 1589. Все исследованные акридиновые мутации нарушали активность *B*-гена независимо от того, были они плюсовыми или минусовыми, однако комбинации из двух таких мутаций (плюс и минус) сохраняли его активность. С другой стороны, при использовании других мутаций (которые, как мы считаем, состоят не в добавлении или выпадении какого-нибудь основания, а в замене его на другое) в половине случаев *B*-ген функционировал, а в половине — нет. Мы считаем, что *B*-ген переставал работать из-за появления бессмысленных троек, и недавно Бензер использовал этот метод для определения бессмысленных комбинаций.

Мы, конечно, не знаем точно, что происходит с биохимической точки зрения. Мы только предполагаем, что в этих случаях два гена производят не две отдельные молекулы РНК-посредника, а одну, что приводит к образованию длинной полипептидной цепи молекулы белка, один конец которой имеет последовательность аминокислот, закодированную в части *A*-гена, а другой конец полностью соответствует *B*-гену, так что вся

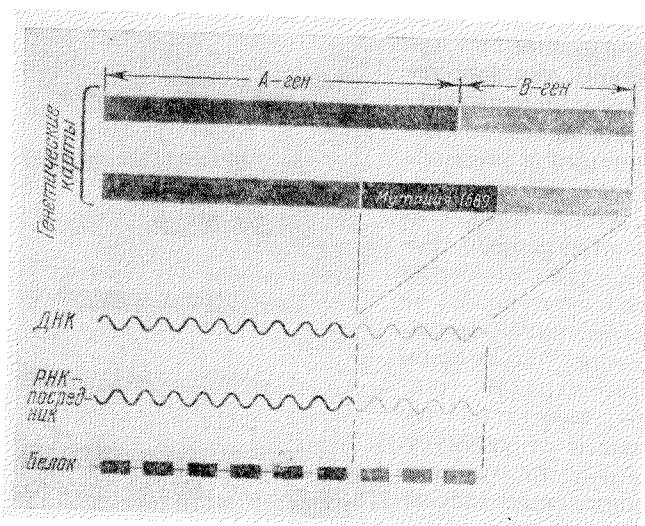


Рис. 8. Вероятный результат мутации 1589 состоит в том, что синтезируется белок, строение которого определяется в основном информацией, заключенной в *B*-гене.

Хотя на генетической карте мутации 1589 изображена в виде разрыва, молекула ДНК, очевидно, в результате этой мутации остается непрерывной, но укорачивается. При воспроизведении вируса генетическая информация, заключенная в ДНК, передается в молекулу рибонуклеиновой кислоты, называемую РНК-посредником. Эта молекула передает информацию в клеточные частицы, рибосомы, где происходит синтез белка строго в соответствии с инструкциями, закодированными в ДНК.

молекула белка обладает многими свойствами белка, синтезируемого *B*-геном. Это положение схематически проиллюстрировано на рис. 8. Возможно, что это предсказание удастся проверить экспериментально.

КАК СЧИТЫВАЕТСЯ ИНФОРМАЦИЯ

До сих пор все опытные данные хорошо согласовывались с общим предположением о том, что информация считывается тройками оснований, начиная с одного конца гена. Однако мы получили бы те же результаты, если бы информация считывалась группами в четыре или даже более оснований. Для проверки этого мы вызывали в одном гене не две, а три акридиновые мутации. Нас особенно интересовали случаи трех мутаций одного знака, таких, как «плюс»—«плюс»—«плюс», расположенных близко друг к другу. Каждая из этих мутаций или две из них нарушают активность *B*-гена. Однако если в гене присутствуют одновременно три мутации, его активность восстанавливается. Это — замечательный результат. Более того, мы получили те же результаты с несколькими различными комбинациями такого типа, а также с комбинациями типа «минус»—«минус» — «минус».

Полученные результаты легко объяснить, исходя из представлений о троичном коде. Одна плюс-мутация нарушит считывание информации.

Вторая плюс-мутация приведет к другому неверному считыванию. Однако если код троичный, третья плюс-мутация восстановит первоначальную информацию и, начиная с третьего плюса и до конца, считывание будет правильным. Информация будет неверной лишь в области между плюс-мутациями (рис. 9).

Отметим, что здесь несущественно, соответствует ли плюс-мутация добавлению основания, а минус-мутация — выпадению; в других случаях выводы останутся прежними. В действительности, даже если некоторые из плюс-мутаций соответствуют добавлению одного основания, другие могут соответствовать выпадению двух оснований; иными словами,

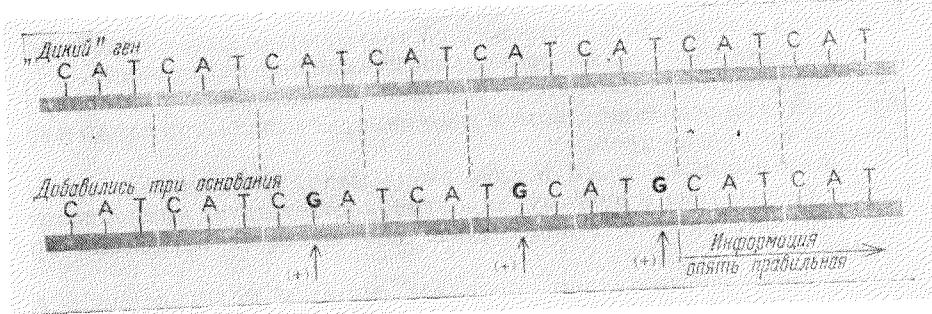


Рис. 9. Тройная мутация, состоящая в добавлении трех оснований на малом расстоянии друг от друга, нарушает генетическую информацию лишь на небольшом участке гена, сохраняя остальную информацию верной.

То же самое получится и при выпадении трех близко расположенных оснований.

плюс может оказаться двумя минусами. Аналогично, некоторые из минус-мутаций могут в действительности быть двумя плюс-мутациями. Однако и в таких случаях наши выводы не изменятся.

Хотя наиболее вероятно, что информация считывается тройками оснований, мы в этом полностью не уверены. Считывание может происходить и группами, содержащими кратное трем число оснований. Предположим, например, что информация считывается шестерками. Тогда единственное изменение, которое необходимо внести в нашу интерпретацию опытных данных, будет состоять в предположении, что все мутации соответствуют выпадению или добавлению четного числа оснований. Мы имеем некоторые слабые доказательства, что это не так. Например, мы можем скрестить мутант 1589 (в котором гены *A* и *B* соединены) с вирусами, у которых в гене *A* выпало небольшое число оснований. Тогда, если число этих выпавших оснований случайно, можно ожидать, что примерно у одной трети скрещенных мутантов *B*-ген должен функционировать, если информация считывается тройками, так как при выпадении одновременно трех оснований информация не нарушается. Если же считывание происходит шестерками, то только 1/6 часть мутантов будет обладать функционирующим *B*-геном. Таких мутантов в эксперименте оказалось более одной трети. Однако, рассматривая весь экспериментальный материал, мы пришли к выводу, что, хотя тройка оснований является наиболее вероятной кодирующей единицей, возможность того, что информация записана группами с кратным трем числом оснований, полностью отбросить нельзя.

Мы можем сделать еще одно общее заключение относительно генетического кода. Если грубо определить длину *B*-гена (сравнивая его с другим геном, длина которого приблизительно известна), то можно оценить

число оснований, лежащих между плюс- и минус-мутациями, при котором *B*-ген еще функционирует. Зная также частоту появления бессмысленных троек в области между плюс- и минус-мутациями, мы можем узнать, много ли таких троек или их всего несколько. Наши вычисления свидетельствуют о том, что бессмысленных троек очень мало. Другими словами, наиболее вероятно, что большинство из 64 возможных троек, или кодонов, не бессмысленно, а соответствует аминокислотам. Из этого следует, что одна аминокислота шифруется несколькими кодонами. Такой код называется вырожденным.

Таким образом, мы можем сделать три общих вывода относительно генетического кода:

1. Информация считывается неперекрывающимися группами оснований, начиная с какой-то фиксированной точки, возможно, с одного конца гена. Начальная точка задает правильное считывание информации.

2. Информация считывается группами определенных размеров, возможно, тройками оснований, хотя и возможно, что число оснований в группе больше, но кратно трем.

3. Бессмысленных троек в коде очень мало. Большинство троек не нарушает работы гена и поэтому, вероятно, соответствует аминокислотам. Таким образом, каждой аминокислоте соответствует несколько троек оснований.

Если *B*-ген действительно управляет синтезом полипептидной цепи молекулы белка, как мы предполагали, то трудно увидеть, каким образом можно избежать первого вывода. Второй вывод также довольно убедительно доказан. Что касается третьего, то он более неопределенный и может оказаться ошибочным.

Наконец, мы должны выяснить, какие дальнейшие опытные факты могут окончательно подтвердить справедливость представленной здесь теории. Мы продолжаем собирать генетические данные, однако я сомневаюсь, что они смогут сделать теорию еще более убедительной. Что нам необходимо — это получить белок, например синтезированный под управлением гена, в котором произошли плюс- и минус-мутации, а затем определить в нем последовательность аминокислот. Согласно общепринятой теории, так как строение гена нарушено в двух местах, последовательность аминокислот в белке должна измениться также в двух соответствующих местах. Согласно же нашей теории последовательность аминокислот изменится не только в этих двух местах, но и во всей области между ними. Другими словами, изменится целый кусок молекулы белка. Существует один белок, лизозим фага T4, который удобно использовать для этой цели, и мы надеемся, что скоро американские исследователи, работающие над изучением этого белка, подтвердят наши выводы.

Этот же эксперимент должен определить, какая из кодовых схем, разработанных Ниренбергом и Маттеи и Очао и его коллегами, верна. Фаговый лизозим, произведенный диким геном, должен отличаться от лизозима, произведенного геном, имеющим плюс- и минус-мутации, на небольшом участке молекулы. На этом участке последовательность аминокислот в этих белках должна соответствовать одинаковым последовательностям оснований в молекуле ДНК, считываемых, однако, разными тройками.

Если последовательности аминокислот на этих участках будут установлены как для дикого, так и для мутантного белков, то можно будет точно установить, действительно ли кодоны, расположенные в области между двумя нарушениями, соответствуют той последовательности аминокислот, которой они приписываются.

II

Десять лет тому назад Джеймс Уотсон и Фрэнсис Крик предложили хорошо известную теперь модель строения ДНК (дезоксирибонуклеиновой кислоты). За эту работу им, вместе с Мориаком Уилкинсом, была присуждена в прошлом году Нобелевская премия. ДНК представляет собой гигантскую спиральную молекулу, являющуюся во всех живых организмах носителем генетического кода. В части I Крик описал общую структуру этого кода. С помощью хитроумных экспериментов с бактериальными вирусами ему и его коллегам удалось установить, что «буквы» кодачитываются в простой последовательности, а «слова» записываются, по всей вероятности, грушами из трех букв. Кодовыми «буквами» в молекуле ДНК являются четыре основания: аденин, гуанин, цитозин и тимин, соответственно обозначаемые через А, Г, С и Т.

В этой части обсуждается вопрос о том, каким образом различные комбинации из этих оснований, или кодовых букв, шифруют определенную биохимическую информацию, используемую клеткой в процессе синтеза белков — гигантских молекул, состоящих из двадцати различных аминокислот. Каждая аминокислота занимает строго определенное место в цепи молекулы белка, положение которого задается последовательностью кодовых букв в молекуле ДНК (или молекулах), наследуемой каждым организмом от своих предков. В процессе эволюции молекула ДНК видоизменяется. Между живыми организмами происходит борьба за существование; иногда случайные изменения генетической информации, заключенной в молекуле ДНК, оказываются для организма выгодными в этой борьбе. Таким способом организм постепенно накапливает свойства, способствующие его выживанию.

Точное количество различных белков, необходимых для жизнедеятельности обычной клетки, неизвестно, оно исчисляется многими сотнями. Большинство из них, если не все, являются ферментами, или биологическими катализаторами, управляющими сотнями различных химических реакций, которые одновременно протекают внутри каждой клетки. Типичный белок представляет собой молекулярную цепь, состоящую примерно из двухсот аминокислотных единиц, связанных в определенной последовательности. Обычно каждый белок содержит все или большинство из 20 различных аминокислот. Код, необходимый для синтеза одного белка, заключен в гене, который в свою очередь является определенной частью линейной молекулы ДНК. Для записи 200 аминокислотных единиц ген должен содержать по крайней мере 200 кодовых слов, представленных последовательностью, вероятно, из 600 оснований. Мы еще не знаем полной последовательности оснований в отдельном гене. Вирусы, или мельчайшие структуры, обладающие механизмом для собственного воспроизведения, могут содержать от нескольких генов до нескольких сотен. Бактерии могут содержать 1000 генов, а клетка человека — миллион. Гены в клетках человека не образуют одной длинной цепи, а разбиты по крайней мере на 46 отдельных молекул ДНК. Минимальное число молекул определяется количеством хромосом в человеческой клетке (46), содержащих весь наследственный материал. В действительности каждая хромосома несет, по всей вероятности, не одну или две, а несколько копий одной и той же генетической информации. Если бы удалось соединить все молекулы ДНК, находящиеся в клетке человека, в одну непрерывную нить, то длина этой нити оказалась бы порядка метра. Этот метровый набор инструкций образуется во время зачатия при слиянии яйцеклетки со сперматозоидом и в дальнейшем при развитии зародыша точно воспроизводится миллиарды раз. Из рис. 10 видно, что основания в молекуле

ДНК образуют поперечные связи, соединяющие две спиральные нити, которые состоят из попеременно следующих друг за другом дезоксирибозы (обычный сахар) и фосфата. Основания прикрепляются к сахарным группам, образуя дополнительные пары: А с Т и Г с С. В результате каждая из нитей молекулы ДНК может служить шаблоном для образования новой нити с дополнительным набором оснований. Точное копирование генов при делении клетки объясняется, очевидно, этим механизмом.

Центральный вопрос проблемы кода заключается в следующем: каким образом с помощью четырехбуквенного алфавита (основания А, Г, С и Т) составляется словарь из двадцати слов, соответствующих двадцати аминокислотам?

В 1954 г. физик-теоретик Гамов указал на то, что кодовые слова в таком словаре должны состоять, по крайней мере, из трех оснований. Совершенно очевидно, что однобуквенных кодовых слов можно составить только четыре. Двухбуквенных — 4×4 , т. е. 16, а трехбуквенных — $4 \times 4 \times 4$, т. е. 64, что более чем достаточно для составления словаря, содержащего 20 слов, соответствующих 20 аминокислотам (см. табл. I на стр. 163). Впоследствии было высказано много предположений о природе генетического кода, однако основные экспериментальные знания о коде были получены только в последние 18 месяцев.

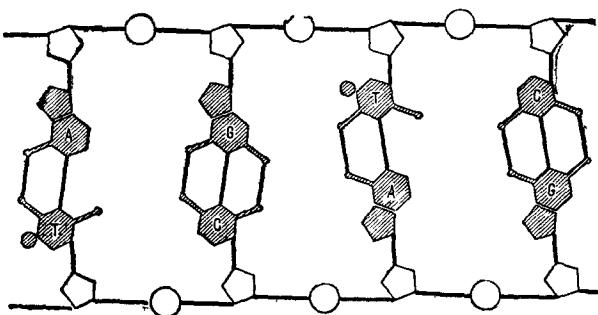


Рис. 10. Молекула ДНК напоминает лестницу, состоявшую из отдельных звеньев (в действительности молекула скручена в спираль), в которой пары оснований соединяют между собой пары линейных звеньев, состоящих из дезоксирибозы и фосфатных групп.

Основания всегда соединены между собой так, что А связано с Т, а Г с С. Генетический код есть последовательность оснований, считываемых вдоль одного края лестницы. Соединения из дезоксирибозы и фосфатной группы в двух линейных звеньях молекулы следуют в противоположных направлениях. Молекулы ДНК содержат тысячи пар оснований.

венных — $4 \times 4 \times 4$, т. е. 64, что более чем достаточно для составления словаря, содержащего 20 слов, соответствующих 20 аминокислотам (см. табл. I на стр. 163). Впоследствии было высказано много предположений о природе генетического кода, однако основные экспериментальные знания о коде были получены только в последние 18 месяцев.

ГЕНЕТИЧЕСКИЙ ПОСРЕДНИК

Вскоре после появления модели ДНК Уотсона — Крика стало ясно, что сама молекула ДНК не может принимать непосредственное участие в процессе синтеза белка, промежуточным звеном между ними должна быть шаблонная молекула РНК (рибонуклеиновая кислота). Синтез белка происходит в клеточных частичках, рибосомах, состоящих наполовину из белка и наполовину из РНК (рибосомная РНК). Несколько лет тому назад Жак Моно и Франсуа Жакоб из Института Пастера в Париже ввели термин «РНК-посредник» для обозначения шаблонной молекулы РНК, переносящей генетическую информацию от ДНК в рибосому.

Примерно в это же время Жерар Гурвиц из Университетской школы медицины в Нью-Йорке, Самуэль Вейсс из Чикагского университета, Одри Стивенс из Университета в Сан-Луисе и их сотрудники экспериментально наблюдали ферментативный синтез молекулы РНК, дополнительной к молекуле ДНК. Эти и другие исследователи показали, что фермент, полимераза РНК, вызывает образование молекул РНК, причем шаблоном служат нити молекулы ДНК.

РНК весьма сходна с ДНК за исключением того, что РНК содержит глюкозу вместо дезоксирибозы и основание урацил вместо тимина.

Когда на шаблонной нити молекулы ДНК синтезируется РНК, то в тех местах, где в дополнительной нити ДНК стоял аденин, в молекуле РНК появляется урацил. Одна из фракций РНК, образующаяся в результате такого процесса, есть РНК-посредник, она управляет синтезом белка. РНК-посредник покидает ядро клетки и направляется в рибосому. Последовательность оснований в РНК-посреднике определяет последовательность аминокислот в синтезируемом белке.

Аминокислоты переносятся в соответствующие места на молекуле РНК-посредника с помощью другой фракции РНК, названной РНК-переносчиком. В каждой клетке имеется особый активирующий фермент, который прикрепляет определенную аминокислоту к строго определенному виду РНК-переносчика. Более того, каждая клетка, очевидно, содержит не один вид РНК-переносчика, способного различать одну и ту же аминокислоту, а несколько. Важность этого факта станет очевидной впоследствии. Хотя до сих пор еще не удалось экспериментально показать, что РНК-переносчик «опознает» кодовые слова, записанные в молекуле РНК-посредника, ясно, что молекулы РНК-переносчика выполняют по крайней мере часть работы по правильному расположению аминокислот в цепи молекулы белка. Когда аминокислоты занимают свои места в этой цепи, они соединяются между собой под действием ферментативного процесса, природа которого пока еще полностью не выяснена. В результате образуются пептидные связи — химические связи, возникающие при удалении молекулы воды из двух смежных аминокислот. Для этого необходимо иметь переносящий фермент: хотя бы один фермент и действующий совместно с ним гуанозин-трифосфат. Оказывается, что аминокислотные группы соединяются в растущую цепь молекулы белка по порядку, начиная с того конца цепи, который содержит аминную группу (NH_2), и далее вплоть до другого, завершающегося карбоксильной группой (COOH).

Синтез белка обычно изучают в бесклеточных экстрактах, приготовленных из кишечных палочек (*Escherichia coli*). Бактерии выращивают в подходящей питательной среде, а затем осаждают из раствора с помощью центрифуги. После этого бактериальные клетки осторожно разрушают, растирая их вместе с мелкораспыленным алюминием (рис. 11); в результате освобождается внутриклеточная жидкость, содержащая ДНК, РНК-посредник, ферменты рибосомы и другие компоненты. Такие экстракти называются бесклеточными системами; когда в такие системы добавляют химически активные вещества (главным образом аденоzin-трифосфат), сразу же из имеющихся аминокислот начинается синтез молекулы белка.



Рис. 11. Эксперимент начинается с того, что клетки кишечных палочек перетираются в ступке вместе с мелкораспыленным порошком окиси алюминия. «Сок», вышедший из разрушенных клеток, способен еще синтезировать белок.

За этим процессом можно проследить, если использовать аминокислоты, содержащие радиоактивный изотоп кальция Ca^{14} .

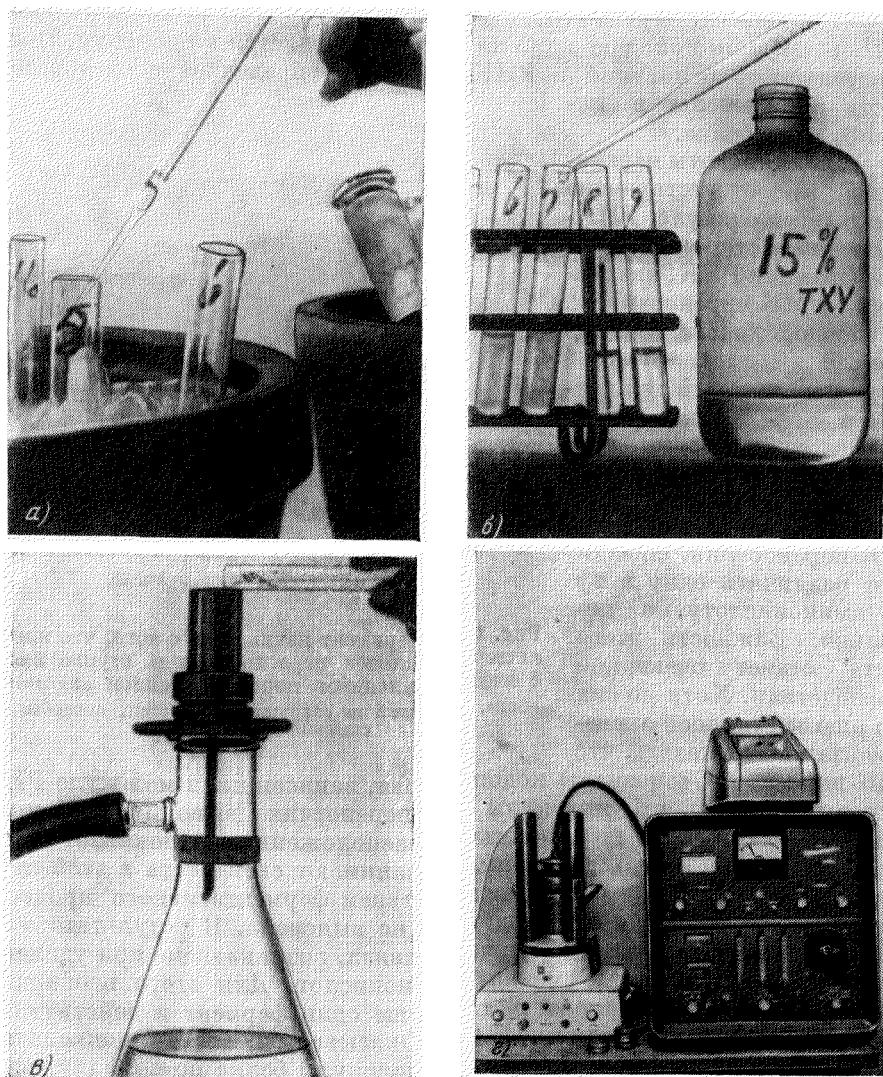
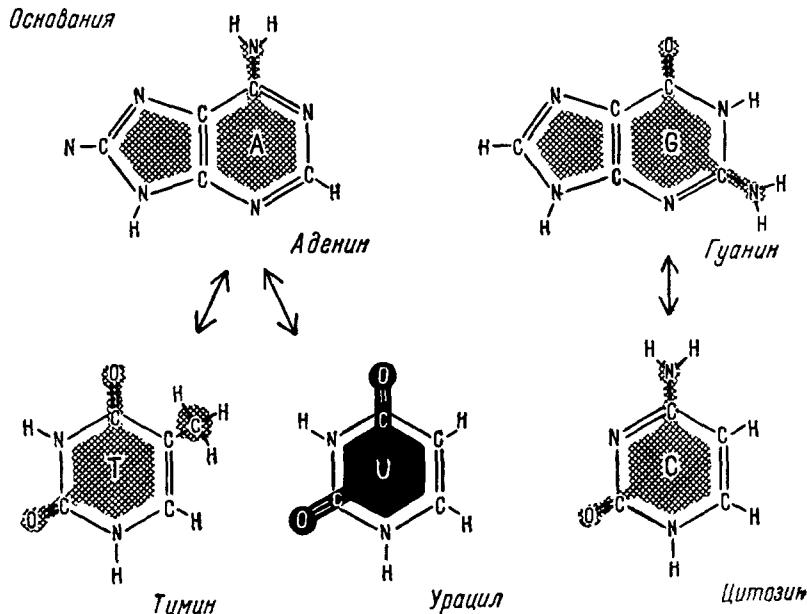


Рис. 12. На этой серии снимков, сделанных в лаборатории автора в Национальном институте здоровья, показаны последовательные этапы расшифровки генетического кода.

В открытых пробирках на снимке а) находятся образцы бактериальных бесклеточных систем, способных при соответствующей стимуляции синтезировать белок. На фотографии как раз показано добавление стимуляторов. Они состоят из искусственного РНК-посредника (рибонуклеиновой кислоты) и аминокислот, одна из которых радиоактивная. Белок образуется за время от 10 до 90 минут. Затем он осаждается (б), с помощью трихлоруксусной кислоты (ТХУ). На снимке в) показано, как полученный осадок переносится на диски из фильтровальной бумаги, которые затем помещены в держатели, называющиеся планшетами. На рисунке г) планшеты находятся в счетчике. Радиационные измерения свидетельствуют об эффективности данного образца РНК-посредника в управлении синтезом белка.

Наилучшие условия для синтеза белков в бактериальных бесклеточных системах были установлены исследователями во многих лабораториях.

риях (рис. 12). Когда мы начали свои исследования, прогресс в работе был сначала весьма незначительным, так как перед каждым опытом нам приходилось готовить свежие ферментные экстракты. Однако потом был найден способ сохранения этих экстрактов в течение длительного



Компоненты цепи молекулы белка

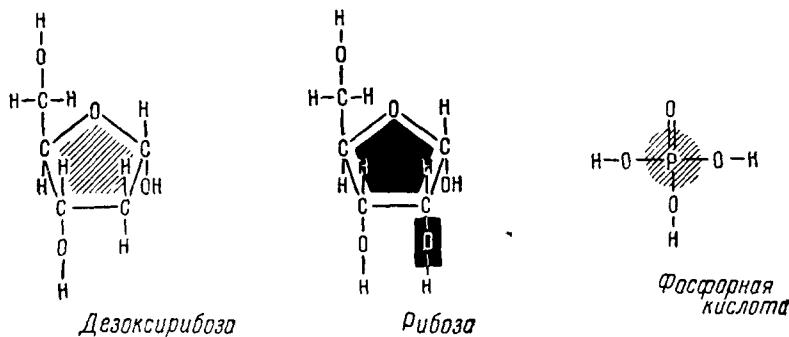


Рис. 13. Компонентами молекулы ДНК (дезоксирибонуклеиновая кислота) являются четыре основания: аденин, гуанин, тимин и цитозин (обозначенных через A, G, T, C), которые играют роль кодовых букв.

Другие составляющие, дезоксирибоза и фосфорная кислота, образуют цепи, к которым прикрепляются основания (см. рис. 10). В молекуле РНК, сходной с молекулой ДНК, тимин заменяется на урацил (U), а дезоксирибоза — на рибозу.

времени (порядка нескольких недель) без существенного ослабления их активности.

Обычно строение белков, синтезирующихся в таких экстрактах, определяется ДНК бактериальных клеток. Если бы удалось установить последовательность оснований одного из клеточных генов, или части гена, а затем сравнить ее с последовательностью аминокислот в соответствующем белке, генетический код был бы расшифрован. Хотя для многих белков последовательность аминокислот известна, никому еще не удалось

установить последовательность оснований в гене, поэтому такое сравнение пока невозможно.

Исследование синтеза белков в бесклеточных системах дает косвенный метод решения проблемы кода. Тисье, Новелли, Бернард Нисман, а затем и сотрудники из Института Пастера установили, что синтез белков в бесклеточных системах приостанавливается, когда в раствор добавляют фермент дезоксирибонуклеазу (ДНКазу), разрушающий молекулы ДНК. Этот эффект был также наблюден и изучен Маттеи и мною. Оказалось, что синтез белков прекращается, по-видимому, после того, как исчерпывается весь запас РНК-посредника. Когда в такие экстракты мы добавляли грубые фракции РНК-посредника, синтез белка возобновлялся. Развитие этого метода исследования явилось основой нашей дальнейшей работы.

При различных естественных источниках, включая вирусы, мы получили много фракций РНК и обнаружили, что большинство из них активно управляет синтезом белков в бесклеточных системах, изготовленных из кишечных палочек. Оказалось, что рибосомы кишечных палочек воспринимают инструкции, содержащиеся в РНК, которые были взяты от инородных организмов, в том числе и от вирусов. Необходимо подчеркнуть, что в этих экспериментах синтезировалось весьма незначительное количество белка.

Мы предполагали, что искусственная РНК, содержащая одно или два основания, может синтезировать простые белки, составленные лишь из нескольких аминокислот. Молекулы синтетической РНК могут быть получены с помощью фермента, полинуклеотид фосфорилазы, открытого в 1955 г. Марианной Грунберг-Манаго и Северо Очоа. В отличие от РНК-полимераз этот фермент не копирует строение ДНК, а синтезирует РНК-полимеры, соединяя основания в случайном порядке.

Нами был синтезирован искусственный РНК-полимер, состоящий только из урацила (названный полиуридилиевой кислотой, или поли-U), после чего он вместе со смесями из двадцати аминокислот был добавлен в активную бесклеточную систему. В каждой смеси одна из аминокислот содержала радиоактивный Ca^{14} , остальные 19 были нерадиоактивные. Таким способом можно было установить, какая именно из аминокислот участвует в синтезе белка, управляемом поли-U.

Оказалось, что это — фенилаланин. Таким образом, можно утверждать, что аминокислота фенилаланин закодирована последовательностью оснований U, содержащихся в поли-U. Было также установлено, что кодом другой аминокислоты, пролина, является последовательность оснований C, содержащихся в полизитидиловой кислоте, или поли-C. Таким образом, синтез белков в бесклеточных системах под управлением РНК с известным химическим строением дает нам простой метод расшифровки генетического кода (рис. 13).

СЛОВАРЬ КОДОВЫХ СЛОВ

Очоа и его сотрудниками, а также независимо нашей группой были синтезированы и исследованы полимеры, состоящие из всех возможных комбинаций, составленных из четырех оснований: A, G, C и U. В первых экспериментах изучались лишь РНК-полимеры, содержащие U, однако недавно Бретцером и Грунберг-Манаго, а также Оливером Джонсом и мною было синтезировано много активных полимеров, построенных из других оснований. Все полученные до настоящего времени результаты приведены в табл. I. В этой таблице перечислены РНК-полимеры, содержащие минимальное число оснований, при котором еще происходит синтез белков. Добавление еще одного основания в РНК-полимер приводит обычно к тому, что синтезируются белки с добавочным набором аминокислот.

Таблица I

Основания, имевшиеся в молекулах искусственных РНК

Из этой таблицы видна специфичность кода, заключающаяся в том, что 18 из 20 аминокислот кодируются искусственными РНК-полимерами, содержащими либо только одно либо два основания. Единственными аминокислотами, требующими для своего кодирования более двух оснований (U, A и G), являются аспарагиновая кислота и метионин. Относительные количества аминокислот, направленных в белок РНК-полимеров, содержащих два основания, определяются соотношением между числом этих оснований. Если полимеры содержат три или четыре основания, в молекуле белка направляются еще и другие аминокислоты. Таким образом активность поли-UCG (РНК-полимер, содержащий U, C и G) сходна с активностью поли-UC плюс поли-UG. Для поли-G не было найдено соответствующей аминокислоты. Несомненно, дальнейшие исследования внесут некоторые изменения в эту таблицу. Словарь кодовых слов, составленный с помощью этой таблицы, представлен в табл. III.

	U	A	C	G	UA	UC	UG	AC	AG	CG
Кодируемые аминокислоты	Фенилаланин	Лизин	Пролин**)		Фенилаланин***)	Фенилаланин***)	Фенилаланин***)	Лизин***)	Лизин***)	Пролин***)
Лейцин*)					Лизин****)	Пролин***)	Лейцин	Пролин****)	Глутаминовая кислота	Аргинин**)
	Тирозин	Лейцин	Валин				Гистидин	Аргинин**)	Аланин**)	
	Лейцин	Серин	Цистеин	Аспарагин			Аспарагин	Глутамин**)		
	Изолейцин			Триптофан			Глутамин	Глицин**)		
	Аспарагин**)			Глицин			Треонин			

*) Поли-U кодирует в основном фенилаланин.

**) Получено только в одной лаборатории; требует проверки.

***) Требует наличия только первого из перечисленных оснований.

****) Требует наличия только второго из перечисленных оснований.

С помощью двух оснований можно составить шесть различных РНК-полимеров: поли-AC, поли-AG, поли-AU, поли-CG, поли-CU и поли-GU. Если отношение между числом оснований поддерживать строго постоянным, то, как нетрудно показать, каждому такому отношению будет соответствовать свой, строго определенный набор аминокислот (рис. 14). Относительное содержание одной аминокислоты в белке по сравнению с другой зависит от соотношения между числом различных оснований в РНК. Предполагая, что последовательность оснований в молекуле РНК случайная, и зная отношение между числом оснований, можно легко подсчитать вероятные количества последовательностей из двух, трех и более оснований. Например, если поли-UC содержит 70% U и 30% C, то вероятность появления триплетной последовательности UUU равна $0,7 \times 0,7 \times 0,7$, или 0,34, т. е. UUU будет составлять около 34% всех триплетов, имеющихся в полимере. Вероятность появления последовательности

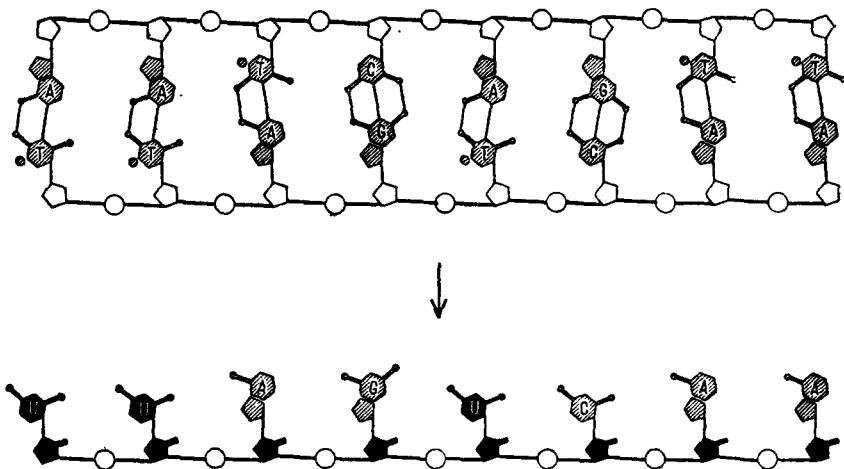


Рис. 14. РНК-посредник является молекулярным агентом,читывающим генетическую информацию, заключенную в молекуле ДНК и переносящую ее в те области клетки (рибосомы), где происходит синтез белков.

Буквы РНК-посредника являются дополнительными к буквам, записанным на одной нити молекулы ДНК. В нашем примере последовательность UUAGUCAA дополнительна к AATGAGTT. Точный механизм считывания неизвестен.

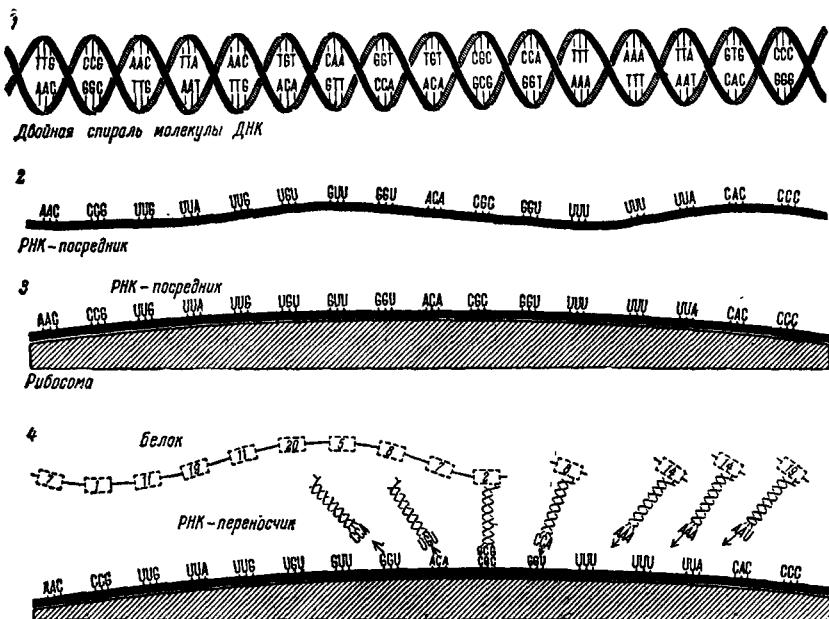


Рис. 15. Синтез белка начинается с генетического кода, заключенного в молекуле ДНК (1).

Этот код считывается РНК-посредником (2). На рисунке нить, с которой считывается код, отмечена жирной линией. РНК-посредник направляется к рибосоме (3) — месту синтеза белка. Аминокислоты, представленные в виде пронумерованных прямоугольников (4), доставляются в нужные места на молекуле РНК-посредника с помощью РНК-переносчиков (см. рис. 16). В действительности расстояния между основаниями одинаковые, т. е. они не располагаются отдельными триплетами, и механизм взаимодействия между РНК-переносчиком и РНК-посредником гипотетический. Соединения аминокислотных единиц образуют молекулу белка.

тельности UUC равна $0,7 \times 0,7 \times 0,3$, или 0,147, т. е. триплетов UUC будет порядка 14,7 %. Однако все эти расчеты предполагают случайное распределение оснований в молекуле РНК, что может оказаться неверным для некоторых полимеров (рис. 15, 16).

Гамовым и Криком было предсказано, что каждой аминокислоте может соответствовать несколько кодовых слов, так как возможных триплетов 64, а аминокислот 20. Код, в котором каждому кодирующему объекту соответствует несколько обозначений, называется вырожденным. Наши опыты показывают, что генетический код является, безусловно, вырожденным. Например, аминокислота лейцин (leucine) кодируется РНК-полимерами, содержащими или только U, или U и A, или U и C, или U и G.

Необходимо подчеркнуть, что такого рода вырождение не говорит о какой-то неопределенности в построении молекулы белка. Оно лишь обозначает, что определенная аминокислота может быть направлена в соответствующее место цепи молекулы белка с помощью нескольких кодовых слов. Более того, такая своеобразная гибкость каким-то еще не совсем понятным способом выгодно используется клеткой (рис. 17, 18).

Недавно было найдено оригинальное молекулярное объяснение явления вырождения. Было известно, что некоторые организмы имеют несколько видов РНК-переносчиков одной и той же аминокислоты. Например, в клетке кишечной палочки существуют два хорошо различимых переносчика лейцина. Вейсблому и Бензеру, а также Роберту Холли удалось выделить по отдельности каждый из этих переносчиков и затем исследовать их в бесклеточных системах. Оказалось, что один переносчик используется РНК-полимером поли-UC, но не используется поли-UG. Другой же используется поли-UG, но не поли-UC (рис. 19). Хотя общее число видов РНК-переносчиков в клетке неизвестно, возможно, что каждому виду соответствует свое кодовое слово.

Рис. 16. РНК-переносчик является особой фракцией РНК (ее молекула свернута в спираль), переносящей аминокислоты в соответствующие места цепи молекулы белка.

Каждой из 20 аминокислот соответствует по крайней мере одна молекула РНК-переносчика. Однако все они, по-видимому, содержат последовательность оснований ACC в том месте, где прикрепляется аминокислота, и основание G на противоположном конце. Для прикрепления аминокислоты требуется как присутствие особого фермента, так и добавочная энергия, которой снабжает аденоzin-трифосфат. Непарные основания в молекуле РНК-переносчика (в нашем примере AAU) могут служить тем механизмом, который «узнает» место в цепи молекулы белка, куда следует направлять аминокислоту.

Однако все же имеется одна реальная возможность появления неопределенности при синтезе белка. Эта неопределенность могла бы возникнуть, если бы одно кодовое слово соответствовало нескольким аминокислотам. До настоящего времени был отмечен только один случай такой неопределенности. Белок, синтезируемый поли-U, состоит не только из лейцина, но и из фенилаланина, причем на каждую молекулу лейцина



приходится 20—30 молекул фенилаланина. При отсутствии в растворе фенилаланина поли-У использует лейцин в количестве, равном половине обычно используемого количества фенилаланина. Молекулярное объяснение этой неопределенности неизвестно. Неизвестно также, встречается ли такой двойственный код не в бесклеточных системах, а в живых организмах (рис. 20).

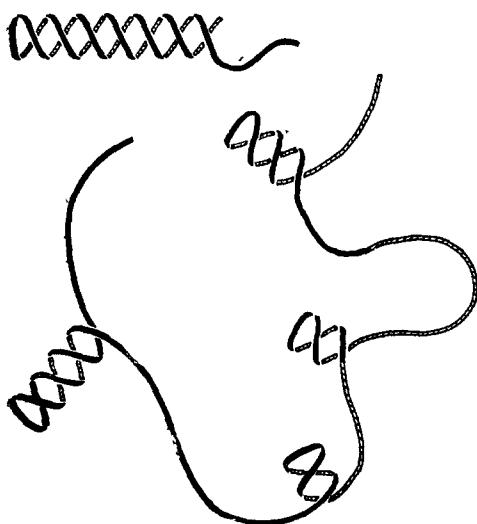


Рис. 17. Молекулы РНК могут иметь разнообразные формы.

По-видимому, молекула РНК-переносчика (вверху) имеет просто вид короткой двойной спирали (возможно, и не такой идеальной, как на рисунке), замкнутой на одном конце. Некоторые молекулы РНК содержат смесь закрученных и незакрученных областей (внизу).

эффективность, если учесть сильную том, что «бессмысленных» слов всего несколько (табл. II).

Таблица II

Число комбинаций из кодовых букв резко возрастает с увеличением длины кодовых слов. Так как нам необходимо закодировать 20 аминокислот, то, предполагая, что все слова имеют одинаковую длину, получим, что минимальное количество букв в кодовом слове равно трем

Единичный код (4 слова)		Двоичный код (16 слов)					Троичный код (64 слова)				
							AAA	AAG	AAC	AAU	
A	AA	AG	AC	AU			AGA	AGG	AGC	AGU	
G	GA	GG	GC	GU			ACA	ACG	ACC	ACU	
C	CA	CG	CC	CU			AUA	AUG	AUC	AUU	
U	UA	UG	UC	UU			GAA	GAG	GAC	GAU	
							GGA	GGG	GGC	GGU	
							GCA	GCG	GCC	GCU	
							GUA	GUG	GUC	GUU	
							CAA	CAG	CAC	CAU	
							CGA	CGG	CGC	CGU	
							CCA	CCG	CCC	CCU	
							CUA	CUG	CUC	CUU	
							UAA	UAG	UAC	UAU	
							UGA	UGG	UGC	UGU	
							UCA	UCG	UCC	UCU	
							UUA	UUG	UUC	UUU	

Крик приводит доказательства того, что число оснований в кодовом слове равно или кратно трем. Недавно мы определили относительные белок

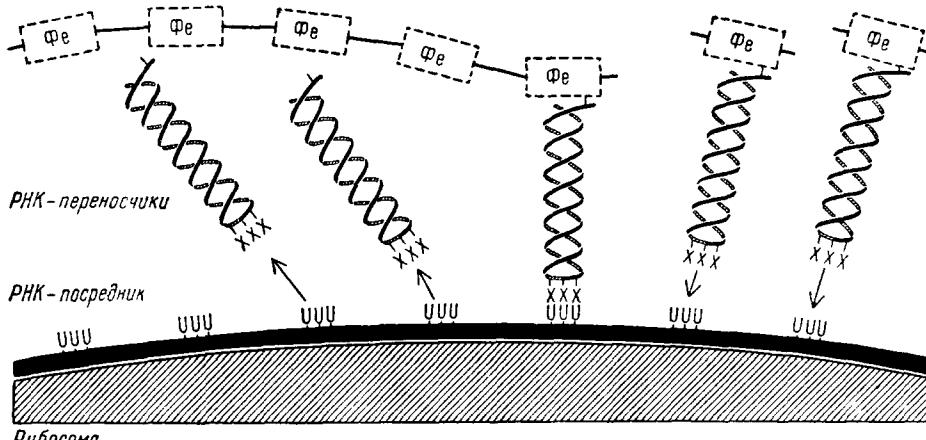


Рис. 18. Первой брешью в разгадке генетического кода явилось открытие того факта, что искусственный РНК-переносчик, содержащий только урацил (поли-*U*), синтезирует белок, фенилаланин (Фе), молекула которого составлена лишь из одной аминокислоты.

Это открытие было сделано автором вместе с Г. Маттеи. XXX на молекуле РНК-переносчика означают то, что основания, опознающие кодовые слова и записанные в молекуле РНК-посредника, неизвестны.

содержания различных аминокислот в белке, синтезированном искусственными РНК-полимерами с заданными отношениями между числом

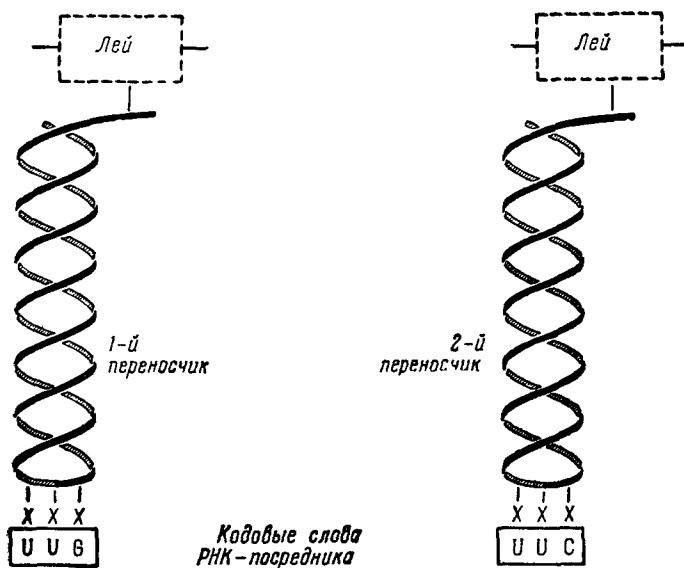


Рис. 19. Были открыты два вида РНК-переносчиков, способных переносить лейцин (лей).
Один вид (слева) «опознает» кодовое слово UUG, другой—UUC (справа).

оснований. Оказалось, что некоторые кодовые слова почти наверняка состоят из трех оснований. Однако, как видно из табл. I, 18 из 20 амино-

кислот могут быть закодированы словами, содержащими только два различных основания. Исключение составляют аспарагиновая кислота и метионин, которые кодируются, по-видимому, комбинациями из U, G и A. (Здесь еще имеется некоторая неясность относительно кодовых слов этих аминокислот, так как при синтезе белка даже полимер поли-UGA

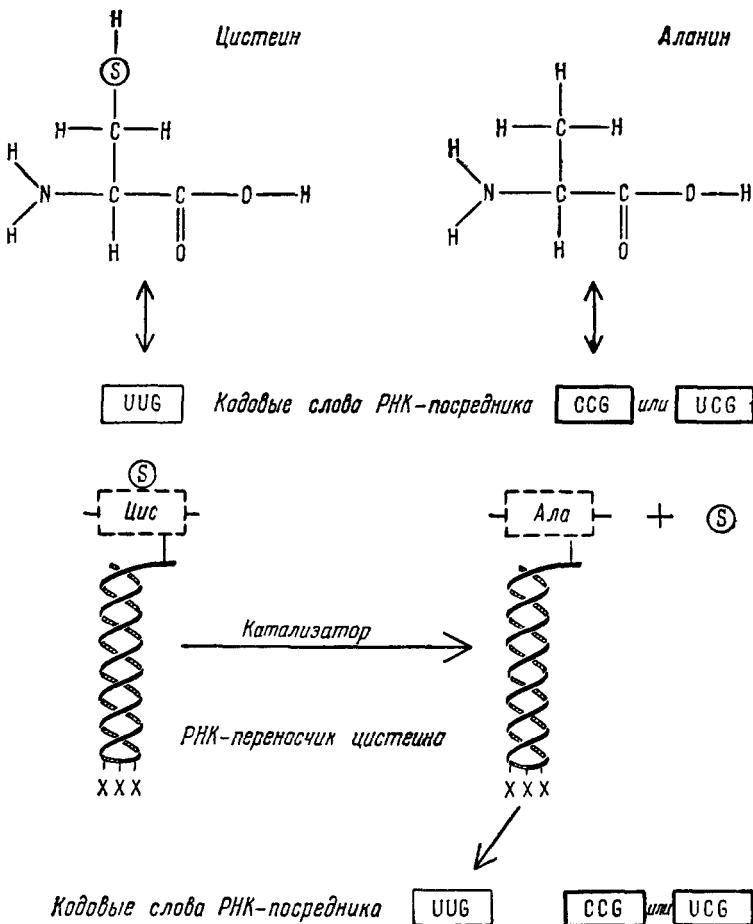


Рис. 20. С помощью хитроумного эксперимента было доказано, что «опознавание» кодовых слов зависит не от структуры переносимой аминокислоты, а от свойств РНК-переносчика.

Цистеин кодируется словом UUG, аланин — CCG или UCG. Цистеин прикреплялся к своему РНК-переносчику, после чего с помощью катализатора (Raney nickel) из его молекулы удалялся атом серы. В результате цистеин превращался в аланин. Тем не менее РНК-переносчики направляли его в то место цепи молекулы белка, где должен находиться цистеин.

использует аспарагиновую кислоту или метионин в очень небольших количествах.) Если же код все-таки троичный, то возможно, что в некоторых случаях правильное кодирование будет иметь место при условии, что из трех оснований считаются только два. Возможно, что такое несовершенство случается более часто в синтетических РНК-полимерах, содержащих одно или два основания, чем в естественных РНК-посредниках, которые всегда состоят из смеси всех четырех оснований. Поэтому результаты, полученные с помощью искусственных РНК, свидетельствуют

Таблица III

Словарь генетического кода содержит кодовые слова, соответствующие 20 аминокислотам. При составлении его предполагалось, что все слова состоят из трех букв. Так как порядок букв в словах еще не установлен, в нашей таблице он произвольный. Хотя более половины аминокислот записываются несколькими кодовыми словами, считается, что каждому триплету соответствует одна, строго определенная аминокислота. Было обнаружено только одно исключение из этого правила: триплет UUU кодирует как фенилаланин, так и лейцин, правда гораздо менее эффективно

Аминокислота	Кодовые слова из оснований молекулы РНК			
Аланин	CCG	UCG**)		
Аргинин	CGC	AGA	UCG**))	
Аспарагин	ACA	AUA		
Аспарагиновая кислота	GUA			
Цистеин	UUG*)			
Глутаминовая кислота	UAA	AGU**))		
Глутамин	ACA	AGA	AGU**))	
Глицин	UGG	AGG		
Гистидин	ACC			
Изолейцин	UAG	UAA		
Лейцин	UUG	UUC	UUA	UUU***)
Лизин	AAA	AAG****)	AAU****)	
Метионин	UGA**))			
Фенилаланин	UUU			
Пролин	CCC	CCU****)	CCA****)	CCG****)
Серин	UCU	UCC	UCG	
Тreonин	CAC	CAA		
Триптофан	GGU			
Тирозин	AUU			
Валин	UGU			

*) Неясно, будет ли кодовым словом UUG или GGU.

**) Неясно, нужно ли основание U.

***) Кодирует преимущественно фенилаланин.

****) Неясно, нужны ли основания G и U.

*****) Неясно, нужна ли комбинация UAG.

лишь о кодовых возможностях клетки, т. е. синтетическая РНК выявляет кодовые слова, обычно имеющиеся в живой клетке, и возможные слова, которые были бы обнаружены, если бы в клеточной ДНК произошли соответствующие мутации. В табл. III приведен словарь кодовых слов, полученных в предположении, что код троичный.

УНИВЕРСАЛЬНОСТЬ КОДА

Обладает ли каждое растение или животное своим собственным генетическим кодом или он одинаков для всех живых организмов нашей планеты? Предварительные исследования показывают, что код существенно универсален и что даже особи, находящиеся на противоположных ступенях эволюционного развития, используют один и тот же код. Например, сотрудниками целого ряда лабораторий в США и Англии недавно сообщалось, что синтетические РНК-полимеры ведут себя в бесклеточных системах, приготовленных из клеток млекопитающих, так же, как и в бактериальных бесклеточных системах. В генетическом коде млекопитающих до настоящего времени удалось лишь определить комбинации оснований, соответствующих шести аминокислотам. Тем не менее вполне вероятно,

что в будущем будут обнаружены некоторые расхождения. Так как каждая аминокислота кодируется несколькими словами, вполне возможно, что один организм использует одно слово, а другой — другое.

Косвенным указанием на правильность кодовых слов, полученных при исследованиях в бесклеточных системах, может служить изучение естественных белков, имеющих в одном и том же месте своих молекул различные аминокислоты. Например, гемоглобин человека, страдающего серповидной анемией, отличается от нормального гемоглобина тем, что в одном месте цепи молекулы вместо глутаминовой кислоты находится валин. Имеется еще один необычный гемоглобин, который содержит в том же месте лизин. Из словаря кодовых слов видно, что все эти три аминокислоты — глутаминовая кислота, валин и лизин — записываются сходными комбинациями оснований. Можно предположить, что два вида необычного гемоглобина образуются в результате мутаций, заменяющих одно основание на другое, в гене, ответственном за синтез гемоглобина. Действительно, из словаря видно, что кодовые слова этих аминокислот настолько сходны, что такое предположение проходит. Одна из записей глутаминовой кислоты есть AGU. Заменяя A на U, получаем UGU, или кодовую запись валина. Заменяя U на A, получаем AGA, или одну из кодовых записей лизина. Такой же анализ проводился и для других белков, в которых были известны замененные аминокислоты, и оказалось, что в большинстве случаев подобные замены могут быть объяснены изменением одного основания в кодовых триплетах. Вероятно, в будущем удастся определить и другие кодовые слова, и тогда сравнение между последовательностями оснований и последовательностями аминокислот можно будет провести более строго.

ПРИРОДА РНК-ПОСРЕДНИКА

Участвует ли каждая молекула РНК-посредника в синтезе белка один раз или много раз? Этот вопрос оказался весьма трудным, так как в экспериментальных системах большая часть поли-U разрушается прежде, чем он начинает функционировать как посредник. Тем не менее мы обнаружили, что на каждую молекулу фенилаланина, участвовавшую в синтезе белка, приходится 1,5 U в молекуле поли-U. Спирисдом и Липманом была получена величина 0,75 С. Если код троичный, то при условии, что посредник действует всего один раз, на каждую аминокислоту приходилось бы три U. Очевидно, каждая молекула поли-U управляет синтезом нескольких длинных молекул полифенилаланина. Сходные результаты были получены и для живых клеток. Левинталь и его сотрудники с помощью антибиотика актиномицина вызывали синтез белка в клетке живой бактерии и обнаружили, что каждая молекула РНК-посредника, оставшаяся после прекращения синтеза, участвовала в построении 10—20 молекул белка.

Кроме последовательности оснований, на активность РНК-посредника сильно влияют еще два фактора: длина цепи молекулы РНК и ее полная структура. Молекулы поли-U, содержащие 100 оснований, гораздо активнее, чем молекулы из менее чем 50 оснований. Мартин и Эймс установили, что оптимальное число оснований U для синтеза белка равно 450—700.

Много еще неясного в вопросе о влиянии пространственной структуры молекулы РНК на ее активность. В отличие от ДНК молекулы РНК почти всегда состоят из одной нити. Однако часто наблюдаются случаи, когда часть молекулы РНК изгибается и соединяется водородными связями с другой частью той же молекулы. Степень подобной закрученности

зависит от последовательности оснований в молекуле. Обычно молекулы поли-У, находящегося в растворе, имеют слабо выраженную вторичную структуру, т. е. они представляют собой простую цепь с несколькими петлями и узлами. Молекулы же других РНК-полимеров характеризуются значительной вторичной структурой (см. рис. 17).

Мы установили, что вторичная структура сильно влияет на активность РНК-посредника. Если смешать растворы поли-У и поли-А, то образуются двухнитевые (У — А) и трехнитевые (У — А — У) спирали, которые совершенно не участвуют в управлении синтезом полифенилаланина. В сотрудничестве с Зингером мы показали, что поли-UG, обладающий ярко выраженной упорядоченной вторичной структурой (возможно благодаря G — G-водородным связям), неспособен кодировать аминокислоты.

Возможно, естественный РНК-посредник имеет на некоторых расстояниях друг от друга небольшие области со вторичной структурой, напоминающие узлы на веревке. Эти области могут обозначать начало или конец белка. Согласно другой гипотезе начало или конец белка обозначаются определенными последовательностями оснований. В любом случае вполне вероятно, что вторичная структура некоторых видов РНК-полимеров оказывает существенное влияние на их биологические функции.

СЧИТЫВАЮЩИЙ МЕХАНИЗМ

Еще не совсем ясно, каким образом данная аминокислота находит нужное место в цепи молекулы белка. Хотя мы установили, что для синтеза полифенилаланина необходимо присутствие РНК-переносчика, не исключена возможность, что кодовые слова, зашифрованные в молекуле поли-У РНК-посредника, опознавались не РНК-переносчиком, а самой аминокислотой.

Чтобы выяснить, какое из этих двух альтернативных предположений справедливо, был проведен следующий блестящий эксперимент. Его осуществили совместными усилиями Шапевилль и Липман, Гюнтер фон Эренштейн, а также Бензер, Вейсблом и Вильям Рэй. Одна аминокислота, цистеин, направлялась в белок с помощью поли-UG. Другая, аланин, молекула которой совпадает с молекулой цистеина, за исключением того, что в ней отсутствует атом серы,— с помощью поли-CG или поли-UCG. Цистеин и аланин переносились разными видами РНК-переносчиков. С помощью ферментов цистеин, отмеченный радиоактивным Ca^{14} , прикреплялся к своему РНК-переносчику. Затем весь этот молекулярный комплекс подвергался воздействию никелевого катализатора (Raney nickel), который удалял атом серы из молекулы цистеина, превращая ее в молекулу аланина, причем сама молекула оставалась на своем РНК-переносчике. Спрашивается теперь, будет ли полученная аминокислота кодироваться как аланин или как цистеин? Оказалось, что она кодируется с помощью поли-UG, т. е. как если бы она была цистеином (см. рис. 20). Этот эксперимент показывает, что после прикрепления к молекуле РНК-переносчика аминокислота утрачивает свои специфические свойства и насилино транспортируется затем к нужному кодовому слову, которое опознается РНК-переносчиком.

За прошедший год было проведено дальнейшее изучение свойств вторичной структуры молекул РНК-переносчиков. Из результатов экспериментов по дифракции рентгеновских лучей они установили, что молекула РНК-переносчика представляет собой двойную спираль, напоминающую спираль молекулы ДНК. Отличие состоит в том, что молекула РНК-переносчика закручена сама на себя, подобно заколке для волос. Молекула содержит целый ряд непарных оснований; возможно, что эта

особенность придает ей способность опознавать кодовые слова, записанные в молекуле РНК-посредника (см. рис. 16).

Мы еще совершенно не знаем, каким образом РНК-посредник прикрепляется к рибосомам и какую роль играют последние в процессе синтеза белков. Было известно, что рибосомы кишечной палочки бывают по крайней мере двух видов, а при определенных условиях они образуют агрегаты из двух (димеры) и четырех (тетрамеры) подобластей. Совместно с Самуэлем Баронде мы установили, что добавление в раствор поли-U вызывает дальнейшую агрегацию рибосом. В ранних экспериментах синтез полифенилаланина происходил только в тетрамерах или в еще более сложных агрегатах. Спирисом и Липманом было показано, что добавление поли-U приводит к агрегации лишь определенных, «активных» рибосом, в остальных же мономерах и димерах синтез полифенилаланина не происходит.

Подобное явление, вероятно, наблюдалось Ричем и его сотрудниками в живых клетках. Они установили, что синтез белка в ретикулоцитах (*reticulocytes*), взятых из крови кролика, происходит в основном в агрегатах, состоящих из пяти рибосом и охваченных одной нитью молекулы РНК-посредника. Такие агрегаты были названы ими полисомами.

Перед нами стоит еще много нерешенных проблем. Одна из них заключается в выяснении действительных последовательностей оснований в кодовых словах. Код сейчас напоминает анаграмму. Мы знаем буквы, но не знаем их порядок в большинстве слов.

Другая захватывающая проблема состоит в следующем: служат ли в живых клетках шаблоном для образования одной нити РНК-посредника обе нити молекулы ДНК или на них образуются две различные, дополнительные молекулы РНК-посредника? Если имеет место последнее, а об этом свидетельствуют имеющиеся опытные данные, то необходимо выяснить функции каждой из этих нитей.

В конце концов мы надеемся, что изучение процессов в бесклеточных системах прольет свет на механизм генетического контроля. Этот механизм, еще совершенно не изученный, позволяет селективно исправлять генетическую информацию. Две клетки могут содержать одинаковые наборы генов, однако в одной клетке строго определенные гены будут функционировать, а в другой нет. Использование бесклеточных систем является мощным методом изучения тех или иных вопросов ферментологии, позволяющим надеяться, что в ближайшем будущем наши знания о молекулярной генетике значительно расширятся.

ЛИТЕРАТУРА

1. S. Benzer, *Sci. American* **206** (1), 70 (1962).
2. F. H. Crick, L. Barnett, S. Brenner and R. J. Watts-Tobin, *Nature* **192**, No. 4809, 1227 (1961).
3. J. Hug with and J. J. Furth, *Sci. American* **206** (2), 41 (1962).
4. The Nucleic Acids, vol. 3, ed. by E. Chargaff and J. N. Davidson, Academic Press, N. Y., 1960.
5. M. W. Nierenberg and J. H. Mattaei, *Proc. Natl. Acad. Sci. USA* **47** (10), 1588 (1961).
6. B. Weissblum, S. Benzer and R. W. Holley, *Proc. Natl. Acad. Sci. USA* **48** (8), 1449 (1962).
7. M. S. Bretscher and M. Grunberg-Manago, *Nature* **195**, No. 4838, 283 (1962).
8. O. W. Jones, Jr., and M. W. Nierenberg, *Proc. Natl. Acad. Sci. USA* **48** (12), 2115 (1962).
9. J. E. Speyer, P. Lenguel, C. Basilio and S. Ochoa, *Proc. Natl. Acad. Sci. USA* **48** (3), 441 (1962).