

АКАДЕМИЯ НАУК СССР

1966

БИОФИЗИКА

ТОМ

VIII

ВЫПУСК ПЯТЫЙ

1963

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР

МОСКВА

11-252

577.3

О ГЕНЕТИЧЕСКОМ КОДЕ*

ВЫВОДЫ ОБ ОБЩЕЙ ПРИРОДЕ КОДА,
ВЫТЕКАЮЩИЕ ИЗ БИОХИМИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Ф. Г. К. КРИК

Сейчас можно считать доказанным, что последовательность аминокислот любого белка определяется последовательностью оснований в некоторой области особой молекулы нуклеиновой кислоты. В белках находят двадцать различных видов аминокислот, в нуклеиновых кислотах — четыре основных вида оснований. Генетический код — это правило, по которому последовательность из 20 или более элементов определяется последовательностью из четырех элементов другого типа.

Вряд ли необходимо особо подчеркивать биологическое значение этой проблемы. Кажется вероятным, что большую часть (если не всю) генетической информации в любом организме несут нуклеиновые кислоты — обычно ДНК, хотя некоторые малые вирусы используют в качестве генетического материала РНК. Можно допустить, что большая часть этой информации используется, чтобы задать последовательность аминокислот в белках этих организмов. (Имеет ли генетическая информация какую-либо другую важную функцию, мы пока не знаем.) Эта идея выражена в классическом изречении Дж. Бидля: «Один ген — один фермент», или в менее удобоваримой современной терминологии: «Один дистрон — одна полипептидная цепь».

Одно из самых поразительных обобщений биохимии, которое (как это ни удивительно) едва ли даже упоминается в биохимических книгах, — то, что 20 аминокислот и четыре основания за немногими исключениями одинаковы для всей природы. Насколько я помню, принятый сейчас набор из 20 аминокислот был впервые составлен Дж. Д. Уотсоном и мной летом 1963 г. в ответе на письмо Г. А. Гамова.

Я не буду стремиться осветить здесь детали проблемы по той причине, что я недавно написал обзор на эту тему [1], который скоро будет опубликован. Не буду также касаться биохимических тонкостей белкового синтеза и синтеза РНК-переносчика. Я лишь постараюсь поставить основные общие вопросы о генетическом коде и ответить на них постольку, поскольку это сегодня возможно.

Давайте примем, что генетический код имеет простой характер, и зададимся вопросом — сколько оснований кодируют одну аминокислоту? Такое кодирование вряд ли может быть проведено парой оснований, так как из различных четырех элементов можно образовать только $4 \times 4 = 16$ различных пар, тогда как необходимо по крайней мере 20 для синтеза аминокислот и, вероятно, еще 1—2 для иных целей, например

* Эта статья представляет собой выдержки из лекции, прочитанной Ф. Криком в Стокгольме, Швеция, 11 декабря 1962 г., при получении Нобелевской премии по медицине и физиологии, которую он разделил с Дж. Д. Уотсоном и М. Уилкинсом. Речь публикуется с разрешения Нобелевского комитета. Этот текст будет также включен в полный том Нобелевских лекций, которые ежегодно публикует Издательство Эльзевир, Амстердам — Нью-Йорк.

для заполнения промежутков. Однако триплеты оснований дадут нам 64 возможности. Я буду использовать слово *кодон* для обозначения группы оснований, которая кодирует одну аминокислоту.

Это приводит нас к первому вопросу — перекрываются ли кодоны? Другими словами, найдем ли мы, читая генетическую информацию, основание, которое является членом двух или более кодонов? Сейчас можно считать вполне доказанным, что кодоны не перекрываются. Если бы это было не так, изменение одного основания в результате мутации вызывало бы изменение двух или более (соседних) аминокислот. Однако типичным является изменение одной аминокислоты как в случае спонтанных мутаций, так и в случае патологических изменений в гемоглобине человека и в химически индуцированных мутациях (вызванных, например, действием азотистой кислоты и других химических веществ на вирус табачной мозаики) [2]. Таким образом, по всей видимости, кодоны не перекрываются.

Это приводит нас к следующей проблеме. Как последовательность оснований делится на кодоны? В остове нуклеиновой кислоты, который является совершенно регулярным, ничто не может подсказать нам, как сгруппировать основания в кодоны. Например, если все кодоны являются триплетами, то кроме правильного чтения информации будут иметь место два неправильных чтения, которые мы получим, если будем начинать группировку в тройки с ошибочного места. Мои коллеги и я [3] недавно получили экспериментальные доказательства того, что каждый участок генетического сообщения действительно читается с фиксированной точки, возможно, с одного конца.

Это весьма хорошо согласуется с экспериментальными данными (наиболее убедительны — результаты Динтиса [4]), из которых следует, что аминокислоты объединяются в полипептидную цепь в линейном порядке, начиная с аминного конца цепи.

РАЗМЕР КОДОНА

Это приводит нас к основному вопросу — о величине кодона. Сколько оснований в кодоне? Эксперименты, на которые я только что ссылался [3], доказывают, что все (или почти все) кодоны состоят из троек оснований, хотя небольшие величины, кратные трем, такие как 6 или 9, как это следует из наших данных, не могут быть полностью исключены. Мы пришли к этому заключению, исследуя А и Б цистроны r_{11} локуса бактериофага T4. Мутации в нашем случае возникали из-за добавления или удаления одного или нескольких оснований из последовательности (генетического текста). Мутации такого рода характерны для акридина, и они никогда не реверсируют к «дикому типу» под действием тех мутагенов, которые только изменяют одно основание на другое. Более того, такие мутации делают ген почти всегда полностью неактивным.

Исследуя такие мутации попарно, мы смогли отнести их все без исключения к двум классам, которые мы назвали «плюс-класс» и «минус-класс». Для простоты будем считать, что плюс-класс имеет лишнее основание в той или иной точке генетического сообщения, а минус-класс характеризуется утратой одного основания. Узловой эксперимент состоял во введении при помощи генетической рекомбинации трех мутаций одного типа в один ген, т. е. либо (+с+ и с—), либо (—с— и с—). Хотя один + или их пара (+с+) делают ген полностью неактивным, группа из трех+, подходящим образом выбранных, имеет такую же активность, как дикий тип. Детальное рассмотрение этих экспериментальных результатов показывает, что они именно такие, каких следовало бы ожидать, если информация читается по тройкам, начиная с одного конца.

Нас иногда спрашивали, каков будет результат, если поместить четыре + в один ген. Чтобы ответить на это, мои коллеги поместили недавно вместе не только четыре, но и шесть +. Как и ожидалось на основе нашей теории, такая комбинация активна, хотя группа из четырех или пяти + неактивна. Мы также проделали большую работу в направлении объяснения так называемых слабых мутантов, т. е. комбинаций, в которых ген работает с очень низкой эффективностью. Результаты тщательных исследований укладываются в гипотезу, что в некоторых случаях, когда считывающий механизм встречает бессмысленный триплет (так называемый нонсенс-триплет), могут совсем случайно происходить описки и читаться, скажем, только два основания вместо обычных трех. Эти результаты позволили нам также вывести направление чтения генетического сообщения, которое в этом случае читается слева направо для области r_{11} , как ее обычно рисуют. Окончательное доказательство наших гипотез может быть получено только при детальном изучении изменений, вызываемых в последовательности аминокислот в белках мутациями описанного типа.

Наши результаты позволяют сделать следующий вывод общего характера. Кажется, что число нонсенов — бессмысленных триплетов — скорее мало, так как мы лишь случайно сталкиваемся с ними. Однако это заключение не столь надежно, как другие наши выводы об общей природе генетического кода.

КОЛИНЕАРНОСТЬ

До сих пор еще не было показано прямым способом, что генетическое сообщение коллинеарно со своим продуктом — белком, т. е. что один конец гена кодирует аминный конец полипептидной цепи, а другой — карбоксильный, и что по мере следования вдоль гена встречаются кодоны, расположенные по отношению друг к другу в том линейном порядке, в котором следуют аминокислоты в полипептидной цепи. Кажется весьма правдоподобным (и это должно быть показано), что несколько мутаций, сказывающихся на одной и той же аминокислоте, чрезвычайно близки друг к другу на генетической карте. Экспериментальных доказательств коллинеарности гена и полипептидной цепи можно с уверенностью ожидать в ближайшие несколько лет.

УНИВЕРСАЛЬНОСТЬ

Существует следующий общий вопрос о генетическом коде, который мы можем теперь поставить. Является ли код универсальным, т. е. одинаков ли он во всех организмах? Предварительные данные позволяют предположить, что это должно быть так. Например, в бесклеточной системе, часть которой взята из ретикулоцитов кролика и часть от *Escherichia coli* [5], может синтезироваться белок, очень похожий на гемоглобин кролика. Этого трудно было бы ожидать, если бы код этих двух организмов сильно различался. Однако, как мы покажем сейчас, имеются новые возможности проверить универсальность кода экспериментами более прямого характера.

ШТУРМ ПРОБЛЕМЫ ГЕНЕТИЧЕСКОГО КОДА

Считается, что если генетическим материалом клетки является ДНК, она не сама непосредственно контролирует синтез белка, а последовательность оснований ДНК (возможно, только одной ее цепи) копируется на РНК, и эта особая РНК затем действует как генетический посредник и фактически направляет процесс соединения аминокислот в полипептидную цепь. Прорыв в проблеме кода начался открытием Ниренберга

и Маттен [6], которые использовали синтетическую РНК в качестве посредника. В частности, они нашли, что полиуридиловая кислота (РНК, в которой каждое основание — урацил), будучи добавлена к уже известной в то время бесклеточной системе, синтезирующей белок, вызывает синтез полифенилаланина. Итак, один кодон для фенилаланина представляется группой УУУ, где У обозначает урацил (точно так же мы используем А, Г и Ц для обозначения аденина, гуанина и цитозина соответственно). Этот результат открыл дорогу быстрой, хотя иногда приводящей к противоречивым результатам атаке на код.

Мы не будем здесь детально обсуждать эти работы. Первые из них я уже критически проанализировал в упомянутом обзоре [1], но специфика деятельности на этом поприще такова, что недавние эксперименты, которые составляли предмет дискуссии, в известной степени уже устарели. Однако некоторые общие заключения можно с уверенностью сделать.

Методика, в основном использованная не так давно и Ниренбергом с коллегами [6], и Очао и его группой [7], заключалась в ферментативном синтезе «беспорядочных полимеров» из двух, трех или четырех оснований. Например, используя полинуклеотид (который я в дальнейшем буду называть поли-У,Ц), содержащий примерно равные доли урацила и цитозина в случайном порядке (предположительно), можно получить увеличение включения аминокислот фенилаланина, серина, лейцина, пролина и, с меньшей достоверностью, треонина. Используя полимеры с разным составом и принимая триплетный код, можно получить некоторую информацию о составе определенных триплетов.

Из такого рода работ стало ясно, что, за немногими исключениями, каждый полинуклеотид включает характерную серию аминокислот. Кроме того, четыре основания дают совершенно различные эффекты. Сравнение между триплетами, созданными опытным путем при помощи этих методов, показывает вполне хорошее согласие с данными по заменам в аминокислотной последовательности, вызванным мутациями. Более того, включение требует тех же самых компонент, которые необходимы для белкового синтеза, и ингибируется теми же самыми ингибиторами. Таким образом, система совсем не похожа на полный артефакт и весьма вероятно, что процесс родствен истинному белковому синтезу.

Сначала думали, что, возможно, каждый триплет включает урацил, но это не является правдоподобным с теоретической точки зрения и не подтверждается экспериментальными данными. Первое прямое доказательство того, что это не так, было получено моими коллегами Бретчером и Грюнберг-Манаго [8], которые показали, что поли-Ц,А стимулирует включение нескольких аминокислот. Недавно другие исследователи [9; 10] сообщили дальнейшие свидетельства такого рода в пользу других полинуклеотидов, не содержащих урацила. Сейчас представляется весьма вероятным, что многие из 64 триплетов (если не большинство) могут кодировать ту или иную аминокислоту и что вообще несколько различающихся триплетов могут кодировать одну аминокислоту. В частности, очень остроумный эксперимент [11] позволяет считать, что и УУС, и УУГ кодируют лейцин (заметим, что порядок внутри триплетов еще неизвестен). Эта общая идея подтверждается несколькими косвенными доказательствами, которые здесь не могут быть детально представлены. К сожалению, однозначное определение триплетов подобными методами много труднее, чем в случае, когда каждую аминокислоту кодирует только один триплет. Кроме того, используя полинуклеотиды с «беспорядочной» последовательностью, невозможно определить порядок оснований в триплете. Начало положено синтезом полинуклеотидов, точная последовательность которых с одного конца известна, но полученные до сего времени результаты скорее обещающие, чем

убедительные [12]. Представляется вероятным, однако, из этих и других (неопубликованных) данных, что аминный конец полипептидной цепи соответствует «правому» концу полинуклеотидной цепи, т. е. цепи с 2', 3' гидроксилами сахара.

Можно считать фактически доказанным, что одиночная цепь РНК может работать как РНК-посредник; такова полиуридиловая кислота — одиночная цепь без вторичной структуры. Если к поли-У добавить поли-А, что приводит к образованию двойной или тройной спирали, то такая комбинация неактивна. Более того, имеются предварительные данные [9], которые позволяют предположить, что вторичная структура полинуклеотидов ингибирует интенсивность белкового синтеза.

Прямыми биохимическими методами еще должно быть доказано (в противоположность косвенным генетическим доказательствам, упомянутым выше), что код в самом деле триплетен.

На основе изучения изменений, вызванных мутациями, были сделаны попытки получить относительный порядок оснований внутри различных триплетов, но, по-моему, такие попытки преждевременны до тех пор, пока не будет более полных и надежных данных о составе триплетов.

Данные, представленные несколькими группами исследователей [8; 9; 11], показывают, что поли-У стимулирует включение как фенилаланина, так и (в меньшей мере) лейцина. Объяснить этот результат пока нельзя, но из него возникает неприятная возможность двусмысленных триплетов, т. е. триплетов, которые могут кодировать более чем одну аминокислоту. Однако можно ожидать с уверенностью, что такие триплеты находятся в меньшинстве.

ВОЗНИКНОВЕНИЕ ГРУППИРОВАНИЯ

Похоже на то, что большинство из 64 триплетов можно сгруппировать в 20 групп. Сумма сведений, полученных как в экспериментах с бесклеточной системой, так и при изучении мутаций, позволяет предположить, что такое группирование не является случайным и что триплеты, кодирующие одну и ту же аминокислоту, должны быть в какой-то мере сходными. Это приводит к главной теоретической проблеме, по сей день не разрешенной. Может ли это группирование быть выведено из теоретических постулатов? К сожалению, нетрудно видеть, что группирование могло возникнуть посредством случайных мутаций на чрезвычайно ранних ступенях эволюции, так что определенный код, возможно, является результатом серий исторических случайностей. Эта точка зрения представляет не только абстрактный интерес. Если код в самом деле имеет некоторые логические основы, тогда закономерно рассматривать все построения, удачные и неудачные, в любой попытке вывести их. В то же время это будет неправильным, если кодоны не имеют простой логической взаимосвязи. В таком случае мало смысла отгадывать кодон; самое важное — обеспечить достаточно данных, чтобы доказать каждый кодон независимо. Еще неясно, какие данные можно считать надежно устанавливающими кодон. Но ясно, что большинство не так давно представленных экспериментальных данных — почти во всех случаях не очень убедительные доказательства.

ОШИБОЧНЫЕ ПРЕДПОЛОЖЕНИЯ

В противовес неопределенности некоторых экспериментальных данных многие коды, предложенные в прошлом, мы можем с уверенностью отвергнуть.

1. *Триплетный код без запятых.* Не только на основе генетических данных, но и на основе результатов тщательных экспериментов с бесклеточной системой такие коды можно считать мало вероятными.

2. *Двухбуквенные или трехбуквенные коды.* Например код, в котором А эквивалентно Ц и Г эквивалентно У. Как уже отмечалось, результаты, полученные на бесклеточной системе, противоречат всем кодам такого характера.

3. *Комбинационный триплетный код.* В этом коде все перемены в данной комбинации кодируют одну и ту же аминокислоту.

Только путем особых ухищрений экспериментальные результаты можно согласовать с таким кодом.

4. *Комплементарный код.* Их имеется несколько классов.

Рассматривается определенный триплет в связи с триплетом, который комплементарен ему в другой цепи двойной спирали. Второй триплет может рассматриваться читаемым либо в том же направлении, что и первый, либо в противоположном направлении. Так, если первый триплет — УЦЦ, мы рассмотрим его в связи либо с АГГ, либо (чтение в противоположном направлении) с ГГА.

Предполагалось, что если триплет обозначает аминокислоту, то комплементарный к нему триплет с необходимостью должен обозначать ту же аминокислоту или в других классах кодов, напротив, комплементарный триплет не должен соответствовать ни одной аминокислоте, т. е. быть нонсенсом.

Группой Очоа недавно было показано, что поли-А стимулирует включение лизина [10]. Таким образом, лизин предположительно кодирует ААА. Однако, так как УУУ кодирует фенилаланин, это отвергает все вышеупомянутые коды. Было также найдено, что поли-У,Г включает совсем другие аминокислоты, чем поли-А,С. Подобным образом поли-У,С отличается от поли-А,Г [9; 10]. Итак, мало шансов, что какая-либо из этих теорий окажется правильной. Кроме того, они все, по-моему, неудовлетворительны из общих теоретических соображений.

Уже положено начало исследованию роли одинаковых полинуклеотидов в бесклеточных системах различных видов, цель которых проверить, одинаков ли код во всех организмах. Возможно, будет относительно легко показать таким способом, является ли код универсальным и если нет, то как он отличается от организма к организму. Предварительные данные, полученные недавно, не обнаружили ясного различия в отношении кода между *E. coli* и млекопитающими, и это ободряет нас [10; 13].

ОБЩИЕ СВОЙСТВА

Таким образом, в настоящее время представляется, что генетический код должен иметь следующие общие свойства.

1. Большинство (если не все) кодонов должно состоять из трех (прилежащих) оснований.

2. Соседние кодоны не перекрываются.

3. Информация читается правильными группами (тройками), начиная с некоторой фиксированной точки.

4. Кодовая последовательность в гене коллинеарна аминокислотной последовательности полипептидной цепи, синтезируемой почленно с аминного конца.

5. Вообще говоря, каждую аминокислоту кодирует более чем один триплет.

6. Возможно, что некоторые триплеты могут кодировать более чем одну аминокислоту, т. е. они могут быть двусмысленными.

7. Триплеты, которые кодируют одну аминокислоту, вероятно, в некоторой степени сходны.

8. Неизвестно, имеется ли общее правило, в согласии с которым такие триплеты группируются вместе, или группирование есть в основном результат исторических случайностей.

9. Число триплетов, которые не кодируют ни одной аминокислоты, вероятно, мало.

10. Определенные коды, предложенные ранее — такие как коды без запятых, двух- или трехбуквенные коды, комбинационный код и различные перестановочные коды, — являются неправильными.

11. Коды различных организмов, вероятно, имеют много общего. Возможно, что они одинаковы, но это пока неизвестно.

Наконец, можно добавить, что, несмотря на огромную сложность синтеза белка, несмотря на значительные экспериментальные трудности в синтезе полинуклеотидов с определенной последовательностью, можно надеяться, что все эти вопросы будут выяснены в недалеком будущем и генетический код будет полностью установлен на твердой экспериментальной основе в течение ближайших нескольких лет.

ЛИТЕРАТУРА

1. Crick F. H. C., *Progress in Nucleic Acid Research*, J. N. Davidson and W. E. Cohn, Eds. Acad. Press, N. Y. (в печати)
2. Wittmann H. G., *Z. Vererbungslehre* (в печати); Tsugita A., *J. Mol. Biol.* 5, 284, 293, 1962
3. Crick F. H. C., Barnett L., Brenner S., Watts-Tobin R. J., *Nature* 192, 1227, 1961
4. Naughton M. A., Dintzis H. M., *Proc. Natl. Acad. Sci. USA* 48, 1822, 1962
5. von Ehrenstein G. a. Lipmann F., там же 47, 941, 1961
6. Matthaei J. H. a. Nirenberg M. W., там же 47, 1580, 1961; Nirenberg M. W. a. Matthaei J. H. там же 47, 1588, 1961; Nirenberg M. W., Matthaei J. H., Jones O. W., там же 48, 104, 1962; Matthaei J. H., Jones O. W., Martin R. G., Nirenberg M. W., там же 48, 666, 1962
7. Lengyel P., Speyer J. F., Ochoa S., там же 47, 1936, 1961; Speyer J. F., Lengyel P., Basilio C., Ochoa S., там же 48, 63, 1962; Lengyel P., Speyer J. F., Basilio C., Ochoa S., там же 48, 282, 1962; Speyer J. F., Lengyel P., Basilio C., Ochoa S., там же 48, 441, 1961; Basilio C., Wahba A. J., Lengyel P., Speyer J. F., Ochoa S., там же 48, 631, 1962
8. Bretscher M. S. a. Grunberg-Manago M., *Nature* 195, 283, 1962
9. Jones O. W. a. Nirenberg M. W., *Proc. Natl. Acad. Sci. USA* (в печати)
10. Gardner R. S., Wahba A. J., Basilio C., Miller R. S., Lengyel P., Speyer J. F., там же (в печати)
11. Weisblum B., Benzer S., Holley R. W., там же 48, 1449, 1962
12. Wahba A. J., Basilio C., Speyer J. F., Lengyel P., Miller R. S., Ochoa S., там же 48, 1683, 1962
13. Arnstein H. R. V., Cox R. A., Hunt J. A., *Nature* 194, 1042, 1962; Maxwell E. S., *Proc. Natl. Acad. Sci. USA* 48, 1639, 1962; Weinstein B. a. Schechter A. N., там же 48, 1686, 1962

ФРЕНСИС ГАРРИ КОМПТОН КРИК

Френсис Гарри Комптон Крик родился в 1916 г. в Нортхэмптоне (Англия). В 1937 г. он получил степень бакалавра наук по физике в Университетском колледже Лондона. Во время войны Ф. Крик проводит исследовательскую работу при Морском министерстве. В 1947 г. он работает в лаборатории Кавэндиша в Кембридже, в группе, занимающейся изучением строения белка. Впоследствии эта группа (1962 г.) выделлась в отдельный Институт молекулярной биологии Совета по медицинским исследованиям при Кембриджском университете. Ф. Крик продолжает работать в этом институте и руководит исследованиями в области молекулярной генетики.

С 1959 г. Ф. Крик — член Королевского общества, с 1962 г. — иностранный почетный член Американской Академии искусства и науки. Он получил следующие знаки отличия: премию Уоррена вместе с Дж. Уотсоном, премию Ласкера совместно с Дж. Уотсоном и М. Вилкинсом, премию Чарльза Леопольда Мейса, премию Исследовательской корпорации вместе с Дж. Уотсоном, премию Гарднеровского фонда и, наконец, Нобелевскую премию совместно с Дж. Уотсоном и М. Вилкинсом в 1962 г.