

Генетические коды

Доктор физико-математических наук
М. Д. Франк-Каменецкий

Химия и Жизнь №5, 1980 г., с. 34-41

ПРОБЛЕМА КОДА

Какие бы захватывающие дух открытия ни были сделаны в будущие двадцать лет, выяснение механизма биосинтеза белка и расшифровка генетического кода останутся главными событиями в биологии XX века.

Как только в 1953 г. появилась знаменитая двойная спираль Уотсона и Крика и стало ясно, что наследственная информация записана в виде последовательности нуклеотидов ДНК, сразу же возник вопрос: как эта запись прочитывается и используется в клетке? Ясной постановкой проблемы и привлечением к ней широкого внимания наука обязана Георгию Гамову. Этот выдающийся физик впервые сформулировал в 1954 году проблему генетического кода. Предстояло понять, как ДНК-овая последовательность, состоящая из четырех сортов нуклеотидов, переводится в белковую последовательность, состоящую из 20 сортов аминокислотных остатков. Гамов предложил вариант словаря, переводящего тексты с четырехбуквенного языка ДНК на двадцати-буквенный язык белка. Это был первый, пока чисто умозрительный, вариант генетического кода.

Генетический код - ключ к пониманию того, как в клетке реализуется наследственная информация. Код определяет правила синтеза белка - наиболее фундаментального и универсального процесса, происходящего в живой клетке. Ведь именно от набора белков зависит работа клетки. Все ферменты - это белковые молекулы. Все небелковые компоненты клетки, включая молекулу ДНК, синтезируются, а в случае ДНК еще и копируются при помощи ферментов.

Естественно, сам биосинтез белка идет с участием многочисленных белков-ферментов. Прежде всего, с помощью фермента, называемого РНК-полимеразой, с какого-то отрезка ДНК снимается копия в виде молекулы РНК. На участке, с которого снята копия, записаны сведения о строении одного белка. Этот участок и называют геном. РНК-овая копия гена получила название матричной РНК (ее сокращенно обозначают мРНК, чтобы не путать с некоторыми другими РНК, также участвующими в биосинтезе белка). Последовательность нуклеотидов в мРНК идентична последовательности нуклеотидов в скопированном участке ДНК, то есть это точный отпечаток гена.

РНК-овый текст также состоит из четырех букв. Именно по нему и синтезируется белок. Происходит это в чрезвычайно сложно устроенном агрегате, включающем десятки различных белковых молекул и называемом рибосомой. Рибосома, кроме белка, состоит также из специальных, рибосомальных РНК (рРНК). Химическая структура этих РНК в принципе такая же, как у мРНК. Однако у них нуклеотиды расположены совсем в другом порядке. Конечно, эти рРНК также копируются с ДНК, то есть в ДНК есть специальные гены рибосомальных РНК. Так что гену вовсе не обязательно должен соответствовать какой-то белок.

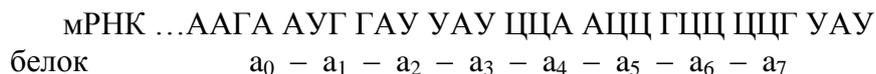
Рибосома пропускает через себя нить мРНК, читает текст, записанный в виде последовательности нуклеотидов на этой нити, и синтезирует соответствующую белковую, то есть аминокислотную, последовательность. Таким образом, именно на рибосоме язык нуклеиновых кислот (ДНК и РНК) переводится на язык аминокислотной последовательности белков.

Мы не будем здесь углубляться в детали этого очень интересного процесса, о котором теперь известно достаточно много. Отметим только, что важную роль в декодировании текста играет еще третий вид РНК - транспортные, или тРНК, которые переносят аминокислоты из цитоплазмы к рибосомам, то есть подтаскивают материал для построения белка.

Итак, в клетке существует очень сложный механизм биосинтеза белка, реализующий генетический код. Но каков же сам код?

На рубеже 50-х и 60-х годов Фрэнсис Крик и его сотрудники выяснили основные свойства

кода. Было доказано, что код триплетный, то есть одной аминокислоте соответствует последовательность из трех нуклеотидов на мРНК. Эта тройка нуклеотидов была названа кодоном. Было показано, что текст, записанный в мРНК, считывается рибосомой последовательно, кодон за кодоном, начиная с некоторого начального, иницирующего кодона, по следующей схеме:



На этой схеме a_0, a_1, \dots обозначают аминокислотные остатки белка. Напомним, что их может быть 20 сортов. А сколько сортов кодонов? Легко подсчитать, что всего существует $4^3 = 64$ различных кодона. Так что же, не всякому кодону соответствует аминокислота? Да, не всякому. Но таких бессмысленных или незначащих кодонов очень немного, и они выполняют специальную функцию - служат стоп-сигналами, обозначают конец белковой цепи. Поэтому их еще называют терминирующими кодонами. Подавляющее же большинство из 64 кодонов соответствует какому-либо аминокислотному остатку. А это означает, что код вырожден - большинству, если не всем, аминокислотным остаткам должно отвечать несколько кодонов.

КАК РАСШИФРОВАЛИ КОД

К 1961 году стало ясно, что код триплетный, вырожденный и неперекрывающийся (то есть считывание происходит кодон за кодоном) и что он содержит иницирующие и терминирующие кодоны. Дело было лишь за тем, чтобы установить соответствие каждого аминокислотного остатка конкретным кодоном и узнать, какие кодоны обозначают начало и конец синтеза белковой цепи. Было совершенно ясно, что именно для этого требуется. Нужно «только» прочесть параллельно два текста - ДНК-овый (или РНК-овый) текст гена и аминокислотный текст соответствующего этому гену белка. Затем сравнить эти два текста - и дело в шляпе. Вспомним, что именно так были когда-то расшифрованы египетские письмена. Но беда в том, что если белковые последовательности к этому времени умели расшифровывать, то ни последовательности ДНК, ни последовательности РНК читать не умели. Поэтому пришлось пойти по иному пути.

Представьте себе, что вместо Розетского камня, на котором один и тот же текст был написан египетскими иероглифами и по-гречески, откопали бы во время наполеоновского похода в Египет живого древнего египтянина. Тогда не потребовался бы гений Шампольона, чтобы составить французско-древнеегипетский словарь. Достаточно было бы показывать египтянину различные предметы, а он рисовал бы соответствующие иероглифы.

Именно этим принципом дешифровки кода и воспользовались М. Ниренберг и Дж. Маттеи в 1961 году. Ведь клетки-то знают код! Значит, надо предложить клеткам распознавать разные последовательности нуклеотидов, лишь бы было точно известно, что это за последовательности. К этому времени как раз научились синтезировать кое-какие искусственные РНК (но отнюдь еще не любые!). Но, конечно, живой клетке такую РНК предлагать бесполезно - она ее просто-напросто съест, то есть расщепит до отдельных нуклеотидов, а их использует для строительства собственных РНК. Поэтому Ниренберг и Маттеи использовали не живые клетки, а клеточные экстракты, которые сохраняли способность синтезировать белок на РНК, но не содержали ферментов, расщепляющих РНК. Эти экстракты не умели, разумеется, многого другого, что умеет делать клетка, но важно лишь одно - они были способны синтезировать белок по внесенной извне РНК. Такие экстракты называли бесклеточной системой.

Ниренберг и Маттеи получили экстракт из кишечной палочки и добавили к нему гомополимер, состоящий только из урацилов. Так бесклеточной системе был задан первый вопрос: какой аминокислоте соответствует кодон УУУ? Ответ был однозначен: кодону УУУ отвечает фенилаланин. Этот ответ произвел настоящую сенсацию. Путь к расшифровке кода был открыт!

Очень быстро удалось сделать подобный перевод для многих аминокислот. Однако определять последовательность нуклеотидов в искусственных кодонах было довольно трудно. В то время еще не умели синтезировать даже короткие фрагменты с заданной последовательностью. Умели лишь получать полинуклеотиды со случайной последовательностью из смеси мономеров, да и то не из любой смеси. Начали думать, как попытаться иными способами расшифровывать

различных дублетов может быть всего $4^2 = 16$. Поэтому третий нуклеотид в кодоне должен нести некоторую смысловую нагрузку.

Существует, однако, правило, которому код подчиняется почти строго. Чтобы его сформулировать, нам надо вспомнить, что четыре нуклеотида - урациловый, цитозинный, адениновый и гуаниновый - принадлежат по строению к двум разным классам - пиримидиновому (У и Ц) и пуриновому (А и Г) (см. рис. 2 и 3). Так вот, правило вырожденности кода состоит в следующем: если два кодона имеют два одинаковых первых нуклеотида и их третьи нуклеотиды принадлежат к одному классу (пуриновому или пиримидиновому), то они кодируют одну и ту же аминокислоту.

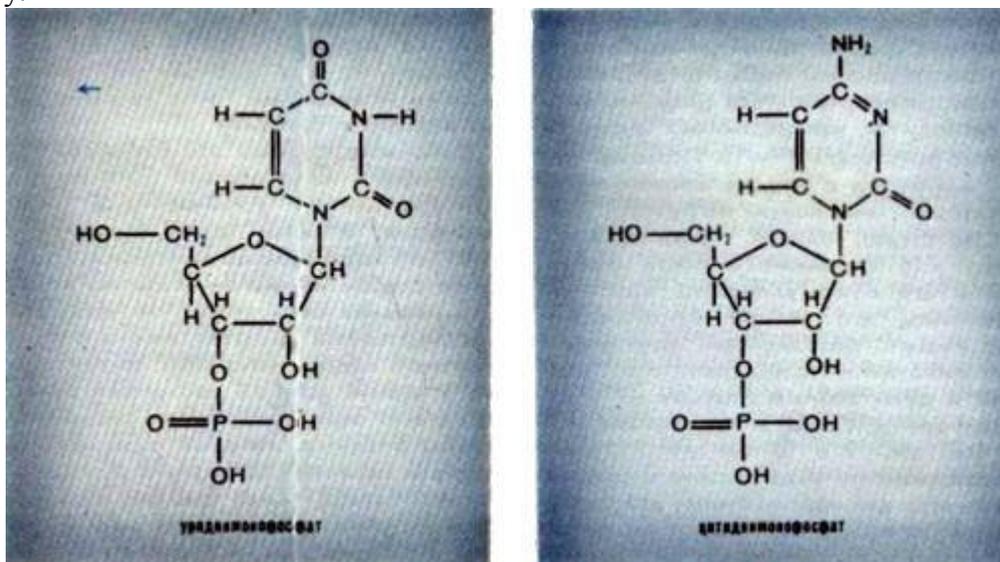


Рис. 2. Пиримидиновый класс

Взгляните еще раз на таблицу кода, и вы убедитесь, что это правило выполняется очень хорошо. Но два исключения все же есть. Если бы сформулированное правило выполнялось совсем строго, то кодон АУА должен был бы отвечать метионину, а не изолейцину, а кодон УГА - триптофану, а не быть сигналом окончания синтеза.

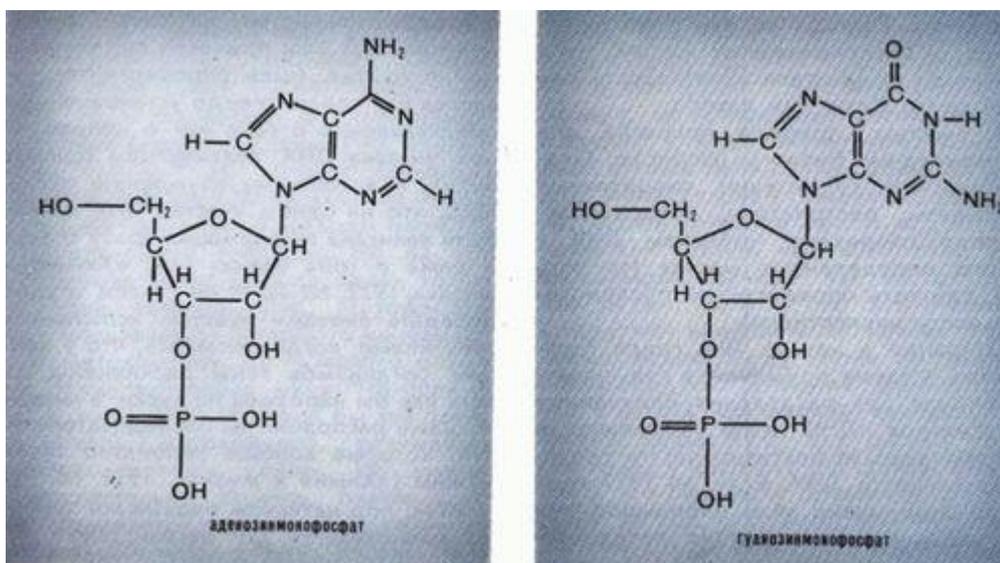


Рис. 3. Пуриновый класс

УНИВЕРСАЛЕН ЛИ КОД?

Но позвольте, вправе спросить читатель, ведь бесклеточная система получена из конкретного организма. Где гарантия, что опыты по расшифровке кода в бесклеточной системе, взятой из другого организма, дадут тот же результат? Вопрос совершенно резонный. И естест-

венно, он возник уже в ходе работ по расшифровке кода.

Первоначально авторы исследований аккуратно оговаривали, что речь идет не о коде вообще, а о коде *E. coli* (кишечной палочки). Именно из этой бактерии была впервые получена бесклеточная система и именно с ней вели работы, о которых было рассказано выше. Однако все свидетельствовало о том, что код других организмов не отличается от кода *E. coli*.

М. Ниренберг повторил опыты, взяв бесклеточные системы из организмов жабы и морской свинки. Никаких отличий от кода *E. coli* эти исследования не выявили. Итак, сомнений как будто бы не оставалось - код универсален.

Правда, были получены мутанты кишечной палочки с некоторыми отклонениями в коде: отдельные терминирующие кодоны читались в них как значащие, то есть отвечали определенным аминокислотам. Такое явление было названо супрессией.

Было ясно, однако, что структура генетического кода должна быть весьма консервативной, устойчивой в ходе эволюции. В самом деле, представим себе, что код внезапно изменился. Пусть даже совсем немного - один из кодонов поменял свой смысл, то есть стал соответствовать другой аминокислоте. Но этот кодон наверняка встречается не в одном гене, а во многих генах. И на всех этих генах будут синтезироваться белки, в которых одна аминокислота заменена на другую. Для некоторых белков такая замена пройдет безнаказанно, они сохранят свои функции. Но очень трудно представить себе, что ни в одном случае не произойдет порча какого-то важного белка. Ведь хорошо известно, что замена одной аминокислоты в одном белке может полностью нарушить его функции и как следствие привести к гибели всего организма. Напомним, например, о серповидноклеточной анемии - тяжелой болезни, вызванной заменой всего одной аминокислоты в молекуле гемоглобина. Эти соображения оставляют мало сомнений в том, что код должен сохраняться неизменным в ходе эволюции.

Вот как оценивал ситуацию в исследованиях кода крупнейший специалист М. Ичас в монографии «Биологический код», опубликованной в 1969 г. (русский перевод вышел в издательстве «Мир» в 1971 г.): «Возможность каких-либо изменений в словаре, и тем более изменений существенных, кажется маловероятной. Имеются, впрочем, некоторые косвенные указания на то, что незначительные изменения в словаре все же происходят. Помимо такого явления, как супрессия, можно указать и еще на один факт значения некоторых кодонов не укладываются в рамки общего правила, определяющего вырожденность кода». Далее Ичас упоминает отличия, о которых мы уже говорили выше, и заключает: «Эти исключения из правила позволяют думать, что некоторые изменения словаря, обусловившие отклонение от «идеальной» структуры, в свое время действительно произошли». Иными словами, Ичас не исключает того, что может быть обнаружен наряду со стандартным кодом еще и «идеальный» код. Однако в то время мало кто обратил внимание на это предвидение. Факты свидетельствовали об одном - код одинаков во всей живой природе.

ДЕСЯТЬ ЛЕТ СПУСТЯ

Итак, код расшифровали и - о нем забыли. Нет, разумеется, о коде читали лекции студентам, о нем писали в учебниках. Но никто им не пользовался. В чем же дело? Очень просто. Ведь код - это словарь, позволяющий переводить нуклеиновые (ДНК-овые или РНК-овые) тексты на аминокислотный (белковый) язык. Этот словарь может быть нужен только тем, кто располагает нуклеиновыми текстами. А таких людей в то время не было.

Но неужели, все-таки, никого не интересовало, что же записано в виде последовательностей нуклеотидов в ДНК и РНК? Еще как интересовало. Важность определения последовательности нуклеотидов в ДНК и РНК стала ясна еще в начале пятидесятых годов, сразу же после работ Уотсона, Крика и Гамова. Но одно дело осознать важность задачи, а совсем другое дело - ее решить.

Вспомните, в те же годы стала ясна и важность проблемы управляемого термоядерного синтеза. А воз и ныне почти что там! В чем главная загвоздка с УТС, все знают - нужны громадные температуры и высокие плотности, чтобы преодолеть кулоновское отталкивание одноименно заряженных ядер. А в чем была загвоздка в прочтении нуклеиновых последовательностей? Уж очень они длинные, эти молекулы нуклеиновых кислот. Как-то надо

было научиться разделять их на более короткие куски. Причем не как попало, конечно, а в определенных местах. Чего только не предлагали использовать для этой цели. Даже луч лазера.

К счастью, природа припасла для молекулярных биологов сюрприз. Оказалось, что существуют специальные ферменты, рестриктазы, которые аккуратно разрезают молекулы ДНК в строго определенных местах (то есть там, где есть определенная последовательность нуклеотидов).

Открытие в начале 70-х годов рестриктаз, а также применение гель-электрофореза для разделения полученных фрагментов ДНК дало возможность чрезвычайно эффективно читать последовательности ДНК. Если до 1977 года было определено лишь небольшое число сравнительно коротких последовательностей, то начиная с 1977 г. расшифрованные последовательности посыпались как из рога изобилия. И вот тут-то код, лежавший без дела целых десять лет, заработал вовсю. Наконец представилась возможность напрямую сравнить между собой естественные нуклеиновые и белковые тексты и окончательно проверить, действительно ли код таков, каким его привыкли считать.

Никаких поправок делать не пришлось - код оказался расшифрованным абсолютно верно.

Новой, очень критической проверке подвергся и тезис об универсальности кода. В самом деле, ведь сама идея генной инженерии, то есть возможности переносить гены из одного организма в другой, предполагает универсальность кода. Оказалось, что гены, перенесенные в кишечную палочку из самых разных организмов, прекрасно в ней работают, то есть синтезируют те же белки, что и в исходном, родном организме. Генные инженеры без кода, как без рук. Вспомним, как были сделаны наиболее впечатляющие работы - синтез в кишечной палочке гормонов высших организмов, включая человеческий инсулин (об этих работах «Химия и жизнь» рассказывала в № 7 за 1978 г. и в № 1 за 1979 г.). По известной аминокислотной последовательности гормона с помощью кода «сочиняли» последовательность нуклеотидов в гене этого гормона. Заметим, что из-за вырожденности кода можно «сочинить» несколько вариантов гена, отвечающих одной и той же аминокислотной последовательности. Затем ген синтезировали химическим путем и встраивали его в кишечную палочку.

Хотя эра расшифровки нуклеиновых последовательностей началась совсем недавно, она уже принесла поразительные открытия. Были опровергнуты, казалось бы, самые твердо установленные представления о гене как о непрерывном участке ДНК, отвечающем одному белку. К всеобщему изумлению оказалось, что на одном участке ДНК может быть записана информация сразу о двух и даже о трех белках (см. «Химию и жизнь», 1977, № 4). И уж совсем трудно передать словами чувства, испытанные генетиками, когда оказалось, что у высших организмов гены расчленены, то есть как бы нарезаны на куски, а между кусками расположены последовательности, которые вообще непонятно зачем нужны («Химия и жизнь», 1978, № 11). Итак, под натиском первых же фактов, ознаменовавших эру расшифровки нуклеиновых последовательностей, само понятие гена потеряло универсальность, а код выстоял. Его универсальность казалась окончательно установленной.

Но не тут-то было. В 1979 г. выяснилось, что у митохондрий код другой.

КОД МИТОХОНДРИЙ

Что это такое, митохондрии? Это не бактерии и не вирусы, не одноклеточные, это просто тельца, плавающие в цитоплазме животных клеток. Просто, да не совсем. Вообще-то митохондрии выполняют очень важную для клетки функцию - в них идет процесс окислительного фосфорилирования, то есть происходит переработка энергии, образующейся при «сгорании» пищи в энергию АТФ. Иными словами, митохондрии - это энергетическая станция клетки. Подобно тому как электричество - универсальный источник энергии у нас в быту, так и АТФ - универсальный источник энергии для клеточных ферментов. Забирая энергию у АТФ, фермент отщепляет у него одну фосфатную группу, делая из него АДФ. В митохондриях происходит «подзарядка» - к АДФ вновь присоединяется фосфатная группа. О том, как идет процесс окислительного фосфорилирования, недавно рассказал на страницах «Химии и жизни» (1979, № 10 и 11) В. П. Скулачев. Но для нашего рассказа все это не имеет никакого значения. Для нас важно другое: митохондрии имеют свою собственную ДНК.

Как, ДНК в цитоплазме? Ведь всем известно, что у животных ДНК находится в хромосомах, а хромосомы расположены в ядре. Да, это верно. Подавляющая часть ДНК находится в ядре. Но митохондрии имеют свою собственную, митохондриальную ДНК. Более того, митохондрии имеют свою собственную РНК-полимеразу, которая снимает мРНК-овую копию с митохондриальной ДНК. Но и это не все. В митохондриях есть свои рибосомы, свои тРНК, короче, свой собственный аппарат белкового синтеза. Это уж совсем странно - ведь в той же цитоплазме навалом нормальных клеточных рибосом. Но на этих рибосомах синтезируется белок только с копий ядерной ДНК. Митохондрии ими пользоваться почему-то не желают. У митохондрии все - малого размера. Мини-рибосомы, мини-РНК-полимераза, мини-ДНК. И вроде бы это понятно - ведь митохондрия, разумеется, гораздо меньше клетки. Но умение самостоятельно строить белок вовсе не означает, что митохондрия - это автономная часть клетки, не зависящая от ядерной ДНК. ДНК митохондрии столь мала по размеру, что на ней никак не может уместиться вся информация о молекулах белков, рРНК и тРНК, необходимая для автономного существования митохондрии. Большая часть этой информации находится в ядре клетки, то есть записана в виде последовательности нуклеотидов в ядерной ДНК. И вот ко всем странностям митохондрий добавилась еще одна, самая удивительная, - у митохондрий свой собственный генетический код.

Обнаружилось все это, по-видимому, случайно. Б. Беррел и его сотрудники из Лаборатории молекулярной биологии в Кембридже (Англия) занимались расшифровкой последовательности митохондриальной ДНК человека. Кстати, это тот самый Беррел, который обнаружил впервые, что гены могут налезать друг на друга. Биологи сравнили последовательность гена, кодирующего одну из субъединиц цитохромоксидазы, с белковой последовательностью, правда, не человеческой, а бычьей цитохромоксидазы.

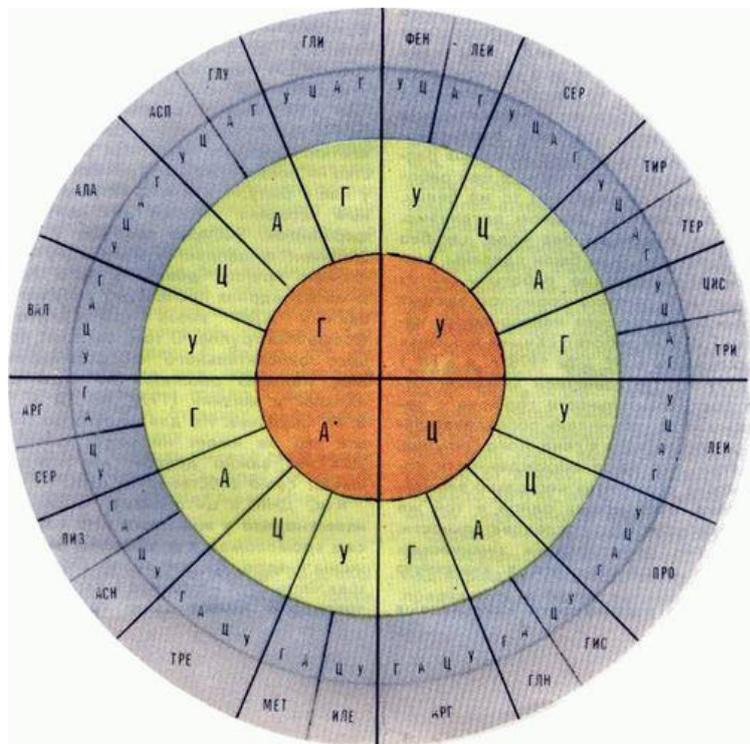


Рис. 4. Код митохондрий

Последнее обстоятельство не помешало совершенно точно определить код митохондрий человека. Он изображен на рис. 4. Видно, что он очень похож на код, уже известный ранее. Исключение только в двух случаях: кодон УГА отвечает триптофану, а АУА - метионину.

Постойте, но ведь именно так должен выглядеть «идеальный» код! Тот самый код, который предсказывал Ичас еще в 1969 г.

ЧТО ЖЕ ДАЛЬШЕ!

Как же оценивать открытие Беррела? Безусловно, возможны разные точки зрения. Можно сказать, что собственно ничего особенного и не произошло. Ведь называл же Ичас предсказываемые им изменения «незначительными». Правда, за прошедшие десять лет отношение к коду значительно изменилось. Десять лет - очень большой срок для молекулярной биологии, которой нет еще и тридцати лет. И за эти годы привыкли считать код абсолютно универсальным, данным как бы от бога. Кто осудит теперь тех, кто метнется в другую крайность и начнет поиск новых кодов? Ведь шутка ли сказать, оказалось, что в одной клетке, причем в нашей собственной, человеческой клетке, существуют два разных кода! Год назад у научного работника, заявившего, что он решил проверить, универсален ли код, допустим, для лососевых рыб, не заподозрили бы никаких других намерений, кроме желания вдоволь полакомиться деликатесами. А теперь? Теперь

расшифровка кода у любого вида, особенно если он достаточно далек от тех, для которых код уже твердо установлен, не представляется уже совершенно бессмысленным делом.

Нет, открытие нового кода не следует недооценивать. Ведь получено первое доказательство того, что код эволюционировал, что он не сразу возник таким, каким мы его видим теперь. Неоднократно высказывалась точка зрения, что митохондрия - это остатки одноклеточного организма, очень давно образовавшего симбиоз с клеткой. То, что у митохондрии даже код другой, служит еще одним очень веским доводом в пользу такого предположения. Более того, «идеальность» кода митохондрий, то есть его большая приближенность к квазидублетной структуре, может свидетельствовать о том, что это более древний код.

Тогда возникает гипотеза, что когда-то у всех клеток был такой же код, как у нынешних митохондрий, а затем в коде произошли небольшие изменения. И может быть, далеко не все живое на Земле произошло от клеток с уже изменившимся кодом? Может быть, часть видов - это прямые потомки древних клеток, имевших митохондриальный, «идеальный» код? А может быть, есть виды, которые эволюционировали от клеток, получившихся после каких-то других, пусть небольших, изменений «идеального» кода?

Словом, неуниверсальность кода может оказаться исключением, которое подтверждает общее правило, а может предвещать и более важные открытия.