

BIOCHEMISTRY OF  
NUCLEIC ACIDS

D. G. KNORRE

*The article deals with the present state of knowledge of proteins and nucleic acids, the two most important classes of biopolymers. The main functions of nucleic acids are described with emphasis on their role as carriers of genetic information. The biotechnological applications are presented, as well as some new possibilities of applications which are going to appear in this field of science.*

**В статье изложены основы принципов строения двух важнейших классов биополимеров – белков и нуклеиновых кислот, описано функционирование нуклеиновых кислот как носителей наследственной информации в живых организмах. Изложены основные методические достижения в химии, биохимии и биотехнологии нуклеиновых кислот и дан обзор некоторых новых нарождающихся областей применения этих знаний.**

© Кнорре Д.Г., 1996

БИОХИМИЯ НУКЛЕИНОВЫХ  
КИСЛОТ

Д. Г. КНОРРЕ

Новосибирский государственный университет

## ВВЕДЕНИЕ

Нуклеиновые кислоты как один из компонентов живой материи были открыты в 1869 году швейцарским ученым Иоганом Мишером. Однако бурное развитие химии и биохимии нуклеиновых кислот началось в конце 40-х – начале 50-х годов XX века, когда было установлено, что один из двух главных типов нуклеиновых кислот – дезоксирибонуклеиновая кислота (ДНК) является носителем наследственной информации. На протяжении второй половины XX века нуклеиновые кислоты были важнейшим объектом исследований химии и биологии. Химики создали методы установления детальной химической структуры нуклеиновых кислот, их искусственного синтеза, изучили их поведение при разных химических воздействиях. Биохимики направили свои усилия на выяснение многочисленных аспектов функционирования нуклеиновых кислот в живых организмах или выделенных из них системах. Поскольку было выяснено, что строение молекул нуклеиновых кислот специально приспособлено для выполнения некоторых основополагающих биологических функций, область биохимии, изучающая поведение нуклеиновых кислот в живых и модельных системах, обособилась в автономную область знания, получившую название молекулярная биология. Многие вопросы биохимии нуклеиновых кислот стали азбукой естествознания, вошли не только в вузовские, но и в школьные учебники. Между тем изучение нуклеиновых кислот продолжает оставаться одной из самых горячих точек на переднем крае современной науки. Продолжается оттачивание инструментов для этих исследований. Одними из наиболее прецизионных инструментов становятся разнообразные химические методы. В связи с этим автор решил попытаться в трех статьях изложить современное положение в области изучения нуклеиновых кислот в живых организмах и модельных системах, проблемы, являющиеся предметом наибольшего внимания исследователей, а также основные направления совершенствования химических методов исследования и открывающиеся в этой связи новые перспективы. Для цельности изложения и облегчения восприятия материала в первой статье в сжатой форме излагаются и положения, ставшие азбукой молекулярной биологии. Однако в первую очередь автор стремился на этом фоне дать почувствовать читателям, где находится и куда движется эта

важнейшая область человеческого знания. Данная статья должна также стать основной для всех желающих ознакомиться с более специальной литературой, в связи с чем с соответствующими пояснениями в нее введены некоторые специальные термины, укоренившиеся в этой области знания. В первой статье автор постарался избегать химических формул, которые, во-первых, можно найти в любом отечественном курсе биохимии, а во-вторых, более органично связаны с материалом следующих двух статей.

## **ДНК – ПЕРВИЧНЫЙ НОСИТЕЛЬ НАСЛЕДУЕМОЙ ИНФОРМАЦИИ**

Среди бесчисленного разнообразия химических веществ, из которых построены живые организмы, особое положение занимают два типа биологических полимеров – белки и нуклеиновые кислоты. Как и любые другие полимеры, они построены из большого числа небольших органических молекул, мономеров, в случае белков – из аминокислот, в случае нуклеиновых кислот – из нуклеотидов. В образовании нуклеиновых кислот могут участвовать две группы нуклеотидов – рибонуклеотиды и дезоксирибонуклеотиды. Первые образуют рибонуклеиновые кислоты (РНК), вторые – ДНК. Однако, в отличие от всех других полимеров, как природных, так и синтетических, созданных химиками, они построены из нескольких разнотипных мономеров – четырех в случае нуклеиновых кислот и 20 в случае белков. Важнейшей группой белков являются ферменты (по английской терминологии энзимы), которые катализируют в живых организмах разнообразные химические реакции. Функции белков чрезвычайно многообразны. При этом число остатков каждого мономера в полимерной цепи и порядок их расположения (последовательность мономеров) имеют глубокий биологический смысл. В этой статье речь будет идти в основном о биологическом смысле последовательностей нуклеотидов.

В отдельных случаях замена одной мономерной единицы в огромной последовательности приводит к серьезным биологическим последствиям. Известно, например, что повышенная чувствительность к алкоголю, характерная для многих представителей восточных народов, связана с заменой одной аминокислоты (лизина на глутамат) из 487 расположенных последовательно аминокислот в ферменте альдегиддегидрогеназе, ответственном за удаление из организма уксусного альдегида, который накапливается при окислении этилового спирта. В то же время в некоторых случаях большое число замен не лишает биополимер его главной функции.

Для каждого живого организма характерен свой набор белков с определенными последовательностями аминокислот и соответственно свой набор нуклеиновых кислот с определенными последовательностями нуклеотидов. Этот набор может быть достаточно большим. По грубым оценкам в челове-

ческом организме содержатся многие десятки тысяч разных белков. Это, однако, ничтожная часть от практически неисчерпаемого мыслимого числа белков. Ведь из 20 аминокислот можно построить  $20 \times 20 = 400$  разных димеров (то есть попарно соединенных аминокислот),  $20 \times 20 \times 20 = 20^3 = 8000$  тримеров, а сравнительно коротких белков длиной всего в 100 аминокислотных остатков –  $20^{100} = 10^{130}$ , что гораздо больше, чем число атомных ядер во всей доступной наблюдению части Вселенной (последнее оценивается как  $10^{80}$ ). Определенные, характерные для данного организма белки не могут возникать случайно. При образовании в живом организме (биосинтезе) белков должна существовать некоторая управляющая система, которая содержит информацию о том, какие именно последовательности аминокислот нужно собирать в данном организме. Первичным материальным носителем такой информации является ДНК.

Главным свойством живых организмов является размножение, то есть воспроизведение себе подобных. У одноклеточных организмов (бактерий, дрожжей, инфузорий) размножение происходит путем деления клеток. Чтобы из одной родительской клетки образовалось две одинаковых дочерних, в основном идентичных родительской, делению должно предшествовать удвоение всех основных компонентов клетки, в том числе ее белков и нуклеиновых кислот. И в первую очередь должна удвоиться ДНК, чтобы в обеих дочерних клетках оказались программы для образования всех свойственных родительской клетке белков и нуклеиновых кислот. Способность ДНК к самоудвоению обеспечивается ее строением. Она построена из четырех дезоксирибонуклеотидов, которые по первым буквам их химических названий будут в дальнейшем обозначаться латинскими буквами dA, dG, dC и dT (префикс “d” введен для того, чтобы отличить их от рибонуклеотидов, из которых построены молекулы другой группы нуклеиновых кислот – рибонуклеиновых, сокращенно РНК). Замечательным свойством этих нуклеотидов является то, что в составе полимерной цепи ДНК dA обладает способностью избирательно (селективно) связываться (образовывать комплекс) с dT, а dG – с dC. Любая последовательность нуклеотидов имеет определенное направление, а, следовательно, вся полимерная цепь имеет два различных конца. В соответствии с деталями химического строения нуклеотидов один из концов обозначают как 5'-конец, а противоположный – как 3'-конец. В дальнейшем направление цепи либо будет указываться, либо просто будет подразумеваться, что слева находится 5'-конец. В соответствии со сказанным могут существовать протяженные последовательности, которые могут расположиться так, что против каждого нуклеотида одной окажется селективно взаимодействующий с ней другой нуклеотид. Такие последовательности называются **комплементарными**. При этом для такого взаимодействия

направления цепей должны быть противоположными (антипараллельными). Примерами двух комплементарных последовательностей могут служить фрагменты ДНК (5')dAdTdGdGdCdTdA(3') и (3')dTdAdCdCdGdAdT(5').

Принцип комплементарности был сформулирован и обоснован в 1953 году американским ученым Джеймсом Уотсоном и английским ученым Френсисом Криком. Важное значение для формулирования этого принципа имели появившиеся незадолго до этого работы американского ученого Эрвина Чаргаффа, который показал, что ДНК из разных биологических источников содержит равное количество dT и dA и равное количество dC и dG, в то время как соотношение количеств этих пар бывает очень различным для разных ДНК. Сам принцип явился одним из основополагающих звеньев в установлении пространственной структуры ДНК, которая оказалась построенной из двух комплементарных цепей. Такую структуру принято называть **дуплексом**. Принцип комплементарности обеспечивает дублирование информации о последовательности нуклеотидов одной цепи в другой комплементарной цепи, поскольку между последовательностями нуклеотидов в комплементарных цепях существует взаимно однозначное соответствие. Поэтому если две цепи исходного (родительского) дуплекса разошлись, то каждая из них способна управлять построением из мономеров комплементарной (дочерней) цепи, что в итоге приведет к воссозданию двух дуплексов, идентичных исходному (см. схему 1).

Вертикальными черточками обозначены так называемые водородные связи, скрепляющие комплементарные пары нуклеотидов, причем между dT и dA существует две, а между dC и dG – три водородных связи.

Утверждение о способности ДНК к самоудвоению требует некоторого уточнения в связи с понятием об информации. Для аналогии представим себе детальнейший проект нового здания, совершенно достаточный для строителей, то есть содержащий всю информацию о будущем здании. Однако из сколь угодно объемистого проекта здания не получится. Нужны строительные материалы и специалисты – строители, которые могут воспользоваться этой информацией. Нужны, наконец, различные механизмы, с помощью которых материалы будут доставляться строителям. Совершенно аналогично обстоит дело и с воспроизведением новых молекул ДНК. Последовательность нуклеотидов в каждой из цепей – это только чертеж для создания новых молекул ДНК. Сколько ни выдерживайте молекулу ДНК – никакие новые копии сами по себе не получатся. Нужен, во-первых, строительный материал – достаточный запас молекул мономеров, подготовленных для сборки новых полимерных цепей. И нужно специальное устройство, которое будет осуществлять шаг за шагом присоединение мономеров к растущей новой полимерной цепи. Поскольку присоединение каждой следующей молекулы мономера представляет собой химическую реакцию, то и для образования новых цепей ДНК необходимо участие большого числа специальных белков (ферментов и других белков), которые образуют сложную структуру, называемую полимеризационным комплексом. Первоначально из этого комплекса был выделен фермент, названный **ДНК-полимеразой**, но затем, во-первых, нашли, что в любых клетках оперируют несколько типов ДНК-полимераз, а во-вторых, обнаружили, что вместе с полимеразой в полимеризационный (или репликационный) комплекс входит несколько десятков других белков, часть из которых будет описана

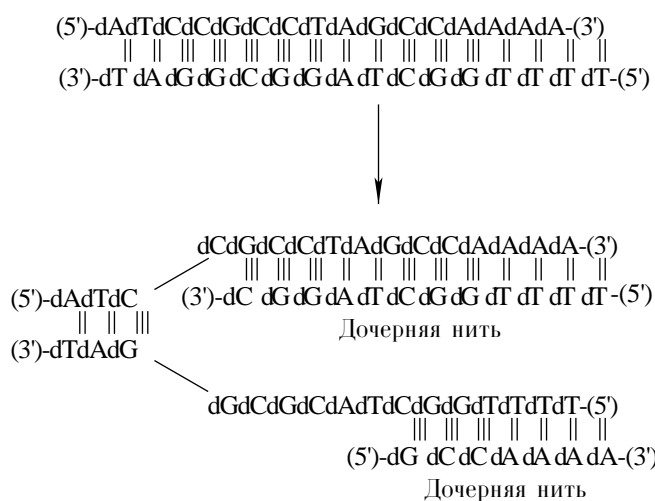


Схема 1.

ниже. ДНК-полимеразы относятся к категории наиболее сложных, так называемых **матричных ферментов**. Они не только способны катализировать реакцию роста новой цепи ДНК, но на каждом шаге выбирают из четырех мономеров тот, который комплементарен звену управляющей ДНК, как бы стоящему в очереди на подачу команды на синтез. Процесс синтеза комплементарной дочерней цепи ДНК на одной из родительских цепей называют **репликацией**. Термин “матричный” остался по наследству от первых наивных представлений, когда сборку новой копии ДНК представляли как печатание на бумаге текстов, набранных на типографских матрицах. На самом деле ДНК, управляющая синтезом новой, комплементарной ей цепи, протягивается через ДНК-полимеразный комплекс, и системе скорее следует уподобить магнитофону, который протягивает через себя магнитофонную ленту, превращая информацию, записанную на ленте, в целенаправленные действия, например звуковые сигналы. Резюмируя сказанное, еще раз подчеркнем, что, говоря о любой информации, в том числе биологической, мы подразумеваем, что имеется некоторый материальный носитель, на котором эта информация записана, безотносительно к тому, с помощью каких устройств и для каких целей эта информация может быть использована.

Заканчивая этот раздел, следует сказать несколько слов о формах организации ДНК в живых организмах и дать представление о ее размерах. Все живые организмы разделяют на две группы – прокариоты и эукариоты. У прокариот ДНК в основном представлена одной двунитевой структурой, как правило кольцевой (таким образом, для каждой из двух комплементарных цепей нельзя указать начало или конец), которая практически не обособлена от остальной части клетки. К прокариотам относятся разнообразные бактерии. Эукариотами являются многие более сложно организованные одноклеточные организмы, такие, как дрожжи и инфузории, и все без исключения многоклеточные, вплоть до человека. У эукариот в клетке имеется четко оформленное клеточное ядро, в котором сосредоточена подавляющая часть ДНК. При этом ДНК распределена по нескольким структурам, число которых зависит от природы живого организма и которые называют хромосомами. Большинство клеток содержит двойной набор хромосом. Одинарный набор характерен для половых клеток – сперматозоидов и яйцеклеток, и эти наборы сливаются в двойной набор при оплодотворении. После этого все развитие организма идет с сохранением двойного набора. Исключением является лишь последний этап созревания половых клеток, когда двойной набор распределяется пополам по двум созревшим сперматозоидам или яйцеклеткам. По-видимому, в пределах каждой хромосомы ДНК представлена одним непрерывным дуплексом гигантского раз-

мера – каждая из нитей может содержать сотни миллионов нуклеотидов.

Клетки человека содержат 23 пары хромосом, причем самые крупные хромосомы содержат ДНК, построенную из более чем 200 миллионов пар нуклеотидов. Огромные успехи, достигнутые в создании высокоэффективных методов определения последовательностей нуклеотидов,<sup>1</sup> вдохновили ученых на работу по полному секвенированию всего набора ДНК человека (генома человека). Программа “Геном человека” является сегодня одним из основных приоритетов естествознания. Задача по масштабу грандиозная, так как речь идет об установлении последовательности размером три миллиарда пар нуклеотидов для одинарного набора. Чтобы записать такую последовательность на обычных листах бумаги, вмещающих 2500 знаков, нужно более миллиона страниц, или более тысячи томов по 1000 страниц в каждом томе.

#### **ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ДНК. ТРАНСКРИПЦИЯ И ТРАНСЛЯЦИЯ**

Очень скоро после установления принципа строения ДНК было сформулировано представление о генетическом коде, то есть о том, как на молекуле ДНК записаны аминокислотные последовательности программируемых ею белков. Из очевидных арифметических соображений американский физик русского происхождения Г. Гамов пришел к выводу, что минимум три нуклеотида ДНК должны кодировать одну аминокислоту. Двух еще не хватало – из четырех нуклеотидов можно построить всего  $4 \times 4 = 16$  упорядоченных комбинаций, этого недостаточно для 20 аминокислот. Тринуклеотидов можно построить  $4^3 = 64$ , что создает даже определенную избыточность. За короткий срок весь генетический код был расшифрован. Приводить его в этой статье нецелесообразно, но важно сказать, что сборкой белков из аминокислот ДНК непосредственно не управляет. Для этого существуют посредники в виде молекул РНК, которые синтезируются при непосредственном участии ДНК.

РНК состоит из четырех рибонуклеотидов, которые аналогично тому, как это было сделано для ДНК, можно обозначить буквами А, G, C и U. По своей химической природе они очень близки к нуклеотидам, составляющим ДНК, причем они сохраняют способность к избирательному взаимодействию с соответствующими партнерами, например G с C. Нуклеотид U по своим свойствам в этом смысле сходен с dT. Синтез новых молекул РНК осуществляется с помощью специального матричного фермента – **РНК-полимеразы** (или более точно: с помощью комплекса белков, в которых важнейшую роль играет РНК-полимераза). Присоединение каждого

<sup>1</sup> Эта задача называется **секвенированием** (от английского слова sequence – последовательность).

следующего мономера к растущей полимерной цепи РНК осуществляется с участием очередного нуклеотида в программирующей ДНК, который обеспечивает отбор комплементарного рибонуклеотида. В том участке РНК-полимеразы, который непосредственно участвует в происходящем химическом превращении, — ее активном центре — одна из нитей ДНК отстраняется от участия в отборе, проис-

ходит временное разделение нитей ДНК и образуется короткий дуплекс, состоящий из кусочка программирующей ДНК и кусочка вновь синтезированной РНК. Затем, однако, по мере удаления от активного центра РНК отделяется от ДНК, а использованный участок ДНК воссоединяется с временно устранным комплементарным участком (схема 2):

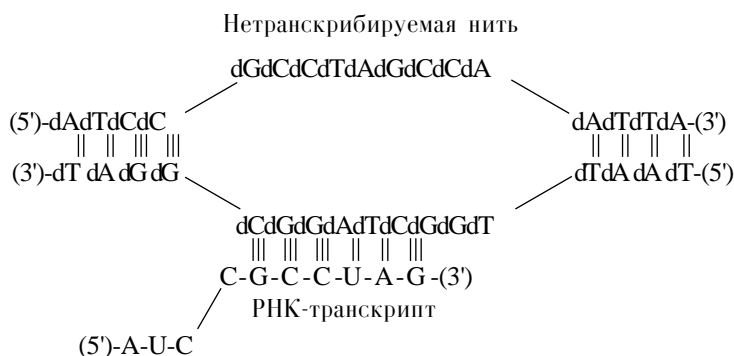


Схема 2.

Нужный кусок ДНК как бы переписывается в виде РНК, в связи с чем процесс получил название **транскрипции** (от англ. transcribe — переписывать). Если пренебречь небольшими химическими различиями между U и dT, C и dC и т.д., то можно сказать, что полученная при транскрипции цепь совпадает с соответствующим участком нити ДНК, не участвовавшим непосредственно в транскрипции (эту нить иногда называют нетранскрибируемой).

Итак, участок ДНК, содержащий информацию о структуре определенного белка, переписывается в виде РНК, имеющей ту же последовательность, что и нетранскрибируемая нить дуплекса, и в таком виде посылается к месту, в котором происходит формирование новых молекул белка. Эта РНК называется обычно мессенджер РНК (от англ. “messenger” — посыльный), или сокращенно мРНК. Новые молекулы белка собираются на довольно сложно устроенных частицах, называемых **рибосомами**. Они состоят из нескольких молекул РНК и нескольких десятков разных белков, которые соответственно называют рибосомными РНК (сокращенно рРНК) и рибосомными белками. Интересно и очень важно для работы рибосом, что их составляющие сгруппированы в две легко отделяемые друг от друга частицы (**субъединицы**) разного размера. Их соответственно называют большой и малой субъединицами. Что делает в рибосоме огромное количество разных белков, еще далеко не полностью установлено, и их чаще всего просто обозначают номерами, снабженными буквой S (small), если они находятся в малой, и буквой L (large), если они принадлежат большой

субъединице. Каждая из субъединиц содержит по одной большой молекуле РНК, которые характеризуются цифрами, отражающими в определенных единицах их молекулярную массу. Например, для рибосом из кишечной палочки — любимого объекта молекулярных биологов (знаменитая *Escherichia coli*) малая и большая субъединицы содержат соответственно 16S и 23S рРНК ( в этом случае буква S обозначает так называемую единицу Сведберга, с помощью которой характеризуют, как быстро частица перемещается в центробежном поле, создаваемом центрифугой)<sup>1</sup>. Роль этих РНК совершенно отлична от роли мРНК и, так же как для рибосомных белков, до конца еще не установлена. Существование, однако, что ДНК должна содержать программы и для синтеза тех РНК, которые нужны сами по себе, а не только как промежуточные посыльные для программирования синтеза белков.

Из давно устоявшихся положений то же следует сказать и об еще одной группе РНК, так называемых транспортных (тРНК). На заре становления молекулярной биологии было установлено, что в биосинтезе белка на рибосомах участвуют не сами аминокислоты, а продукты их присоединения к сравнительно небольшим специальным молекулам РНК. При этом каждой из двадцати аминокислот соответствует своя тРНК, а иногда и несколько

<sup>1</sup> Первую ультрацентрифугу, способную вращать ротор с огромным числом оборотов в минуту, построил шведский физико-химик Теодор Сведберг (1884 — 1971), в честь которого единицу осаждения (седиментации) назвали *сведбергом*.

разных тРНК. Следовательно, в ДНК должны быть запрограммированы и все последовательности, соответствующие необходимому для данного живого организма набору транспортных РНК. Молекула мРНК на каждой стадии удлинения создаваемой на рибосоме белковой цепи непосредственно участвует не в отборе самой аминокислоты, а в отборе той тРНК, к которой аминокислота успела присоединиться перед поступлением на рибосому. Следует подчеркнуть, что термин “транспортная РНК” не совсем удачно передает функциональную роль этого класса РНК, сводя ее к транспорту аминокислот к рибосомам. То же относится и к английскому эквиваленту этого термина transfer RNA, что переводится как “переносящая тРНК”. На самом деле основные две функции каждой тРНК заключаются в их способности присоединять определенную аминокислоту (акцепторная функция) и приспособливать (адаптировать) эту аминокислоту к соответствующему кодону (адапторная функция).

Теперь можно коротко резюмировать, что происходит при биосинтезе белка. Рибосома представляет собой довольно просто устроенную молекулярную машину, которая способна создавать новые молекулы белка. При этом не имеет существенного значения, какие именно молекулы белка предстоит создавать: все рибосомы устроены однотипно. Для создания белков необходимо сырье – аминокислоты, предварительно прикрепленные к определенным тРНК. Но этого для эффективной работы рибосом недостаточно: для этого нужна программа в виде мРНК. Каждая тройка нуклеотидов (ее часто называют **кодоном**) последовательно связывается с соответствующей данному кодону тРНК, к концу которой прикреплена кодируемая аминокислота. Например, тринуклеотид UUU в составе мРНК отберет тРНК, на которой предварительно посажена аминокислота фенилаланин. То же самое произойдет и в случае кодона UUC. Изменение одного нуклеотида в кодоне может изменить аминокислоту в белке. Например, описанная выше замена в ферменте альдегиддегидрогеназе аминокислоты лизина на глутамин происходит в результате замены в соответствующей мРНК кодона AAA, кодирующего лизин, на кодон GAA, кодирующий глутамат. Существенное изменение функции фермента происходит всего-навсего в результате замены одного А на G. В отборе кодона также участвуют комплементарные взаимодействия. В составе той тРНК, которая присоединяет именно фенилаланин, имеется участок, часто называемый **антикодоном**, с последовательностью (3')AAG(5'), строго комплементарный кодону UUC и частично комплементарный UUU. Именно антикодон опознается и отбирается участком кодона, вошедшим в рибосому. Взаимодействие кодона мРНК и антикодона тРНК является ключевым информационным процессом, при котором происходит перевод информации, запи-

санной в виде троек нуклеотидов, в соответствующую последовательность аминокислот в синтезируемом белке. Поэтому синтез белка на рибосоме получил название **трансляции** (от англ. translation – перевод). Вопрос, почему же на тРНК с таким антикодом оказался именно фенилаланин, относится уже к совершенно другой проблеме, рассмотрение которой выходит за рамки этой статьи. Можно лишь вкратце упомянуть, что фенилаланин, как и любая другая аминокислота, участвующая в синтезе белка, присоединяется к соответствующей тРНК вне рибосом с помощью специальных ферментов, которые называют аминокислота: тРНК-лигазами (от англ. ligate – соединять). Наряду с этим названием используют название аминоацил-тРНК-синтетазы, поскольку фактически к тРНК присоединяется не вся молекула аминокислоты, а ее часть, в которой вместо свойственной всем карбоновым кислотам карбоксильной группы COOH остается лишь фрагмент C=O. Такие фрагменты карбоновых кислот называют ацилами, а аминокислот, соответственно, аминокислотами. Аминоацил-тРНК-синтетазы существуют для каждой аминокислоты. В частности, для фенилаланина таким ферментом является фенилаланин-тРНК-лигаза или фенилаланил-тРНК-синтетаза. Эти ферменты обладают уникальной избирательностью: из природных аминокислот они способны иметь дело только с фенилаланином, а из транспортных РНК – только с теми, которые существуют для переноса фенилаланина. В связи с этим они являются одним из главных объектов исследований проблемы специфичности нематричных ферментов.

Репликация, транскрипция и трансляция – три основополагающих процесса, на которых зиждется любая жизнедеятельность. Информация о структуре белков, свойственных каждому живому организму, и структуре РНК, программирующих синтез этих белков, а также тех РНК, которые принимают участие в расшифровке генетического кода, может рассматриваться как главная в структуре ДНК. Однако информационное содержание ДНК этим не исчерпывается. Некоторые другие аспекты информации, запрограммированной в ДНК, будут рассмотрены в следующей статье.

## РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

*Кнорре Д.Г., Мызина С.Д.* Биологическая химия. М.: Высшая школа, 1992.

\* \* \*

Дмитрий Георгиевич Кнорре, доктор химических наук, профессор, действительный член РАН, директор Новосибирского Института биоорганической химии Сибирского отделения РАН, лауреат Ленинской премии, автор более 250 публикаций, двух монографий и трех учебников.