

## BIOCHEMISTRY OF NUCLEIC ACIDS

D. G. KNORRE

*The paper deals with some fine aspects of the informational content of DNA, namely the regulator regions of DNA and the existence of the intervening sequences (introns). The latter are transcribed by RNA polymerases but are further excised in the course of the RNA maturation by the process termed splicing. The main achievements in methods of investigations and treatment of nucleic acids are described.*

**Описаны некоторые аспекты информационного содержания ДНК – регуляторные районы и наличие вставочных последовательностей (интронов). Последние переписываются в ходе транскрипции, но вырезаются из образовавшейся РНК с помощью процесса, называемого сплайсингом, при окончательном созревании мРНК. На этом фоне излагаются основные методические и биотехнологические достижения химии и биохимии нуклеиновых кислот.**

© Кнорре Д.Г., 1998

## БИОХИМИЯ НУКЛЕИНОВЫХ КИСЛОТ

Д. Г. КНОРРЕ

Новосибирский государственный университет

### ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ДНК. РЕГУЛЯТОРНЫЕ РАЙОНЫ И ИНТРОНЫ

В предыдущей статье<sup>1</sup> было рассказано лишь о самом главном информационном содержании ДНК: ДНК содержит программы для синтеза свойственных данному организму белков, которые представляют собой последовательности троек нуклеотидов на транскрибируемой нити. Непосредственно они программируют синтез соответствующего кодона мРНК. Например, будущий остаток фенилаланина должен быть записан на транскрибируемой нити ДНК как (3')dAdAdA(5'), если программируется кодон UUU, или как (3')dAdAdG(5'), если кодируется кодон UUC. Участок ДНК, содержащий всю информацию о программируемом белке, называют геном для данного белка. Генами называют и участки ДНК, которые содержат всю необходимую информацию о структуре РНК (рибосомные и транспортные РНК), а также вспомогательные участки, необходимые для транскрипции гена. В действительности последовательностью кодирующих три-нуклеотидов для синтеза белков или последовательностью нуклеотидов, кодирующих структуру тРНК или рРНК, содержащаяся в ДНК информация не исчерпывается. Сегодня уже ясно, что оно неизмеримо богаче, а, по-видимому, многое еще и вообще неизвестно.

Для создания определенной РНК необходимо, чтобы транскрипция началась в определенной точке огромной молекулы ДНК. Здесь должна оказаться и расположиться нужным для транскрипции образом РНК-полимераза. Для этого существует специальная область, чаще всего находящаяся перед кодирующей последовательностью, роль которой состоит в том, чтобы связать и ориентировать определенным образом РНК-полимеразу. Эту область, если она расположена вдоль цепи близко от начала старта транскрипции, называют промотором.

Новое неожиданное оказалось связанным со структурой генов эукариот. Выяснилось, что смысловая последовательность, кодирующая определенную последовательность аминокислот в белке или определенную последовательность нуклеотидов в РНК, необязательно является непрерывной. Она может содержать некоторые вставочные последовательности, которые получили название интронов.

<sup>1</sup> Кнорре Д.Г. Биохимия нуклеиновых кислот // Соросовский Образовательный Журнал. 1996. № 3. С. 11–16.

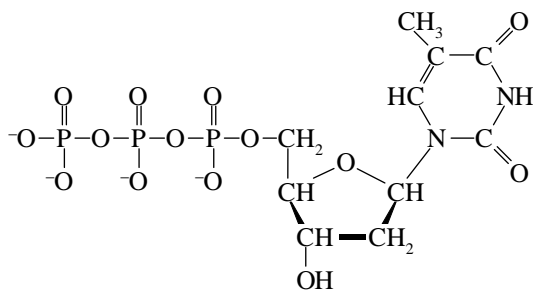
Кодирующие последовательности, разделенные интронами, в этом случае называют экзонами. При транскрипции получаются РНК-копии, которые содержат и нужные в конечной (зрелой) молекуле РНК фрагменты, и лишние, которые должны быть удалены (вырезаны) из зрелой молекулы. Каждое такое вырезание должно сопровождаться воссоединением концов последовательных экзонов, между которыми находился интрон. Этот процесс получил название «сплайсинг». Сплайсинг происходит с избирательностью, характерной для процессов, катализируемых ферментами. До открытия сплайсинга считалось общепринятым, что все ферменты имеют белковую природу. Однако никакие белки в некоторых случаях в сплайсинге не участвуют, то есть РНК, содержащая интроны, может сама осуществлять селективную реакцию, причем с довольно большой скоростью. В результате пришлось признать, что биологическими катализаторами могут служить не только белки, но и рибонуклеиновые кислоты. По аналогии с ферментами такие катализаторы получили название рибозимов.

Говоря об информационном содержании ДНК, нельзя не сказать о том, что информационное содержание ДНК может изменяться как в результате ошибок при репликации, так и при действии различных внешних факторов — облучения и обработки некоторыми химическими реагентами. Изменения в структуре ДНК какого-либо организма носят название мутаций. Мутация может заключаться в том, что одна пара нуклеотидов в двух цепях превращается в другую комплементарную пару. Это, естественно, происходит не одновременно: сначала возникает одна ошибка, например dC заменяется на dT. При репликации dT отберет для введения в новую цепь dA, а не dG, которое находилось в этом месте цепи у родительской ДНК. В итоге вместо пары dC · dG в дочерней молекуле, а тем самым и во всех последующих поколениях на этом месте будет находиться пара dT · dA. Такие мутации получили название точечных. Происходят и более значительные по масштабу мутации. Часто, хотя далеко не всегда, мутации имеют серьезные биологические последствия. Известны, например, гены, которые отвечают за контроль над размножением клеток, стимулируя его до определенного предела и останавливая в нужной для организма фазе. Некоторые единичные изменения в таком гене приводят к тому, что способность к контролю за процессом деления клеток теряется и ген превращается в онкоген, то есть ген, способствующий неограниченному размножению клеток — клетки становятся злокачественными и возникает раковая опухоль.

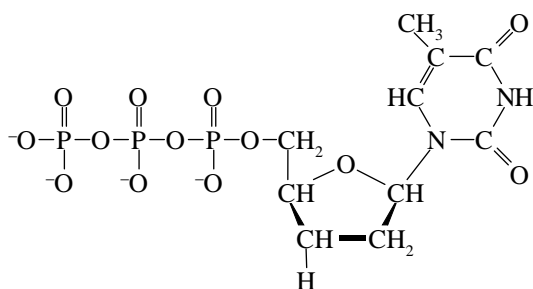
## ЧТО СОВРЕМЕННАЯ ХИМИЯ И БИОХИМИЯ УМЕЮТ ДЕЛАТЬ С НУКЛЕИНОВЫМИ КИСЛОТАМИ

Химия и биохимия нуклеиновых кислот не только углубили наши представления об огромной группе важнейших биологических процессов, связанных с сохранением, размножением и использованием наследственной информации, заложенной в генетическом материале клеток, но дали мощный импульс для становления современной биотехнологии с важными практическими выходами. Для того чтобы не вслепую манипулировать нуклеиновыми кислотами, было необходимо научиться изучать их структуру, прежде всего составляющие их последовательности нуклеотидов. Впервые ценной воистину героически усилий это было сделано для нескольких тРНК, состоящих из нескольких десятков нуклеотидов. Приятно отметить, что на первом этапе соревнования за установление первичной структуры тРНК в лидирующей пятерке была и группа советских ученых, возглавляемая А.А. Баевым. Однако созданные в ходе этих работ методы были настолько трудоемки, что для прорыва в исследование ДНК и больших молекул РНК они были непригодны. Революционные решения, принципиально отличающиеся как друг от друга, так и от первых методов, были найдены в США Алленом Максамом и Уолтером Гилбертом и в Англии Фредериком Сенгером. Эти методы, в особенности метод Сенгера, сделали возможным секвенирование нуклеиновых кислот общим размером в миллионы нуклеотидных пар.

Конечно, в рамках одной статьи рассказать в деталях об этих методах невозможно. Поэтому придется ограничиться изложением наиболее принципиальных особенностей метода, причем только метода Сенгера. Особенностью проблемы является необходимость определить последовательное расположение очень большого числа остатков нуклеотидов (за один проход — несколько сот). Зато свойства каждого остатка уже хорошо изучены. Основополагающая идея Сенгера состояла в том, чтобы исследовать последовательность не самой ДНК или, точнее, ее достаточно длинного фрагмента, а исследовать структуру новосинтезированной ДНК, полученной на исследуемой ДНК как матрице с помощью ДНК-полимеразы. Из-за комплементарности новой ДНК по ее последовательности нуклеотидов нетрудно восстановить структуру матрицы. При этом оказалось возможным частично заменить в смеси мономеров, из которой синтезируется новая цепь, один из них на так называемый дидезоксинуклеотид. На рис. 1 приведена химическая формула нуклеотида, а именно тимидинфосфата, обозначенного символом dT. Остатком фосфорной кислоты он связан в цепи с предыдущим нуклеотидом. Его ОН-группа должна присоединить следующий нуклеотид при продолжении роста цепи. Оказалось, что при репликации можно совершить небольшой обман ДНК-полимеразы — подмешать к смеси нуклеотидов дидезокси-



Тимидин-5'-трифосфат



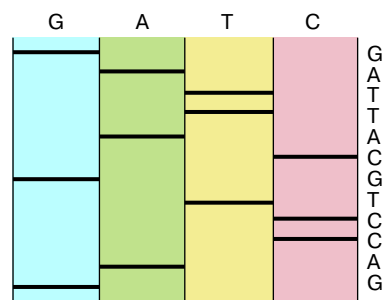
Дидезокситимидин-5'-трифосфат

**Рис. 1.** Структуры тимидин-5'-трифосфата и его дидезоксианалога

рибонуклеотид, у которого эта OH-группа отсутствует. Его можно обозначить в представленном случае символом ddT. С какой-то вероятностью ДНК-полимераза узнает и присоединит к растущей цепи очень похожий ddT вместо dT. Так как синтез идет в строгом соответствии с принципом комплементарности, то это произойдет напротив той точки матричной ДНК, где находится комплементарный dA. Но из-за отсутствия у ddT OH-группы эта цепочка не сможет расти дальше, произойдет обрыв цепи. Обрыв с определенной вероятностью может произойти в любой точке напротив dA. Таким образом, получится смесь новосинтезированных цепей разной протяженности, причем длины (числа нуклеотидных остатков) всех этих цепей точно соответствуют номерам остатков dA матрицы. Следовательно, в таком эксперименте определяются точки расположения всех остатков dA в исследуемой ДНК. Если три аналогичных эксперимента провести со смесями, содержащими примеси других дидезоксинуклеотидов: ddA, ddC и ddG, то аналогично будут расставлены по номерам все остальные три нуклеотида на исследуемой ДНК. Разделить же полученные смеси по длинам не представляет труда с помощью электрофореза – метода, основанного на перемещении заряженных молекул под действием постоянного электрического поля. Дело в том, что каждый остаток нуклеотида содержит остаток фосфорной кислоты, который несет отрицательный заряд. Поэтому, будучи помещены в электрическое поле, фрагменты разной длины будут перемещаться

к аноду с разными скоростями. Так как электрофорез обычно проводят в очень вязкой среде (геле), то эта среда оказывает сопротивление перемещению фрагментов, тем большее, чем больше размер фрагмента. Этот фактор пересиливает действие поля, поэтому, чем длиннее фрагмент, тем медленнее он движется, но все они располагаются в порядке, соответствующем их длинам. Остается только “увидеть” место каждого фрагмента. До сих пор для этой цели чаще всего используют радиоактивную метку: в нуклеотиды, из которых синтезируется новая цепь, вводят радиоактивный фосфор. Поэтому после окончания гель-электрофореза гель прикладывают к рентгеновской пленке и на месте нахождения фрагментов после проявления радиоавтографа появляются темные пятна. На рис. 2 схематично показаны такой радиоавтограф и читаемая с него последовательность маленького кусочка ДНК.

Следует отметить, что метод Сенгера, равно как и метод Максама–Гилберта, совершившие революцию в изучении структуры ДНК, оказался успешным благодаря нарушению основного канона аналитической химии, который можно сформулировать так: если нельзя напрямую определить структуру некоторой молекулы, нужно ее количественно превратить в



- а (3') X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>X<sub>4</sub>X<sub>5</sub>X<sub>6</sub>X X X A X X A X X X A A XA  
 б X̄<sub>1</sub>X̄<sub>2</sub>X̄<sub>3</sub>X̄<sub>4</sub>X̄<sub>5</sub>X̄<sub>6</sub>X̄ X̄ X̄ ddT  
 в X̄<sub>1</sub>X̄<sub>2</sub>X̄<sub>3</sub>X̄<sub>4</sub>X̄<sub>5</sub>X̄<sub>6</sub>X̄ X̄ X̄ T X̄ X̄ ddT  
 г X̄<sub>1</sub>X̄<sub>2</sub>X̄<sub>3</sub>X̄<sub>4</sub>X̄<sub>5</sub>X̄<sub>6</sub>X̄ X̄ X̄ T X̄ X̄ T X̄ X̄ X̄ ddT  
 д X̄<sub>1</sub>X̄<sub>2</sub>X̄<sub>3</sub>X̄<sub>4</sub>X̄<sub>5</sub>X̄<sub>6</sub>X̄ X̄ X̄ T X̄ X̄ T X̄ X̄ X̄ T ddT  
 е X̄<sub>1</sub>X̄<sub>2</sub>X̄<sub>3</sub>X̄<sub>4</sub>X̄<sub>5</sub>X̄<sub>6</sub>X̄ X̄ X̄ T X̄ X̄ T X̄ X̄ X̄ T T XddT

**Рис. 2.** Схема метода Сенгера и схематичное представление фрагмента радиоавтографа геля с соответствующей ему последовательностью нуклеотидов. X и X̄ – произвольные комплементарные остатки нуклеотидов. Пронумерованы индексами остатки, входившие в состав праймера. dA – остатки дезоксиаденозинового нуклеотида в составе анализируемой ДНК. dT и ddT – остатки тимидин монофосфата и дидезокситимидин монофосфата, включившиеся в новые цепи ДНК, полученные при репликации с помощью ДНК-полимеразы

другую молекулу, строение которой известно. На этом принципе построен и замечательный метод Эдмана для установления последовательности аминокислот в полипептидах (фрагментах белков). Автор нашел химическую реакцию, которая позволяет отщепить с одного, определенного конца полипептида один остаток аминокислоты, превратив его в так называемый тиогидантоин и не разрушая при этом остальную цепочку. Разным аминокислотам соответствуют разные тиогидантоины, так что легко определить, какая именно из 20 аминокислот, входящих в состав белков, была отщеплена. Укороченный на одну аминокислоту полипептид может быть повторно использован для определения природы следующего аминокислотного остатка. Однако каждый найденный аминокислотный остаток — результат отдельной процедуры, и таким путем в лучших случаях удастся проделать несколько десятков шагов. В методе Сенгера анализируется сложная смесь продуктов репликации, так что с одного геля (точнее, с четырех дорожек одного геля) сразу расставляются по своим местам сотни нуклеотидных остатков.

Успешно развиты и методы химического синтеза больших фрагментов нуклеиновых кислот — олигонуклеотидов. Как и в случае секвенирования, все началось с чрезвычайно трудоемких методов. Разработав и применив на практике эти методы, американский ученый Гобинд Корана в 1970 году завершил синтез ДНК, кодирующей последовательность одной из тРНК дрожжей, специфичной к аминокислоте аланину. Однако из процедуры, требовавшей многолетнего труда большого числа специалистов высшей экспериментальной квалификации, в рутинный метод, выполняемый на специальных автоматических синтезаторах, синтез превратился существенно позднее. Это было связано как с изменением химических процессов, положенных в основу последовательного присоединения к растущей цепи необходимых нуклеотидов, так и в огромной мере с введением в практику так называемого твердофазного синтеза, общий принцип которого был предложен американским ученым Робертом Меррифилдом для синтеза белков из аминокислот.

Как в белках, так и в нуклеиновых кислотах связь между мономерными звеньями, сколько бы их ни было, идентичны. В нуклеиновых кислотах это всегда связь между остатком фосфорной кислоты, принадлежащей одному звену мономера, и гидроксильной группой (ОН) соседнего мономера. Чтобы построить, например, фрагмент нуклеиновой кислоты длиной в 100 нуклеотидов, нужно последовательно 99 раз осуществить реакцию образования этой связи, поочередно вводя в реакцию один мономер за другим. Сегодня проблемы образования такой связи в химии, можно сказать, не существует. Вопрос в том, чтобы каждый раз проводить эту реакцию как можно более количественно и с минимальным накоплением ненужных (побочных) ве-

ществ. Химик считает, что, проведя реакцию с выходом 80%, он имеет неплохой результат. Но общий выход стостадийного процесса при выходе 80% на каждую стадию составит  $0,8^{100} = 10^{-9}$ , то есть одну миллиардную часть от введенного в реакцию первого нуклеотида. Очевидно, что такой выход не может быть основой для прогрессивной технологии. Между тем, по крайней мере частично, неколичественные выходы происходят от того, что по химическим традициям при проведении реакции в несколько стадий нужно после каждой стадии выделить полученный продукт, очистить его от не вступивших в химическую реакцию реагентов, от других примесей и только потом приступать к следующей стадии. Потери, причем вполне ощутимые, на каждой стадии неизбежны.

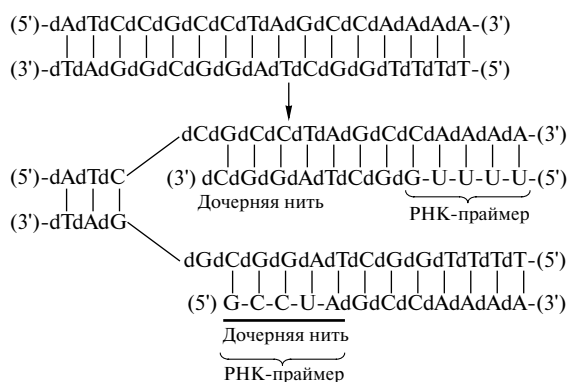
Меррифилд предложил отказаться от этой классической химической традиции. Согласно его идее, первое звено создаваемой цепи полимера прочно, до самого конца синтеза, закрепляется химической связью на твердом веществе-носителе, который помещается в специальную трубку-реактор (колонку). Через реактор пропускается второй мономер, причем в достаточном избытке, чтобы превращение закрепленного мономера было как можно более полным. Затем избыток второго мономера и все вспомогательные вещества, нужные для образования химической связи между мономерами, отмываются пропусканием через реактор подходящего растворителя. Заодно отмываются и побочные вещества, по крайней мере те, которые не связаны с носителем. После этого приступают к присоединению третьего звена. Процедуру можно повторять много раз, хотя, конечно, и не до бесконечности — какие-то потери и загрязнения неизбежны. Ясно, что такая процедура легко поддается автоматизации. Синтезатор олигонуклеотидов (иногда его называют ген-синтезатором) имеет несколько сосудов, содержащих отдельные мономеры, вспомогательные вещества, нужные для синтеза, и растворители для промежуточных промывок реактора. Все эти вещества подаются поочередно в реактор, содержащий носитель и растущую на нем полимерную цепь с помощью насоса, который выбирает каждый раз нужный сосуд по введенной в прибор программе. В программу вводится и та последовательность, которую следует синтезировать. Этим задается структура будущего продукта. Остается только по окончании синтеза разрушить химическую связь, с помощью которой первое мономерное звено было закреплено на носителе, и провести очистку синтезированного продукта от неизбежно накопившихся примесей. В умелых руках на надежно работающих ген-синтезаторах удастся получать фрагменты длиной до ста нуклеотидов и более. Конечно же, биологически интересные гены значительно длиннее. Однако уже на заре становления искусственного синтеза нуклеиновых кислот Корана разработал способ соединения достаточно длинных олигонуклеотидов между



собой с помощью существующего у всех живых организмов фермента ДНК-лигазы. Так что в настоящее время синтез генов является хотя и трудоемкой, но вполне доступной процедурой.

Замечательной способностью ДНК является возможность ее размножения. В принципе для этого нужны лишь должным образом подготовленные нуклеотиды и ДНК-полимераза. Поэтому, получив некоторое небольшое количество ДНК, ее можно размножить. В настоящее время эта процедура для небольших молекул ДНК или определенных кусков большой молекулы хорошо разработана. Она получила название амплификации, или полимеразной цепной реакции (ПЦР). Имеется лишь одна осложняющая, но без особого труда преодолимая проблема. Дело в том, что особенностью всех известных ДНК-полимераз является их неспособность начать синтез новых полимерных цепей. Они могут только удлинять уже существующую цепь, связанную с комплементарной матрицей.

В живых организмах эта трудность преодолевается довольно сложным образом. Когда приходит время для удвоения ядерной или бактериальной ДНК, вступает в действие специальный фермент, выбирающий определенные участки ДНК, с которых должна начаться репликация, и синтезирует на них короткие затравочные фрагменты РНК – праймеры. Фермент, по сути дела, является РНК-полимеразой, но выполняющей специальную функцию. Его часто называют ДНК-праймазой. После образования праймера ДНК-полимераза продолжает рост цепи уже с помощью дезоксирибонуклеотидов. Таким образом, вначале образуется ДНК, несущая на 5'-конце небольшой фрагмент РНК:



Из неспособности ДНК-полимераз начинать синтез новых цепей ДНК следует, что для осуществления ПЦР помимо ДНК-полимеразы и мономеров необходимы праймеры. Если речь идет об амплификации определенного участка ДНК, то достаточно ввести в систему праймеры, комплементарные 3'-концам выбранных участков обеих нитей. Это делают при умеренной температуре, чтобы дуплекс матрицы с праймером был достаточно прочным. После проведения синтеза дочерних копий выбран-

ного участка новые дуплексы разрушают нагреванием до достаточно высокой температуры, чтобы развести исходную и новосинтезированную нити, затем снова охлаждают до температуры, благоприятствующей образованию дуплексов с праймером и последующей репликации. Ясно, что при каждом таком цикле количество копий амплифицируемого участка удваивается, то есть за 20 циклов можно увеличить количество интересующей исследователя последовательностей в  $2^{20} = 10^6$  раз.

Среди многочисленных применений ПЦР следует в первую очередь отметить, что она нашла применение для выявления мутантных генов, что имеет огромное практическое значение для выявления наследуемых заболеваний. При проведении ПЦР при этом вводятся такие праймеры, чтобы они помогли ему простимулировать амплификацию мутантного гена, но не дали возможность амплифицировать неизмененный ген. Если выявляемая мутация присутствует, то на последней стадии амплификации на фоне разнообразных фрагментов ДНК, которые не подверглись амплификации, получится большое количество мутантного фрагмента, который легко обнаружить и определить, что это именно искомый фрагмент, поскольку он имеет определенный заранее известный размер. При наличии хороших реактивов и приборов удастся зарегистрировать изначально ничтожные количества определенной последовательности. ПЦР позволяет проводить такие тонкие анализы, как пренатальную диагностику – определение нежелательной последовательности у плода на первых неделях беременности. В зависимости от характера заболевания, провоцируемого таким мутантным геном, можно либо принять специальные меры для нейтрализации негативного эффекта, вызываемого этим геном, либо в некоторых не поддающихся профилактике случаях рекомендовать женщине своевременно избавиться от дефектного плода и тем самым предотвратить появление дефектного ребенка.

Возможность химически синтезировать, а также направленно вырезать из природных ДНК различные гены привела к рождению новой области биотехнологии – генной инженерии. Появились методы, позволившие вставлять произвольные гены в небольшие молекулы ДНК, способные внутри определенных клеток автономно размножаться, или самокопироваться синхронно с репликацией клеточной ДНК. Появилась возможность встроить ген, не свойственный данному организму, в наследственные структуры этого организма. А поскольку ген может программировать образование кодируемого им продукта – белка или РНК, – то такой организм может начать производить несвойственный ему продукт. Оказалось возможным производить в клетках бактерий нужные человеку или какому-либо важному для сельского хозяйства животному белки. Одним из самых впечатляющих примеров является организация генноинженерного производства

человеческого инсулина. Это чрезвычайно важный гормон, производимый поджелудочной железой и совершенно необходимый для усвоения сахара. Недостаток, а тем более полное отсутствие инсулина приводят к тяжелейшему и широко распространенному заболеванию – диабету. Основным средством спасения диабетиков является ежедневное введение инсулина. Большая часть больных может пользоваться бычьим инсулином, который добывают из поджелудочной железы крупного рогатого скота на мясокомбинатах. Однако существуют многие десятки тысяч диабетиков, которые не могут использовать чужеродный инсулин, несколько отличающийся по структуре от человеческого. Создание гена человеческого инсулина позволило поставить производство человеческого инсулина с помощью бактерий, то есть генноинженерными методами. В настоящее время созданы бактерии, производящие многие белки, нужные в медицине или сельском хозяйстве, такие, как гормон роста для повышения продуктивности мясного животноводства и человеческий интерферон для профилактики и лечения некоторых вирусных заболеваний.

Новым, пока еще нарождающимся направлением биотехнологии является антисмысловая технология. Многие заболевания, в первую очередь онкологические и вирусные, связаны с появлением в клетках пациента чужеродной или неблагоприятно измененной наследуемой информации. Вирусы, например, вносят в зараженные ими клетки свои нуклеиновые кислоты и заставляют клетки работать

по их программам, воспроизводить вирусную нуклеиновую кислоту и вирусные белки. Если ввести в зараженные клетки нуклеиновую кислоту или синтетический олигонуклеотид, комплементарный какой-либо жизненно важной части вирусной нуклеиновой кислоты, можно надеяться, что он образует дуплекс с этой частью и заблокирует ее способность программировать работу клетки в нужном для вируса направлении. Так как этот участок вирусной нуклеиновой кислоты имеет определенный биологический смысл, то такие олигонуклеотиды и нуклеиновые кислоты получили название антисмысловых.

## РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Кнорре Д.Г., Мызина С.Д. Биологическая химия. М.: Высш. шк., 1992.
2. Шелкунов С.Н. Генетическая инженерия. Новосибирск: Изд-во Новосиб. ун-та, 1994.
3. Шабарова З.А., Богданов А.А. Химия нуклеиновых кислот и их компонентов. М.: Химия, 1987.
4. Овчинников Ю.А. Биоорганическая химия. М.: Просвещение, 1987.

\* \* \*

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.