

УСПЕХИ ФИЗИЧЕСКИХ НАУК**МОЛЕКУЛЫ-ГИГАНТЫ****В. Л. Брэгг *)**

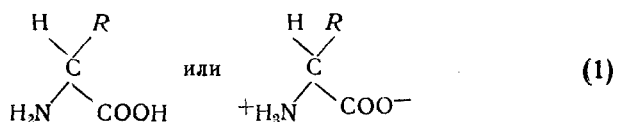
Я займусь здесь главным образом исследованиями структуры кристаллических белков рентгенометрическими методами. Это — грандиозная задача. Рентгенометрический анализ сейчас применяется ко всё более и более сложным молекулам; но в белках мы встречаемся со степенью сложности на несколько порядков выше, чем у самых сложных органических молекул, которые до сих пор удалось успешно расшифровать. Следующая таблица молекулярных весов иллюстрирует сказанное:

Нафталин	128	Глобулин табачных	
Пенициллин	302	семян	300 000
Инсулин	12 000	Гемоцианин	6 800 000
Миоглобин	17 000	Вирус томатных стеб-	
Пепсин	37 000	лей	10 000 000
Гемоглобин	66 700		

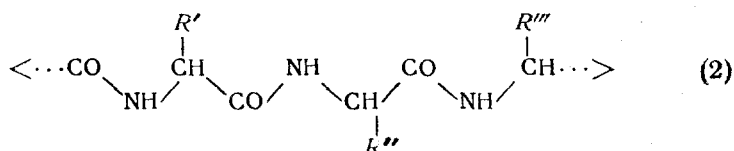
Молекула нафталина была первой органической молекулой, успешно проанализированной рентгенометрически около 25 лет тому назад. Пенициллин является одним из последних триумфов рентгеновского анализа. Инсулин, являющийся одним из простейших белков, имеет молекулярный вес, почти в 40 раз больший, чем у пенициллина, а за ним мы переходим к колоссальным молекулам вирусов и других сложных белков. Все эти вещества образуют достаточно совершенные кристаллы и дают рентгенограммы. Эти рентгенограммы показывают, что протеины построены весьма правильно вплоть до атомных расстояний. Задача рентгеновского анализа заключается в том, чтобы найти способ интерпретировать эти диаграммы. Нетрудно измерить по несколько тысяч рефлексов от каждого кристалла, и мы стоим перед задачей прочесть зашифрованное письмо без ключа.

*) Nature 164, 7—10 (1949) (ср. УФН XXII, 98—104 (1939). Перевод Е. Н. Беловой).

Белки, как это было показано Эмилем Фишером в 1906 г., построены из длинных цепочек или колец, составленных аминокислотными остатками. Типичная аминокислота имеет формулу



Здесь R изображает одновалентную боковую цепь, специфическую для каждой характерной аминокислоты. Кислая группа любой аминокислоты может быть связана с основной группой другой аминокислоты с выделением частицы воды; в результате многократного повторения такой реакции возникает цепь из остатков, в которой различные боковые цепочки R подобны различным окрашенным бусам, подвешенным к ожерелью. Формулой такой пептидной цепочки будет:



Мы сейчас знаем 23 различные аминокислоты, сложность которых возрастает от глицина, в котором R представляет собой атом водорода, и до фенилаланина, в состав которого входит бензольное кольцо, и триптофана, в который входят связанные пяти- и шестичленное кольца. Большинство аминокислот состоит исключительно из углерода, кислорода, азота и водорода. Но некоторые содержат и другие элементы, например серу. В большинстве аминокислот боковая цепь R нейтральна, например в аланине — CH_3 , в валине — $\text{CH}(\text{CH}_3)_2$, фенилаланине — $\text{CH}_2 \cdot \text{C}_6\text{H}_5$. Боковая цепь может быть основной, как, например, в лизине — $\text{CH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{NH}_2$, или же кислой, как в аспарагиновой кислоте — $\text{CH}_2 \cdot \text{COOH}$. Цистин является аминокислотой с двух концов, — $\text{CH}_2 \cdot \text{S} \cdot \text{S} \cdot \text{CH}_2$, — и образует мостики между одной цепочкой и другой. Анализ аминокислотных остатков в каком-либо белке представляет весьма длительную и трудную задачу, и всё же удалось проанализировать с большой степенью полноты многие белки. Инсулин (молекулярный вес 12 000) по Кибналлу и Зангеру составлен из 106 остатков, соединённых в четыре цепочки, между которыми перекинута шесть цистиновых дисульфидных мостиков. Миоглобин (17 000) содержит в себе 140 остатков, а гемоглобин (67 000) — около 540. Если 23 типа аминокислотных остатков уподобить буквам алфавита, то цепочка из этих остатков в миоглобине будет подобна фразе из 20—30 слов, тогда как гемоглобин придётся охарактеризо-

вать целым абзацем примерно из 10 предложений. По каким-то пока неясным причинам природа избрала этот простой способ образования молекул протоплазмы как животного, так и растительного мира. Те сложные и специфические функции, которые молекулы должны выполнять, оказываются возможными не за счёт образования каких-либо особо сложных органических молекул, но путём нанизывания этих относительно простых групп в различном порядке и в разных количествах. Так, при помощи одного и того же небольшого числа букв алфавита может быть написана и мильтоновская поэма и страница из телефонного справочника.

Некоторые характерные черты структуры белков заслуживают особого внимания:

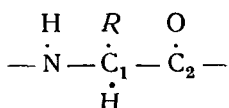
а) Почти все известные аминокислоты входят в почти все известные белки.

б) Средний молекулярный вес остатков почти одинаков во всех белках и лежит между 110 и 120. Так как молекулярный вес той части атомов, которая входит в костяк цепочки $\text{CO}\cdot\text{CH}\cdot\text{NH}$, равен 56, то, таким образом, половина молекулярного веса в белках сосредоточена в этом костяке («хребте») и другая половина — в боковых группах.

в) Почти с достоверностью можно утверждать, что все встречающиеся в природе аминокислоты характеризуются одной и той же стерической конфигурацией вокруг центрального углеродного атома, которая изображена выше (1). Так как центральный атом углерода тетраэдрически окружён четырьмя различными группами, то все аминокислоты возможны либо в виде правой формы, либо левой. Обе они являются зеркальным отражением одна другой, подобно тому как правая рука является отражением левой. Все аминокислоты принадлежат к левым формам за исключением, конечно, глицина, в котором остаток R является вторым водородным атомом, и потому для глицина не может быть различия между правой и левой формами. Повидимому, является простым делом случая, что все формы жизни связаны с одинаковой левой цепочкой. Если бы наш мир отразить в Алисином зеркале, то он, повидимому, функционировал бы с таким же успехом, и только какой-то случай определил «левизну» всех форм живой материи.

г) Кристаллическую структуру нескольких простых аминокислот, а именно дипептидов, составленных из двух связанных между собой остатков, уже удалось определить с помощью рентгеновского анализа (например глицин). Расшифрованные аминокислоты характеризуются двумя существенными особенностями. Во-первых, они достаточно строго подчиняются правилу постоянства расстояний между атомами, а также и постоянства углов между связями, которые были установлены в других органических соединениях, в частности в таких, в которых отсутствуют боль-

шие напряжения или натяжения. Хёгинс дал следующую сводку постоянных для основного составляющего звена:



(они усреднены по многим структурам):

$$\begin{array}{ll} \text{N}-\text{C}_1=1,41 \text{ \AA}; & \text{C}_1-\text{C}_2=1,52 \text{ \AA}; \\ \text{C}_2-\text{O}=1,25 \text{ \AA}; & \text{C}_2-\text{N}=1,33 \text{ \AA}; \\ \text{угол } \text{NC}_1\text{C}_2=112^\circ; & \text{угол } \text{C}_1\text{C}_2\text{N}=118^\circ; \\ & \text{угол } \text{C}_2\text{NC}_1=118^\circ. \end{array} \quad (3)$$

Во-вторых, все белковые структуры содержат многочисленные водородные связи между азотом и кислородом $\text{N}-\text{H}-\text{O}$. Повидимому, именно эта связь имеет весьма большое значение для вопроса о форме структуры; её длина 2,65 Å. Эти числовые данные весьма важны при построении возможных моделей сколько-нибудь протяжённых полипептидных цепочек.

Трудности применения рентгенометрических методов анализа к столь сложным структурам, как белки, с первого взгляда могут показаться непреодолимыми. Вообще говоря, непосредственный переход от рентгенограммы к конкретной структуре возможен только в особенно простых случаях. Одних рентгенограмм самих по себе недостаточно для определения структуры. Они могут послужить для этого лишь в соединении с другими источниками информации. В случае молекул относительно небольших размеров эта информация состоит в знании, что элементарная ячейка содержит известное ограниченное число определённых атомов и что эти атомы связаны друг с другом связями, длина и взаимная ориентация которых более или менее хорошо известны на основании прежних определений химически аналогичных веществ. Не так уж трудно испытать целый ряд подобных конфигураций, с тем, чтобы отобрать ту из них, которая наилучшим образом согласуется с рентгеновскими данными. Но при белковой молекуле, когда она, как в случае гемоглобина, содержит 8000 атомов, этот метод «проб и ошибок» превосходит человеческие возможности.

Уже в течение полутора десятка лет существует метод такой обработки непосредственных данных рентгеновского анализа, который даёт прямые и совершенно точные указания на многие характерные черты структуры. Это — способ кристаллографического патерсоновского синтеза или способ векторной диаграммы. Числа, полученные нами при оценке яркости отдельных рефлексов

рентгенограммы, служат коэффициентами ряда Фурье

$$\sum_h \sum_k \sum_l I_{hkl} \cos 2\pi (hx + ky + lz). \quad (4)$$

В этой строке I_{hkl} есть квадрат амплитуды рефлекса (hkl), а x, y, z — координаты любой точки в элементарной ячейке определяемой структуры. Если вычислить эти тройные суммы для значительного числа точек в ячейке, то мы и получим то, что называется векторной диаграммой. Пусть в точке x_1, y_1, z_1 находится атом A и в точке x_2, y_2, z_2 — атом B . Тогда на векторной диаграмме мы получим пик или сгусток плотности в точке с координатами $x_1 - x_2, y_1 - y_2, z_1 - z_2$. Легко видеть, что расстояние от начала ячейки до этой точки равно по величине и направлению линии, соединяющей в истинной структуре атомы A и B . Другими словами, векторная диаграмма не может сказать нам, где в кристалле лежат атомы A и B , но зато указывает нам, каким образом они расположены в пространстве один относительно другого. Далее, высота соответствующего пика пропорциональна произведению масс A и B . Рис. 1 иллюстрирует этот принцип. Мы не можем войти здесь в подробности построения векторной диаграммы, хотя это делается весьма просто, если исходить из принципов оптической интерференции. Наиболее важно понять физическую сущность этих векторных диаграмм, поскольку они играют чрезвычайно большую роль в современном рентгеноструктурном анализе. Итак, мы получили определённые сведения о структуре, оплачивая их дорогой ценой, ибо легко видеть, что при увеличивающемся числе атомов векторная диаграмма быстро усложняется. В самом деле, если структура содержит n атомов, то между ними может быть проведено n^2 векторов, и все они взаимно накладываются в векторной диаграмме.

При анализе органических молекул не очень больших размеров эта трудность обходится с помощью весьма простого способа. В исследуемую молекулу вводится тяжёлый атом, например брома или иода, и тогда векторы, отвечающие расстояниям между этими тяжёлыми атомами, резко выделяются на векторной диаграмме, и их всегда легко распознать, поскольку их небольшое число, а соответственные пики очень мощные. Далее легко установить, как должны располагаться в кристалле атомы иода или брома, чтобы между ними получились определённые нами векторы. Как только тяжёлые атомы фиксированы, то определить положение и лёгких атомов становится гораздо более простым делом. Мы до некоторой степени ярко окрашиваем некоторые характеристические точки в молекуле, подобно тому как микроскопист окрашивает ядра клетки. Но молекула гемоглобина включает в себе свыше 8000 атомов и, следовательно, в её векторной диаграмме должно быть более семидесяти миллионов взаимно налагающихся

ников. Такому количеству невозможно противопоставить никакой достаточно тяжёлый атом.

Тем не менее, как раз в случае белковых молекул, есть одно облегчающее обстоятельство, а именно факт существования полипептидных цепочек. Если допустить, что цепочки эти взаимно располагаются в определённом порядке, например подобно ряду параллельных палочек, то многие векторы внутри одной какой-либо цепочки будут параллельны не только между собой, но и аналогичным векторам в других параллельных цепочках. На векторной диаграмме они совместно отразятся длинным

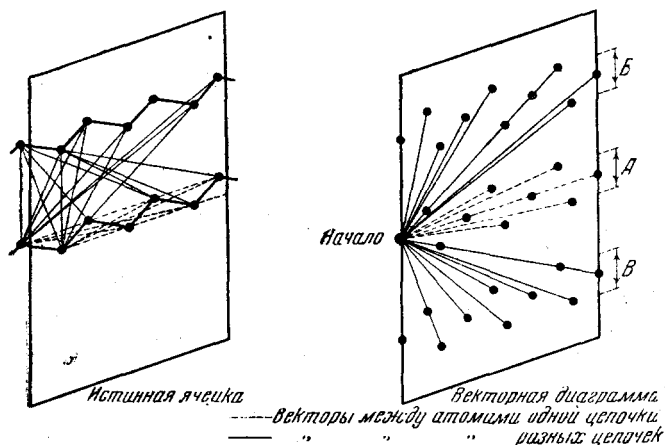


Рис. 1.

стержнем с высокой плотностью, который будет проходить через начало ячейки и будет параллелен всем указанным векторам в самом кристалле. Далее легко видеть, что концы всех векторов между атомами одной цепочки и атомами другой цепочки также образуют стержень, параллельный первому и на надлежащем от него расстоянии.

Рис. 1 иллюстрирует сказанное. Векторы между атомами внутри одних и тех же цепочек даны пунктирными линиями от двух атомов до всех их соседей, но ясно, что такие же векторы могут быть проведены и от любой другой пары атомов цепочки. Все эти векторы своими концами на векторной диаграмме уложатся в узкой области *A*. Подобные же векторы от атомов одной цепочки к атомам другой цепочки проведены тонкими линиями, и на векторной диаграмме концы их попадают в две области *B*. Поэтому если на векторной диаграмме мы увидим высокую гряду, тянущуюся от её начала в некотором определённом направлении, то будет весьма вероятным, что в истинной ячейке также имеются

цепочки из атомов, тянущиеся в том же направлении. Если же мы, кроме того, обнаружим высокие гряды, параллельные гряде, идущей от начала, то это будет веским указанием на то, каким образом параллельные цепочки взаимно уложены в кристалле.

При изучении кристаллического гемоглобина лошадиной крови Перутц обнаружил на частных векторных диаграммах подобные параллельные гряды. Это побудило его построить полный трёхмерный патерсоновский синтез для гемоглобина. Принцип этого синтеза, согласно вышесказанному, весьма прост. Для каждой точки ячейки с координатами xuz мы вычисляем тройную сумму (4) с использованием в качестве коэффициентов всех замеренных интенсивностей I_{hkl} , и получившаяся сумма даёт нам патерсоновскую плотность в точке xuz . Вопрос в трудоёмкости соответствующей работы. В случае кристаллического гемоглобина Перутц замерил около 28 000 рефлексов. Чтобы получить векторную диаграмму с достаточной точностью, необходимо вычислить тройные суммы в точках ячейки, отвечающих делению оси a на 120 частей и осей b и c на 60 частей. Общее число слагаемых во всех суммах (4) равно $28\,000 \times 120 \times 60 \times 60$, т. е. $1,21 \times 10^{10}$.

На практике анализ не столь уже страшен, поскольку существует весьма большое число вычислительных и механических усовершенствований, значительно сокращающих эту работу. И всё же промер 28 000 рефлексов и вычисление всех тройных сумм для одного кристалла гемоглобина потребовали четырёх лет. Когда эта работа была запланирована, не было никаких гарантий, что её результаты оправдают вложенные усилия, но, к счастью, эти опасения не оправдались.

Часть полученных результатов показана на рис. 2, где даны три поперечных разреза через трёхмерную векторную диаграмму. Элементарная ячейка гемоглобина моноклинная, и в изображённых разрезах ось a поднимается слева направо, ось b — перпендикулярна к плоскости диаграммы, а ось c — на чертеже вертикальна.

Левая диаграмма даёт разрез через начало координат O ; само это начало лежит слева на середине оси c . Плотности выражены с помощью контурных линий (изогипс). На разрезе мы ясно видим высокую гряду, проходящую через начало параллельно оси a . Пики на ней имеют интервалы около $5\text{ \AA}$. Та же диаграмма показывает несколько менее выраженные гряды выше и ниже только что указанной. На средней диаграмме дан разрез на высоте одной десятой оси b , и здесь мы видим ещё две высокие гряды, параллельные ранее указанным. На ещё более высоком уровне векторной диаграммы (правый разрез) видна ещё одна гряда, параллельная первой (идущая через начало). На рис. 3 дана проекция этих трёх гряд, если на них смотреть вдоль оси a так, чтобы они были видны своими торцами. Гряда, прохо-

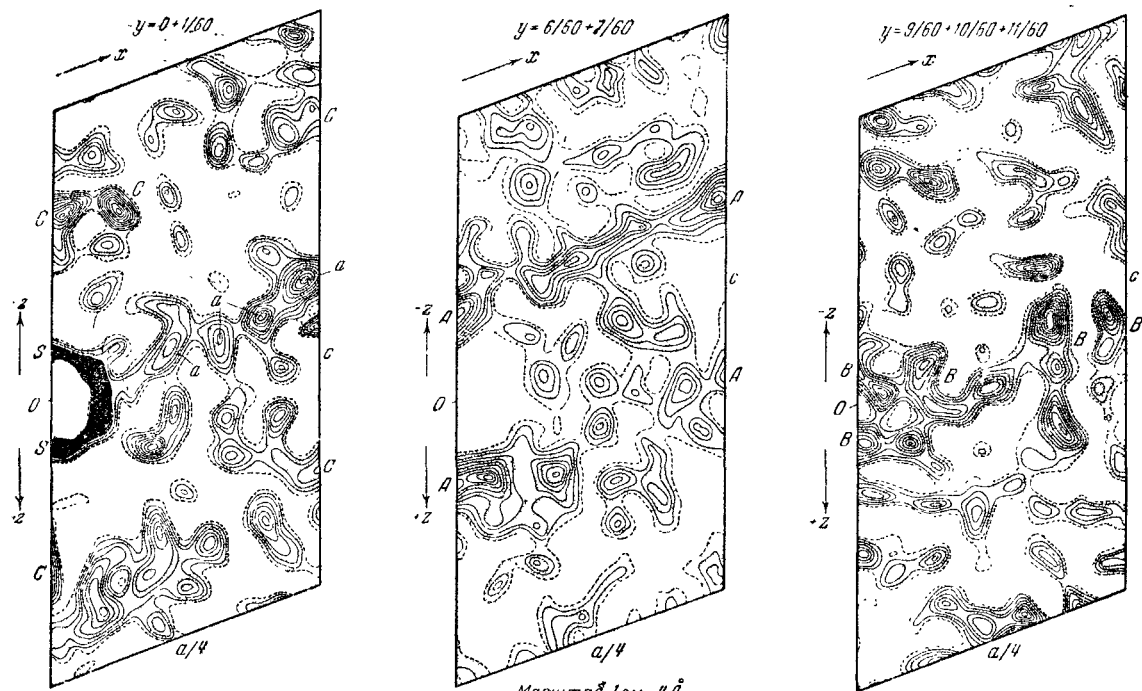


Рис. 2.

дящая через начало координат, здесь является центром фигуры, и мы можем видеть, каким образом размещены в пространстве вокруг неё все другие гряды.

Таким образом, результаты Перутца указывают на то, что в гемоглобине имеются полипептидные цепочки, тянущиеся параллельно оси a кристалла, что они находятся на взаимном расстоянии $10,5 \text{ \AA}$ и уложены примерно по закону плотной упаковки

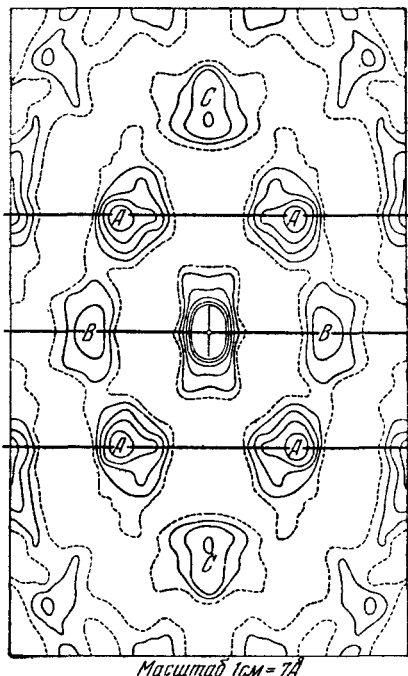


Рис. 3.

круглых стержней. Далее, ряд пиков в каждой цепочке со взаимным расстоянием около $5 \text{ \AA}$ указывает, что этот период является характерным для внутреннего строения цепочек, поскольку именно этим периодом порождается множество векторов, обеспечивающих мощность соответствующих пиков векторной диаграммы. Отсюда выводится дальнейшее важное следствие. Если цепочки характеризуются таким взаимным расстоянием, а каждая цепочка таким внутренним периодом повторяемости, то удельный вес кристалла гемоглобина укажет нам, какова масса, связанная с каждым повторяющимся мотивом. Результат подсчёта показывает, что эта масса весьма близка к средней массе трёх аминокислотных остатков.

Далее, если растянуть до пределов возможности внутри цепочки три остатка так, чтобы в точности повторилась схема (2), то это потребует $10\text{\AA}$. Следовательно, наши полипептидные цепочки должны быть как-то либо сложены, либо закручены. Точно так же связь $\text{N}—\text{H}—\text{O}$, имеющая длину $2,65\text{\AA}$, не может служить связью между различными цепочками, а уже было указано, насколько важны именно эти связи. Очевидно, эти связи располагаются между атомами одних и тех же цепочек и, нужно думать, именно эти связи удерживают цепочку в состоянии складок или закручивают её.

Эти конкретные указания векторной диаграммы подтверждают некоторые заключения, которые значительно ранее были сделаны Астбёри на основании тех весьма диффузных колец, которые получаются на рентгенограммах шерсти и других форм кератина, являющегося «денатурированным» волокнистым белком. Астбёри пришёл к выводу, что α -кератин является много раз сложенной полипептидной цепочкой с внутренним периодом повторяемости трёх остатков около $5\text{\AA}$.

Можно поставить законный вопрос — насколько обоснованы наши предположения, что именно векторы внутри цепочек будут выражены особенно сильно, тогда как существует ещё очень большое число векторов между атомами в цепочках и атомами в боковых группах, а также между атомами одной боковой группы и атомами другой. Не следует ли ожидать, что векторная диаграмма представляет собою вовсе не упорядоченное множество нулей?

Ответом будет, что, во-первых, как выше показано, костяк белка, т. е. его цепочки, составляет ровно половину всех атомов белка. Во-вторых, если цепочки действительно все имеют приблизительно форму стержня и если эти стержни друг другу параллельны, то обязательна значительная концентрация атомов вокруг центральных осей, если смотреть с торца. В качестве примера мы рассмотрим наиболее простую модель цепочки с тремя остатками и периодом повторяемости $5\text{\AA}$ и допустим, что эта цепочка построена по закону спирали с тремя группами R на каждый оборот. Вышеприведённые длины связей, а также величины углов замыкают такую спираль в цилиндр с радиусом $1,5\text{\AA}$. В среднем боковая цепочка содержит около 4 атомов, не считая водорода; типичными являются цепочки лейцина, для которого $R = \text{CH}_2 \cdot \text{CH}(\text{CH}_3)_2$. Рис. 4 изображает в весьма идеализированном виде такую структуру, если на неё смотреть вдоль цепочек. По причинам, изложенным ниже, взаимное расстояние цепочек принято $9,5\text{\AA}$ по горизонтали и $14\text{\AA}$ по вертикали. Кругами очерчены те атомы, которые принадлежат костякам цепочек в количестве 12 на каждый период повторяемости, так как на каждый оборот спирали приходится 3 лейциновые группы с четырьмя атомами в каждой

(на чертеже не показано взаимное расположение этих атомов, поскольку задачей диаграммы является продемонстрировать относительную концентрацию атомов). Внешние очертания лейциновой группы соответствуют вандерваальсовским радиусам, характерным для взаимной упаковки органических групп, не связанных друг с другом непосредственной химической связью. Плотность внутри

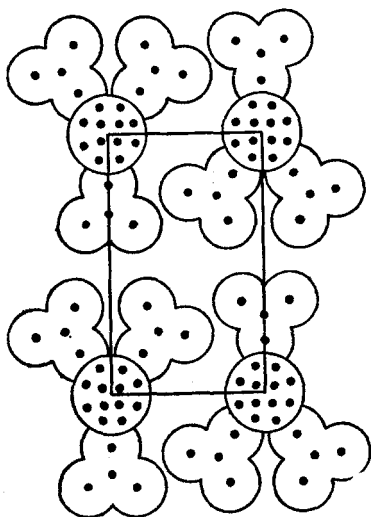


Рис. 4.

кружков, в которых все атомы связаны друг с другом ковалентными связями, приблизительно в 10 раз большая, чем в пространствах между ними. Естественно предположить, что эти массы атомов с высокой их концентрацией будут определять собой максимумы (пики) на патерсоновских проекциях, причём это будет относиться как к коротким векторам внутри каждой цепочки, так и к тем векторам, которые связывают атомы одной цепочки с атомами другой.

Взаимное расположение цепочек, изображённое нами на рис. 4, принято на основе недавней расшифровки структуры миоглобина, выполненной Кендрю. Эта структура значительно более проста, чем

структура гемоглобина, поскольку молекула миоглобина имеет молекулярный вес в четыре раза меньший. Патерсоновские проекции вдоль главных осей элементарной ячейки обнаруживают цепочки, весьма близкие к тем, которые установлены в гемоглобине, и с тем же малым периодом повторяемости внутри цепочек, равным $5 \text{ \AA}$. Эти цепочки лежат в плоскостях, параллельных грани b , и имеют направление $[201]$. Патерсоновская проекция на плоскость, перпендикулярную к этим цепочкам, получилась необычайно простой и выразительной. Она показывает, что расположение цепочек весьма близко к тому, которое дано на рис. 4, причём горизонтальные расстояния равны $9,5 \text{ \AA}$; цепочки расположены пакетами на расстояниях $14 \text{ \AA}$ вдоль вертикали b . Та же проекция показывает, что цепочки располагаются в направлении b не одна над другой, как это показано на рис. 4, но слегка отклоняются. Тот факт, что боковые группы, очевидно, не заполняют собой всего пространства на рис. 4, несомненно, обусловлен необходимостью дать место частицам воды, связанным с каждой молекулой. За подробностями мы отсылаем читателя к оригинальной работе, имеющей появиться в свет в ближайшем будущем, но мы должны

отметить здесь же, что выполненный анализ миоглобина заставляет нас отнестись с большим доверием ко всем вышеприведённым заключениям и, в частности, даёт нам первое приближение к картине распределения цепочек в этом белке.

Все выполненные до сих пор работы по рентгеноструктурному анализу белков всё ещё содержат слишком много умозрительного элемента. Мы должны быть всегда готовы без колебания отказаться от любой предположенной модели, какой бы привлекательной она нам ни казалась, если она будет опровергнута экспериментально. Мы во многом подобны альпинисту, который стремится забраться на трудно достигаемую вершину и который с радостью использует любые расщелины и выступы, на которых он может укрепиться. Достигнутые результаты всё ещё весьма мизерны, но вдали сверкает очень ценный приз. Повидимому, существуют весьма фундаментальные причины, по которым природа выбрала полипептидные цепочки в качестве носителя всех форм живой материи. А эту причину мы сможем выяснить только в том случае, если будут разрешены секреты строения белковых молекул.
