

ЧТО НУЖНО ЗНАТЬ ПРИ ИНТЕРПРЕТАЦИИ РЕНТГЕНОСТРУКТУРНЫХ РАСШИФРОВОК БИОЛОГИЧЕСКИХ МАКРОМОЛЕКУЛ

А.В.Алексеевский



"Подпоручик с ненавистью посмотрел на беззаботное лицо бравого солдата Швейка и зло спросил:

- Вы меня знаете? - Знаю, господин лейтенант.

Подпоручик Дуб вытаращил глаза и затопал ногами.

- А я вам говорю, что вы меня еще не знаете!

Швейк невозмутимо-спокойно, как бы рапортуя, еще раз повторил:

- Я вас знаю, господин лейтенант. Вы, осмелюсь доложить, из нашего маршевого батальона.

- Вы меня не знаете,- снова закричал подпоручик Дуб. Может быть, вы знали меня с хорошей стороны, но теперь узнаете меня и с плохой стороны. Я не такой добрый, как вам кажется. Я любого доведу до слез. Так знаете теперь, с кем имеете дело, или нет?

- Знаю, господин лейтенант.

- В последний раз вам повторяю, вы меня не знаете! Осел!"

Я.Гашек, "Похождения бравого солдата Швейка"

Благодарю моих студентов, воображая их ясные очи сочинялся данный опус.

СОДЕРЖАНИЕ

1. Введение	2
2. Как отличить кристаллическую конфигурацию атомов от некристаллической?	3
3. Немного о движениях	3
3.1. Как представлять себе движение плоскости	3
3.2. Примеры движений плоскости	4
3.3. Что такое ориентация плоскости и пространства	5
3.4. Полный список движений	6
4. Алгебраическая запись движения плоскости	7
5. Определение кристаллической конфигурации	9
6. Кристаллографические симметрии и кристаллографические группы	10
7. Примитивная кристаллическая ячейка	11
8. Асимметрическая единица	12
9. Некристаллографические симметрии	13
10. Биологическая единица	14
11. Банки биологических единиц	15
12. Восстановление соседних асимметрических единиц кристалла и биологических единиц с помощью SwissPDBViewer'a	16
13. Заключение	17

1. ВЕДЕНИЕ

Расшифровки пространственных структур белков, нуклеиновых кислот, сложных макромолекулярных комплексов дают уникальную информацию о белках, их строении, механизмах реакций, роли отдельных аминокислотных остатков и т.п. и т.д.

Эту информацию нужно уметь извлекать из пространственных структур, хранящихся в записях банка PDB.

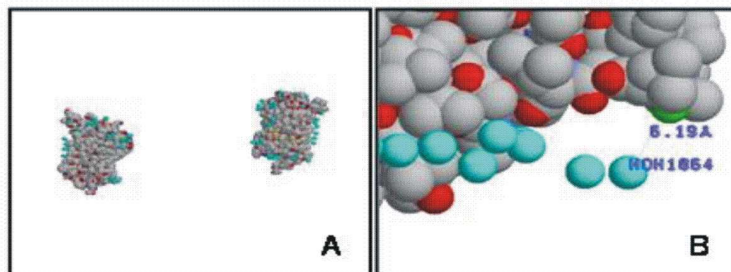
Посмотрим на структуру из PDB-записи 1RC2. Эта структура решена с помощью рентгеноструктурного анализа (РСА) кристалла, выращенного из белка аквапорин Z 1.

Возникают следующие вопросы.

1) Почему в структуре две копии одного и того же белка, удаленные друг от друга на 80 ангстрем??? (Рис. 1a) Это что-то значит или нет?

2) Посмотрим на молекулу воды HON1064 (см. рис.1b). Почему эта молекула, удаленная от белка на 6 ангстрем, и значит, не взаимодействующая с ним, присутствует в структуре? Ведь вокруг белка в структуре много пустого места (нет никаких атомов). Значит, другие молекулы воды, в материальном кристалле окружавшие белок, в расшифрованной структуре отсутствуют. Получается, что HON1064 - какая-то специальная молекула воды??

Рис. 1. Структура аквапорина Z (PDB код 1RC2). А: Общий вид. В: Молекула воды НОН1064 удалена от белка более чем на 6Å. Натуральная модель. Молекулы воды - голубые шарики. Ближайший к НОН1064 атом белка - зеленоватый



Для ответа на эти вопросы необходимо понимать, что такое кристалл, и уметь восстанавливать всю трехмерную информацию, приведенную в записи PDB.

2. КАК ОТЛИЧИТЬ КРИСТАЛЛИЧЕСКУЮ КОНФИГУРАЦИЮ АТОМОВ ОТ НЕКРИСТАЛЛИЧЕСКОЙ?

Нас интересуют "кристаллические конфигурации наборы атомов (объединенные в молекулы, даже макромолекулы) в пространстве, которые следует называть кристаллическими. Кристаллы содержат так много отдельных молекул, что вполне можно считать, что они бесконечны в пространстве во все стороны; так мы и будем считать. Реальные кристаллы белков содержат десятки тысяч белковых молекул по каждой координате, т.е. более 10^{12} молекул. Для сравнения, представьте размеры листа клетчатой бумаги со стороной 10000 клеточек - чем не бесконечный лист!

Вот пример кристаллической структуры на плоскости (Рис.2а). Чтобы восстанавливать части кристалла белка в компьютере по записи PDB, необходимо знать определение кристалла и основные понятия кристаллографии. Математическое описание кристаллов основано на понятии "движение пространства".

3. НЕМНОГО О ДВИЖЕНИЯХ

Нас интересуют движения пространства.

Тем не менее, все основные понятия можно объяснить на примере движений плоскости - так проще.

3.1. Как представлять себе движение плоскости. Движение плоскости надо понимать так. Представим себе, что на плоскости как-то расположены (нарисованы) атомы (см. рис.2а). Возьмем прозрачку (бесконечную во все

стороны - это математика!), положим на плоскость, скопируем конфигурацию атомов, а потом прозрачку передвинем. Тогда для каждого атома можно указать куда он переместился, его образ при движении. Более того, для каждой точки плоскости можно указать ее образ в результате движения. Полученное в итоге соответствие между исходным положением точек плоскости и их новыми положениями - образами, - называется движением. Таким образом, движение в математике - это результат, а не процесс. Важно также понимать, что образы точек принадлежат той же самой, первоначальной плоскости; можно вообразить, что их перевели с прозрачки на плоскость.

¹. Образ любой фигуры при движении - это равная ей фигура, но иначе расположенная.

Важно понимать, что данное движение (обозначим его буквой M) применяется к любой точке плоскости, образ точки A при движении M обозначается так: $M(A)$.

3.2. Примеры движений плоскости.

- Сдвиг, он же параллельный перенос, он же трансляция, на данный вектор \vec{V} (см. рис. 2с).

Вектор - это направленный отрезок на плоскости, который можно приложить к любой точке. Трансляция на вектор \vec{V} ставит в соответствие произвольной точке A конец вектора \vec{V} , приложенного к точке A .

- Вращение вокруг данной точки B на данный угол ϕ в данном направлении (по или против часовой стрелки) (Рис. 3а).
- Зеркальная симметрия относительно данной прямой l (Рис. 3б).
- Скользящая симметрия относительно данной прямой l со сдвигом на вектор \vec{V} , параллельный l (Рис. 3с)

Зеркальная симметрия меняет ориентацию плоскости. Это значит, что если вы нарисуете окружность и на ней стрелочкой покажете направление против часовой стрелки, а затем примените зеркальную симметрию, то на образе окружности стрелочка укажет направление по часовой стрелке. То же верно и в отношении скользящей симметрии.

Представим себе движение как перемещение прозрачки. Тогда зеркальную симметрию придется осуществить так. Снять прозрачку с плоскости, перевернуть ее и положить на плоскость. Вращение же и трансляцию можно реализовать двигая прозрачку по плоскости без отрыва.

Движения плоскости делятся на те, которые сохраняют ее ориентацию, и те, которые меняют ориентацию.

¹Определения. Преобразованием плоскости Π называется взаимнооднозначное отображение M плоскости Π на себя. Это значит, что каждая точка X плоскости имеет образ $M(X)$; при преобразовании; разные точки имеют разные образы; каждая точка Y является образом какой-то точки. Преобразование M называется *движением*, если оно сохраняет расстояние между точками. Это значит, что для любой пары точек X_1, X_2 расстояние между их образами $M(X_1), M(X_2)$ равно расстоянию между ними самими.

Определение. Движения, которые сохраняют ориентацию плоскости (пространства), называются собственными.

Только собственные движения являются физически осуществимыми. Невозможно стать таким, как видишь себя в зеркале: левая рука справа, а правая - слева.

Поэтому в кристаллографии и физике обычно рассматривают только собственные движения. Мы тоже ограничимся собственными движениями.

3.3. Что такое ориентация плоскости и пространства.

Определение. Ориентацией в данной точке плоскости называется выбор направления обхода вокруг нее.

Таким образом, в данной точке возможны две различные ориентации. Иногда их называют "по часовой стрелке" и "против часовой стрелки", хотя в этих названиях есть условность: нужно условиться с какой стороны смотреть! Ведь если часы прозрачные, глядя на стрелки с обратной стороны, обнаружим, что они движутся против часовой стрелки :).

Двигаясь по непрерывной кривой из точки A в точку B мы можем "перенести" ориентацию, заданную в точке A . Это легко себе представить так: идем по дорожке, двигая за собой вращающееся в плоскости колесо.

Определение. Ориентацией плоскости называется согласованный выбор ориентаций в каждой точке плоскости. Ориентации в точках A и B называются согласованным, если для любой непрерывной кривой, соединяющей A с B , ориентация в B совпадает с перенесенной из A вдоль кривой.

Перенос ориентации (направления обхода) вдоль кривой - интуитивно понятная вещь. Чтобы понять, зачем такая формалистика, выполните

Упражнение 3.1. Можно ли определить ориентацию поверхности цилиндра? листа Мебиуса - полоски бумаги, концы которой склеены не в цилиндр, а "наоборот"?

У плоскости, очевидно, есть две различные ориентации, которые опять же условно, называют "против часовой стрелки" и "по часовой стрелке".

Ориентация в точке пространства определяется выбором направления вращения волчка, острие которого расположено в этой точке (Рис.4). Ось волчка может принимать разные положения, но как бы мы не направляли ось, при возврате ее в первоначальное положение направление вращения останется тем же. Таким образом, есть два разных направления вращения волчка: (1) против часовой стрелки, если смотреть со стороны ручки; (2) по часовой стрелке.

Ориентацией пространства называется согласованный выбор ориентаций во всех точках. Согласованность ориентации в точках A и B , как и в случае плоскости, - это согласованность при перенесении вдоль любой непрерывной кривой, соединяющей эти точки. В физическом трехмерном пространстве² можно определить ориентацию, и разных ориентаций - две.

3.4. Полный список движений. Ограничимся собственными движениями - движениями, не меняющими ориентацию плоскости (пространства).

Примером движения пространства, меняющим ориентацию, служит зеркальная симметрия относительно плоскости.

У многих молекул, например, у белков, наблюдается хиральность, т.е. молекула и зеркально симметричная ей молекула отличаются ("правые" и "левые" аминокислоты и т.п.). Таким образом, ограничение собственными движениями физически обосновано.

Теорема 3.1 (Не слишком сложная, но и не очевидная). *Каждое собственное движение плоскости является либо трансляцией, либо вращением.*

В чем нетривиальность этого утверждения: вы можете взять прозрачку, сдвинуть ее, потом поворачивать, еще сдвинуть и т.п., т.е. поочередно применить много разных движений. И тем не менее, результирующее движение - от начального положения прозрачки до конечного, - всегда может быть получено либо как вращение вокруг какой-то одной точки, либо как трансляция на какой-то вектор.

Теорема 3.2 (Не слишком сложная, но и не очевидная). *Каждое собственное движение пространства является либо трансляцией, либо вращением вокруг оси, либо винтовым вращением вокруг оси.*

Винтовое вращение - это вращение вокруг оси l плюс сдвиг на вектор \vec{V} , параллельный l .

Для полноты картины сформулирую теоремы для любых, в том числе, несобственных движений.

Теорема 3.3. *Каждое движение плоскости является либо трансляцией, либо вращением, либо скользящей симметрией*

Зеркальная симметрия является частным случаем скользящей симметрии, в котором вектор сдвига \vec{V} - нулевой.

Теорема 3.4. *Каждое движение пространства является либо трансляцией, либо вращением вокруг оси, либо винтовым вращением, либо зеркальной симметрией, либо зеркальной симметрией со сдвигом на вектор, параллельный зеркалу, либо зеркальной симметрией с вращением вокруг оси, перпендикулярной зеркалу*

²по крайней мере, в той его части, которую удастся наблюдать, - кто знает, не переменяются ли правая и левая рука у человека, пропутешествовавшего по сверхдалекому пути через ненаблюдаемую часть Вселенной!

Зеркальная симметрия пространства определяется плоскостью - "зеркалом", образом точки служит зеркально симметричная ей точка.

4. АЛГЕБРАИЧЕСКАЯ ЗАПИСЬ ДВИЖЕНИЯ ПЛОСКОСТИ

Определение (А.Алексеевский, 2005 (если кто раньше не придумал)). *Алгебра* - это способ перевести математику на язык, понятный компьютерам.

На Рис.5 показано как закодировано движение в PDB-файле. Нетрудно видеть, что движение пространства закодировано 12ю числами. Эти числа позволяют программе (например, SwissPDBviewer'у) из лежащего в PDB-файле белка получить с помощью движения соседнюю в кристалле молекулу белка.

Опять начнем с движений плоскости.

Чтобы иметь возможность записывать движение алгебраическими средствами выберем декартову систему координат на плоскости. Тогда каждая точка A приобретает уникальное имя, состоящее из двух чисел; эти числа называются координатами точки. Ясно, что каждой точке A соответствует вектор \bar{A} с началом координат и концом в этой точке и, наоборот, каждому вектору, приложенному к началу координат, отвечает точка плоскости.

Трансляцию на вектор \bar{V} можно записать так: $\bar{A} \rightarrow \bar{A} + \bar{V}$.

Если быть педантом, то нужно было бы сказать так: возьмем вектор \bar{A} , соответствующий точке A , прибавим к нему вектор \bar{V} (векторы уж точно можно складывать!) и тогда точка - конец вектора $\bar{A} + \bar{V}$, - и будет образом A при трансляции плоскости на вектор V . Конечно, про это все забывают и говорят о точках как о векторах.

Вращение с центром в начале координат можно записать в виде матрицы $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Каждая матрица R определяет отображение плоскости в себя, при котором точка A с координатами (x, y) переходит в точку

$$R(A) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

В правой части равенства написано произведение матрицы и вектора. По определению произведения матриц (вектор $\begin{pmatrix} x \\ y \end{pmatrix}$ - это тоже матрица размера 1×2), имеем:

$$R(A) = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

Пример. $R = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

Упражнение 4.1. Какое движение плоскости задает матрица R ?

Не любая матрица R определяет вращение плоскости вокруг начала координат. Например, если $R = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, то, как легко убедиться, образом любой точки будет точка - начало координат. Это даже не преобразование, а сплошное недоразумение: сжатие плоскости в одну точку (превращение всей Вселенной в черную дыру :)).

Примеры преобразований, заданных матрицами в выбранной декартовой системе координат.

(1) $R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ Тавтологическое (единичное) движение: каждая точка остается на месте

(2) $R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ Несобственное движение - отражение относительно оси абсцисс

(3) $R = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ Преобразование, не являющееся движением: растяжение вдвое вдоль оси абсцисс

(4) $R = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$ Вращение на 45 градусов вокруг начала координат.

(5) $R = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$ Вращение на 30 градусов вокруг начала координат.

Чтобы охарактеризовать матрицы $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, которые определяют движение (а не какое-либо другое преобразование плоскости), представим R как набор из двух векторов-столбцов: $\bar{V}_1 = \begin{pmatrix} a \\ c \end{pmatrix}$, $\bar{V}_2 = \begin{pmatrix} b \\ d \end{pmatrix}$.

Теорема 4.1 (простая). *Матрица R определяет движение плоскости тогда, и только тогда, когда*

(i) *длины вектора \bar{V}_1 и \bar{V}_2 равны единице, что равносильно равенствам: $(\bar{V}_1, \bar{V}_1) = 1$, $(\bar{V}_2, \bar{V}_2) = 1$, где скобки - это скалярные произведения*

(ii) *векторы $(\bar{V}_1$ и $\bar{V}_2)$ ортогональны, или, что то же, $(\bar{V}_1, \bar{V}_2) = 0$.*

Определитель матрицы R , задающей движение, равен +1 или -1; если определитель равен +1, то движение сохраняет ориентацию плоскости, если -1, то меняет ориентацию на противоположную

Доказательство теоремы проистекает из того простого наблюдения, что вектор \bar{V}_1 является образом вектора $(1, 0)$, а вектор \bar{V}_2 - образом вектора $(0, 1)$.

В случае трехмерного пространства все аналогично: R - матрица 3×3 , ее столбцы обозначим $\bar{V}_1, \bar{V}_2, \bar{V}_3$; R задает движение если, и только если, выполнены условия: (i) $(\bar{V}_i, \bar{V}_i) = 1$ ($i = 1, 2, 3$); (ii) $(\bar{V}_i, \bar{V}_j) = 0$ ($i \neq j$).³

³Ясно как написать условия и для многомерного пространства

Матрицы, удовлетворяющие условиям (i) и (ii), называются ортогональными; таким образом, ортогональные матрицы определяют движение, а не ортогональные - определяют преобразование (если определитель не равен нулю!) не являющееся движением - например, растяжение вдоль одной оси и сжатие - вдоль другой.

Теорема 4.2 (простая). *Каждое движение M плоскости (пространства) в выбранной декартовой системе координат можно записать так: $A \rightarrow M(A) = RA + \bar{V}$ где A - точка, она же вектор; R - ортогональная матрица, а \bar{V} - вектор.*

Таким образом, любое движение задается матрицей R и вектором \bar{V} .

Такая форма записи движений используется в PDB-файлах и программе SwissPDBViewer (Рис.5).

Упражнение 4.2. *На Рис.6 приведена копия фрагмента PDB файла, содержащая два движения. В одном из движений - вопиющая ошибка. Найдите ее и предположите как нужно ее исправить.*

5. ОПРЕДЕЛЕНИЕ КРИСТАЛЛИЧЕСКОЙ КОНФИГУРАЦИИ

Пусть у нас есть конфигурация атомов (мыслимая бесконечной, если речь о кристаллических конфигурациях) (Рис.7). Как и раньше, говорим о плоскости.

Существуют движения плоскости, переводящие конфигурацию в себя. Это значит, что образ любого атома конфигурации в результате движения пространства совпадает с другим атомом конфигурации (конечно, при этом углерод должен переходить в углерод, а азот - в азот и т.п.). Мы условились представлять движения с помощью прозрачки. Значит, копируем конфигурацию на прозрачку а потом прозрачку передвигаем (не переворачивая) так, чтобы рисунок на ней совпал с исходным.

На Рис.7 изображены такие движения: 1e - трансляция на вектор \bar{V}_1 ; 2e - трансляция на вектор \bar{V}_2 .

Определение. *Движения, переводящие конфигурацию атомов в себя, называются симметриями конфигурации.*

Обратите внимание на то, что для бесконечных конфигураций, каковыми являются кристаллические конфигурации атомов, трансляции тоже могут быть симметриями. В быту под симметрией редко понимают сдвиг (трансляцию); правда, и бесконечные предметы в быту наблюдаются редко :)

Одна симметрия существует у любой конфигурации - это движение, оставляющее все точки на месте; но это конечно математический выверт (который я оправдываю :)

Определение. Конфигурация атомов на плоскости называется кристаллической, если среди ее симметрий есть две трансляции на неколлинеарные⁴ векторы.

Конфигурация атомов в пространстве называется кристаллической, если среди ее симметрий есть три трансляции на некопланарные⁵ векторы.

У кристаллической конфигурации на плоскости, кроме двух трансляций, всегда есть бесконечно много симметрий. Это трансляции на другие вектора. Например, сдвиги на вектора \bar{V}_1 и \bar{V}_2 можно применять последовательно несколько раз. В результате, как нетрудно проверить, получится сдвиг на вектор $n\bar{V}_1 + m\bar{V}_2$, где n - сколько раз применялся сдвиг на \bar{V}_1 , m - сколько раз применялся сдвиг на \bar{V}_2 . Для трансляций порядок применения не важен⁶. Кроме того, очевидно, трансляции на противоположные вектора тоже будут симметриями. Таким образом, можно считать, что n и m целые, но не обязательно положительные числа.

Кроме трансляций, у некоторых кристаллических конфигураций на плоскости могут быть другие симметрии - вращения вокруг точек на определенные углы (Рис.7)

6. КРИСТАЛЛОГРАФИЧЕСКИЕ СИММЕТРИИ И КРИСТАЛЛОГРАФИЧЕСКИЕ ГРУППЫ

Определение. Множество всех симметрий кристаллической конфигурации атомов называется кристаллографической группой. В записях PDB пишут "Space group", это то же самое.

По определению кристаллической конфигурации, кристаллографическая группа всегда содержит бесконечно много трансляций. Трансляции образуют т.н. подгруппу.

Немного о терминах.

Очевидно, если M_1 и M_2 - две симметрии конфигурации, то последовательное их применение даст новое движение, переводящее конфигурацию в себя, т.е. новую симметрию. Последовательное применение движений записывается как их произведение: M_1M_2 . Это новое движение обозначим одной буквой M_3 . Таким образом, M_3 переводит точку A в точку $C = M_1(M_2(A))$ (более подробно: M_2 переводит A в точку $B = M_2(A)$, а затем M_1 переводит B в точку $C = M_1(B)$).

Кроме того, если M - симметрия конфигурации, то и обратное движение, обозначаемое M^{-1} , - тоже симметрия⁷

⁴коллинеарные векторы - это параллельные векторы

⁵компланарные векторы - это векторы, лежащие в одной плоскости (если их отложить от одной точки)

⁶А для других движений важен!

⁷Обратное к M движение определяется так. Любая точка A есть образ некоторой точки B при движении M : $A = M(B)$. По определению, M^{-1} переводит B в A : $M^{-1}(B) = A$

Обозначим группу симметрий конфигурации одной буквой G . Множество G состоит не из точек, а из движений⁸; один элемент $M \in G$ - это одно движение плоскости.

Определение. Подмножество S симметрий конфигурации называется подгруппой если для любых двух движений $M_1, M_2 \in S$ их произведение $M_1 M_2$ тоже принадлежит S и для каждого $M \in S$ обратное движение M^{-1} тоже принадлежит S .

Так вот, ясно, что множество трансляционных симметрий кристаллической конфигурации атомов образует подгруппу кристаллографической группы.

Важный факт природы состоит в том, что различных кристаллографических групп пространства существует не так уж много, во всяком случае - конечное число⁹. Слово "различных" нужно придать точные смысл. Например, если кристаллическую конфигурацию растянуть во все стороны в 2 раза (применить гомотегию, как, возможно, вас учили в школе), то вектора трансляционных симметрий растянутся в два раза, но очевидно, вся группа симметрий будет в определенном смысле той же самой.

7. ПРИМИТИВНАЯ КРИСТАЛЛИЧЕСКАЯ ЯЧЕЙКА

Симметрия в естествознании (математике, физике, химии, биологии и др.) - способ упростить описание объекта.

Итак, у кристаллической конфигурации атомов есть подгруппа трансляционных симметрий.

Лемма 7.1. Каждая трансляционная симметрия \bar{V} кристаллической конфигурации плоскости (пространства) выражается через две (соотв. три) трансляции на векторы \bar{V}_1, \bar{V}_2 (соотв. $\bar{V}_1, \bar{V}_2, \bar{V}_3$), а именно, $\bar{V} = n\bar{V}_1 + m\bar{V}_2$, где m и n - целые числа (соотв. $\bar{V} = n\bar{V}_1 + m\bar{V}_2 + l\bar{V}_3$).

Выбор трансляционных симметрий $\bar{V}_1, \bar{V}_2, \bar{V}_3$, порождающих все другие трансляционные симметрии, не однозначен - их можно выбрать по-разному.

Упражнение 7.1. Пусть в каждой вершине (бесконечной) клетчатой буллы расположен атом. Предъявите две существенно различные пары векторов, каждая из которых порождает все трансляционные симметрии этой кристаллической конфигурации. Различие должно заключаться в длине и/или угле между векторами.

Для данной кристаллографической конфигурации стараются выбрать порождающие трансляционные симметрии $\bar{V}_1, \bar{V}_2, \bar{V}_3$ так, чтобы углы между

⁸Множество чего? - спросила Алиса. - Просто множество. - ответила гусеница"Л. Керол, "Алиса в стране чудес"

⁹Если разрешить симметрии, меняющие ориентацию пространства, то их 230; если запретить - то меньше; не успел найти сколько именно, кажется около 60... ААл

векторами были прямыми а длины векторов - самыми коротким из возможных. Увы, это удается сделать не для любых кристаллических конфигураций.

Используя трансляционные симметрии, можно всю кристаллическую конфигурацию атомов восстановить из примитивной кристаллографической ячейки - параллелограмма (на плоскости) или параллелипипеда (в пространстве), обладающего свойствами:

(i) каждая точка плоскости (соотв. пространства) может быть перенесена трансляционными симметриями в основную ячейку; это же условие можно переформулировать так: образы основной ячейки при действии всех трансляционных симметрий кристаллической конфигурации покрывают всю плоскость (пространство);

(ii) образы основной ячейки при трансляционных симметриях либо совпадают, либо не пересекаются, либо пересекаются только по стороне (границе в случае 3D) ячейки.

Если \bar{V}_1, \bar{V}_2 - два вектора, порождающие все трансляционные симметрии кристаллической конфигурации, то в качестве примитивной ячейки можно выбрать параллелограмм, построенный на векторах \bar{V}_1, \bar{V}_2 , приложенных к одной точке (см. рис.7)

В кристаллографии иногда используют не примитивную ячейку, а большую. Так поступают в тех случаях, когда большая ячейка, составленная из нескольких примитивных, является прямоугольным параллелипипедом или, по крайней мере, углы между некоторыми ее сторонами прямые. Дело в том, что прямоугольность ячейки упрощает некоторые формулы. Такие ячейки называются не примитивными. Они удовлетворяют условию (i), но не удовлетворяют условию (ii) (см. выше).

8. АСИММЕТРИЧЕСКАЯ ЕДИНИЦА

Кроме трансляционных симметрий, у некоторых кристаллических конфигураций бывают и другие симметрии. В соответствии с теоремой, приведенной выше, любая такая симметрия - либо вращение вокруг оси на определенный угол, либо винтовое движение - трансляция с вращением. Факт, доказываемый в геометрии¹⁰, состоит в том что в кристаллографической группе могут быть вращения и винтовые движения с вращением на углы, равные половине, трети, четверти или одной шестой от 360 градусов и кратные им - т.е. на две трети, три четверти, пять шестых от 360°. Поэтому угол вращения для кристаллографических симметрий обозначают так: 2, 3, 4, 6; соответствующие оси вращения называют осями второго, третьего, четвертого, шестого порядков¹¹

¹⁰и, следовательно, имеющий место в природе - если евклидова геометрия правильно описывает пространство; а это так, что проверено с большой точностью

¹¹Осей пятого порядка в кристаллах не может быть, это теорема; поэтому обнаруженные физиками оси пятого порядка в регулярной структуре привело к открытию так называемых квазикристаллов - регулярных, но не кристаллических конфигураций атомов в пространстве

Вращательные и винтовые симметрии кристаллической конфигурации позволяют еще больше "ужать" информацию о кристалле - вся кристаллическая конфигурация может быть восстановлена с помощью кристаллографической группы из т.н. асимметрической единицы - ячейки пространства, меньшей чем примитивная ячейка.

Определение. Многогранник P в пространстве (многоугольник на плоскости) называется асимметрической ячейкой для данной кристаллической конфигурации атомов если выполнены условия

(i) каждая точка пространства (соотв. плоскости) может быть перенесена подходящей симметрией из кристаллографической группы в P ; это же условие можно переформулировать так: образы P при действии всех симметрий кристаллической конфигурации покрывают все пространство (соотв. плоскость);

(ii) образы P при кристаллографических симметриях либо совпадают, либо не пересекаются, либо пересекаются только по грани (соотв. стороне, в случае 2D) P .

Таким образом, асимметрическая единица по отношению к всей кристаллографической группе играет такую же роль, как примитивная ячейка - по отношению к подгруппе трансляционных симметрий.

В отличие от примитивной ячейки, которая всегда параллелипипед, асимметрическая ячейка может быть (и обычно бывает) не параллелипипедом, а более сложно устроенным многогранником.

В записях PDB обычно приводят координаты атомов из *одной асимметричной ячейки* потому, что этой информации, в совокупности с информацией о кристаллографической группе, также приведенной в записи, достаточно для восстановления, в принципе, всего кристалла. На практике для адекватной интерпретации многих записей PDB координат одних атомов из асимметрической ячейки недостаточной - необходимо достраивать молекулы в соседних асимметрических ячейках.

Упражнение 8.1. Для кристаллической структуры, изображенной на Рис.2

(i) найти все симметрии (отметить центры вращения и указать порядок оси, т.е. точки в данном случае)

(ii) найти и нарисовать асимметрическую ячейку

9. НЕКРИСТАЛЛОГРАФИЧЕСКИЕ СИММЕТРИИ

Отдельная белковая молекула, а особенно комплекс из многих белковых цепей - капсид вируса, какой-нибудь гомотетрамер и т.п., - может обладать собственной симметрией - движением, переводящим данную молекулу и комплекс молекул в себя. Пусть у нас кристалл, составленный из симметричных

молекул. Тогда возможны две ситуации. (1) Симметрия молекулы продолжается до симметрии всей кристаллической конфигурации атомов. Таким образом, это кристаллографическая симметрия, она входит в кристаллографическую группу, о таких симметриях шла речь выше. (2) Симметрия молекулы не продолжается до симметрии всей кристаллической конфигурации атомов. Такая симметрия называется некристаллографической. Примером могут служить рентгеноструктурные расшифровки структур капсидов вирусов, имеющих форму додекаэдра - правильного 12-гранника. У додекаэдра имеется ось симметрии пятого порядка; из написанного выше следует, что такая симметрия не может быть кристаллографической.

Наличие некристаллографической симметрии содержимого асимметрической единицы позволяет еще больше ужать информацию, а именно, написать координаты не всех атомов из асимметрической единицы, а только части, из которой можно восстановить асимметрическую единицу и весь кристалл в два шага: с помощью (1) некристаллографических симметрий (2) кристаллографических симметрий.

Хорошо это или плохо, но иногда именно так описывают структуру в записях PDB, приводя, естественно, в аннотации движения, которые являются некристаллографическими симметриями. Часто так поступают именно с расшифровками белков капсидов вирусов - для экономии места и облегчения анализа отдельного белка капсида.

10. Биологическая единица

То, что говорилось выше, позволяет адекватно описать кристалл с точки зрения кристаллографии, но иногда предлагаемые кристаллографами описания в записях PDB явно неудобны с точки зрения молекулярной биологии.

1. Из теории не следует, что всегда можно выбрать асимметрическую ячейку, содержащую целиком внутри одну молекулу белка (многомолекулярный комплекс), составляющего (-щий) кристалл¹². Может случиться так, что в асимметрической ячейке содержится первая половина от одной копии молекулы и вторая половина - от другой копии, хотя, по правде говоря, мне не приходят на ум примеры из PDB...

2. Бывает так, что соседние в кристалле молекулы белка имеют различающиеся конформации. Особенно часто такие различия в конформации наблюдаются у белков, у которых есть относительно подвижные части. И при этом пара белков с различными конформациями далее в пространстве повторяется образуя кристаллографическую конфигурацию. В таком случае в одну асимметрическую единицу кристалла попадают две химически одинаковые молекулы белка (иногда - не две, а больше). В таком случае биологический смысл имеет каждая отдельная молекула белка из структуры; их сравнение тоже имеет смысл - например, для выявления подвижных частей молекулы.

¹²См. пример с кристаллом из "огурцов" в лекциях Лунина

Поэтому биологически оправданным является разделение структуры на две, так называемые, *биологические единицы*.

3. Часто наблюдается ситуация, противоположная предыдущей: в кристалле лежат гомодимеры биологических макромолекул (иногда - тетрамеры и др.). Эти димеры симметричны, и более того, симметрия является кристаллографической. Таким образом, в асимметрической единице остается одна молекула гомодимера, и картина, получаемая из записи PDB, не отражает биологической ситуации. Например, бывает, что исследуемый белок *in vivo* всегда существует и функционирует только в виде гомодимера. Особенно шокирующими на первый взгляд получаются структуры гомодимера белка, связанного с полиндромной ДНК, представленные в записи PDB мономером на одноцепочечной ДНК...

В описанной ситуации содержательная *биологическая единица* получается добавлением к содержимому асимметрической единицы ее копии из подходящей соседней асимметрической единицы, т.е. применением подходящей кристаллографической симметрии.

Четвертое, пятое и т.п. Для получения биологических единиц из содержимого одной асимметрической ячейки иногда приходится применять и процедуру разделения (как в 2.), и процедуру объединения (как в 3.) одновременно. Во многих записях PDB (но не во всех) в аннотации описана процедура восстановления одной или нескольких биологических единиц по информации, приведенной в записи.

11. Банки биологических единиц

Наряду с PDB, основной единицей хранения в котором является обычно запись, представляющая содержимое асимметрической единицы кристалла, есть банки, хранящие восстановленные биологические единицы. Таким образом, часто именно к ним следует обращаться для получения адекватной структурной информации.

Это такие банки.

- (i) PDB с некоторых пор кроме основных записей предлагает также биологические единицы.
- (ii) Банк PQS (Protein Quaternary Structure)
- (iii) Банк NDB (Nucleic acids Data Bank) - структуры нуклеиновых кислот и их комплексов с белками
- (iv) Банк капсидов вирусов

Необходимо иметь ввиду следующие обстоятельства.

Биологическая единица - понятие из молекулярной биологии, а не из кристаллографии. Таким образом, знание биохимии и молекулярной биологии есть главный источник для восстановления и интерпретации биологической единицы. Таким знанием может не обладать специалист по рентгенооструному анализу, представляющий результат в PDB. И уж наверняка такого знания

лишен "компьютер", пытающийся восстановить биологическую единицу автоматически.

В банках PDB, PQS, NDB биологические единицы восстанавливаются автоматическими программами со всеми вытекающими последствиями (см. предыдущий абзац).

Иногда, наоборот, контакты белка с другими копиями его же в соседних асимметрических ячейках позволяют высказать или подтвердить предположения о возможности димеризации (полимеризации) исследуемых белков, способности к белок-белковым взаимодействиям; позволяют локализовать взаимодействующие поверхности. Действительно, можно предполагать, что в кристалле молекулы белка ориентируются так, чтобы взаимодействовать друг с другом наиболее выгодным способом, и если взаимодействие этих белков функционально значимо *in vivo*, то скорее всего, оно и воспроизведется в кристалле. Приведенное объяснение, конечно, не более, чем разумное предположение и требует независимых проверок.

В кристалле соседние белковые молекулы могут не взаимодействовать непосредственно друг с другом, только через буфер, остающийся, как правило, не расшифрованным в эксперименте по РСА.

Контакты белка с молекулами из соседних асимметрических единиц могут приводить к артефактам. Так, вопрос про молекулу воды, "повисшую в воздухе", заданный во введении, решается просто: данная молекула связана водородными связями (? проверить!!!) с молекулой белка из соседней асимметрической ячейки; такие молекулы часто находятся в одинаковом положении относительно всех белков кристалла и потому могут быть расшифрованы с помощью рентгеноструктурного анализа. В записях PDB встречается множество разных на первый взгляд, странных ситуаций, которые правильно интерпретируются при анализе контактов с молекулами из соседних асимметрических ячеек.

12. ВОССТАНОВЛЕНИЕ СОСЕДНИХ АСИММЕТРИЧЕСКИХ ЕДИНИЦ КРИСТАЛЛА И БИОЛОГИЧЕСКИХ ЕДИНИЦ С ПОМОЩЬЮ SWISSPDBVIEWER'А

В записи PDB приводится следующая информация.

- Размер и форма элементарной ячейки - длины трех векторов - ребер параллелепипеда и углы между ними; элементарная ячейка может не быть примитивной.
- Название кристаллографической группы в соответствии с номенклатурой; как всегда, бывают синонимы, что может затруднить идентификацию группы для некристаллографа.
- Порождающие кристаллографические симметрии. Они задаются двумя способами: (1) в недекартовых кристаллографических координатах¹³ (2) в обычных декартовых координатах - с помощью матрицы и

¹³В этих координатах концы трех ребер элементарной ячейки, выходящих из начала координат, имеют координаты, соответственно, (1, 0, 0), (0, 1, 0) и (0, 0, 1)

вектора. Каждая кристаллографическая симметрия может быть получена последовательным применением этих порождающих симметрий и трансляций на три вектора - ребра элементарной ячейки.

- Некристаллографические симметрии в декартовых координатах (если они имеются)
- Инструкция по восстановлению биологических единиц в виде короткого текста и движений, заданных в декартовых координатах (Например, "применить движение 1 к цепи А, движение 2 - к цепи В ...)

Программа SwissPDBviewer умеет:

- Применить каждое из порождающих движений или все вместе и таким образом получить содержимое соседних асимметрических единиц. Следует подчеркнуть, что таким способом не обязательно находятся все соседи - для получения всех, возможно потребуется применить также трансляции.
- Применить любую из трех трансляций или их комбинацию.
- Применить любое движение, заданное в декартовых координатах матрицей и вектором ¹⁴. Такие действия необходимы, например, при восстановлении биологических единиц, т.к. обычно SwissPDBviewer не в состоянии распознать автоматически и применить инструкцию по восстановлению из записи PDB .

13. ЗАКЛЮЧЕНИЕ

Что успел, то записал.

Буду благодарен за любую критику и предложения (что непонятно или плохо объяснено, чего не хватает). Надеюсь, они помогут мне (если когда-нибудь будет свободное время :-(...) дописать и поправить эти странички для будущих поколений студентов ФББ.

¹⁴Программа не контролирует ортогональность введенной матрицы