

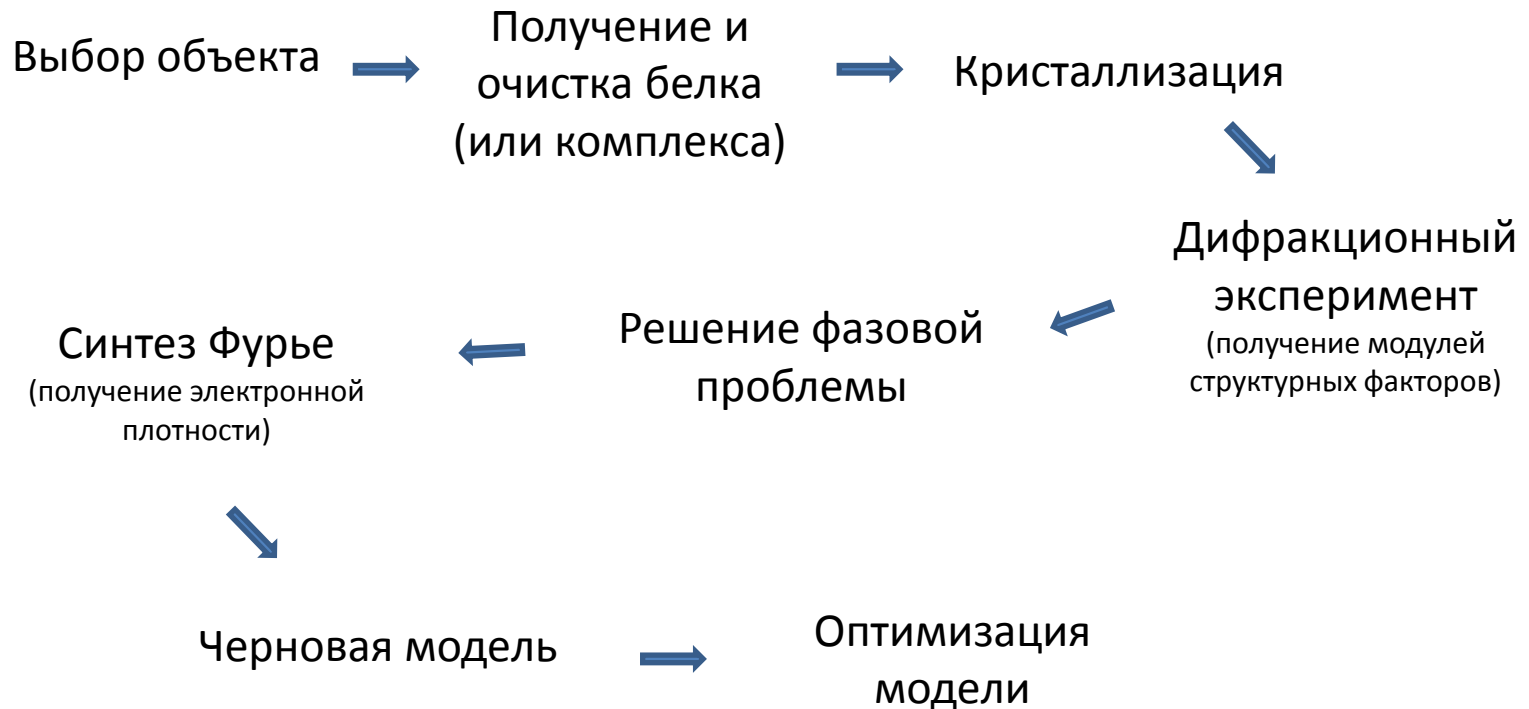
Введение в рентгеноструктурный анализ макромолекул

Часть 2

С.А. Спирин

Использованы материалы лекций В.Ю. Лунина

Процедура получения 3D-модели макромолекулы методом РСА



Вектор рассеяния

Везде далее направление падающего луча (то есть единичный вектор с таким направлением) будем обозначать σ_0 .

Лучи от объекта рассеиваются во все стороны. Интенсивность рассеянных лучей измеряется детектором. Каждой точке детектора соответствует своё направление σ от объекта в эту точку.

Удобно для обозначения точки детектора вместо направления σ использовать вектор рассеяния S

$$S = \frac{\sigma - \sigma_0}{\lambda}$$

Здесь λ — длина волны

Таким образом, **каждой точке детектора соответствует свой вектор S**

Модуль и фаза структурного фактора

В каждую точку детектора s приходит рассеянная волна.

В результате в этой точке зависимость E от времени выглядит так:

$$E(t) \propto F(s) E_0 \sin[2\pi(-\nu t) - \varphi(s)]$$

$F(s)$ называется **модулем структурного фактора** (в точке s)

$\varphi(s)$ называется **фазой структурного фактора**

Детектор измеряет энергию падающей волны, т.е. **квадрат** $F(s)$

Фаза непосредственно не наблюдаема!

Кристаллы

Определение. Кристаллом называется объект, чья электронная плотность ρ удовлетворяет условиям:

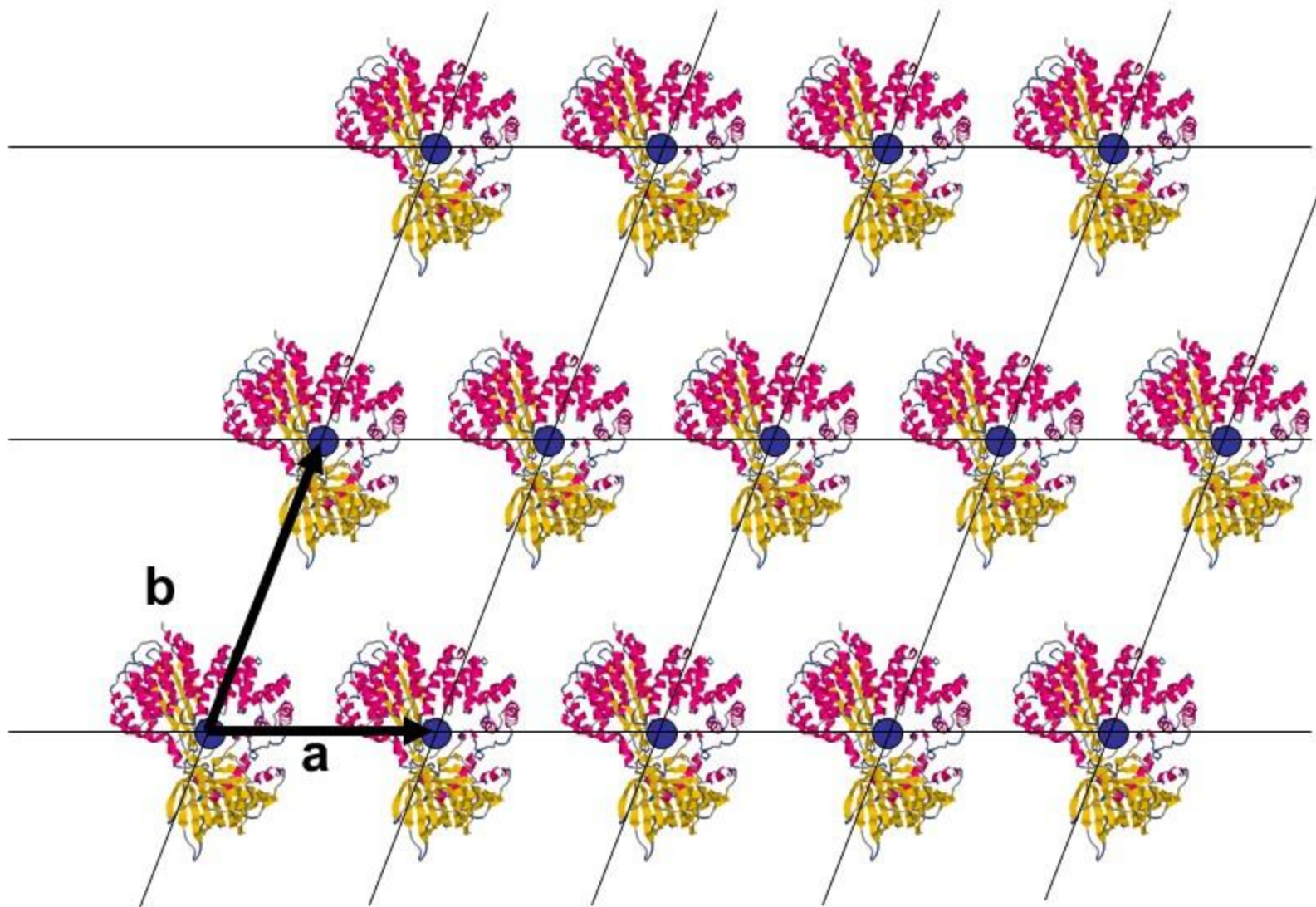
$$\rho(\mathbf{r}) = \rho(\mathbf{r}+\mathbf{a}) = \rho(\mathbf{r}+\mathbf{b}) = \rho(\mathbf{r}+\mathbf{c})$$

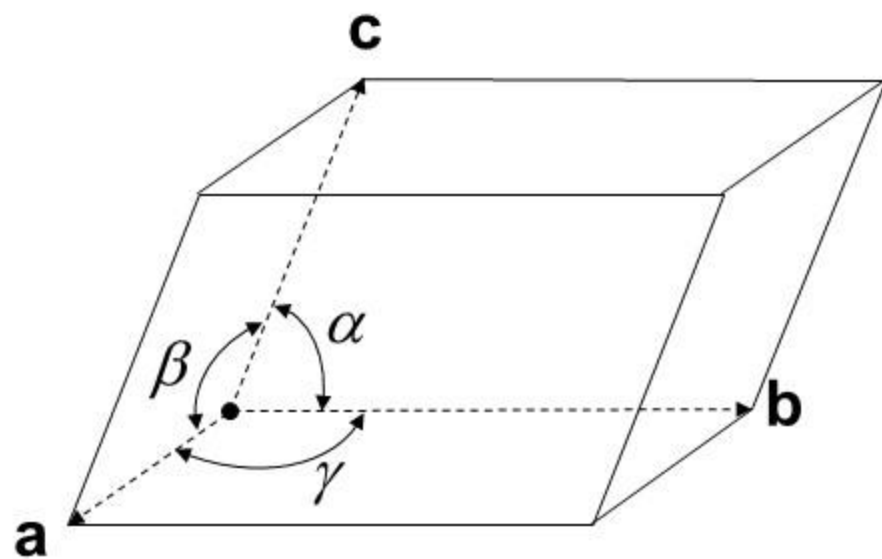
где $\mathbf{r}$ — точка пространства (=вектор из начала координат), а

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ — три **некомпланарных** вектора

(равенства должны быть верны для любой точки кристалла $\mathbf{r}$, для которой сдвиги на $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ не выводят за пределы кристалла)

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ — называются **рёбрами элементарной ячейки** кристалла





элементарная ячейка
(unit cell)

$a, b, c, \alpha, \beta, \gamma$
параметры элементарной
ячейки
(unit cell parameters)

содержимое элементарной ячейки полностью
определяет «внутреннюю» структуру кристалла

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{a}) = \rho(\mathbf{r} + \mathbf{b}) = \rho(\mathbf{r} + \mathbf{c})$$

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

x, y, z – относительные
(кристаллографические)
координаты

$$\tilde{\rho}(x, y, z) = \rho(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$$

элементарная ячейка – куб
 $0 \leq x, y, z \leq 1$

$\tilde{\rho}(x, y, z)$ имеет период 1. по всем переменным.

Параметры ячейки

(кусочек файла формата cif)

```
#
_cell.entry_id          101M
_cell.length_a          91.670
_cell.length_b          91.670
_cell.length_c          45.970
_cell.angle_alpha       90.00
_cell.angle_beta        90.00
_cell.angle_gamma       120.00
_cell.Z_PDB             6
_cell.pdbx_unique_axis  ?
#
```

линейные размеры кристалла $0.1\text{мм}=10^6\text{\AA}$

линейные размеры элементарной ячейки $100\text{\AA}$

количество копий молекулы в кристалле $(10^4)^3=10^{12}$

Кристалл усиливает интенсивность
в 10^{24} раз !!!

Условия дифракции
(Лауэ):

$$(\mathbf{s}, \mathbf{a}) = h$$

$$(\mathbf{s}, \mathbf{b}) = k$$

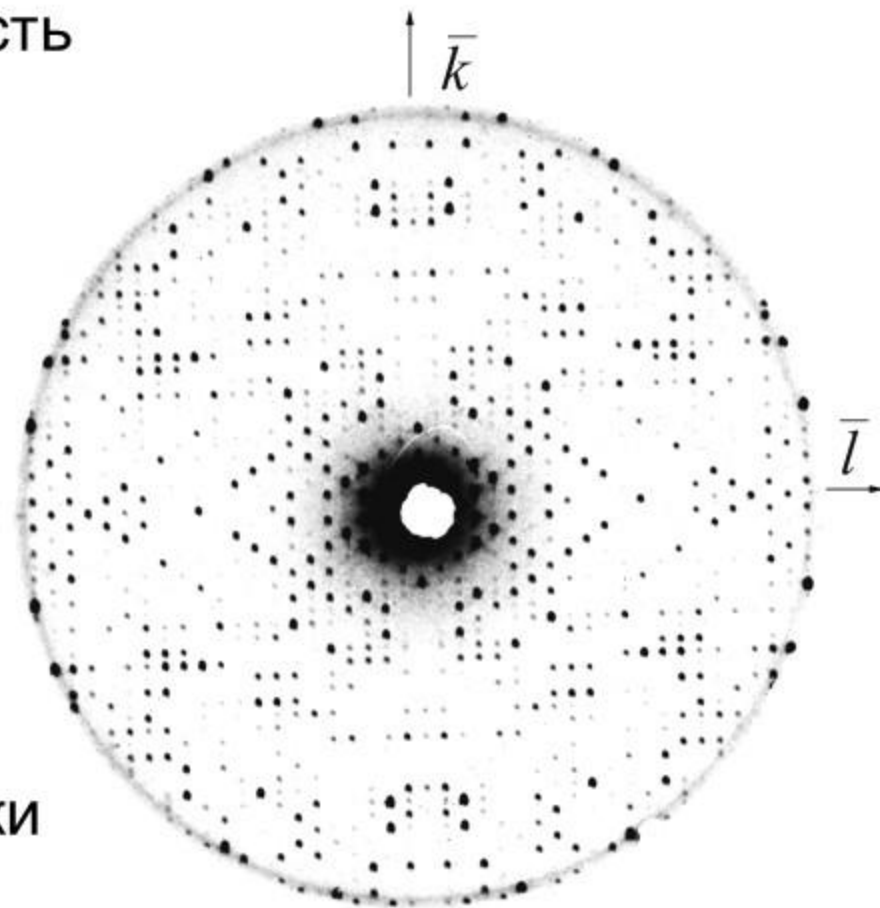
$$(\mathbf{s}, \mathbf{c}) = l$$

h, k, l - целые числа

(индексы рефлекса)

$\mathbf{a}, \mathbf{b}, \mathbf{c}$ - ребра элементарной ячейки

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \text{ - вектор рассеяния}$$



Ряды Фурье для функций нескольких переменных

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

$$\tilde{\rho}(x, y, z) = \rho(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$$

x, y, z – относительные
(кристаллографические)
координаты

элементарная ячейка – куб
 $0 \leq x, y, z \leq 1$

$\tilde{\rho}(x, y, z)$ имеет период 1. по всем переменным.

$$\tilde{\rho}(x, y, z) \approx F_{000} + \sum_{hkl} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

$$F_{hkl} = \sqrt{A_{hkl}^2 + B_{hkl}^2} \quad \operatorname{tg} \varphi_{hkl} = \frac{B_{hkl}}{A_{hkl}}$$

$$A_{hkl} = \int_{0 \leq x, y, z \leq 1} \tilde{\rho}(x, y, z) \cos[2\pi(hx + ky + lz)] dx dy dz$$

$$B_{hkl} = \int_{0 \leq x, y, z \leq 1} \tilde{\rho}(x, y, z) \sin[2\pi(hx + ky + lz)] dx dy dz$$

Интенсивность рассеянного луча пропорциональна квадрату коэффициента ряда Фурье

для разложения в ряд Фурье электронной плотности в **относительных** координатах

$$I(\mathbf{s}) \sim F_{hkl}^2$$

если $(\mathbf{s}, \mathbf{a}) = h$, $(\mathbf{s}, \mathbf{b}) = k$, $(\mathbf{s}, \mathbf{c}) = l$

Ряд Фурье:

$$\rho(x, y, z) = \sum_{\substack{hkl \\ \text{целые}}} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

x, y, z — относительные координаты:

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

Синтез Фурье:
$$\tilde{\rho}_S(x, y, z) \approx \sum_{(hkl) \in S} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

Обратите внимание на диапазоны суммирования для ряда и для синтеза!

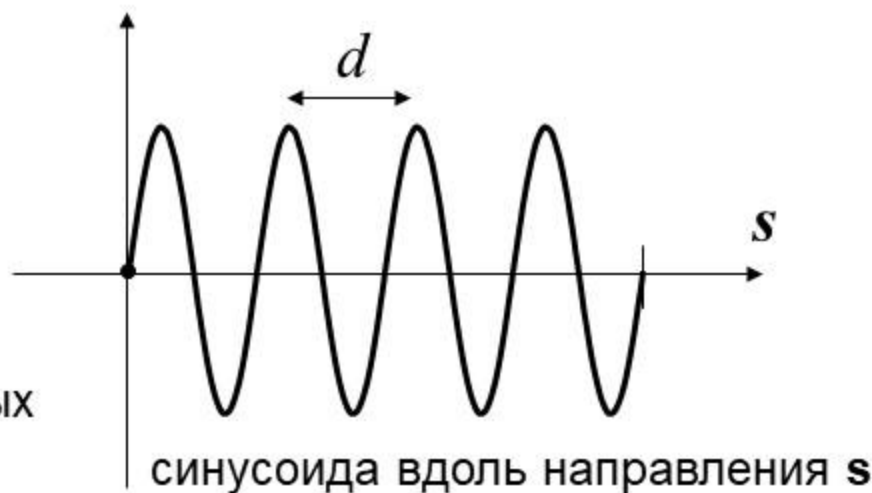
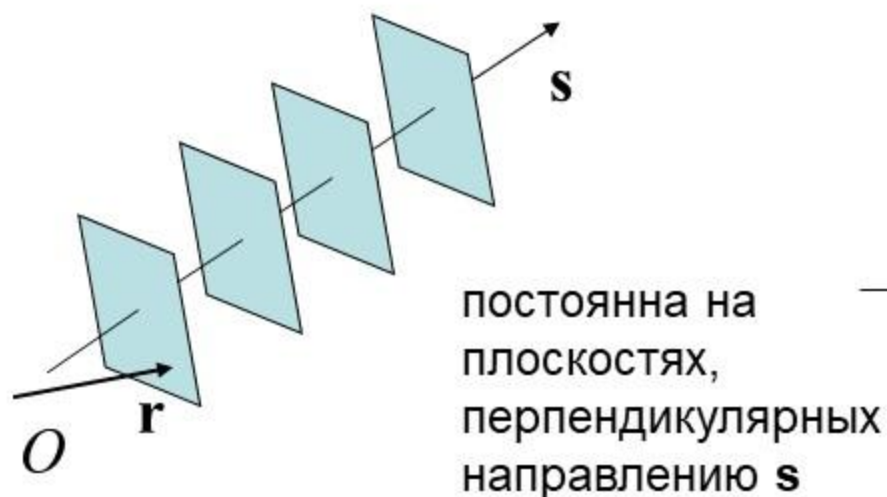
S — это множество рефлексов, для которых удалось: (а) измерить F , (б) получить φ

Гармоника Фурье

$$H_s(\mathbf{r}) = \cos[2\pi(hx + ky + lz)] = \cos[2\pi(\mathbf{s}, \mathbf{r})]$$

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

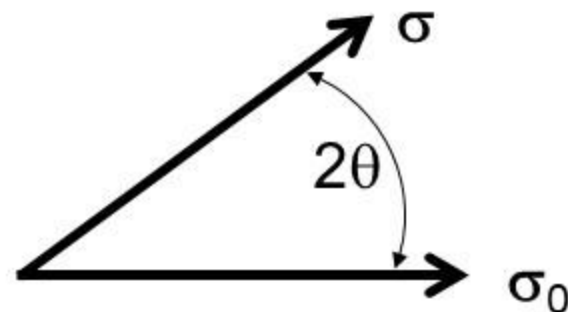
$$h = (\mathbf{s}, \mathbf{a}), k = (\mathbf{s}, \mathbf{b}), l = (\mathbf{s}, \mathbf{c})$$



Разрешение, соответствующее гармонике Фурье

$d = \frac{1}{|\mathbf{s}|}$ - расстояние между соседними максимумами в направлении $\mathbf{s}$;

$$|\mathbf{s}| = \frac{2 \sin \theta}{\lambda} \quad d = \frac{\lambda}{2 \sin \theta}$$



Разрешение структуры

Точнее — разрешение рентгеноструктурного эксперимента

Имеется множество S измеренных рефлексов (h,k,l)

Каждому набору $h,k,l \Rightarrow$ гармоника Фурье и вектор $s \Rightarrow$ разрешение d

Пусть $d_{\min}$ — минимальное из разрешений рефлексов из S

Для каждого $d \geq d_{\min}$ можно сосчитать два числа:

1. Число **всех возможных** троек целых чисел (h,k,l) с разрешением $\geq d$
2. Число **измеренных** рефлексов $(h,k,l) \in S$ с разрешением $\geq d$

Второе число, как правило, меньше первого!

Отношение второго к первому называется **полнотой** набора S , соответствующей разрешению d

На практике: выбирается самое маленькое d с «приличной» полнотой ($>0,9$) и записывается в результат как «разрешение структуры»

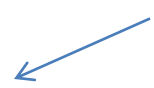
Если всё сделано правильно, разрешение структуры примерно соответствует минимальному расстоянию между пиками электронной плотности, при котором эти пики будут видны на результате синтеза Фурье отдельно.

Некоторые жулики выставляют как разрешение структуры значение с полнотой 0,8 и даже 0,7 и меньше. Это — прямой обман пользователя! Следите за полнотой!

Разрешение и полнота

```
#
_reflns.entry_id          101M
_reflns.observed_criterion_sigma_I  0.
_reflns.observed_criterion_sigma_F  ?
_reflns.d_resolution_low    6.00
_reflns.d_resolution_high   2.07
_reflns.number_obs          13624
_reflns.number_all          ?
_reflns.percent_possible_obs 99.9
_reflns.pdbx_Rmerge_I_obs    0.0700000
_reflns.pdbx_Rsym_value      ?
_reflns.pdbx_netI_over_sigmaI ?
_reflns.B_iso_Wilson_estimate 14.0
_reflns.pdbx_redundancy      7.45
_reflns.pdbx_diffraction_id   1
_reflns.pdbx_ordinal         1
#
```

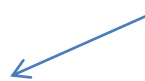
Очень хорошо!



Разрешение и полнота

#	
_reflns.entry_id	1WET
_reflns.observed_criterion_sigma_I	?
_reflns.observed_criterion_sigma_F	?
_reflns.d_resolution_low	10.000
_reflns.d_resolution_high	2.500
_reflns.number_obs	?
_reflns.number_all	?
_reflns.percent_possible_obs	87.000
_reflns.pdbx_Rmerge_I_obs	0.0629000
_reflns.pdbx_Rsym_value	?
_reflns.pdbx_netI_over_sigmaI	7.400
_reflns.B_iso_Wilson_estimate	?
_reflns.pdbx_redundancy	3.000
_reflns.pdbx_diffraction_id	1
_reflns.pdbx_ordinal	1
#	

Нормально

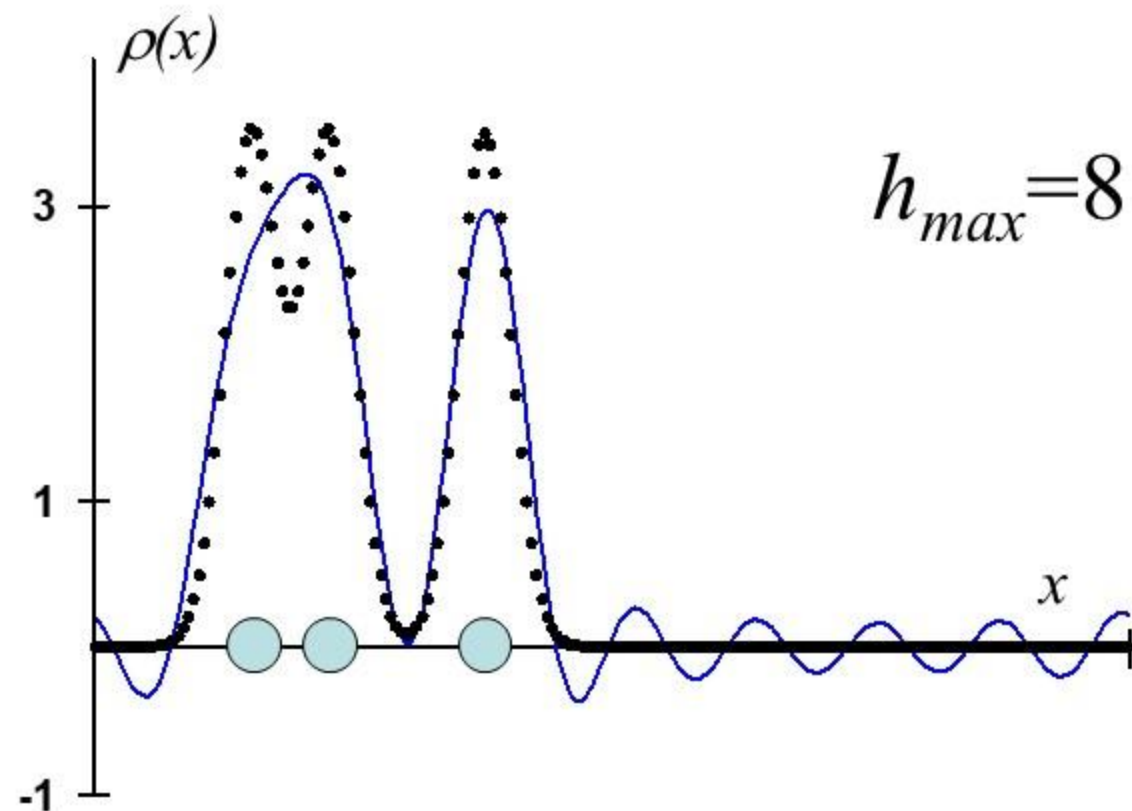


Разрешение и полнота

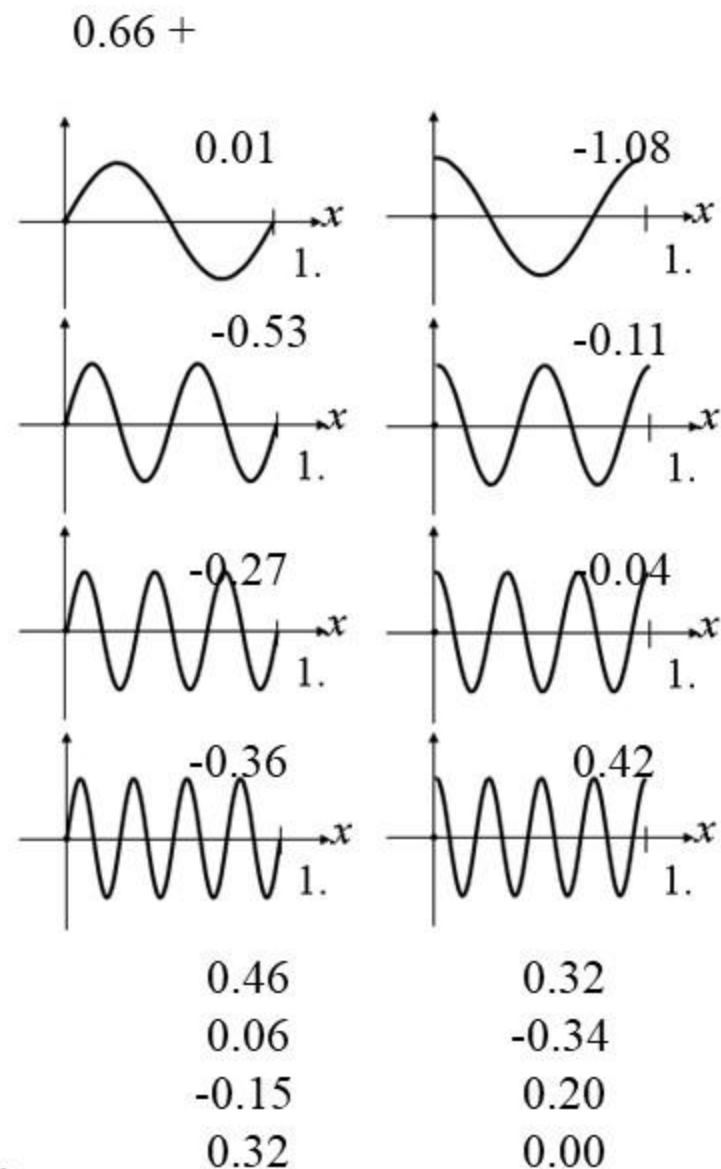
# _reflns.entry_id	10MH	
_reflns.observed_criterion_sigma_I	?	
_reflns.observed_criterion_sigma_F	?	
_reflns.d_resolution_low	30.0	
_reflns.d_resolution_high	2.5	
_reflns.number_obs	15976	
_reflns.number_all	?	
_reflns.percent_possible_obs	77.2	← Плохо!

«Честное» разрешение, вероятно, заметно хуже указанных 2,5Å

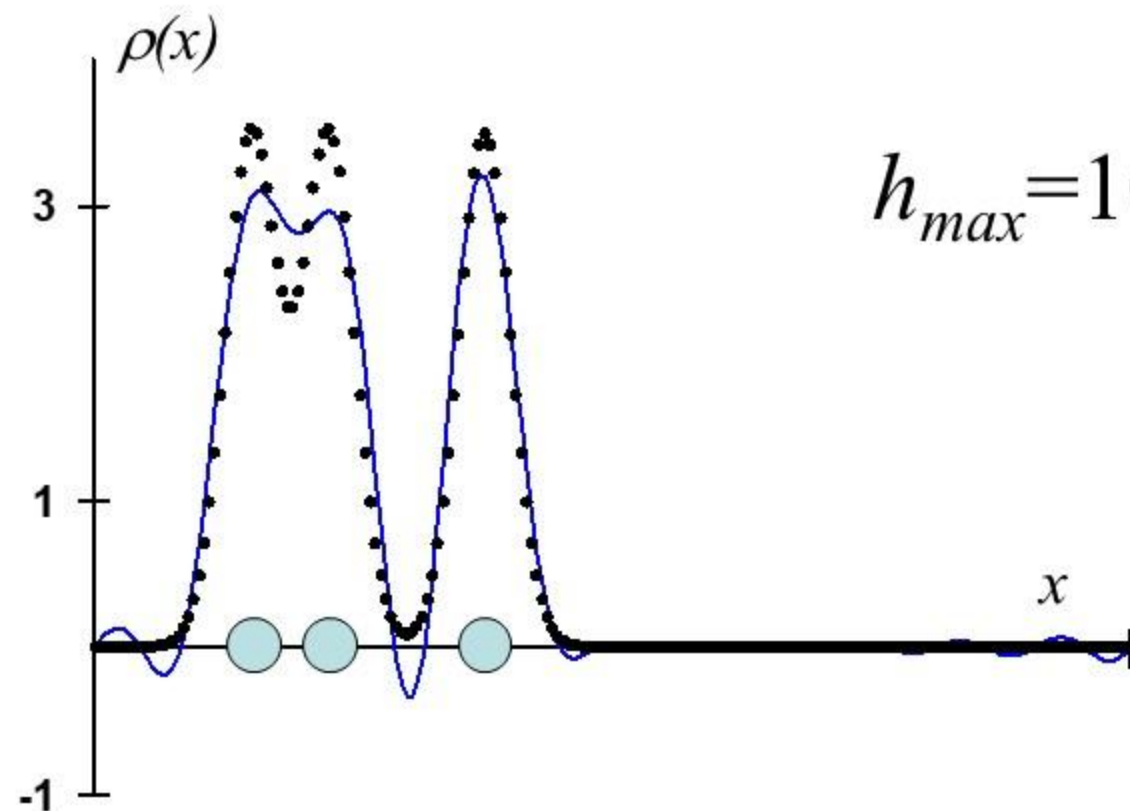
При таком разрешении два левых пика сливаются
(но зная формулу молекулы, координаты атомов всё равно можно указать довольно точно)



$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

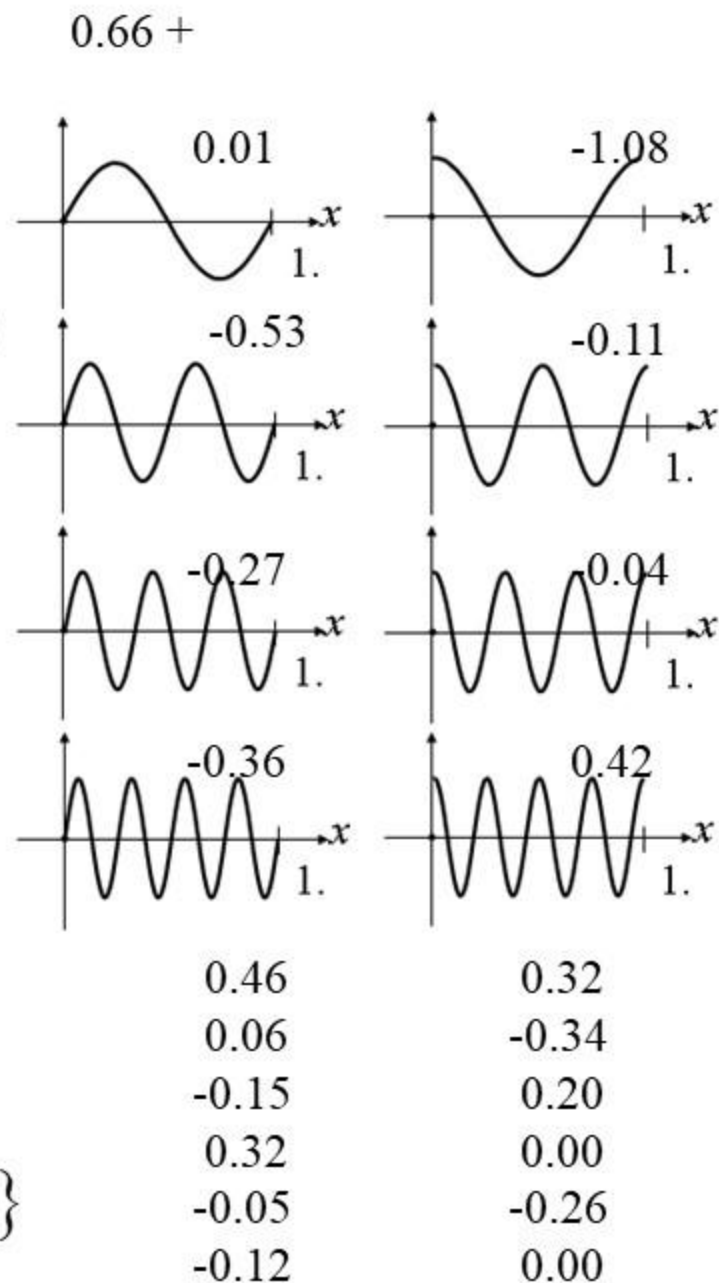


Здесь разрешение примерно равно расстоянию между левыми пиками. Точность предсказания координат атомов заметно лучше разрешения!



$$h_{max}=10$$

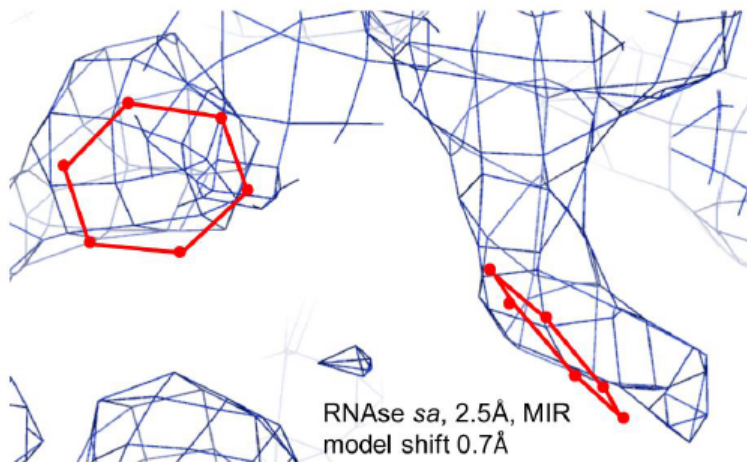
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



Разрешение синтеза Фурье

Разрешение – формальная характеристика. Она учитывает только количество включенных в работу рефлексов и дает теоретическую оценку качества «идеального» синтеза. Ошибки в используемых значениях структурных факторов ухудшают визуальное качество синтеза.

Разрешение дает оценку величины набора собранных экспериментальных данных и визуального качества карт. Оно не совпадает с точностью определения координат.



При работе с набором данных разрешения 2Å, типичной оценкой точности определения координат атомов после уточнения является величина 0.02Å.

Разрешение

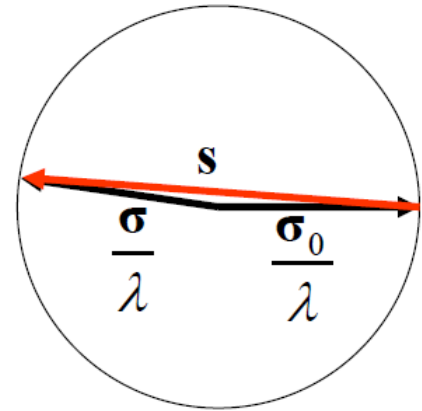
$$|s|_{\max} = \frac{2}{\lambda}$$

$$d_{\min}^{\text{theoretical}} = \frac{\lambda}{2}$$

$$s = \frac{\sigma - \sigma_0}{\lambda}$$

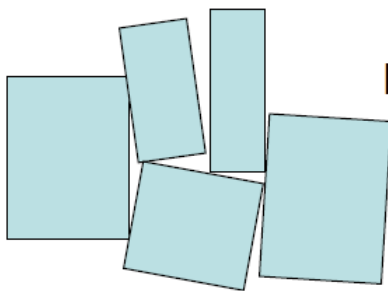
$$|\sigma| = |\sigma_0| = 1$$

$$d = \frac{1}{|s|}$$

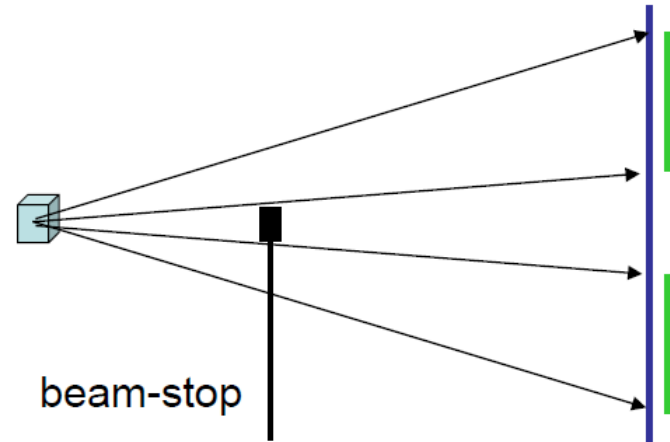


Разрешение:

- теоретический предел $\lambda/2$;
- недостаток чувствительности детектора – ухудшает разрешение;
- дефекты кристалла – ухудшают разрешение;
- “beam-stop” мешает регистрировать низкоугловые рефлексы.



“мозаичность
кристалла”



Фазовая проблема

Молекулярное замещение
Molecular Replacement (MR)

Множественное
изоморфное замещение
*Multiple Isomorphous
Replacement (MIR)*

Многоволновое аномальное
рассеяние
*Multi-wavelength Anomalous
Dispersion / Diffraction (MAD)*

Метод молекулярного замещения

1. Постулаты:

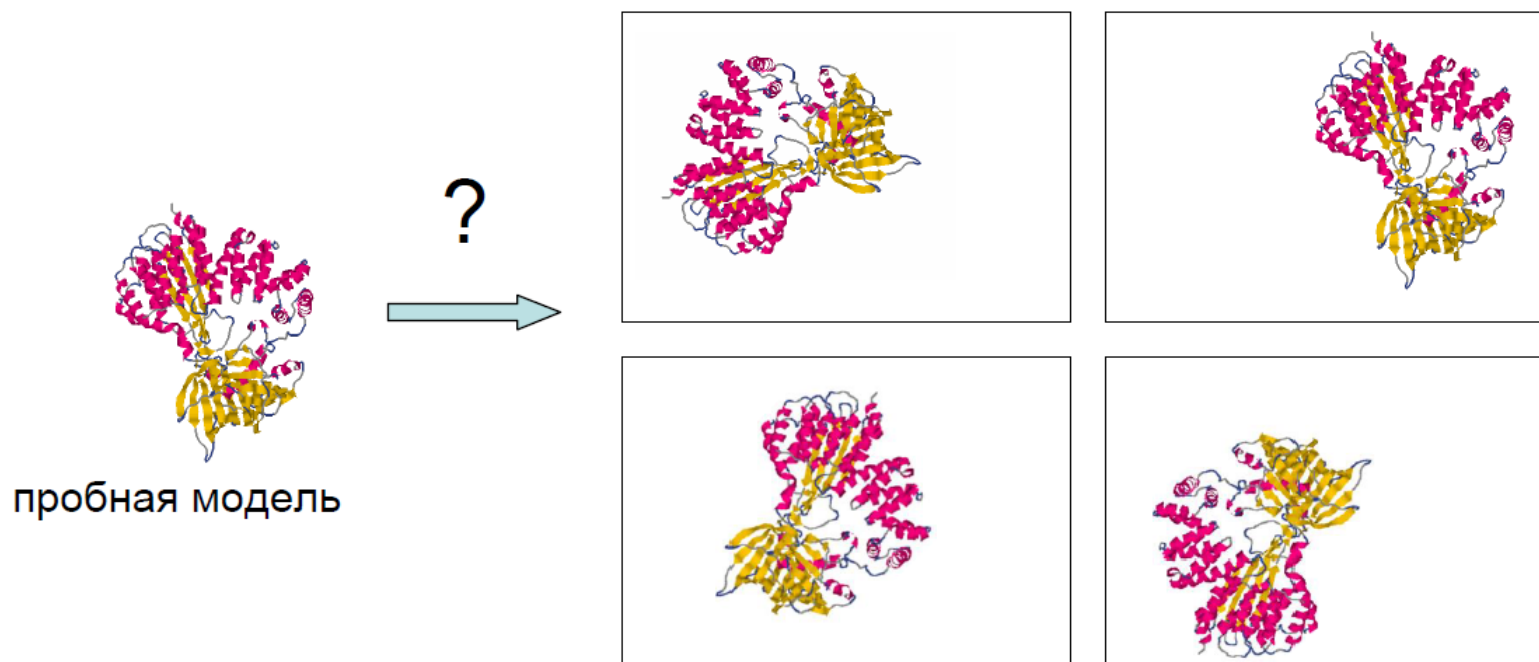
- белок – цепочка разнообразных аминокислотных остатков, свернутая в пространстве;
- последовательность типов аминокислот в цепочке определяет пространственную структуру белка;
- цепочки с похожей (гомологичной) последовательностью имеют похожую пространственную структуру.
- значения фаз, рассчитанные по модели гомологичного белка могут служить приближением к искомым значениям фаз.

Метод молекулярного замещения

2. Программа действий:

- найти в банке структур белок с похожей последовательностью и уже известной структурой; или воспользоваться программой предсказания структуры;
- найденную структуру использовать в качестве первого приближения к искомой структуре; рассчитать по ней фазы; построить синтезы Фурье;
- внести корректировки в модель в соответствии с рассчитанными синтезами.

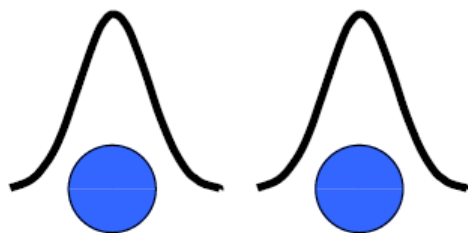
MR - Molecular Replacement Молекулярное замещение;



Находим положение пробной модели в ячейке, при котором среднее отклонение рассчитанных по модели F_{hkl} от экспериментальных минимально. По тому, каково это минимальное отклонение, можно судить о том, годится ли пробная модель для молекулярного замещения.

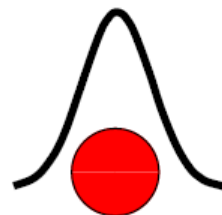
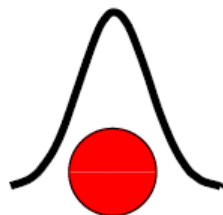
«Комбинированные» синтезы Фурье.

«Точная»
структура



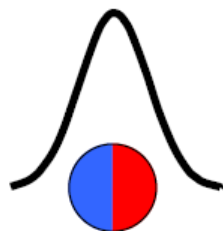
F^{obs} φ^{true}

модель



F^{calc} φ^{calc}

комбинированный
синтез Фурье



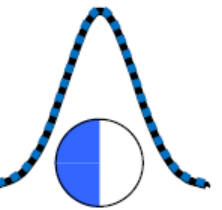
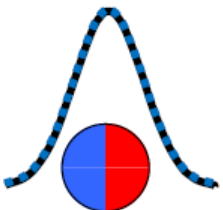
F^{obs} φ^{calc}

разностный
синтез Фурье



$F^{obs} - F^{calc}$ φ^{calc}

«2-1»
 $2F^o - F^c$
синтез Фурье



$2F^{obs} - F^{calc}$ φ^{calc}

MR - Molecular Replacement Молекулярное замещение;

- 😊 • один из наиболее используемых методов решения фазовой проблемы;
- 😊 • чем больше белков исследовано, тем шире возможности метода;
- 😞 • необходимо наличие гомологичной пробной модели;
- 😞 • результат зависит от обоснованности гипотезы о пространственной гомологии между пробной моделью и исследуемой структурой.

Метод изоморфного замещения

"нативный" белок



эксперимент

$\{F^P\}$

изоморфное производное

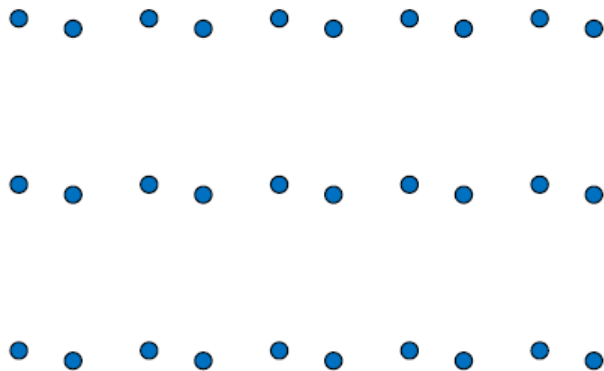
атом тяжёлого
металла



эксперимент

$\{F^{PH}\}$

подструктура тяжёлых атомов



$\{F^H\}$

?

$$\vec{F}_{hkl}^{PH} = \vec{F}_{hkl}^P + \vec{F}_{hkl}^H$$

$$F_{hkl}^H \approx \left| F_{hkl}^{PH} - F_{hkl}^P \right|$$

- 😊 • метод позволяет решать фазовую проблему;
- 😊 • наличие каналов в кристаллах белка позволяет получать производные;
- 😞 • изоморфизм имеет место лишь приближенно;
- 😞 • проблема определения мест присоединения тяжелых меток;
- 😞 • проблема создания тяжелых меток для больших макромолекулярных комплексов
- 😞 • фазы определяются неоднозначно; необходимо несколько производных.

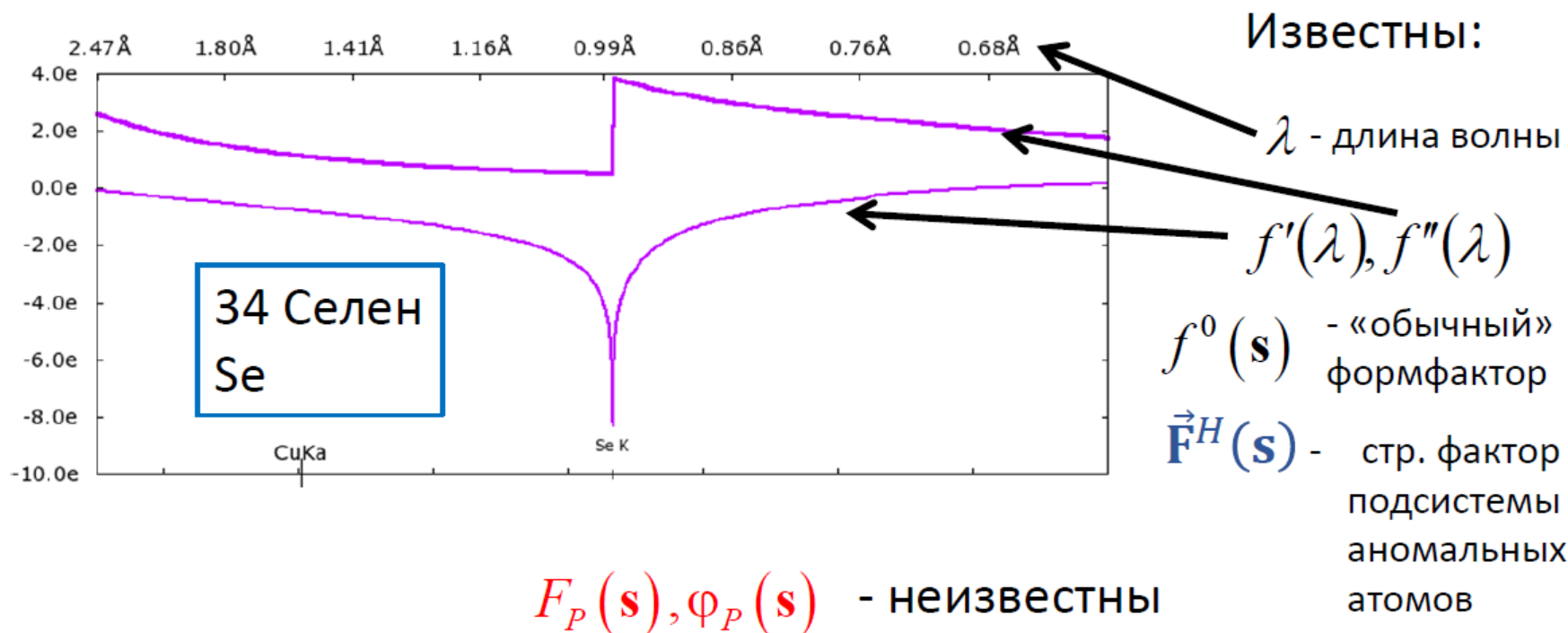
MIR - Multiple Isomorphous Replacement
Множественное изоморфное замещение;

AD - Anomalous Dispersion / Аномальное рассеяние

Кинематическая теория рассеяния: $I^{obs}(\mathbf{s}) = |\vec{\mathbf{F}}_P(\mathbf{s})|^2$

Наличие аномального рассеивающих атомов:

$$I^{obs}(\mathbf{s}; \lambda) = \left| \vec{\mathbf{F}}^P(\mathbf{s}) + \frac{f'(\lambda)}{f^0(\mathbf{s})} \vec{\mathbf{F}}^H(\mathbf{s}) + \frac{f''(\lambda)}{f^0(\mathbf{s})} \vec{\mathbf{F}}^H(\mathbf{s}) \Big|_{rot\ 90^\circ} \right|^2$$

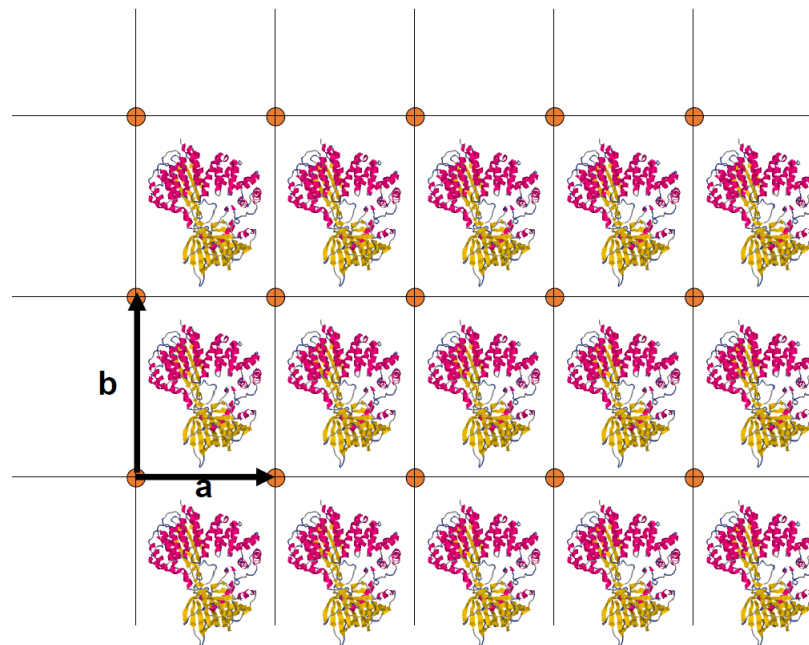
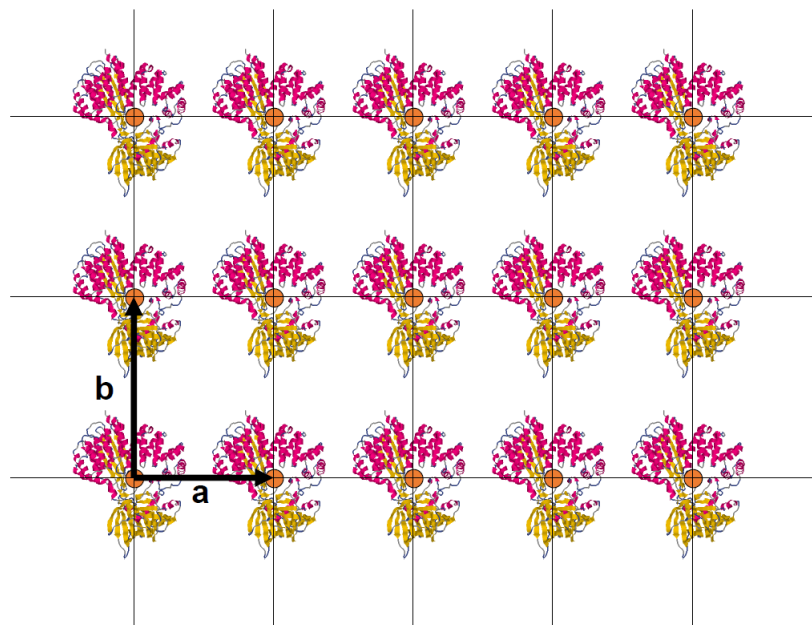


AD - Anomalous Dispersion

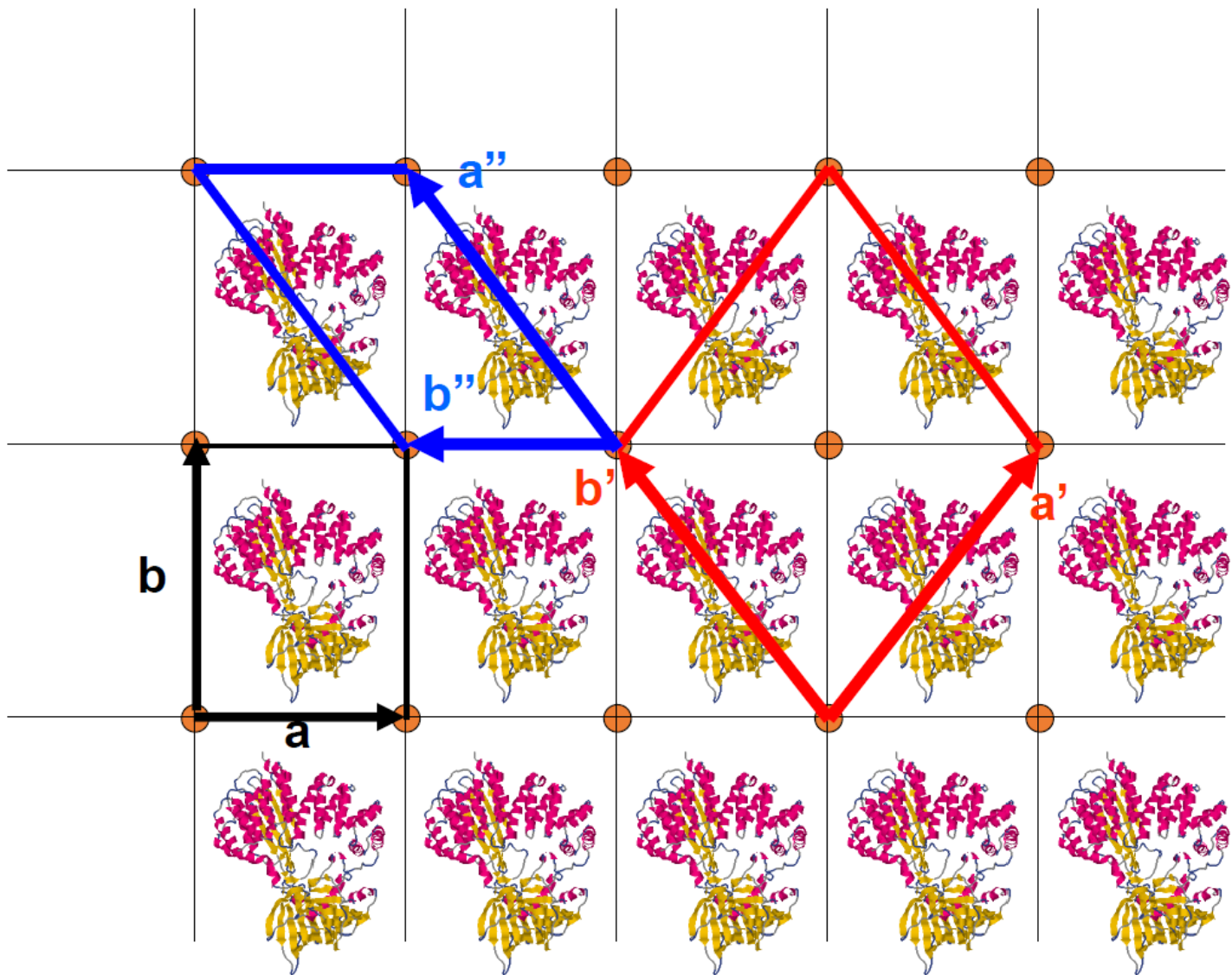
Аномальное рассеяние

- метод позволяет решать фазовую проблему;
- в белках встречаются аномально рассеивающие атомы; можно использовать производные с тяжелыми атомами;
- требуется наличие аномально рассеивающих атомов;
- требуются эксперименты при разных длинах волн (синхротрон);
- проблема определения мест нахождения аномально рассеивающих атомов;
- при использовании одной длины волны фазы определяются неоднозначно;

Выбор элементарной ячейки



Положение ячейки определяется произвольно

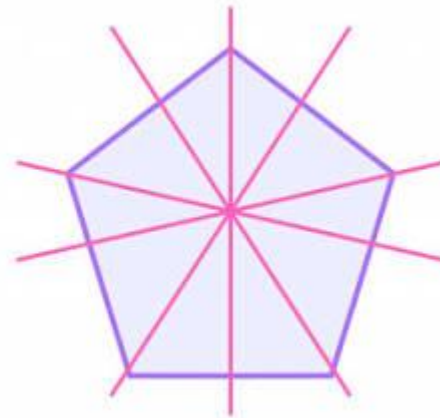
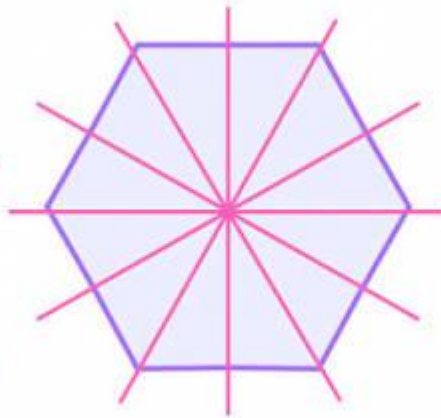
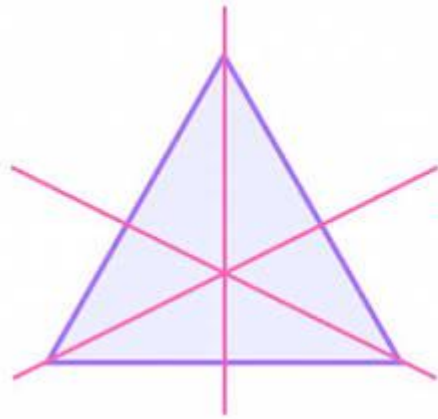


Сами $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ тоже могут быть выбраны по-разному

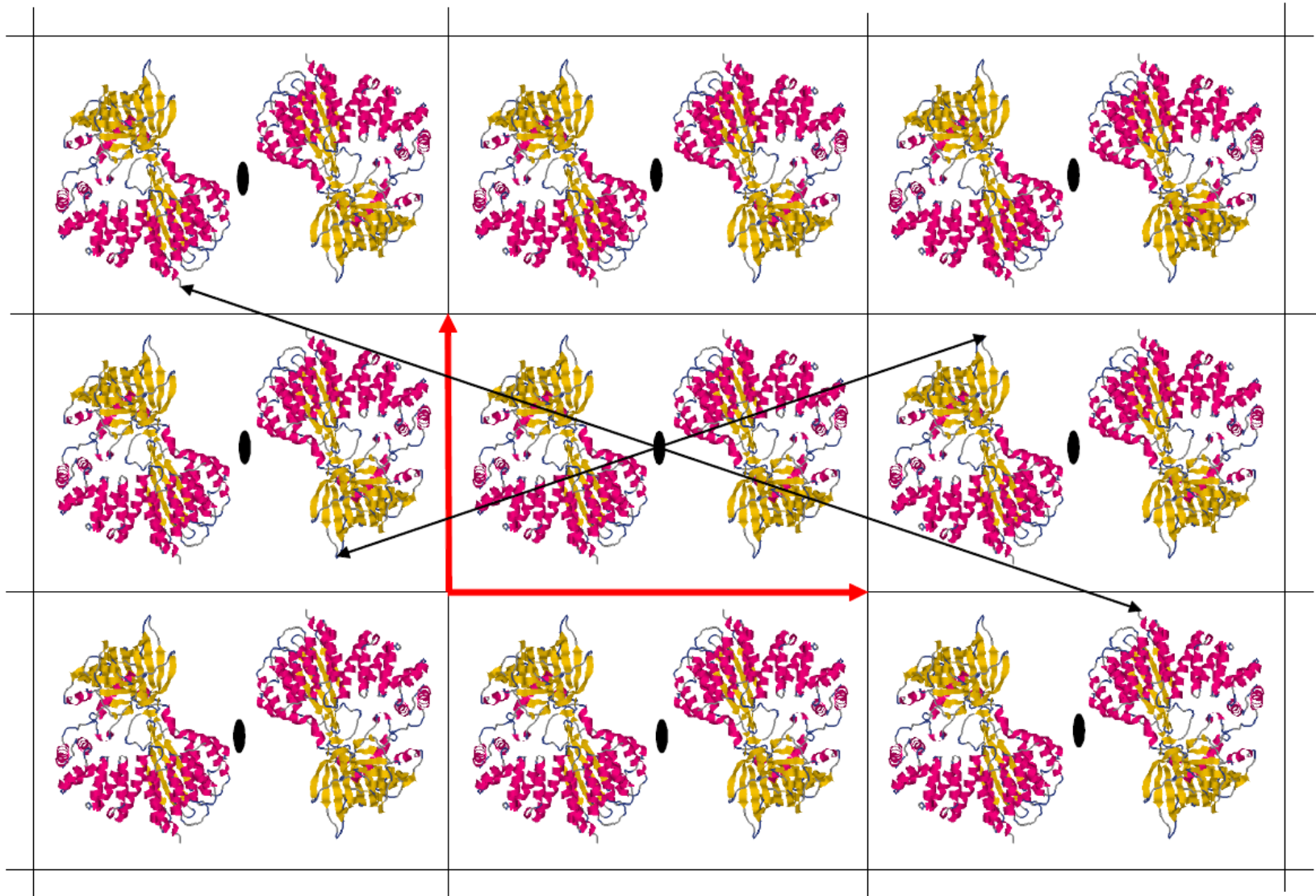
- Элементарная ячейка - математический объект. Она вводится для удобства работы.
- Выбор элементарной ячейки и начала координат в значительной мере произвольны.
- В начале координат может не находиться никакого атома.
- Молекула не всегда лежит в выбранной элементарной ячейке целиком.

Симметрия объекта

Симметрией в математике называется движение пространства, переводящее объект в себя



Кристаллографическая симметрия



Кристаллографическая симметрия “действует” для всех точек кристалла

Симметрия

Элементы симметрии:

- поворотная ось; 2, 3, 4, 6
 - винтовая ось; $2_1, 3_1, 3_2, 4_1, 4_2, 4_3, 6_1, 6_2, 6_3, 6_4, 6_5$
 - центр инверсии; $(x,y,z) \rightarrow (-x,-y,-z)$;
 - зеркальная плоскость; $(x,y,z) \rightarrow (x,y,-z)$;
 - плоскость скольжения.
- } Меняют ориентацию

Симметрия задается:

- матрицей вращения $\mathbf{R}$;
- вектором трансляции $\mathbf{t}$.

$$\mathbf{r}' = \mathbf{R}\mathbf{r} + \mathbf{t}$$

$$x' = r_{11}x + r_{12}y + r_{13}z + t_x$$

$$y' = r_{21}x + r_{22}y + r_{23}z + t_y$$

$$z' = r_{31}x + r_{32}y + r_{33}z + t_z$$

$\mathbf{r}'$ и $\mathbf{r}$ - точки, связанные симметрией

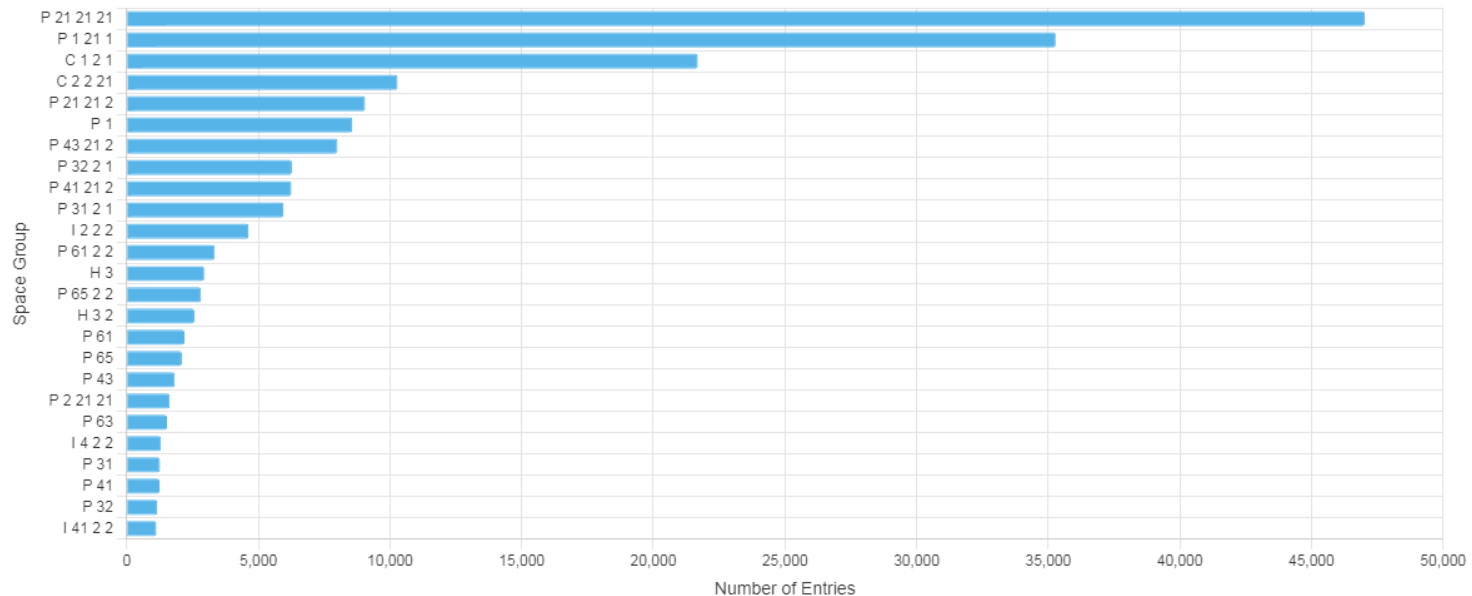
Кристаллографическая группа

Все симметрии конкретной кристаллической структуры образуют группу.

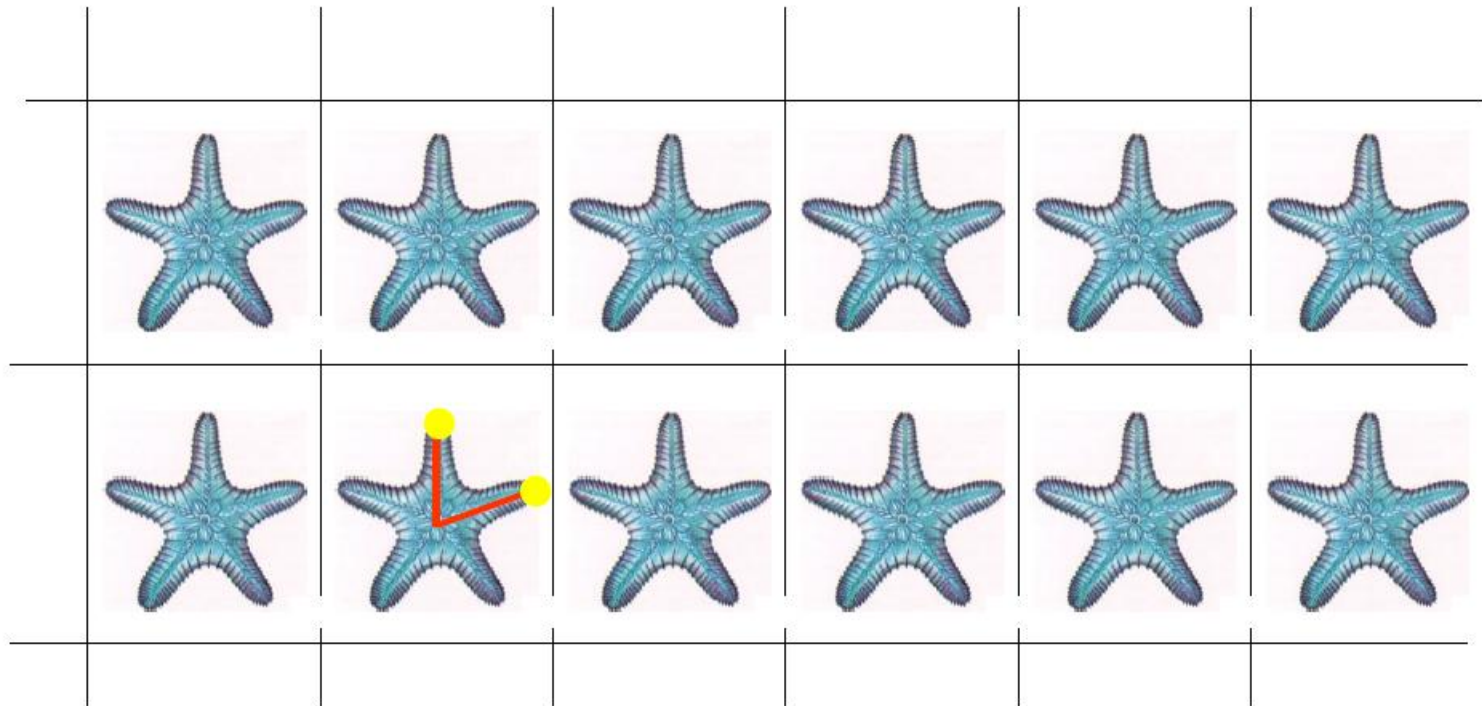
Существует конечное число (230) групп симметрии кристаллов.

Каждая группа имеет свое обозначение.

PDB Data Distribution by Space Group



Некристаллографическая симметрия

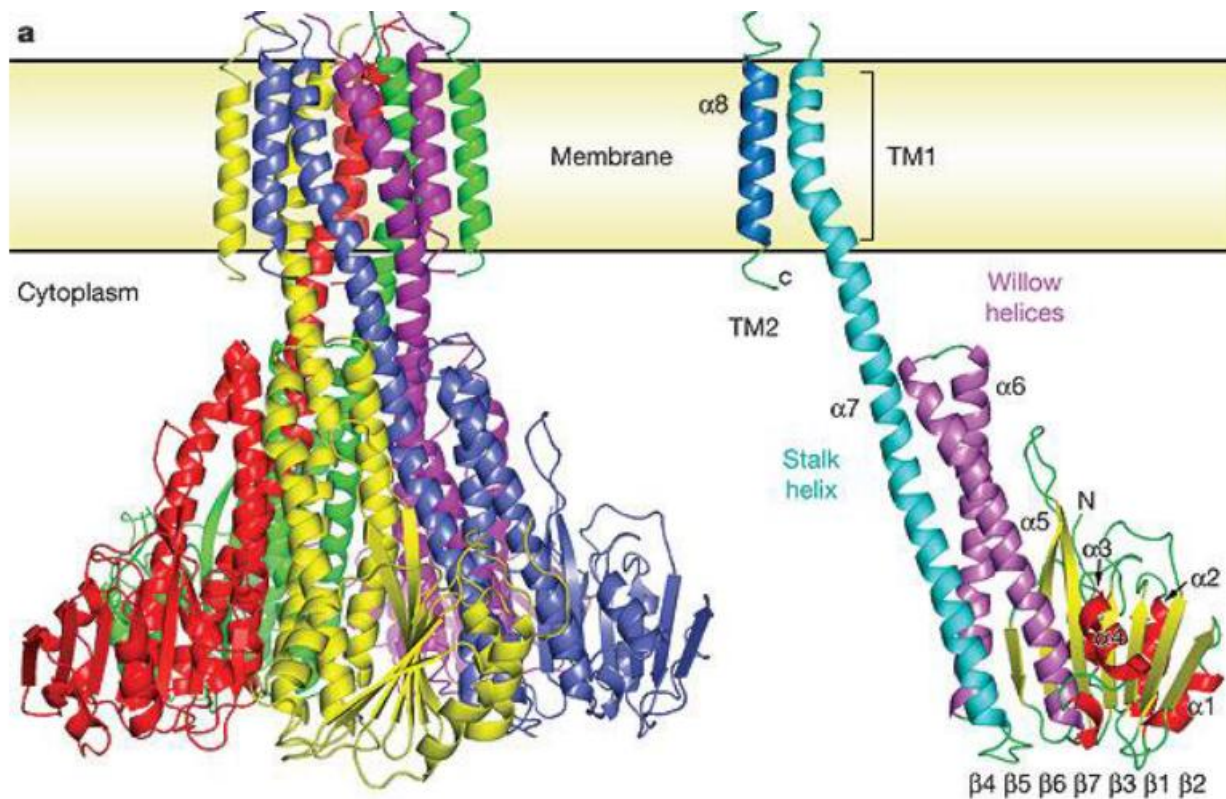


Некристаллографическая (локальная) симметрия имеет место только в ограниченной области пространства и не распространяется на весь кристалл

Элементарная ячейка и асимметрическая ячейка

Элементарная ячейка кристалла: минимальная область пространства, из которой весь кристалл можно получить **трансляционными** симметриями (сдвигами, параллельными переносами)

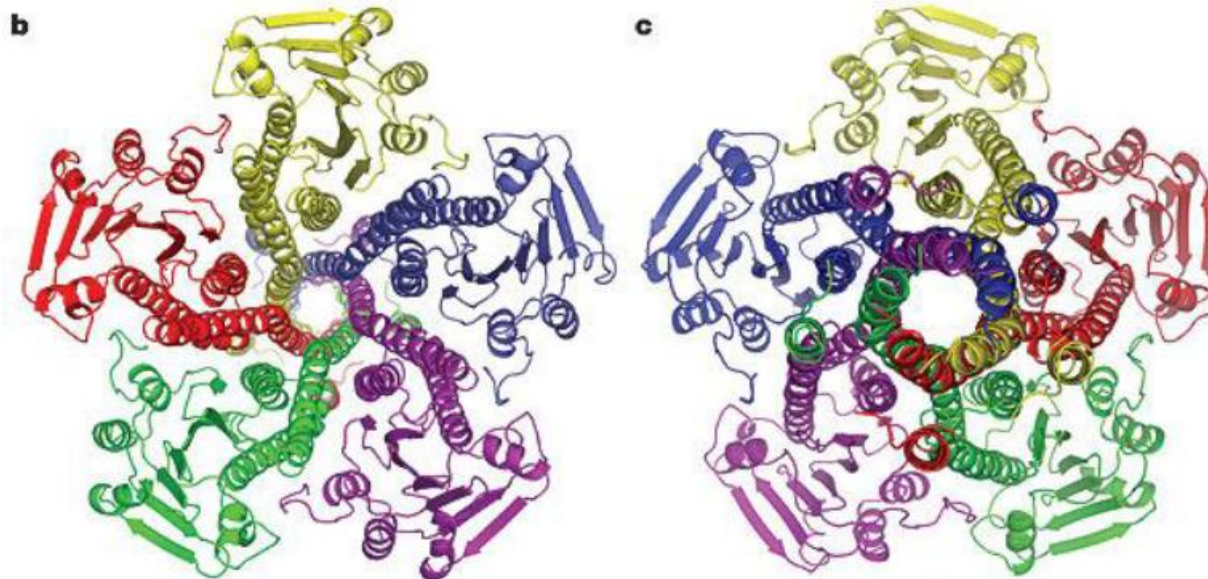
Асимметрическая ячейка: минимальная область пространства, из которой весь кристалл можно получить **всеми кристаллографическими** симметриями



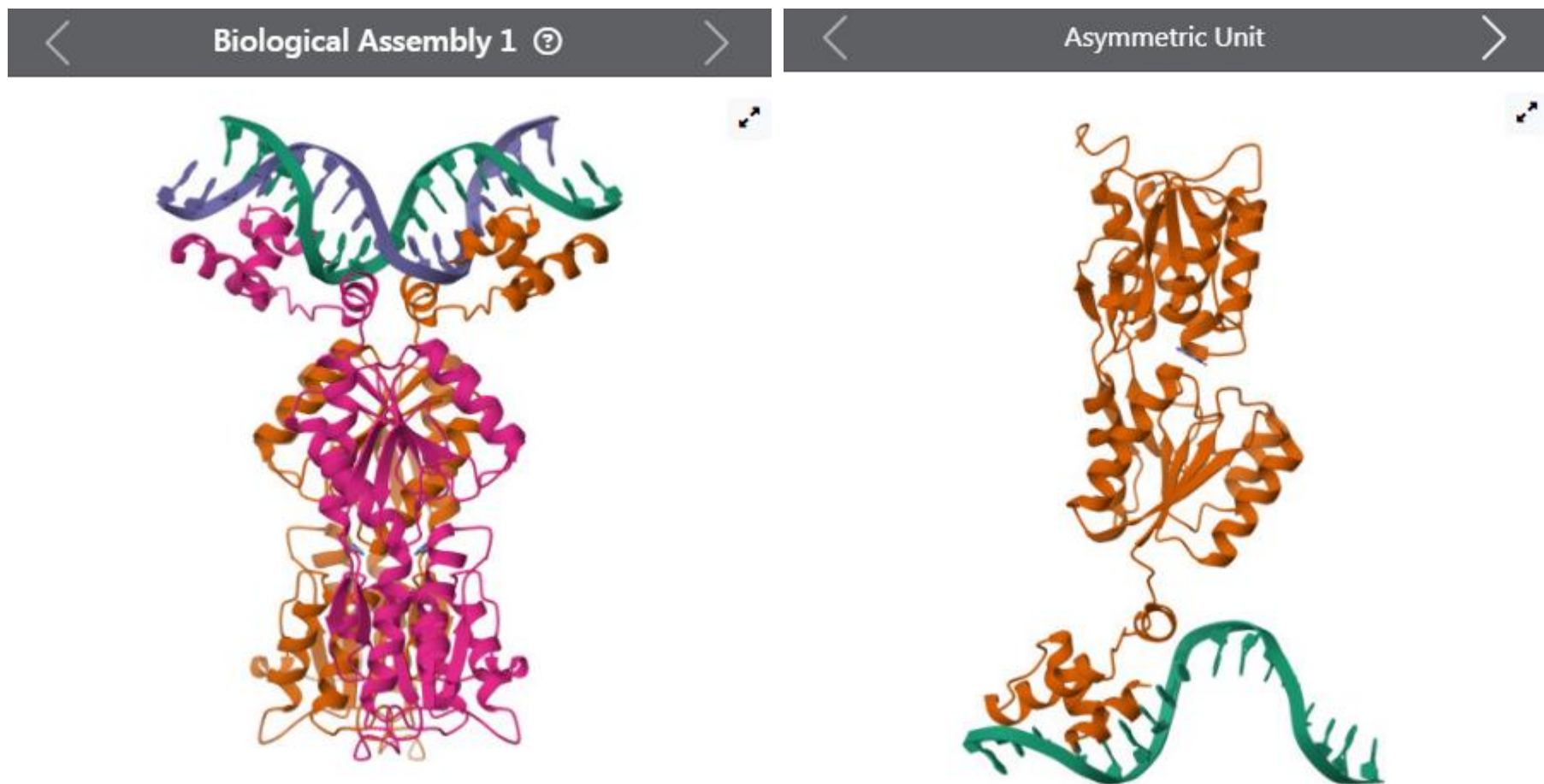
Structure of the
CorA Mg^{2+} channel.

From the following
article: [Crystal
structure of the CorA
 \$Mg^{2+}\$ transporter](#)

Nature 440, 833-837
(6 April 2006)



Что хранится в записи PDB



Файл с записью PDB содержит координаты атомов, попадающих в одну асимметрическую ячейку. Чтобы получить биологически осмысленный комплекс, часто нужно применить к этим координатам операторы симметрии.

Запись симметрий в cif файле

```
loop_  
_pdbx_struct_oper_list.id  
_pdbx_struct_oper_list.type  
_pdbx_struct_oper_list.name  
_pdbx_struct_oper_list.symmetry_operation  
_pdbx_struct_oper_list.matrix[1][1]  
_pdbx_struct_oper_list.matrix[1][2]  
_pdbx_struct_oper_list.matrix[1][3]  
_pdbx_struct_oper_list.vector[1]  
_pdbx_struct_oper_list.matrix[2][1]  
_pdbx_struct_oper_list.matrix[2][2]  
_pdbx_struct_oper_list.matrix[2][3]  
_pdbx_struct_oper_list.vector[2]  
_pdbx_struct_oper_list.matrix[3][1]  
_pdbx_struct_oper_list.matrix[3][2]  
_pdbx_struct_oper_list.matrix[3][3]  
_pdbx_struct_oper_list.vector[3]  
1 'identity operation' 1_555 x,y,z 1.0000000000 0.0000000000  
0.0000000000 0.0000000000 0.0000000000 1.0000000000  
0.0000000000 0.0000000000 0.0000000000 0.0000000000 1.0000000000  
0.0000000000  
2 'crystal symmetry operation' 4_555 x,-y,-z 1.0000000000 0.0000000000  
0.0000000000 0.0000000000 0.0000000000 -1.0000000000  
0.0000000000 0.0000000000 0.0000000000 0.0000000000 -1.0000000000  
0.0000000000  
#
```

Операторы симметрии
(матрица 3×3 и вектор сдвига)



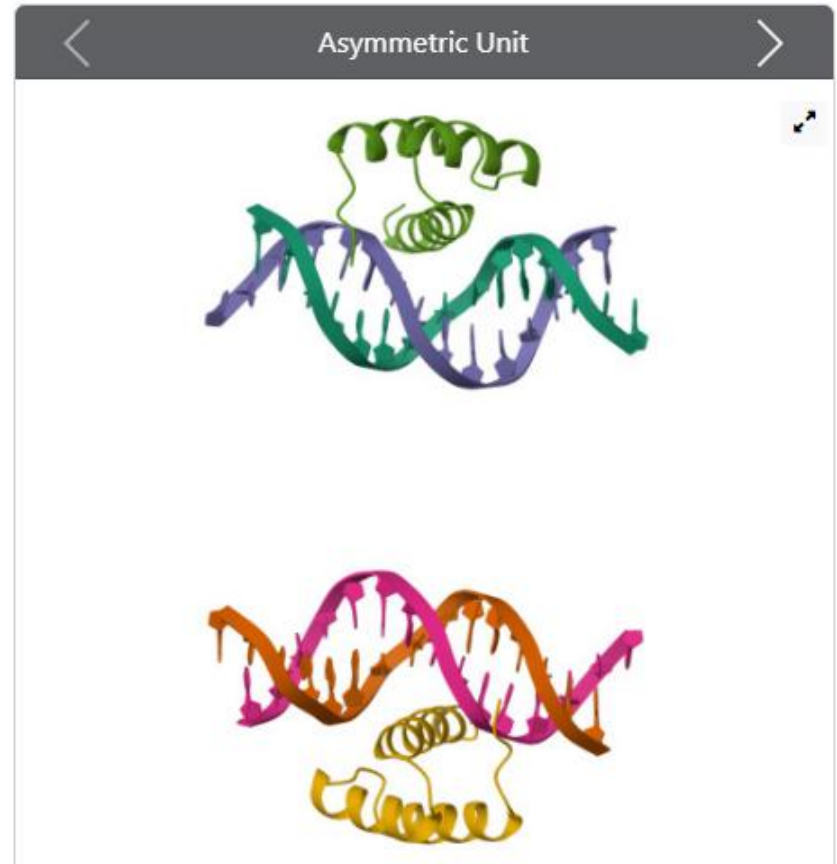
Биологическая единица

Биологическая единица (biological unit, biological assembly) — часть кристалла, которая (как считают авторы модели) соответствует комплексу, существующему в природе.

В большинстве случаев асимметрическая ячейка включает одну биологическую единицу, но часто, чтоб получить биологическую единицу, к асимметрической ячейке нужно применить набор операторов, указанных в файле с записью.

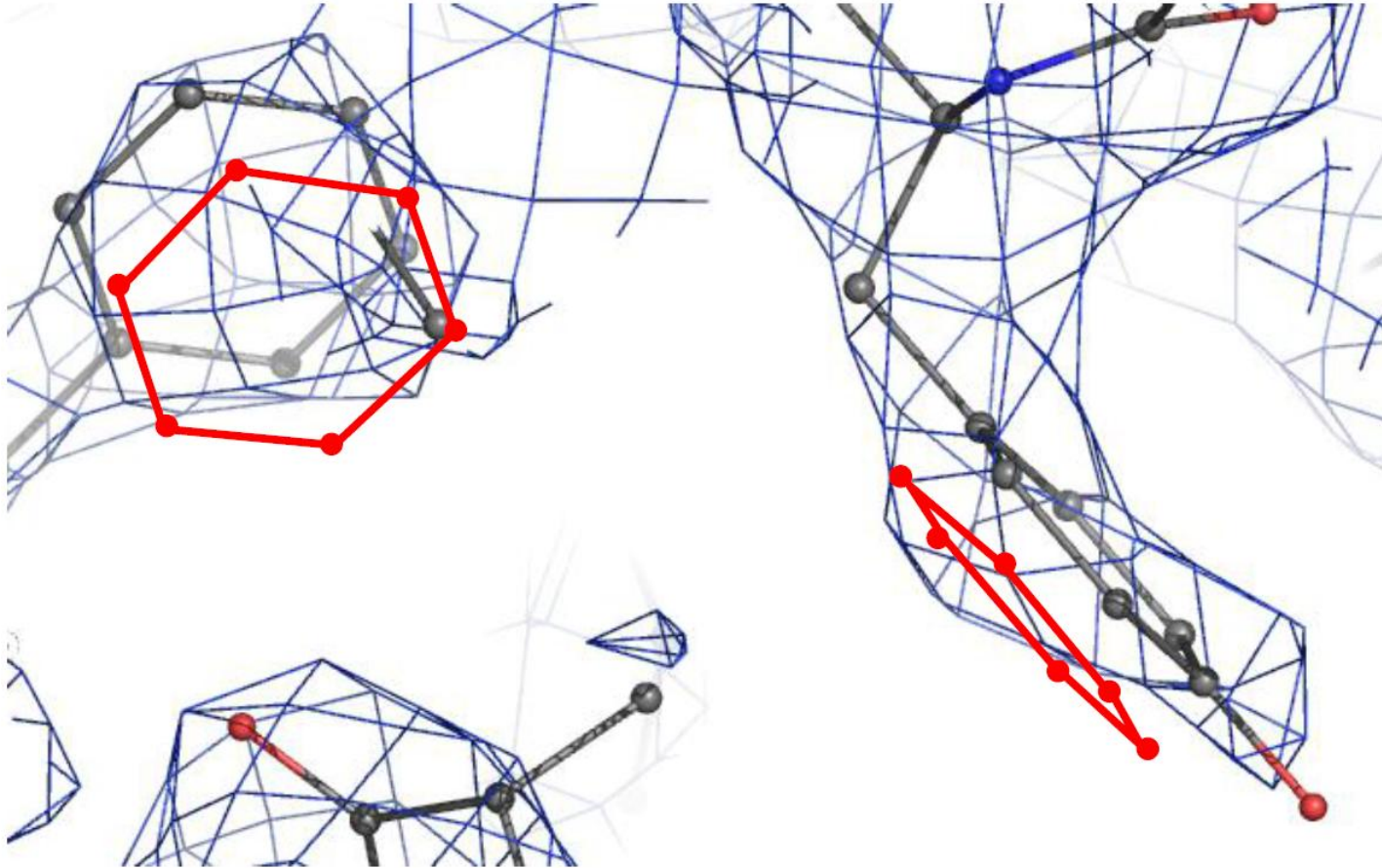
Изредка асимметрическая ячейка включает больше одной биологической единицы.

Когда биологическая единица меньше асимметрической



Так бывает, если белок (или, как в данном случае, комплекс) закристаллизовался парами с немного разной конформацией. Но гораздо чаще такое встречается в записях, авторы которых в погоне за хорошим R-фактором решили увеличить число параметров для подгонки...

Оптимизация модели



Черновая модель получается «вписыванием» известной последовательности белка в полученную в эксперименте электронную плотность. Следующий этап — уточнение модели.

Уточнение параметров модели

$\{x_j, y_j, z_j, B_j, T_j\}$ - параметры модели

$F_{hkl}^{calc}(\{x_j, y_j, z_j, B_j, T_j\})$ - рассчитанные по модели модули структурных факторов

F_{hkl}^{obs} - экспериментально определенные модули структурных факторов

Хотим иметь

$$F_{hkl}^{calc}(\{x_j, y_j, z_j, B_j, T_j\}) = F_{hkl}^{obs} \quad \text{для всех } hkl$$

Стандартный фактор достоверности / R-factor

$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Хотим иметь

$$F_{hkl}^{calc} \left(\{x_j, y_j, z_j, B_j, T_j\} \right) = F_{hkl}^{obs} \quad \text{для всех } hkl$$

Важное соотношение: $\kappa = \frac{\text{число уравнений}}{\text{общее число уточняемых параметров}}$

Число уравнений определяется качеством кристалла (разрешением собранного набора данных). Число параметров модели можем пытаться менять.

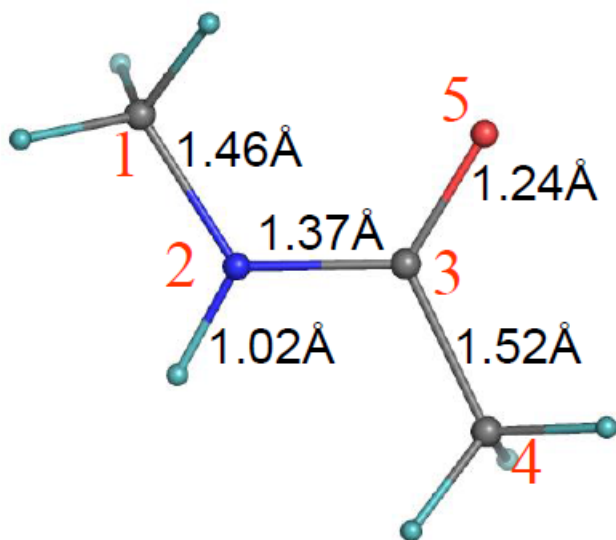
Уточнение, как задача минимизации:

$$R_{X\text{-ray}} = \sum_{hkl} \left(F_{hkl}^{calc} - F_{hkl}^{obs} \right)^2 \Rightarrow \min$$

Проблемы:

- много локальных максимумов; возможно только локальное уточнение; радиус сходимости $\sim 0.7\text{\AA}$;
- модель "рассыпается, если κ мало.

Увеличение числа уравнений: стереохимические ограничения.



$\{x_j, y_j, z_j, B_j, T_j\}$ - параметры модели

d_{ij}^{calc} - расстояние между i -ым и j -ым атомами в модели

d_{ij}^{exact} - идеальное расстояние между i -ым и j -ым атомами в модели

Хотим иметь:

$$d_{12}^{calc} \left(\{x_j, y_j, z_j, B_j, T_j\} \right) = 1.46$$

$$d_{23}^{calc} \left(\{x_j, y_j, z_j, B_j, T_j\} \right) = 1.37$$

. . . .

Можем сформулировать как задачу минимизации

$$R_{dist} = \sum_{ij} \left(d_{ij}^{calc} - d_{ij}^{exact} \right)^2 \Rightarrow \min$$

$$R_{X-ray} = \sum_{hkl} (F_{hkl}^{calc} - F_{hkl}^{obs})^2 \Rightarrow \min$$

$$R_{dist} = \sum_{ij} (d_{ij}^{calc} - d_{ij}^{exact})^2 \Rightarrow \min$$

Составной критерий

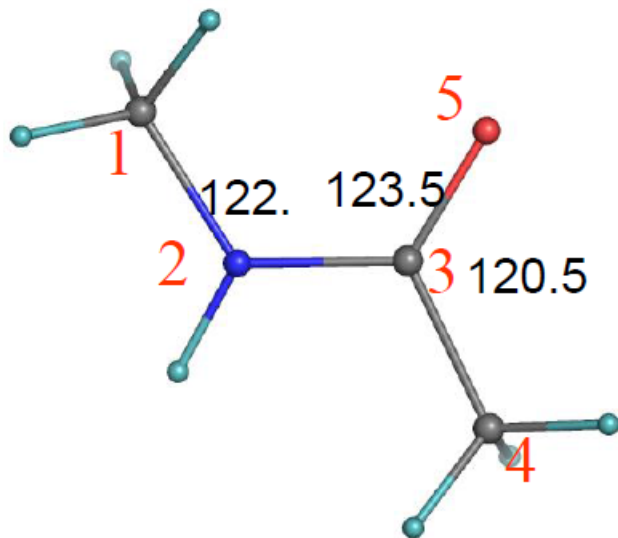
$$R = w_{X-ray} R_{X-ray} + w_{dist} R_{dist}$$

$$= w_{X-ray} \sum_{hkl} (F_{hkl}^{calc} - F_{hkl}^{obs})^2 + w_{dist} \sum_{ij} (d_{ij}^{calc} - d_{ij}^{exact})^2 \Rightarrow \min$$



А как выбрать веса w_{X-ray} , w_{dist} ?

Стереохимические ограничения



$\{x_j, y_j, z_j, B_j, T_j\}$ - параметры модели

α_{ijk}^{calc} - угол между i, j, k -ыми атомами в модели

α_{ijk}^{exact} - идеальный угол между i, j, k -ыми атомами в модели

Хотим иметь

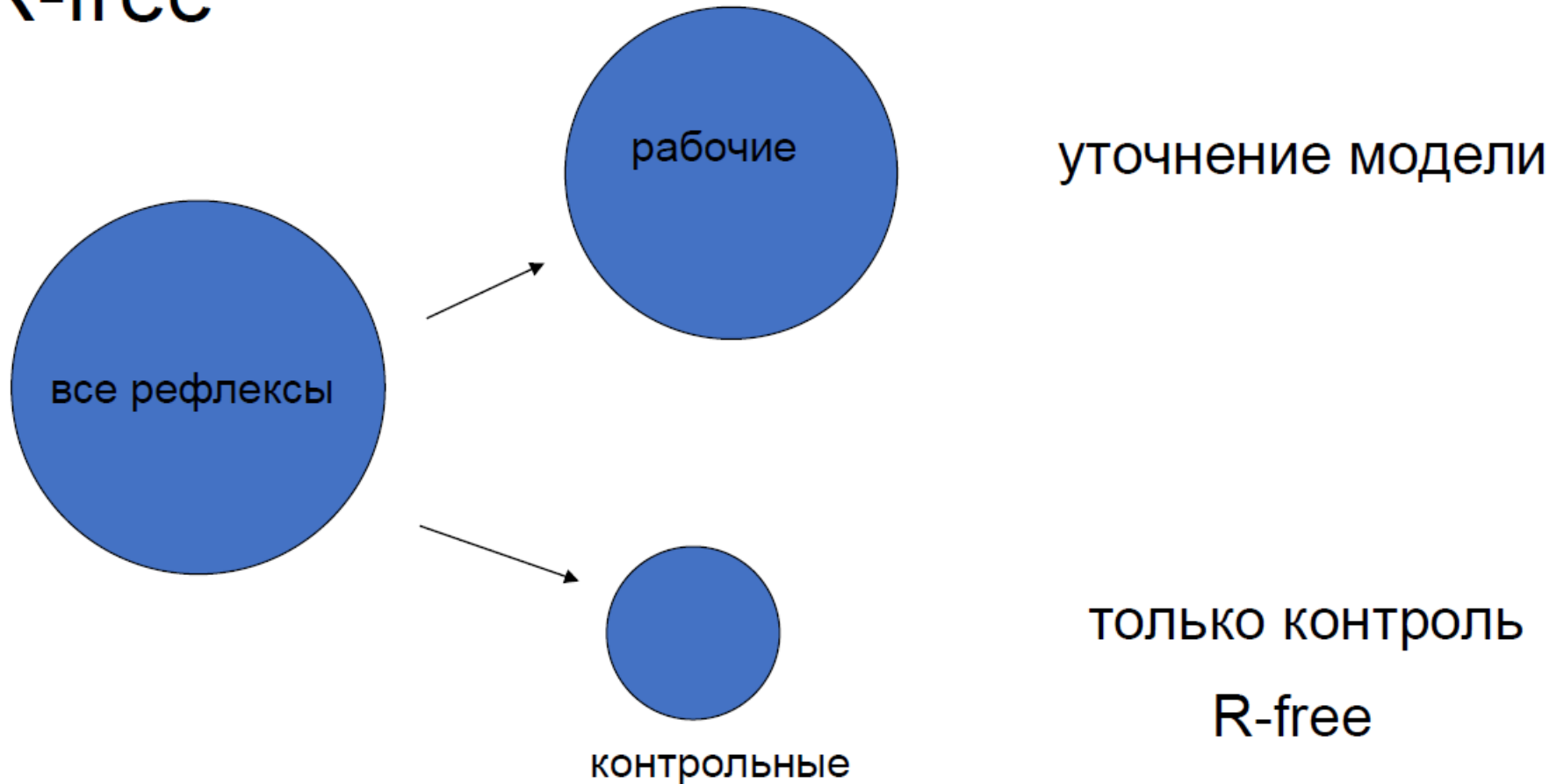
$$(\alpha_{123}^{calc} - 122.)^2 + (\alpha_{235}^{calc} - 123.5)^2 + (\alpha_{534}^{calc} - 120.5)^2 \Rightarrow \min$$

$$R_{angle} = \sum_{ijk} (\alpha_{ijk}^{calc} - \alpha_{ijk}^{exact})^2 \Rightarrow \min$$

$$R = w_{X-ray} R_{X-ray} + w_{dist} R_{dist} + w_{angle} R_{angle}$$

Снижение величины минимизируемого критерия может не сопровождаться улучшением параметров модели.

R-free

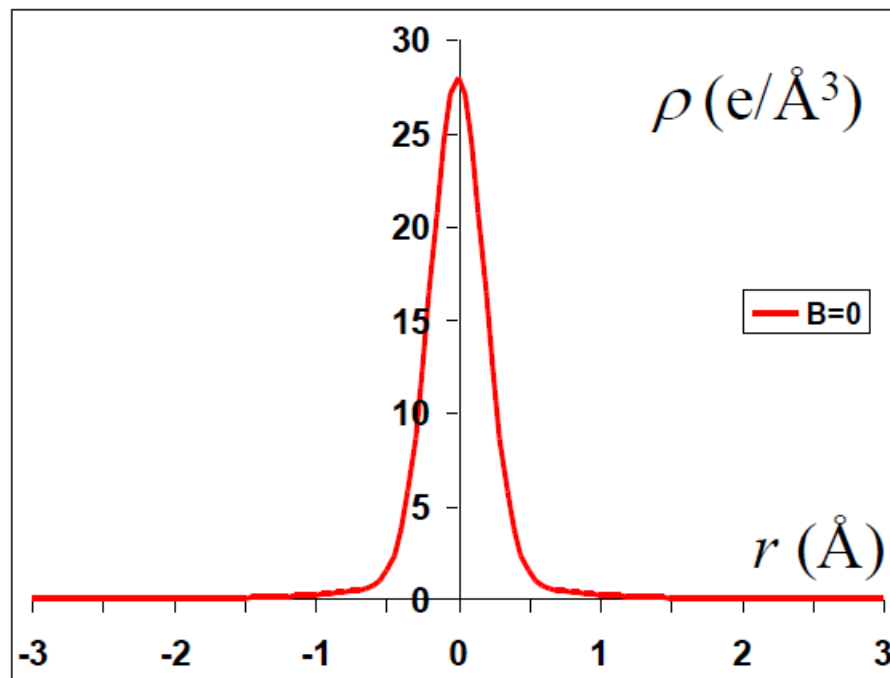


Динамическая и статическая гетерогенность (Dynamic and static disorder)

- Атомы в процессе эксперимента находятся в движении (тепловые колебания). *Динамическая неопределенность.*
- Разные копии молекулы в кристалле могут иметь отличия в координатах атомов. *Статическая неопределенность.*
- Эксперимент дает информацию об усредненном по времени и пространству распределении электронной плотности в молекуле.
- Параметр неопределённости положения атома (Atomic Displacement Parameter, ADP) характеризует его динамическую и статическую неопределенность.
- Прежнее (историческое) название ADP – температурный фактор (temperature factor, B-value)

Температурный фактор (temperature factor / Atomic Displacement Parameter / ADP)

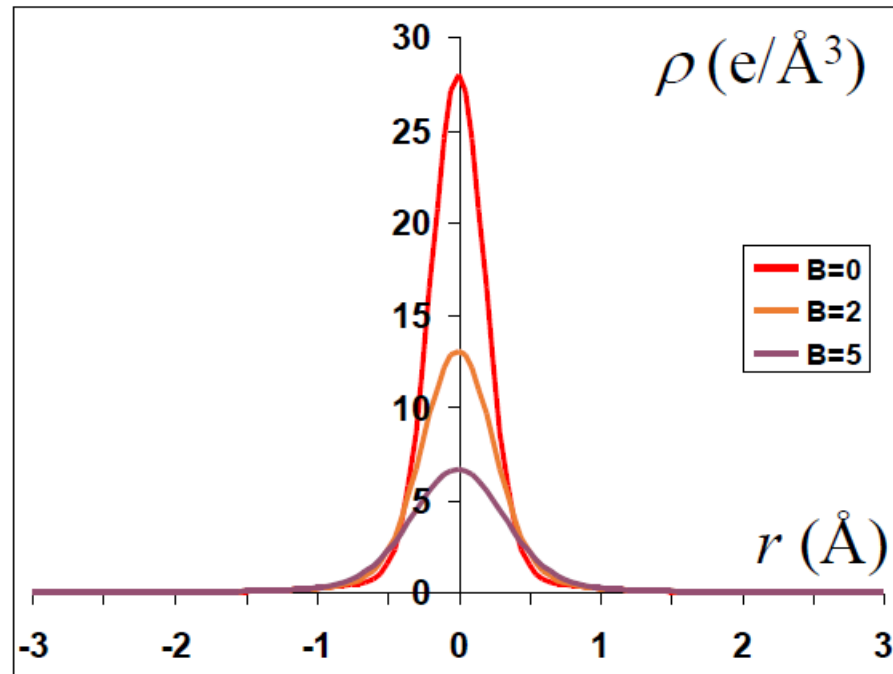
Динамическая и
статическая
неупорядоченность



$$\rho(\mathbf{r}) = \sum_{j=1}^5 C_j \left(\frac{4\pi}{B_j} \right)^{3/2} \exp \left(- \frac{4\pi^2 |\mathbf{r}|^2}{B_j} \right)$$

Движение атома в процессе эксперимента
«размазывает» распределение электронной плотности.

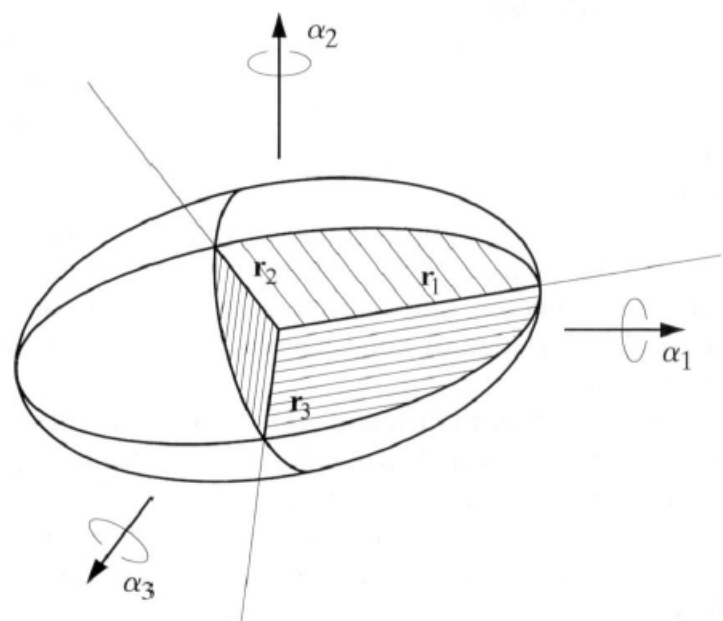
Радиальное распределение электронной плотности в атоме углерода при разных значениях температурного фактора B.



«Расплывание» плотности моделируется увеличением ширины гауссовых пиков.

$$\rho(\mathbf{r}) = \sum_{j=1}^5 C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp \left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

Анизотропный температурный фактор



Введение анизотропных температурных факторов увеличивает число параметров модели до 9 на каждый атом. Применяется при работе с данными высокого разрешения.

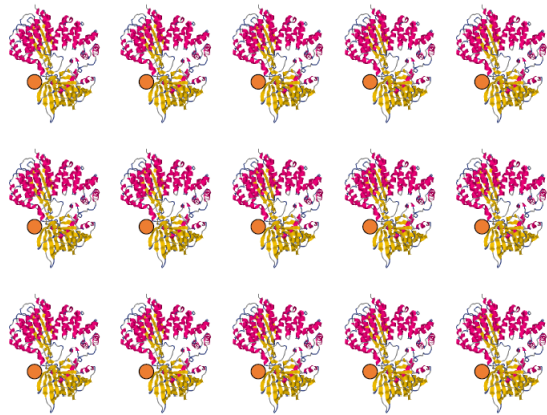
Fig. 18.4.1.1. The thermal-ellipsoid model used to represent anisotropic atomic displacement, with major axes indicated. The ellipsoid is drawn with a specified probability of finding an atom inside its contour. Six parameters are necessary to describe the ellipsoid: three represent the dimensions of the major axes and three the orientation of these axes. These six parameters are expressed in terms of a symmetric U tensor and contribute to atomic scattering through the term $\exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{12}hka^*b^* \cos \gamma^* + 2U_{13}hla^*c^* \cos \beta^* + 2U_{23}klb^*c^* \cos \alpha^*)]$.

Изотропное значение

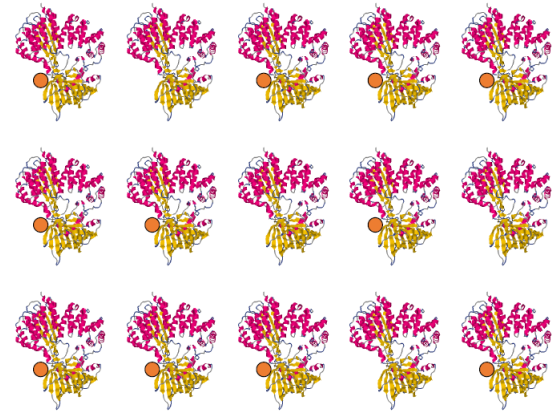
ATOM	30	N	SER	2	13.117	9.840	39.210	1.000	12.49	
ANISOU	30	N	SER	2	1510	2105	1130	447	-393	-1019

Поправки, умноженные на 1000.

Коэффициент заполнения / заселенность / occupancy



Идеальный кристалл - содержимое всех элементарных ячеек идентично.



"Реальный" кристалл - молекула "воды" присутствует не во всех элементарных ячейках.

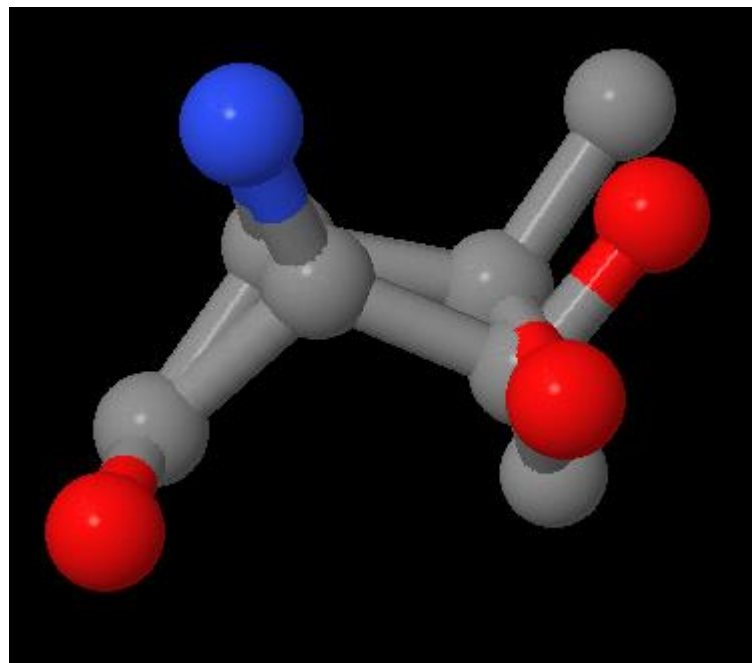
Для каждого атома коэффициент заполнения показывает, какой процент элементарных ячеек кристалла содержат атом в указанной позиции.

Атомы в формате cif

Diagram illustrating the mapping of PDB atom site fields to the output table columns:

- Коэф. заполнения (Occupancy):**
 - Field: `_atom_site.occupancy` maps to the value `1.00`.
- B-фактор (B-factor):**
 - Field: `_atom_site.B_iso_or_equiv` maps to the value `38.41`.

Альтернативные конформации



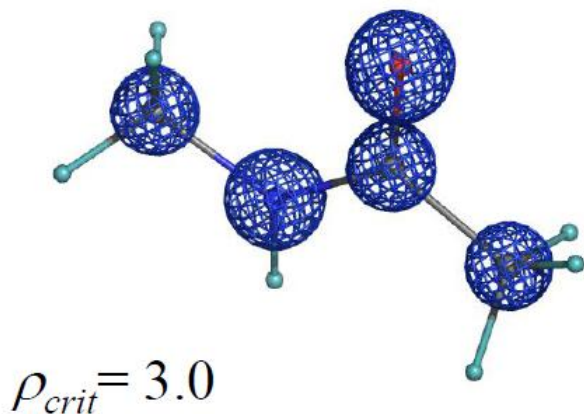
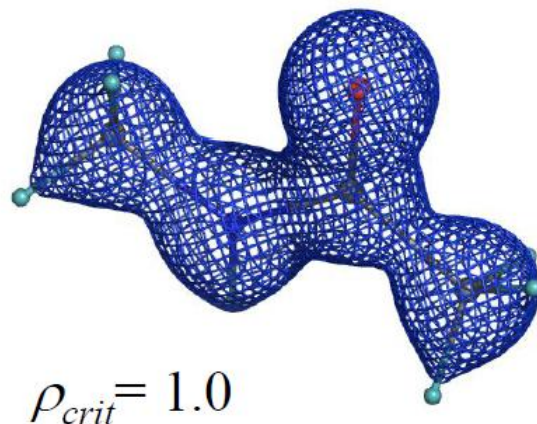
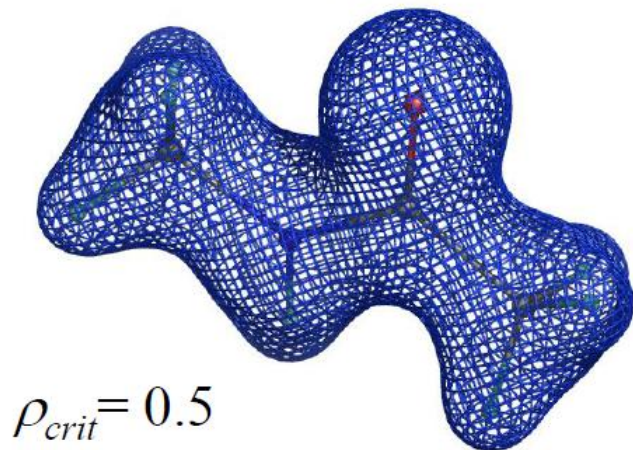
Коэф. заполнения

ATOM	21	C	CB	A	THR	A	1	2	?	12.784	10.610	5.099	0.70	4.35	?	2	THR	A	CB	1
ATOM	22	C	CB	B	THR	A	1	2	?	12.938	9.905	5.108	0.30	1.07	?	2	THR	A	CB	1
ATOM	23	O	OG1	A	THR	A	1	2	?	13.364	9.360	4.756	0.70	5.60	?	2	THR	A	OG1	1
ATOM	24	O	OG1	B	THR	A	1	2	?	12.559	10.450	3.917	0.30	5.43	?	2	THR	A	OG1	1
ATOM	25	C	CG2	A	THR	A	1	2	?	12.451	11.323	3.797	0.70	7.65	?	2	THR	A	CG2	1
ATOM	26	C	CG2	B	THR	A	1	2	?	11.642	9.729	5.862	0.30	2.71	?	2	THR	A	CG2	1

Все атомы этой боковой цепи имеют по две альтернативных конформации: А и В

Как визуализировать электронную плотность?

$\rho(x, y, z) = \rho_{crit}$ - поверхность



ρ_{crit} — «уровень подрезки»

Как визуализировать электронную плотность?

σ – шкала (σ -scale, z-score)

$$\langle \rho \rangle = \frac{1}{|V|} \int_V \rho(\mathbf{r}) dV_{\mathbf{r}} \quad - \text{средняя электронная плотность в объекте}$$

$$\sigma = \sqrt{\frac{1}{|V|} \int_V (\rho(\mathbf{r}) - \langle \rho \rangle)^2 dV_{\mathbf{r}}} \quad \sigma - \text{среднеквадратичное отклонение плотности (от среднего), r.m.s.d;}$$

$$z_{crit} = \frac{\rho_{crit} - \langle \rho \rangle}{\sigma} \quad - \text{уровень срезки в "сигмах"}$$

Комбинированный синтез Фурье

$F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - синтез Фурье

$w_{hkl} F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - взвешенный синтез Фурье (чем надежнее определена фаза, тем больше вес w_{hkl})

$2F_{hkl}^{obs} - F^{calc}, \varphi_{hkl}^{calc}$ - комбинированный синтез Фурье

$2mF_{hkl}^{obs} - DF^{calc}, \varphi_{hkl}^{calc}$ - взвешенный комбинированный синтез Фурье
(σ_A -weighted $2F^o - F^c$ synthesis)

Комбинированный синтез Фурье

Final σ_A -weighted $2|F_o| - |F_c|$ electron density (Read, 1986) for the GSH molecule bound to subunit A of DmGSTe6 contoured at 1σ .

