

# Введение в биологическую кристаллографию

## Лекция 3

Владимир Юрьевич ЛУНИН  
(lunin@impb.ru)

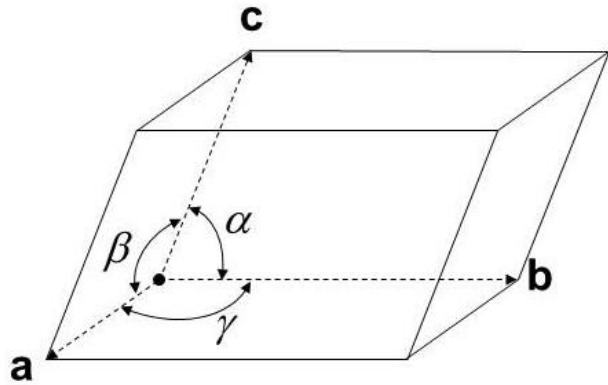
Институт Математических Проблем Биологии РАН

(филиал Института Прикладной Математики  
им. М.В. Келдыша РАН)

Пущино

# Ряды Фурье

$$\rho(\mathbf{r}), \quad \mathbf{r} \in V$$



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

элементарная ячейка

$$0 \leq x, y, z \leq 1$$

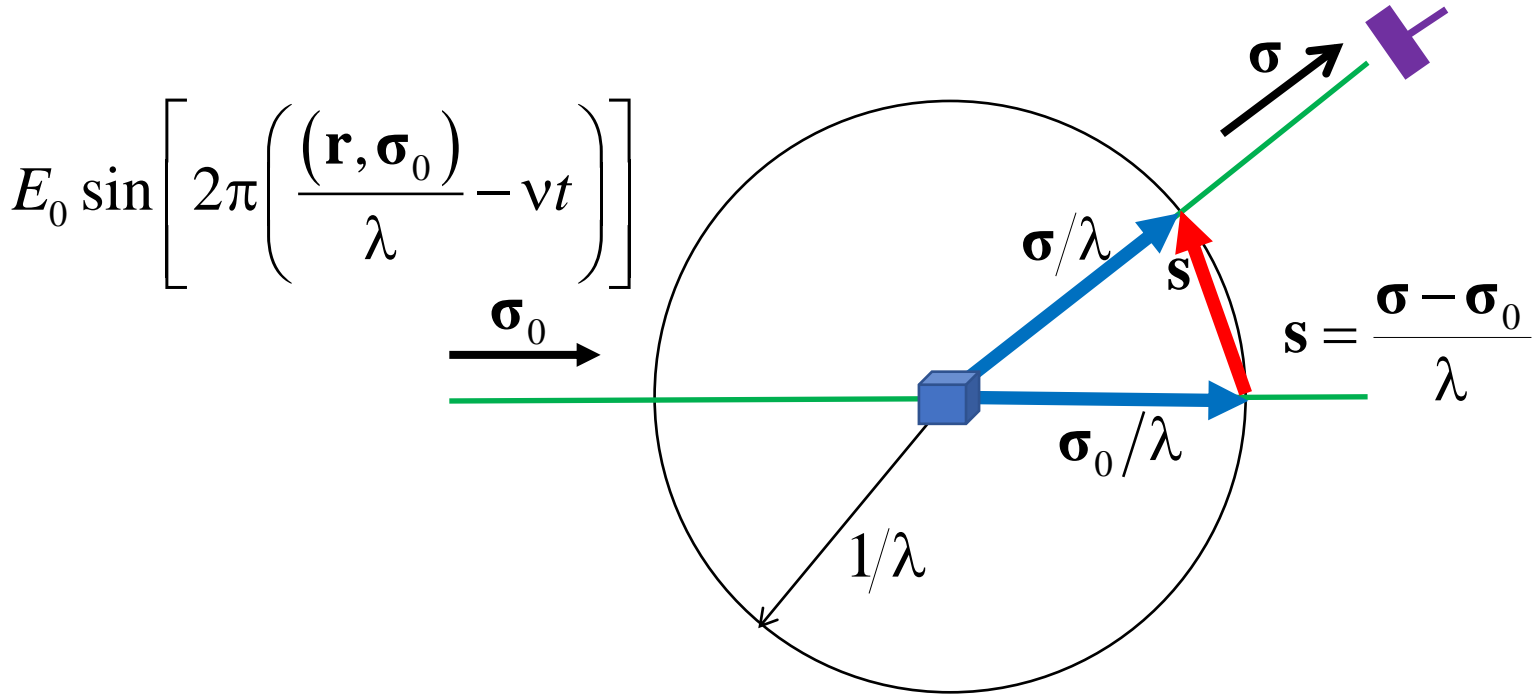
$$\tilde{\rho}(x, y, z) \approx F_{000} + \sum_{\substack{hkl \\ \text{целые}}} F_{hkl} \cos \left[ 2\pi (hx + ky + lz) - \varphi_{hkl} \right]$$

$F_{hkl}$ ,  $\varphi_{hkl}$  – модуль и фаза коэффициента Фурье

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

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

$$\frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[ 2\pi \left( \frac{R_{s-d}}{\lambda} - vt \right) - \varphi(\mathbf{s}) \right]$$



$F(\mathbf{s}), \varphi(\mathbf{s})$  – модуль и фаза структурного фактора

$$F(\mathbf{s}) \cos \phi(\mathbf{s}) = \int \rho(\mathbf{r}) \cos 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$

$$F(\mathbf{s}) \sin \phi(\mathbf{s}) = \int \rho(\mathbf{r}) \sin 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$

Условия дифракции

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

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

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

$h, k, l$  - целые числа

# Прямое пространство direct space

**элементы:**

точки физического пространства ;

**базис:**  $\mathbf{a}, \mathbf{b}, \mathbf{c}$

**координаты:**  $x, y, z$

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

$$(\mathbf{r}, \mathbf{s}) = hx + ky + lz$$

# Обратное пространство reciprocal space

**элементы:**

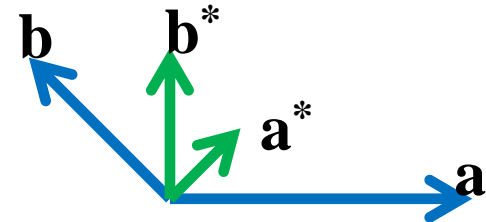
индексы коэффициентов Фурье;  
вектора рассеяния

**базис:**  $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$

$$(\mathbf{a}^*, \mathbf{a}) = 1, (\mathbf{a}^*, \mathbf{b}) = 0, (\mathbf{a}^*, \mathbf{c}) = 0$$

$$(\mathbf{b}^*, \mathbf{a}) = 0, (\mathbf{b}^*, \mathbf{b}) = 1, (\mathbf{b}^*, \mathbf{c}) = 0$$

$$(\mathbf{c}^*, \mathbf{a}) = 0, (\mathbf{c}^*, \mathbf{b}) = 0, (\mathbf{c}^*, \mathbf{c}) = 1$$

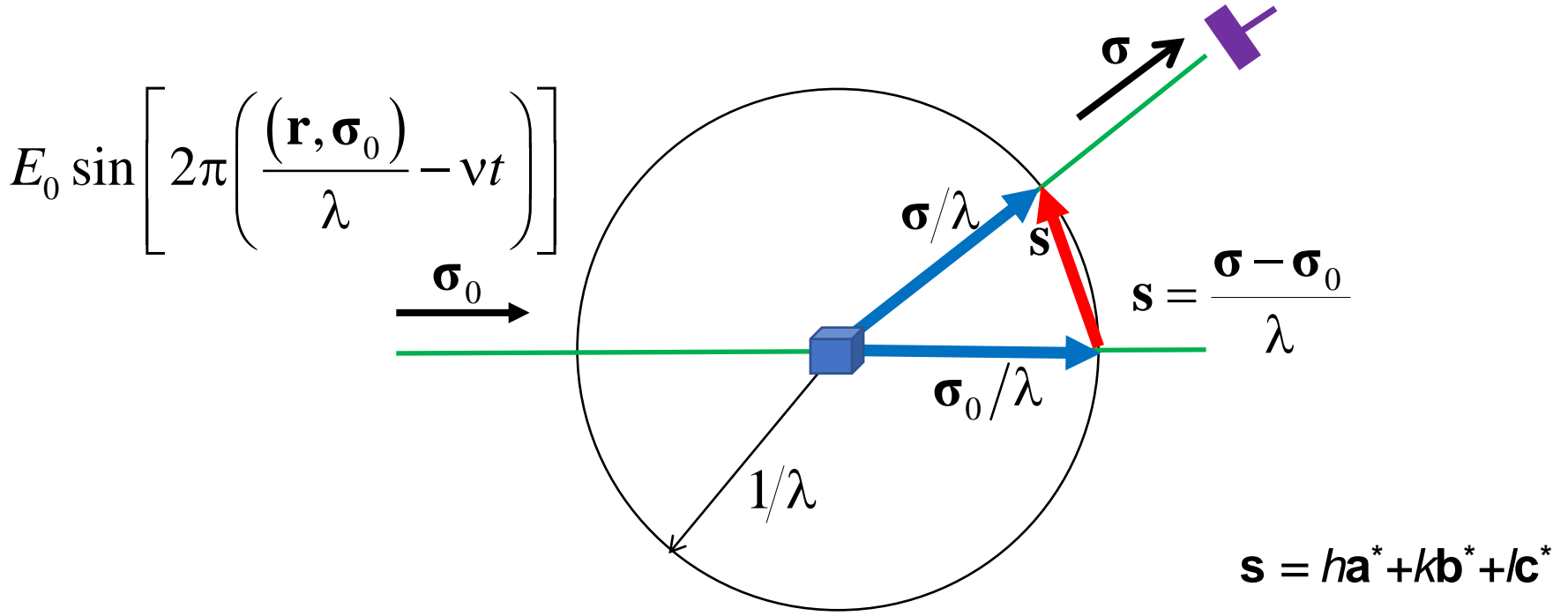


**координаты:**  $h, k, l$

$$\mathbf{s} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$$

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

$$\frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[ 2\pi \left( \frac{R_{s-d}}{\lambda} - vt \right) - \varphi(\mathbf{s}) \right]$$

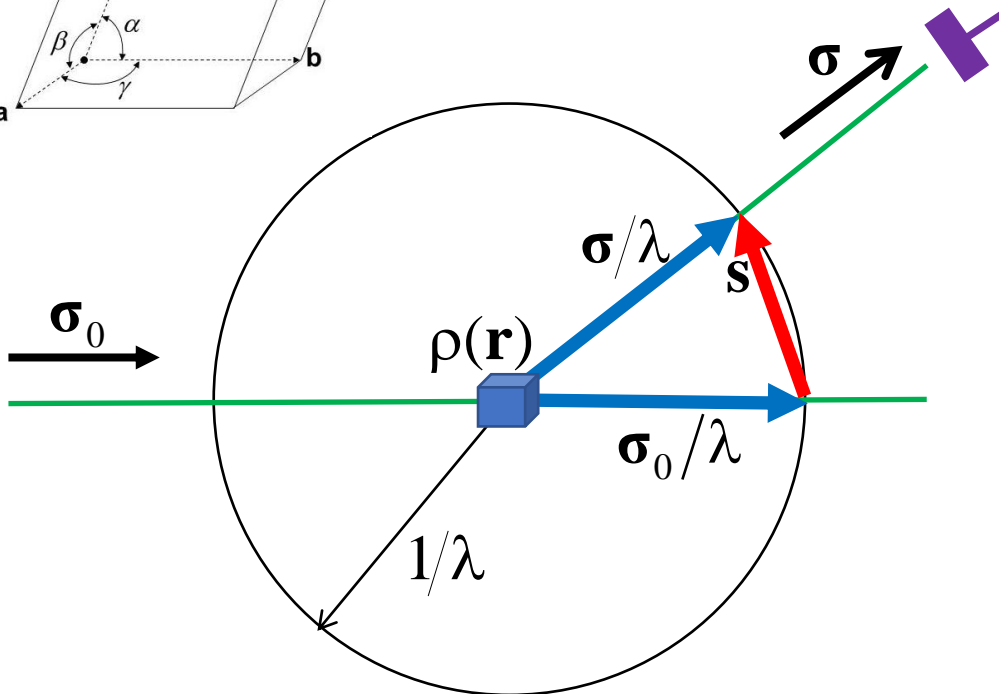
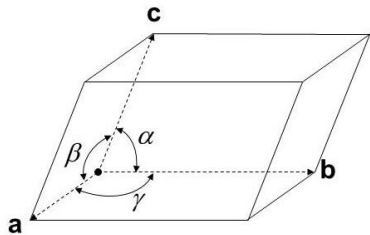


$F_{hkl}$ ,  $\varphi_{hkl}$  – модуль и фаза структурного фактора

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

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

Условия дифракции:  $h, k, l$  - целые числа



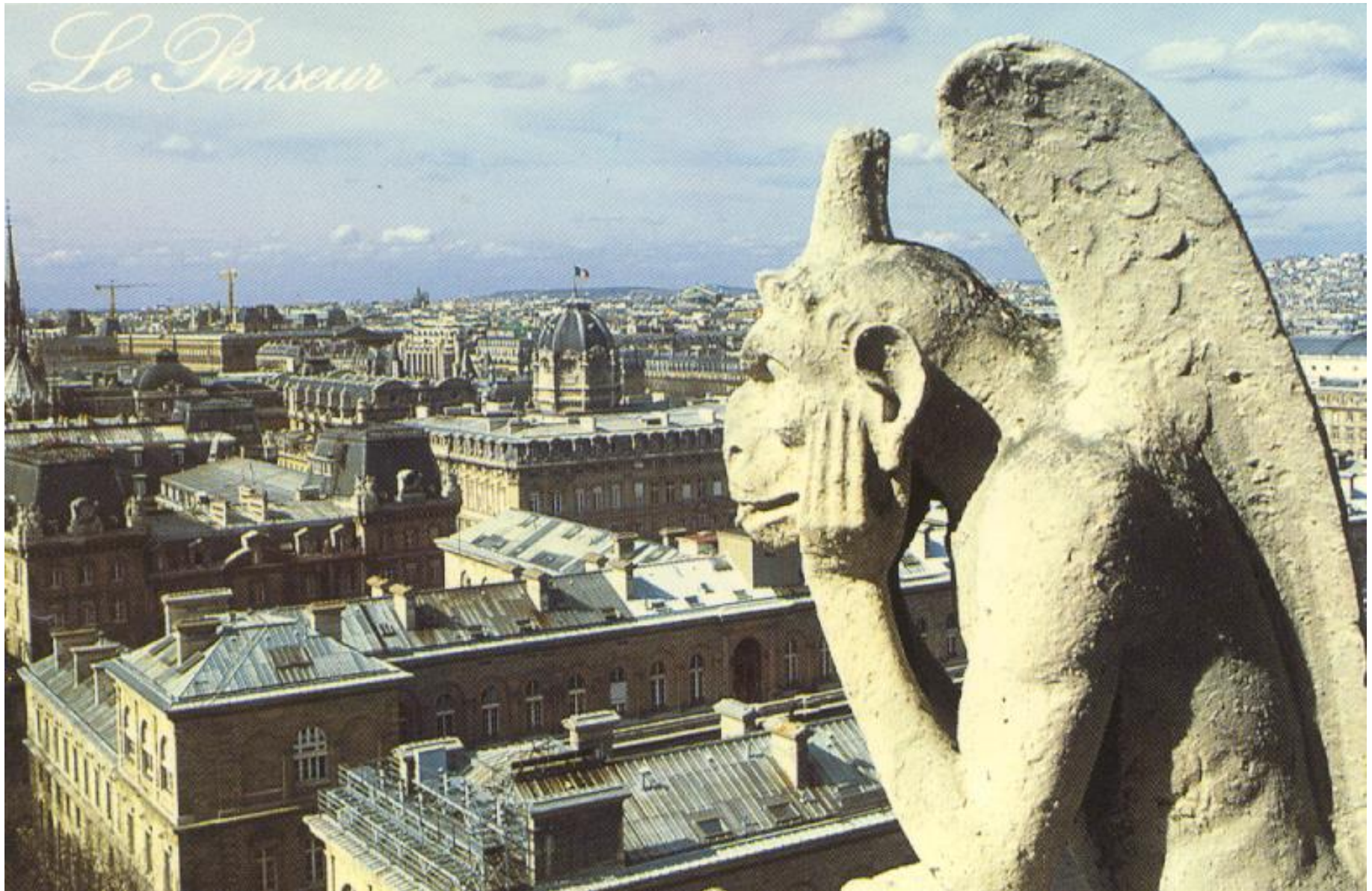
модуль и фаза  
коэффициента Фурье  
 $F_{hkl}, \varphi_{hkl}$

ИНДЕКСЫ  
 $h=(s,a), k=(s,b), l=(s,c)$

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

**Ограничения:**  $|\mathbf{s}| \leq \frac{2}{\lambda}$  , фазы теряем

# Фазовая проблема



# Синтез Фурье электронной плотности

Если фазы найдены, то

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

$S$  – набор рефлексов.

*Разрешение  $d$ ,*  
соответствующее  
гармонике Фурье

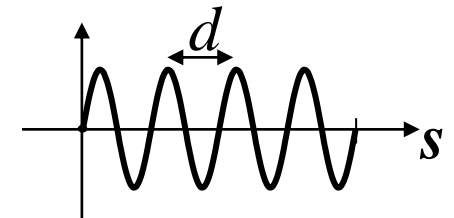
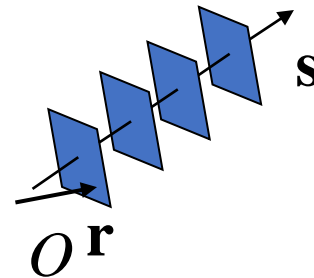
$$\mathbf{s} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$$

*Разрешение* рефлекса  $hkl$

*Разрешение* структурного фактора

*Разрешение* набора  $S$  рефлексов (синтеза Фурье)

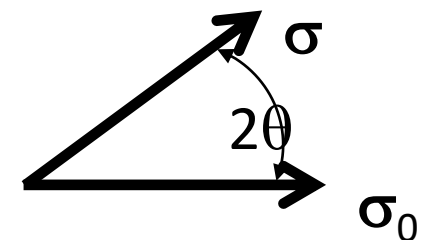
$$H_s(\mathbf{r}) = \cos[2\pi(hx + ky + lz)]$$



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

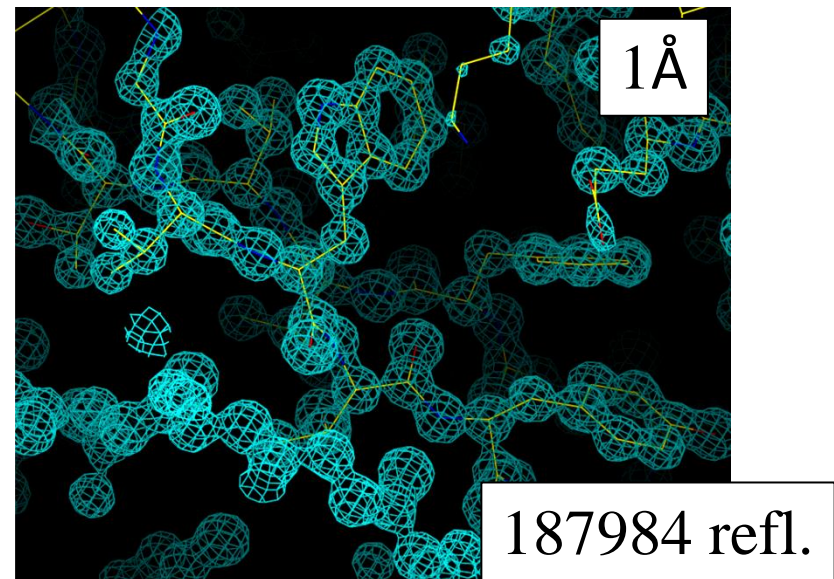
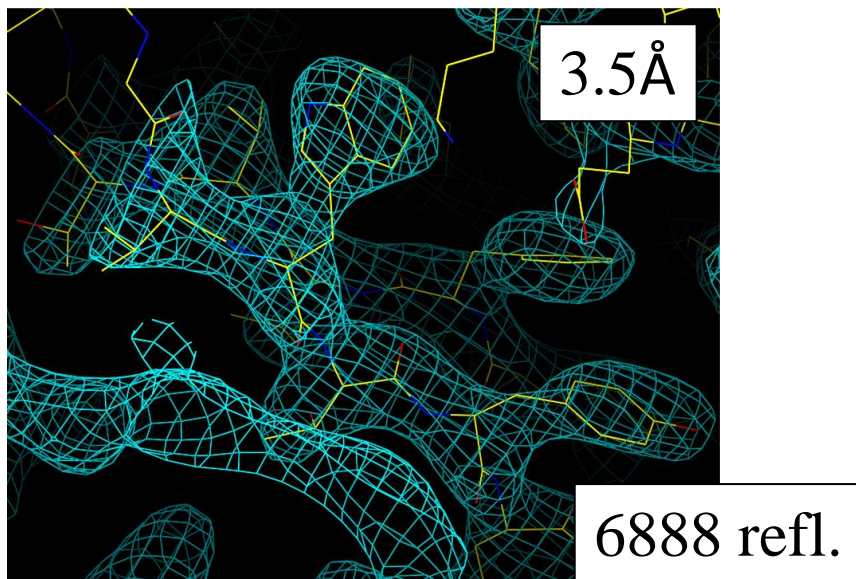
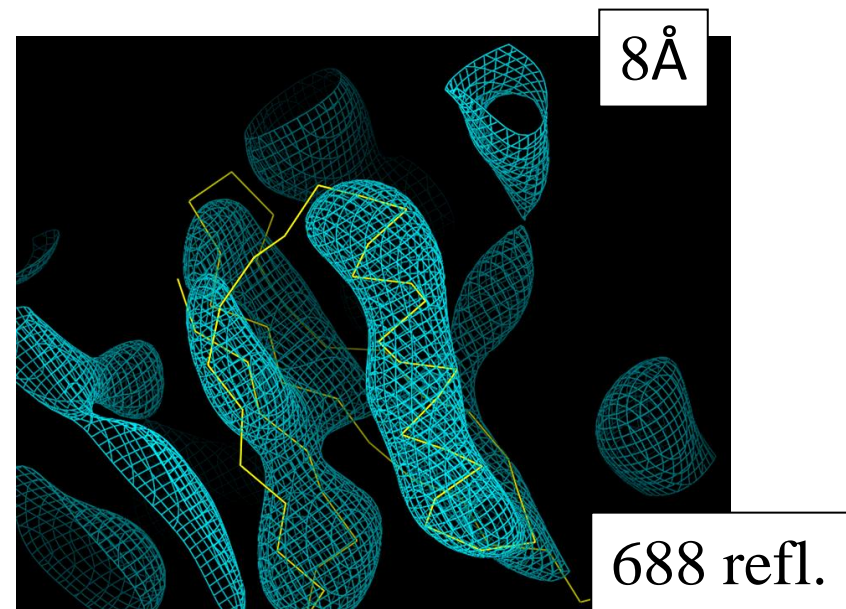
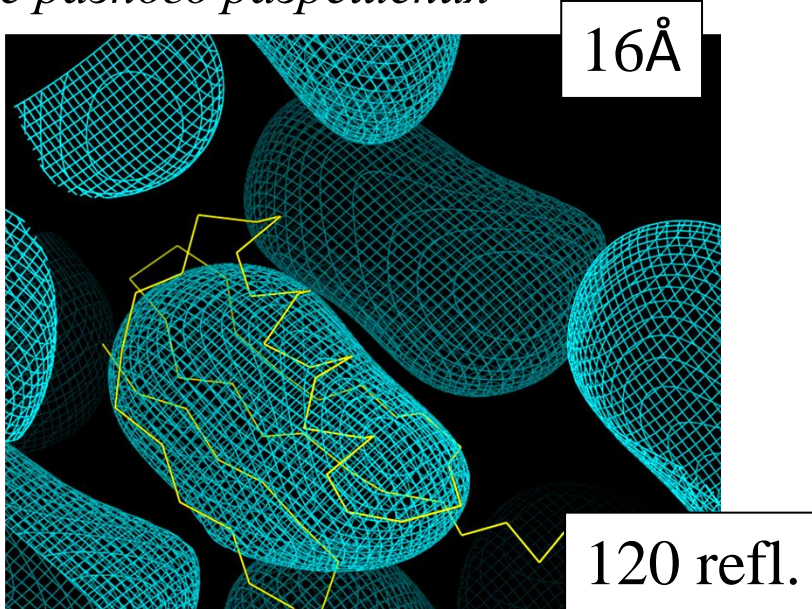
$$|\mathbf{s}| = \frac{2 \sin \theta}{\lambda} \quad d = \frac{\lambda}{2 \sin \theta}$$

$2\theta$  – угол рассеяния





*Изображение областей высокой плотности для белка Protein G, полученные на основе синтезов Фурье разного разрешения*



# Синтез Фурье электронной плотности

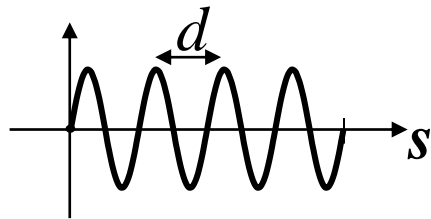
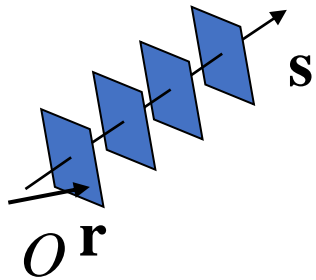
Если фазы найдены, то

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

$S$  – набор рефлексов.

*Разрешение*, соответствующее гармонике Фурье

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



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

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

$$d = \frac{\lambda}{2 \sin \theta}$$

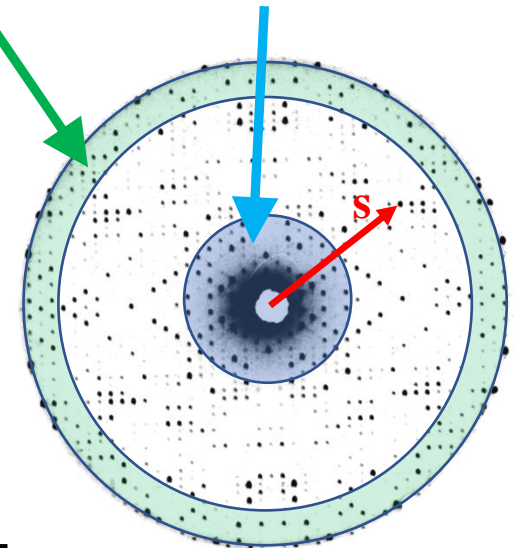
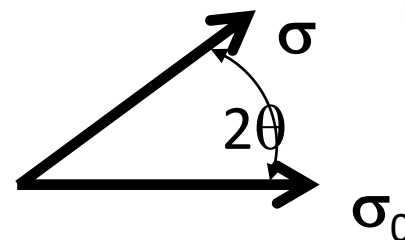
*Высокое разрешение:*

- большие углы;
- малые  $d$ ;
- большие  $s$ .

*Низкое разрешение:*

- малые углы;
- большие  $d$ ;
- малые  $s$ .

$2\theta$  – угол рассеяния



# Фазовая проблема.

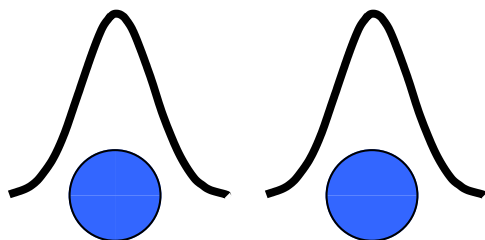
## Метод молекулярного замещения

Программа действий:

- найти в банке структур белок с похожей последовательностью и уже известной структурой;
- для модели из банка подобрать оптимальную ориентацию и положение в ячейке исследуемого кристалла
- известную структуру использовать в качестве первого приближения к искомой структуре; рассчитать фазы; построить комбинированный синтез Фурье ( $F^{\text{obs}}$ ,  $\varphi^{\text{PDB}}$ );
- внести корректировки в модель в соответствии с рассчитанным синтезом.

# «Комбинированные» синтезы Фурье.

«точная»  
структура



$F^{obs}$

$\varphi^{true}$

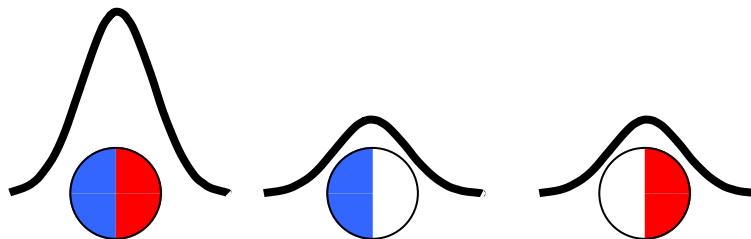
модель



$F^{calc}$

$\varphi^{calc}$

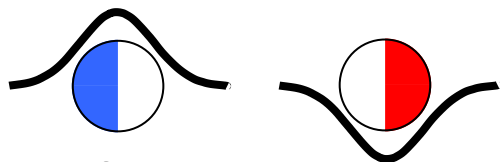
комбинированный  
синтез Фурье



$F^{obs}$

$\varphi^{calc}$

разностный  
синтез Фурье



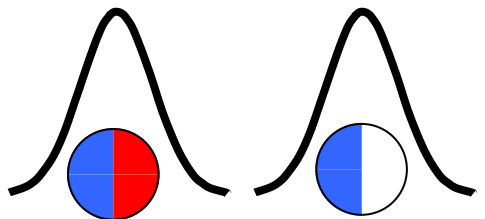
$F^{obs} - F^{calc}$

$\varphi^{calc}$

«2-1»

$2F^o - F^c$

синтез Фурье



$2F^{obs} - F^{calc}$

$\varphi^{calc}$

# Комбинированные синтезы Фурье

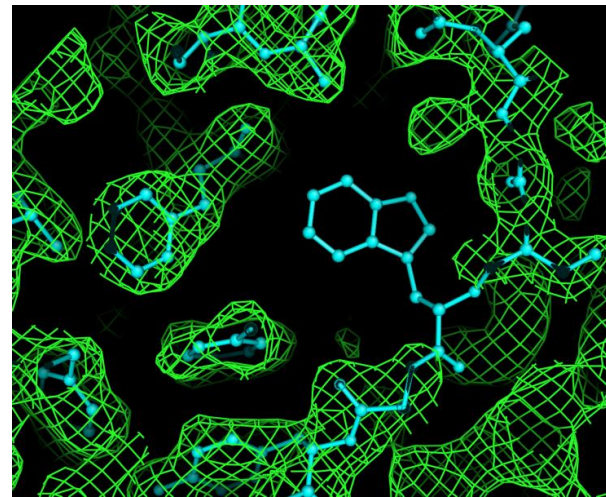
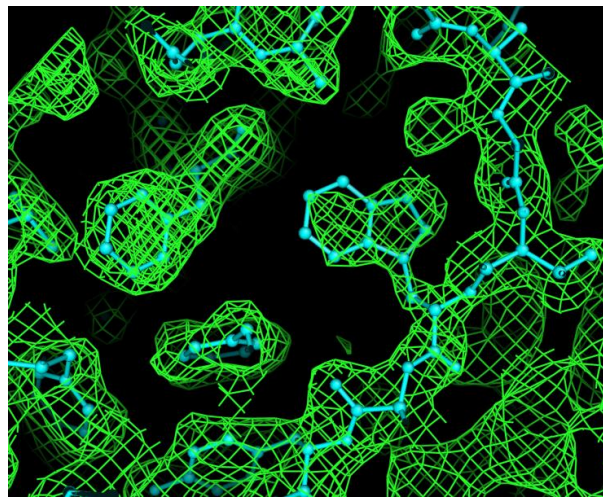
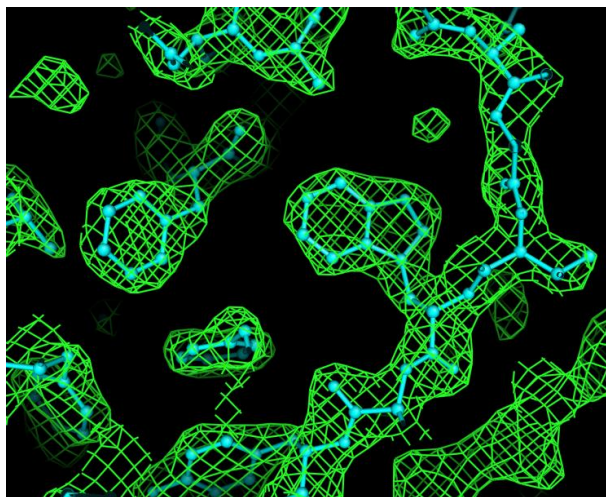
Эксперимент:  $\{F_{hkl}^{obs}\}$

Модель:  $\{F_{hkl}^{model}, \varphi_{hkl}^{model}\}$

$F_{hkl}^{obs}, \varphi_{hkl}^{exact}$

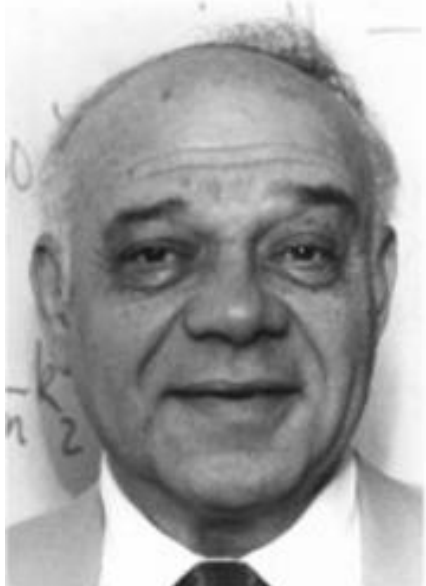
$F_{hkl}^{obs}, \varphi_{hkl}^{model}$

$F_{hkl}^{model}, \varphi_{hkl}^{model}$



Jerome Karle   Gerbert Hauptman

$\mathcal{F}^{\text{Karle}},$   
 $\varphi^{\text{Karle}}$



$\mathcal{F}^{\text{Hauptman}},$   
 $\varphi^{\text{Hauptman}}$

**Что важнее для  
изображения:  
модули или фазы?**

$\mathcal{F}^{\text{Karle}},$   
 $\varphi^{\text{Hauptman}}$



$\mathcal{F}^{\text{Hauptman}},$   
 $\varphi^{\text{Karle}}$

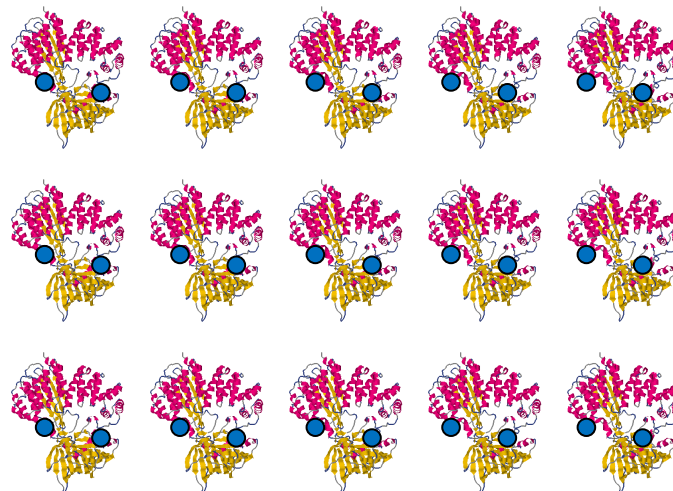
# Метод изоморфного замещения

"нативный" белок



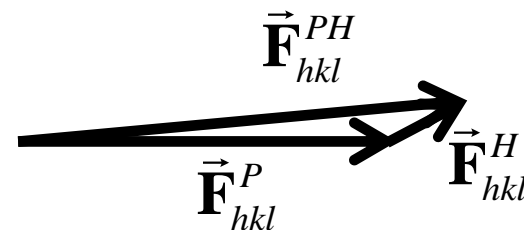
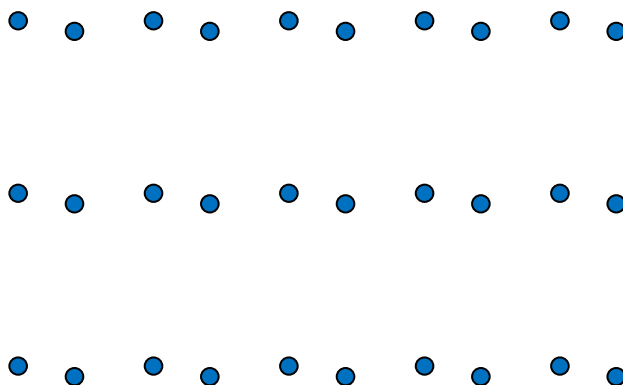
$\{F^P\}$

изоморфное производное



$\{F^{PH}\}$

подструктура тяжелых атомов



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

## Шаг 1. Определение подструктуры «тяжелых атомов»

- небольшое число атомов:  $\{\mathbf{r}_j^H\}, j = 1, \dots, m$

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

- расчет модулей и фаз  $F_{hkl}^H, \varphi_{hkl}^H$

## Шаг 2. Расчет значения фазы для каждого коэффициента Фурье

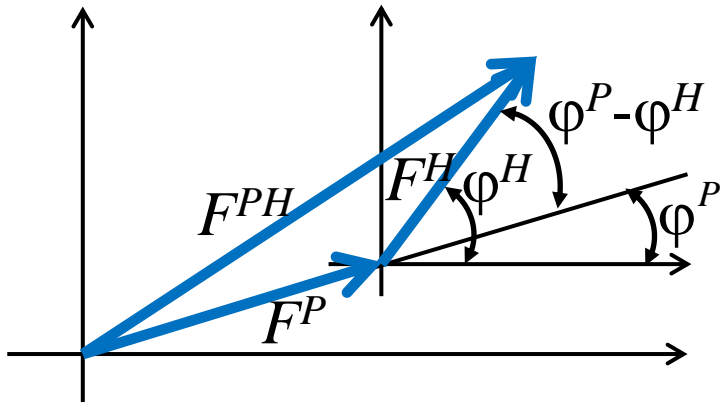
$$\vec{F}_{hkl}^{PH} = \vec{F}_{hkl}^P + \vec{F}_{hkl}^H$$

$$(F^{PH})^2 = (F^P)^2 + (F^H)^2 + 2F^P F^H \cos(\varphi^P - \varphi^H)$$

$F^P, F^{PH}$  - получили из эксперимента

$F^H, \varphi^H$  - определили на Шаге 1

$\varphi^P$  - находим из уравнения; для каждого коэффициента





нативный белок  $F^P, \varphi^P$

изоморфное производное  $F^{PH}, \varphi^{PH}$

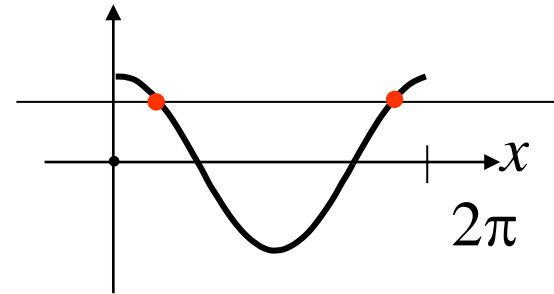
подструктура тяжелых атомов  $F^H, \varphi^H$

уравнение  $(F^{PH})^2 = (F^P)^2 + (F^H)^2 + 2F^P F^H \cos(\varphi^P - \varphi^H)$

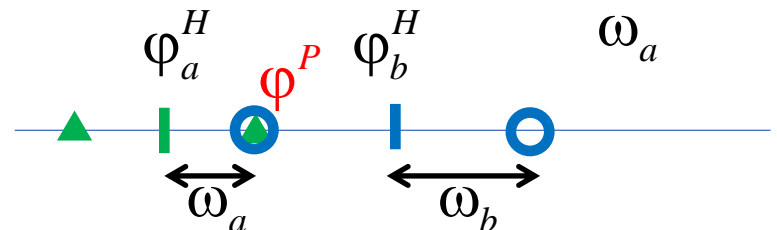
$$\cos(\varphi^P - \varphi^H) = \frac{(F^{PH})^2 - (F^P)^2 - (F^H)^2}{2F^P F^H}$$





$$\varphi^P = \varphi^H \pm \omega$$

$$\omega = \arccos \frac{(F^{PH})^2 - (F^P)^2 - (F^H)^2}{2F^P F^H}$$



решение не единственно  
(проблема снимается при  
наличии двух производных)



-  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

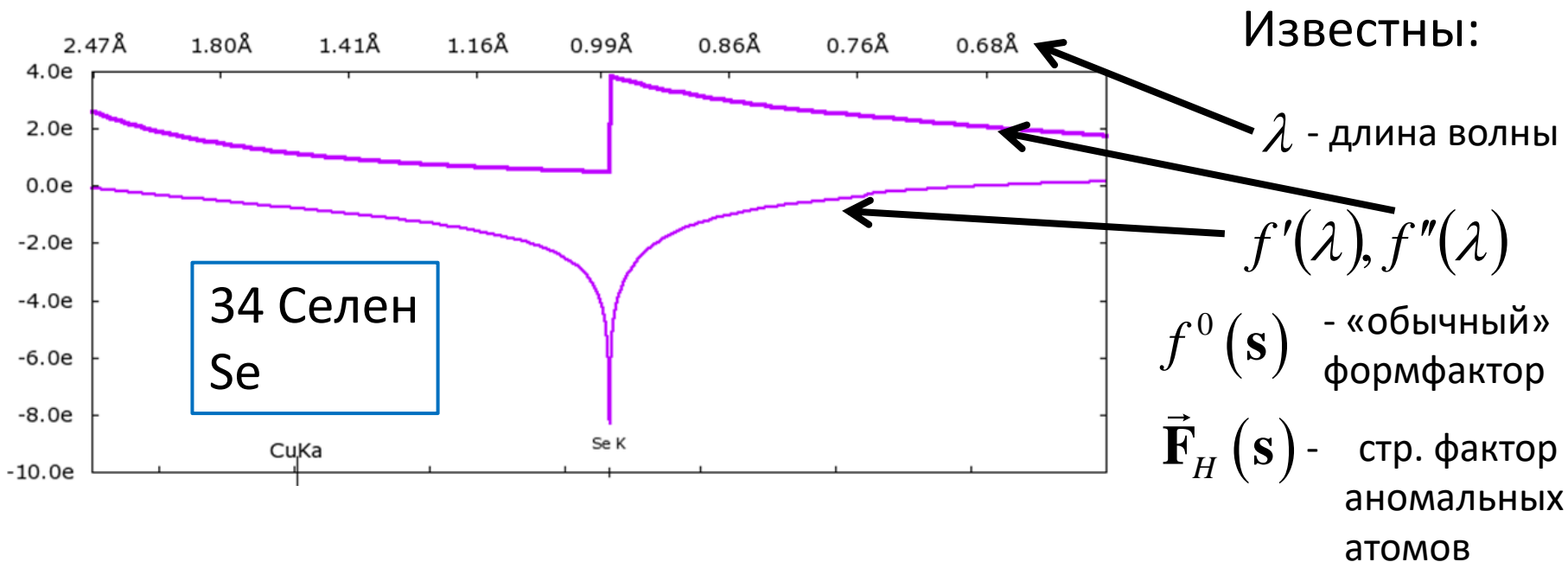
Множественное изоморфное замещение;

# AD - Anomalous Dispersion / Аномальное рассеяние

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

Наличие аномальных рассеивающих атомов:

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



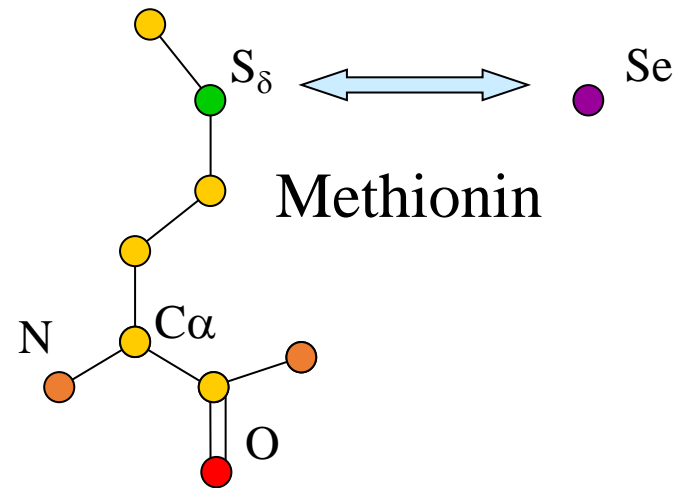
$F_P(\mathbf{s}), \varphi_P(\mathbf{s})$  - неизвестны

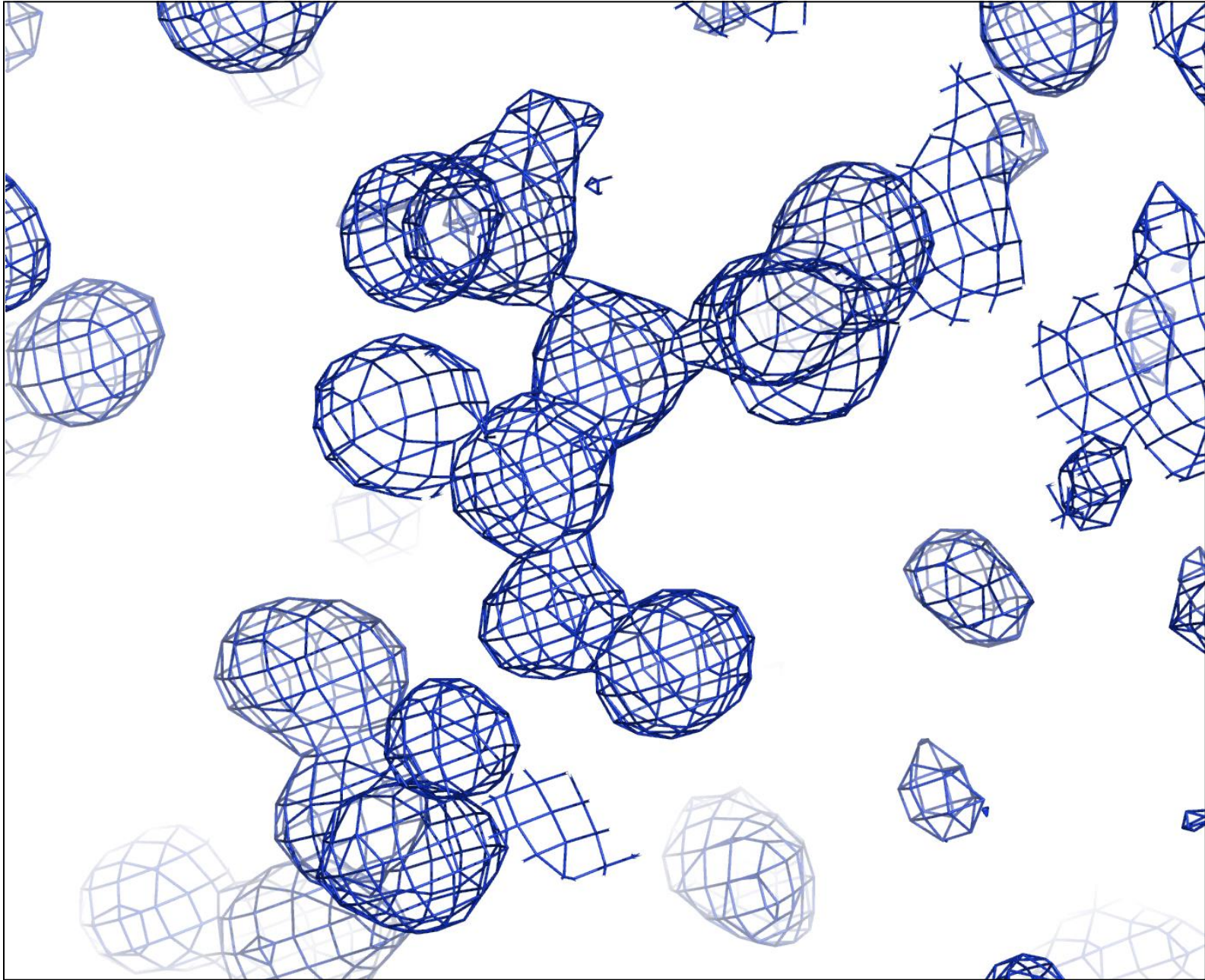
# MAD - Multiwavelength Anomalous Diffraction

## Многоволновое аномальное рассеяние;

В основе подхода лежит изменение интенсивности аномального рассеяния при изменении длины волны.

- ☺ • один из активно используемых методов решения фазовой проблемы;
- ☺ • возможность использования селен-метиониновых производных;
- ☹ • проблема определения мест нахождения аномально рассеивающих атомов;
- ☹ • не всегда получается.



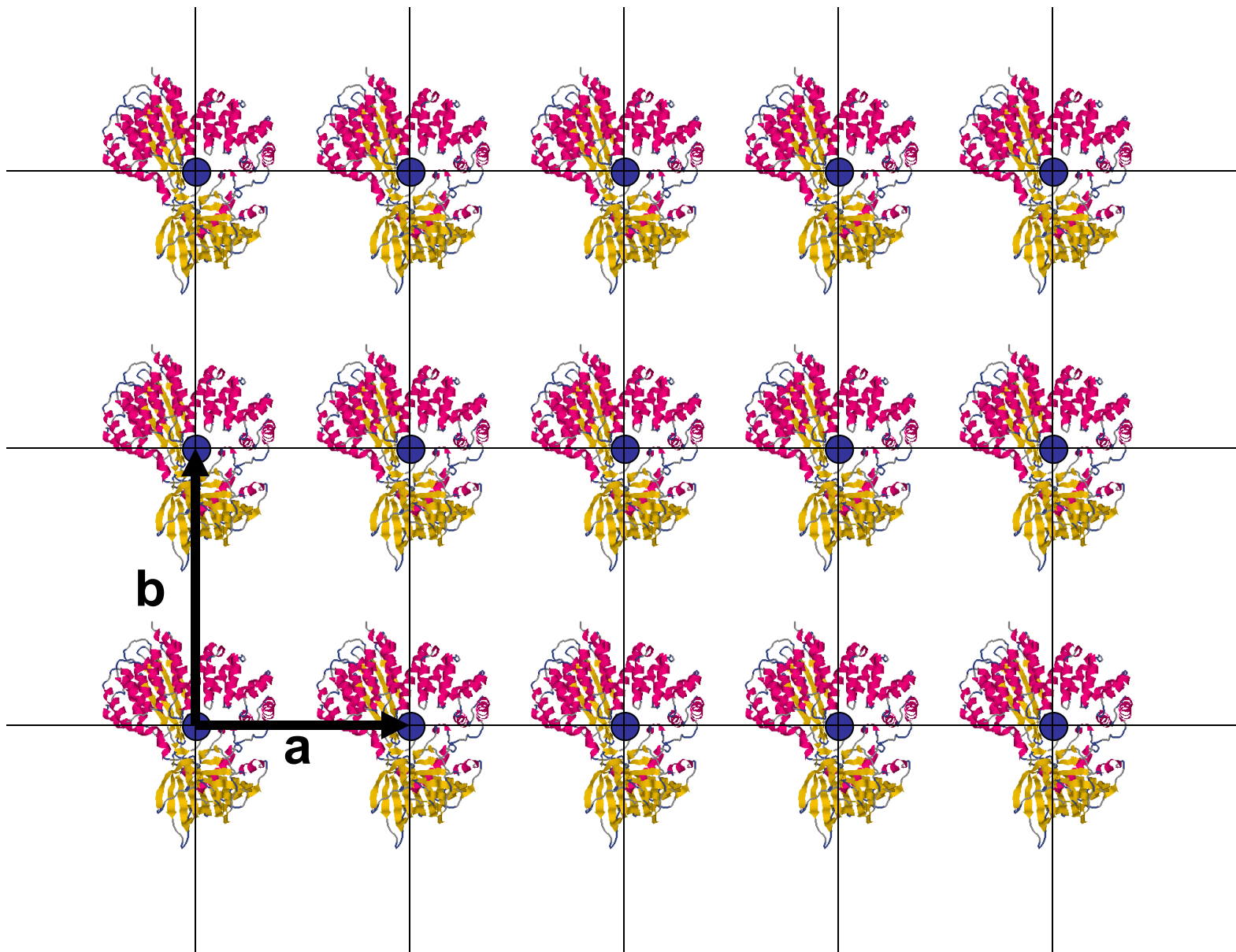


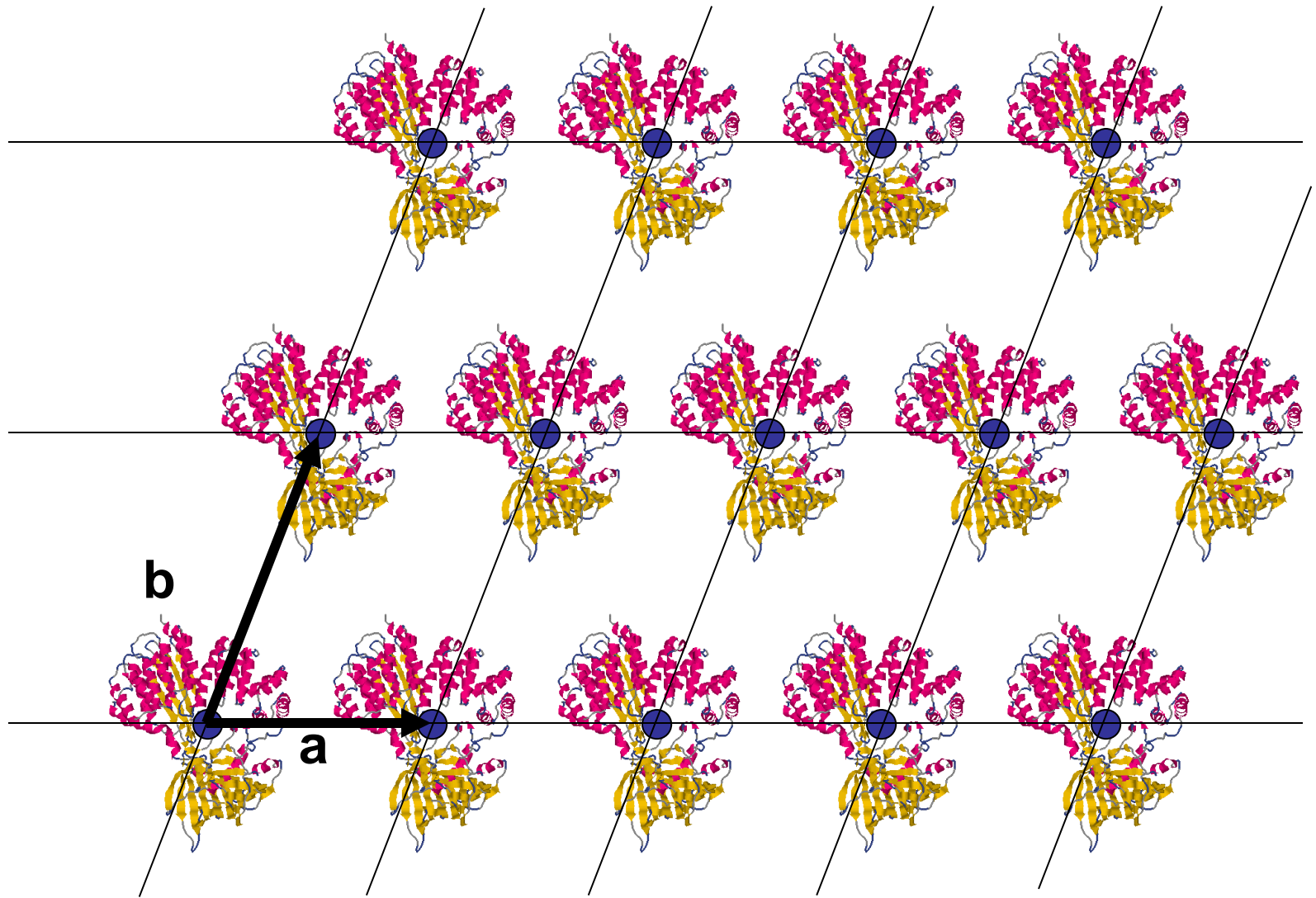
aldose reductase, 0.9Å, MAD

# Кристаллография



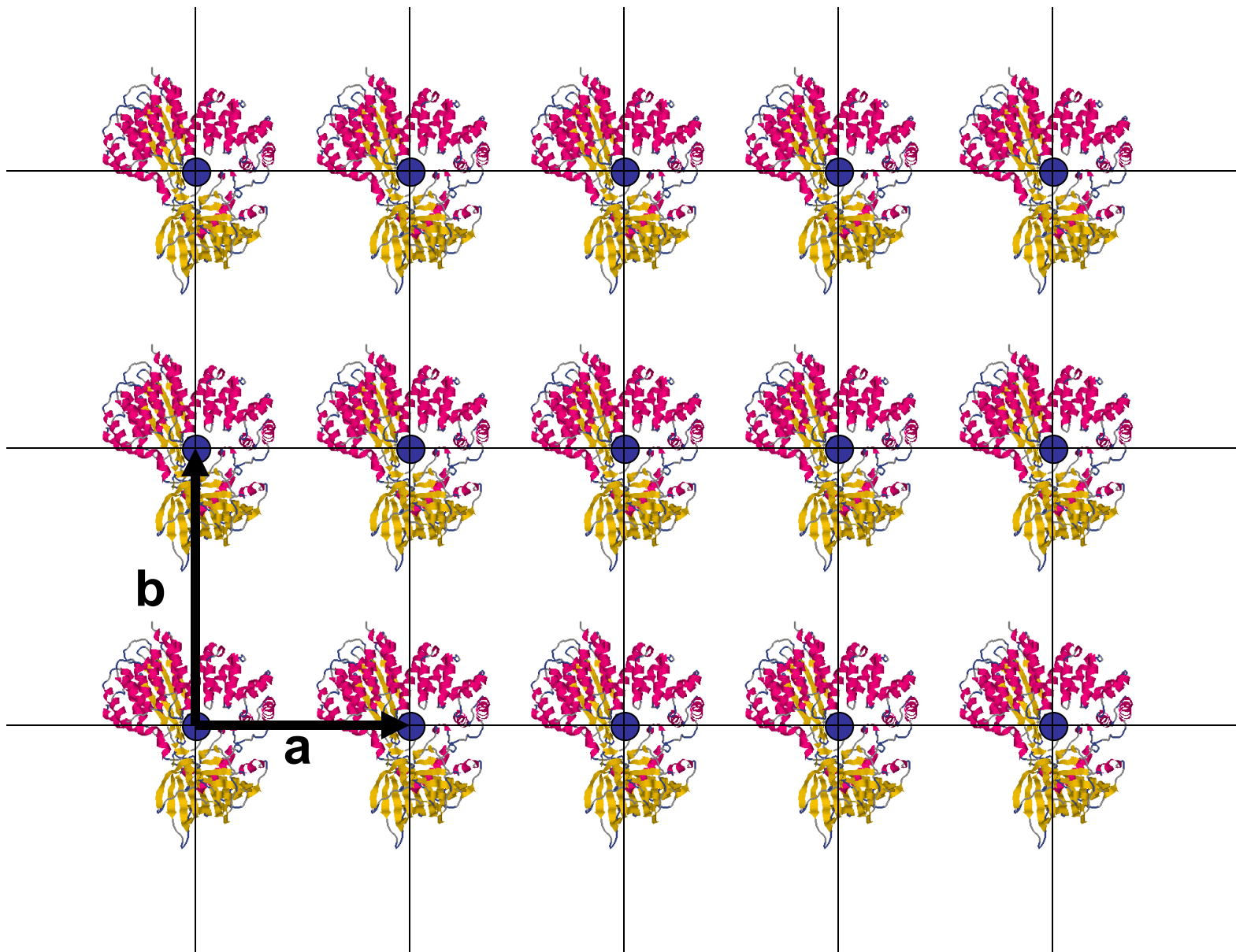
# Кристаллическая решетка

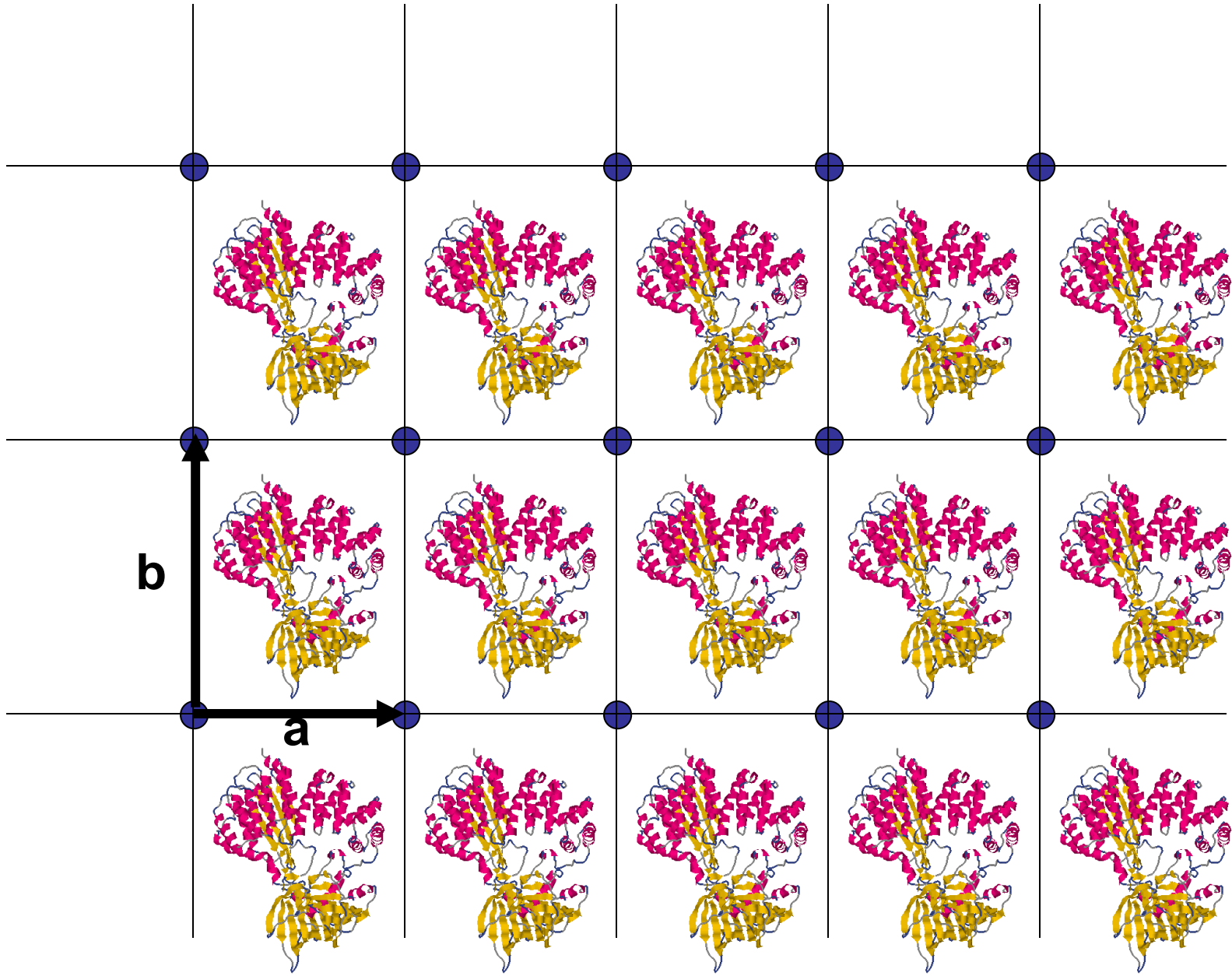


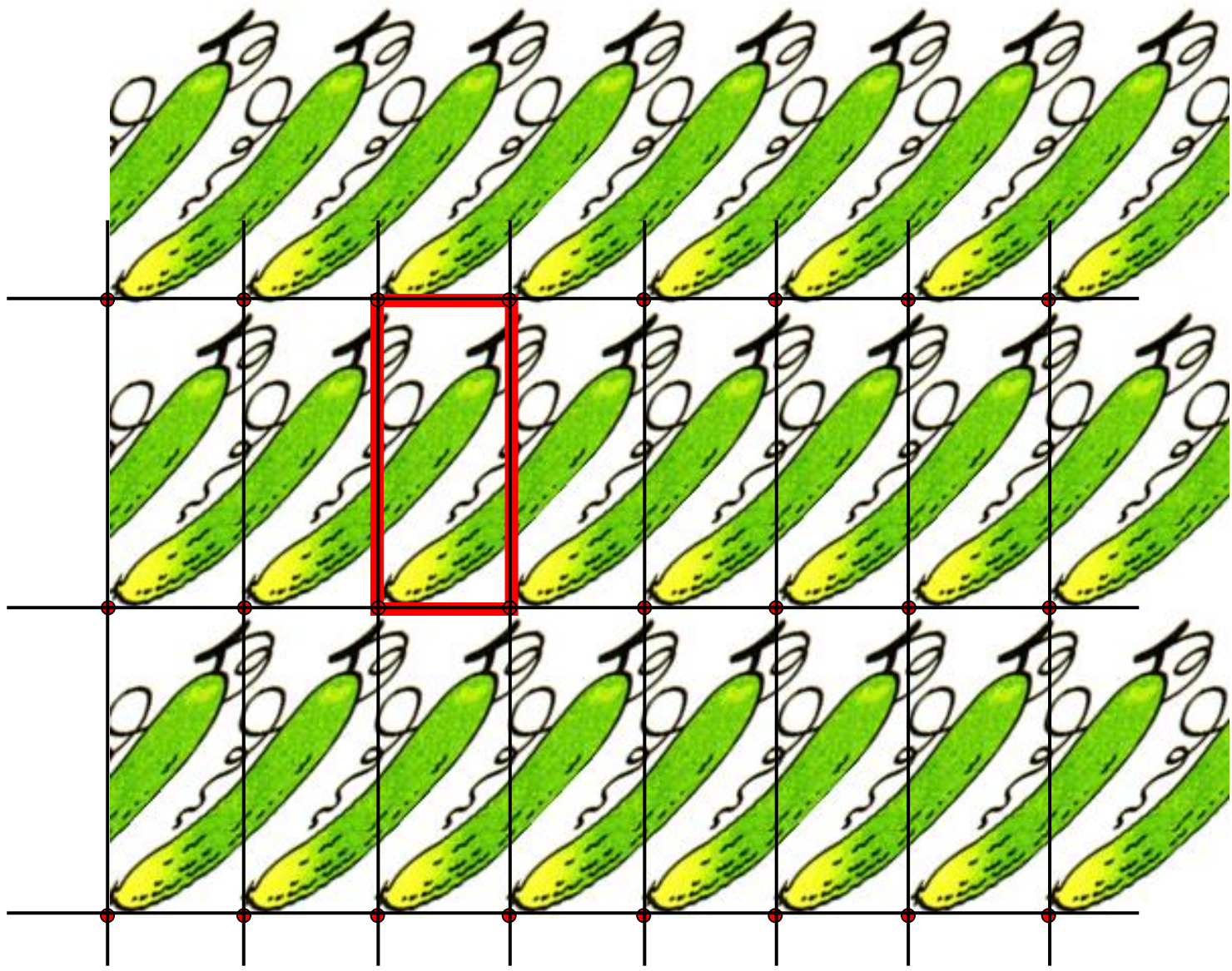


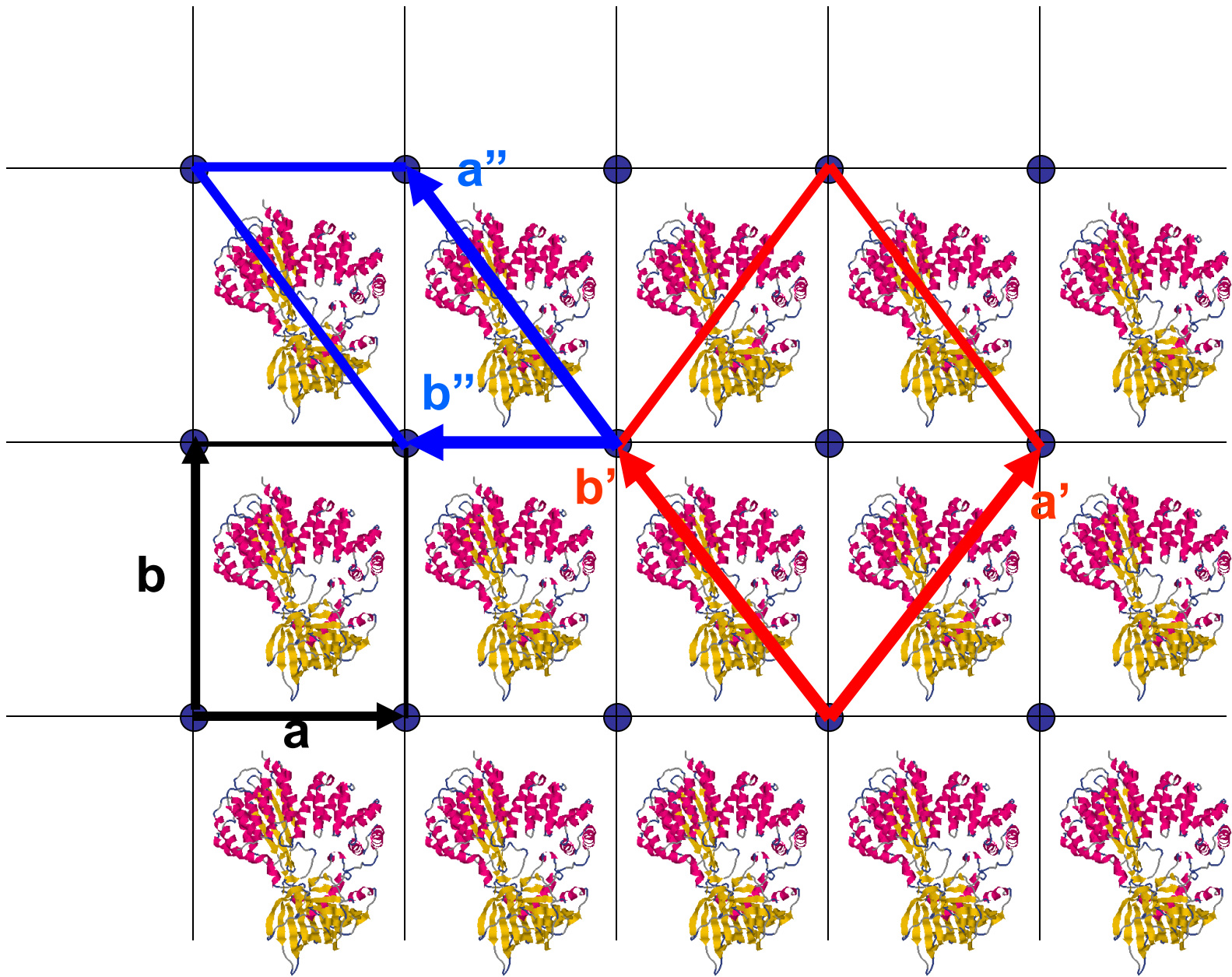


# Выбор элементарной ячейки

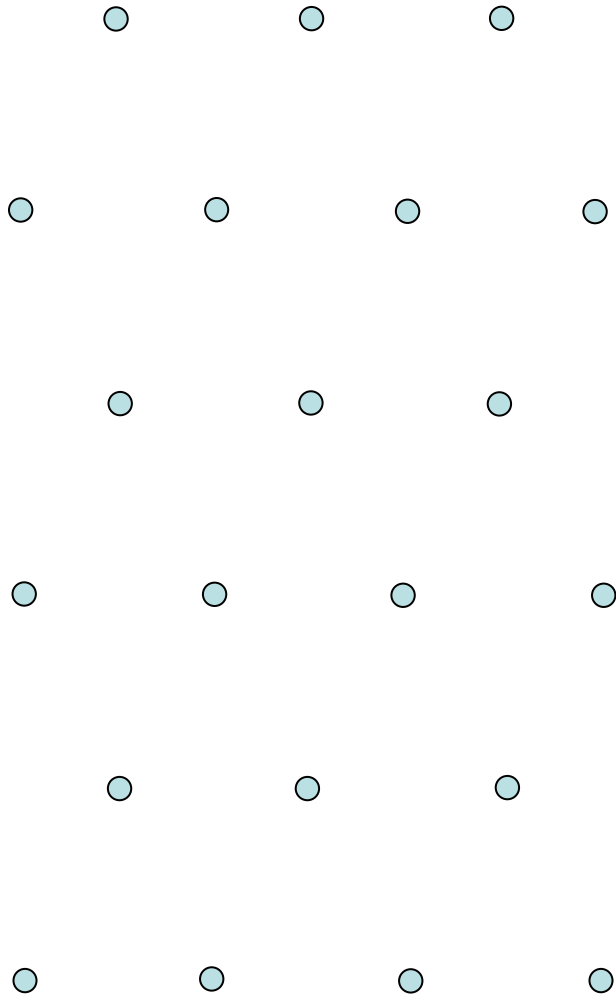




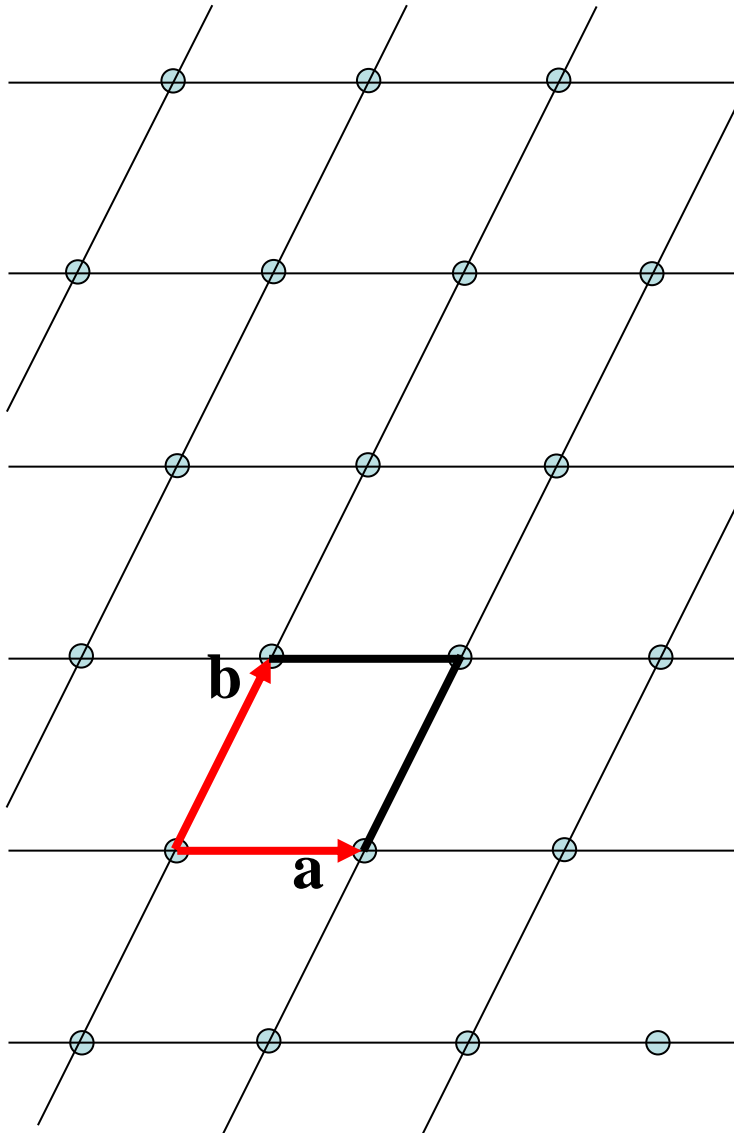




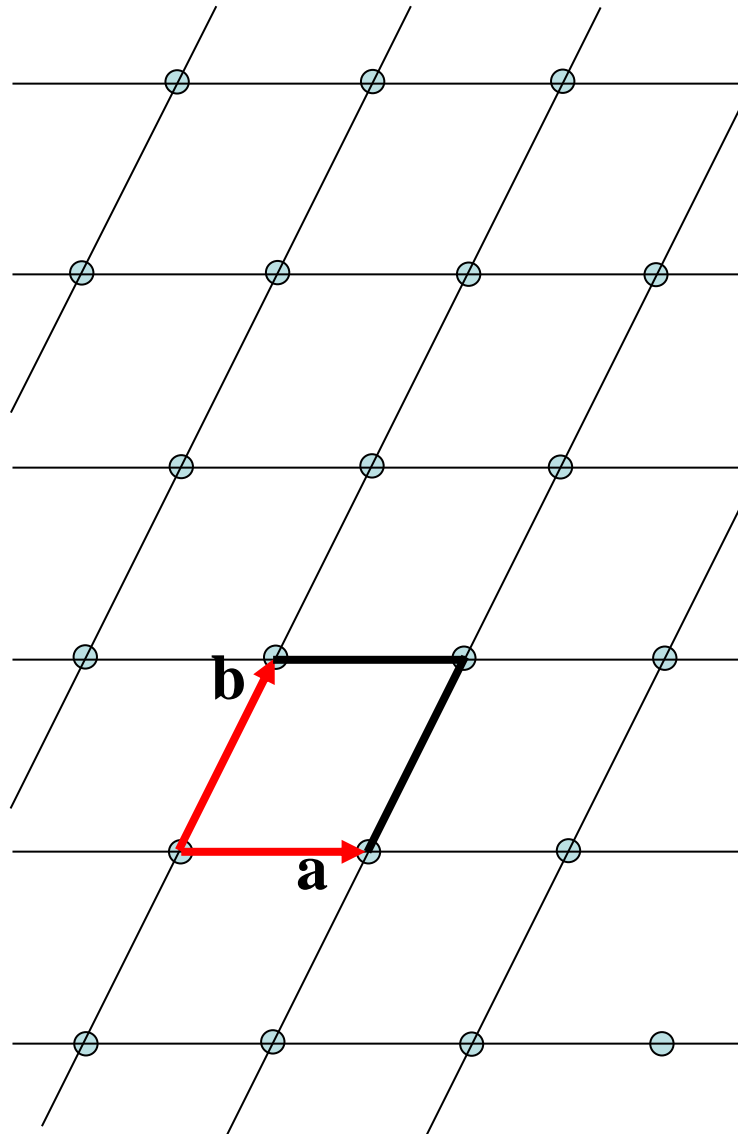
# Центрированные ячейки



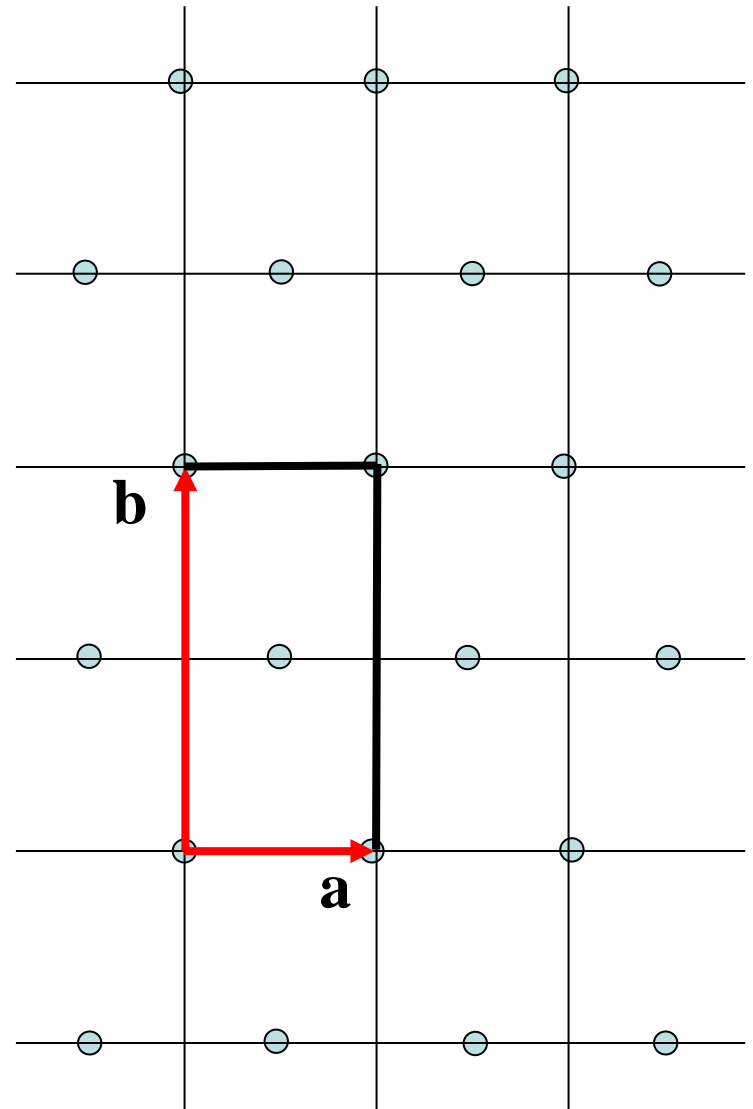
примитивная косоугольная ячейка



примитивная косоугольная ячейка



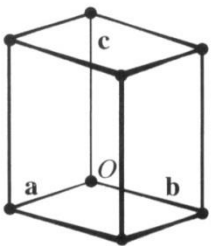
ортогональная ячейка;  
дополнительная  
трансляция  $(1/2, 1/2)$



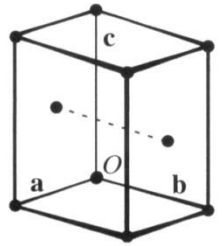
примитивная

базоцентрированная

**P**



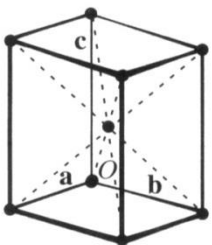
A primitive unit cell (*P*)



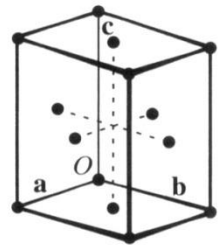
A unit cell centred in the (010) planes (*B*)

**A**  
**B**  
**C**

**I**



A body-centred unit cell (*I*)



A face-centred unit cell (*F*)

**F**

объемоцентрированная

гранецентрированная

дважды – объемоцентрированная (ромбоэдрическая)

**R**

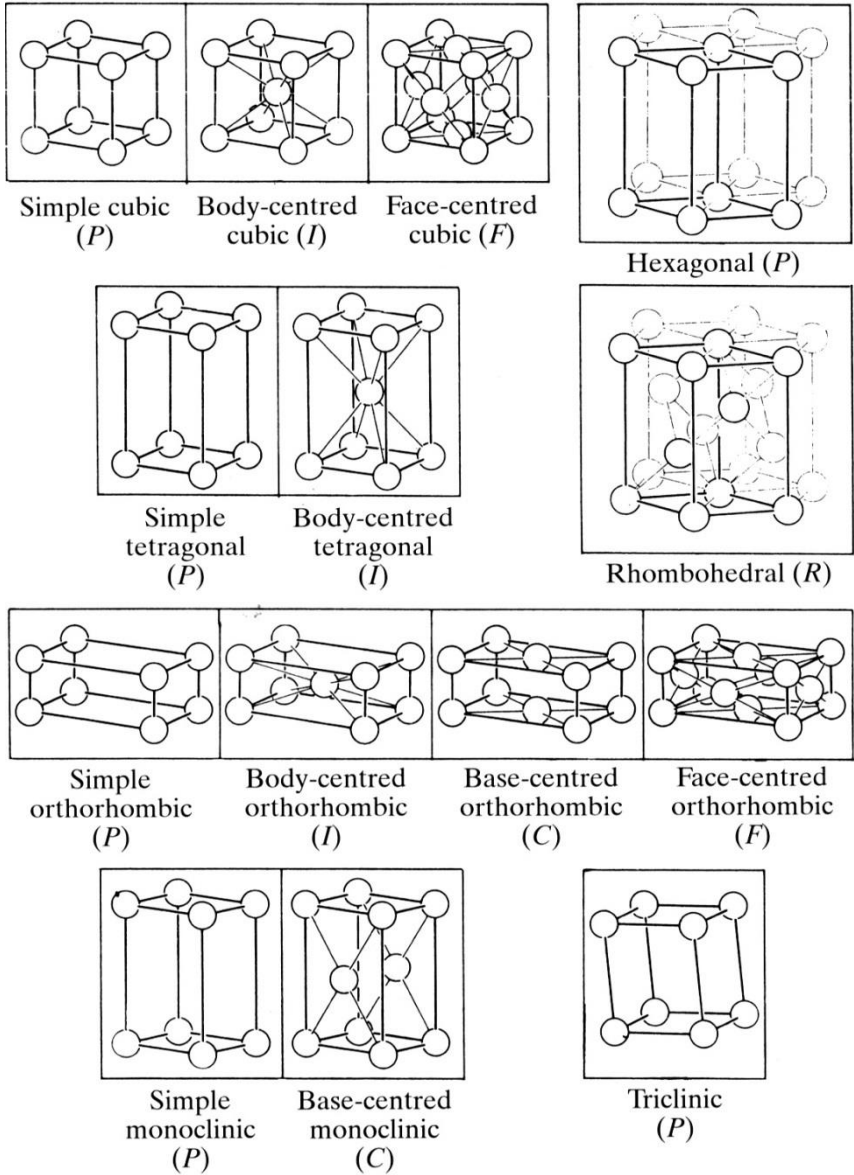


Fig. 2.1.3.3. The 14 Bravais lattices. Reproduced with permission from Burzlaff & Zimmermann (1995). Copyright (1995) International Union of Crystallography.



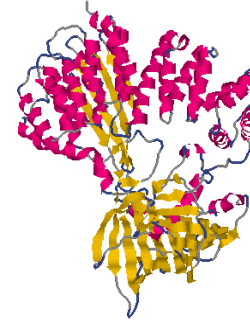
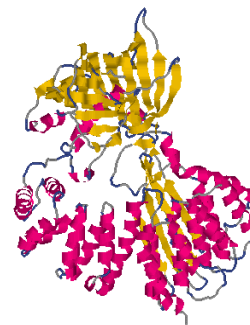
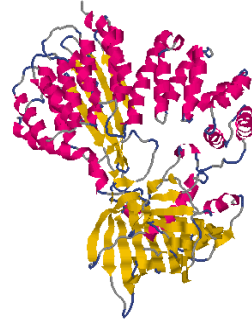
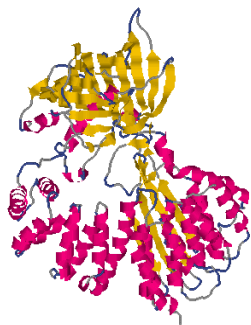
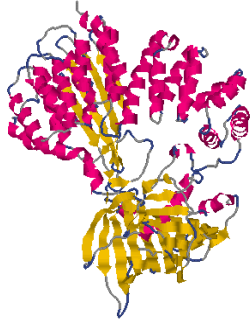
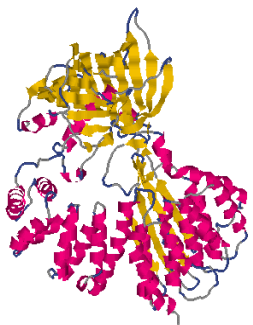
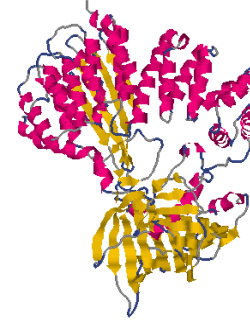
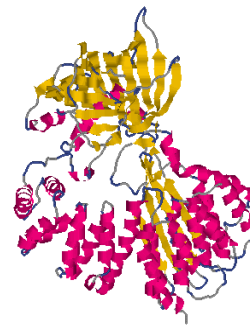
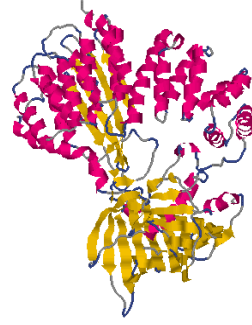
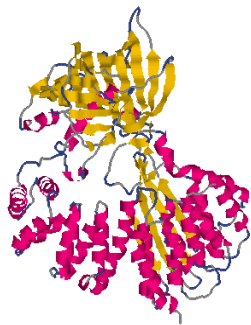
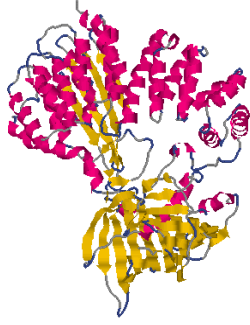
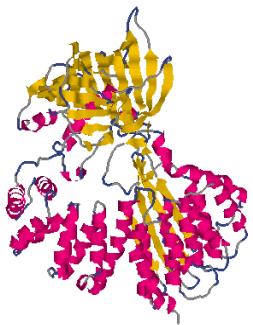
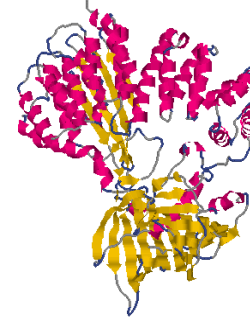
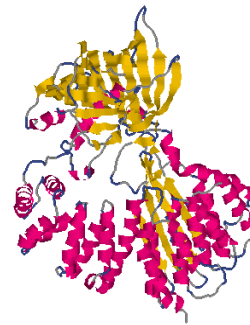
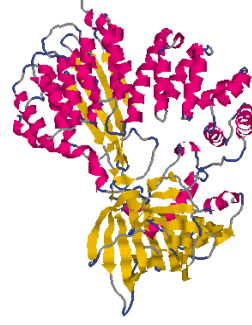
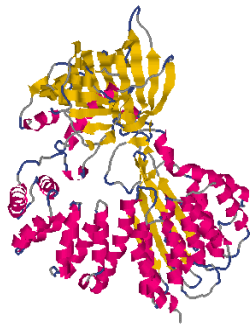
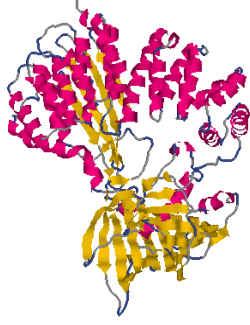
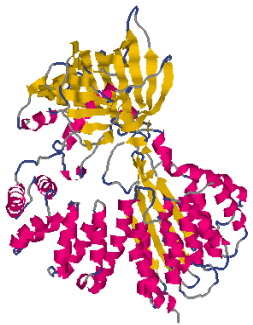
CRYST1 66.224 66.224 40.561 90.00 90.00 120.00 **P** 63 6

тип решетки

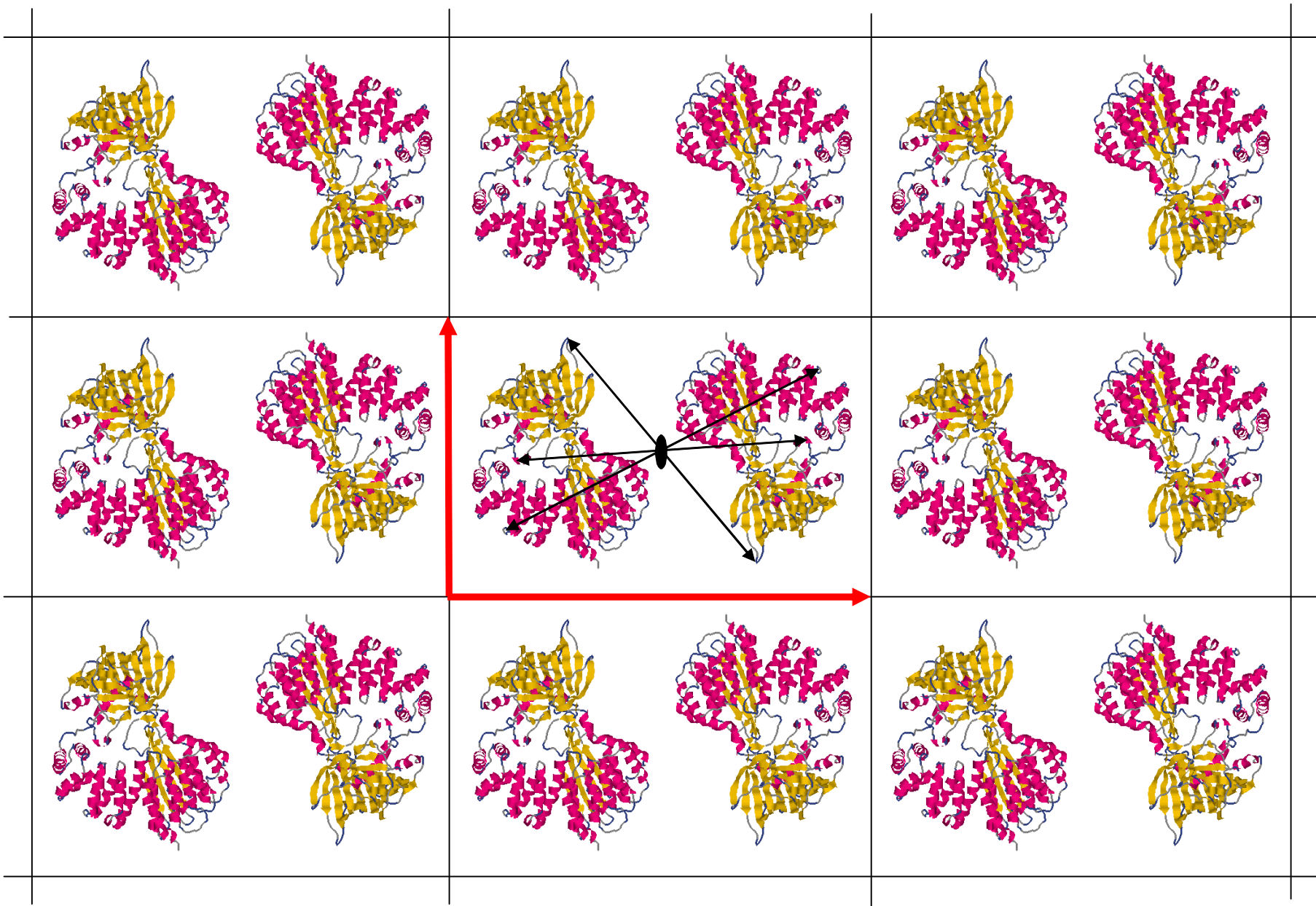


- Элементарная ячейка - математический объект. Она вводится для удобства работы.
- Выбор элементарной ячейки и начала координат в значительной мере произвольны.
- В начале координат может не находиться никакого атома.
- Молекула не всегда лежит в выбранной элементарной ячейке целиком.
- Выбор элементарной ячейки – это выбор системы координат
- При сравнении координат двух структур эти структуры должны быть предварительно "выровнены".

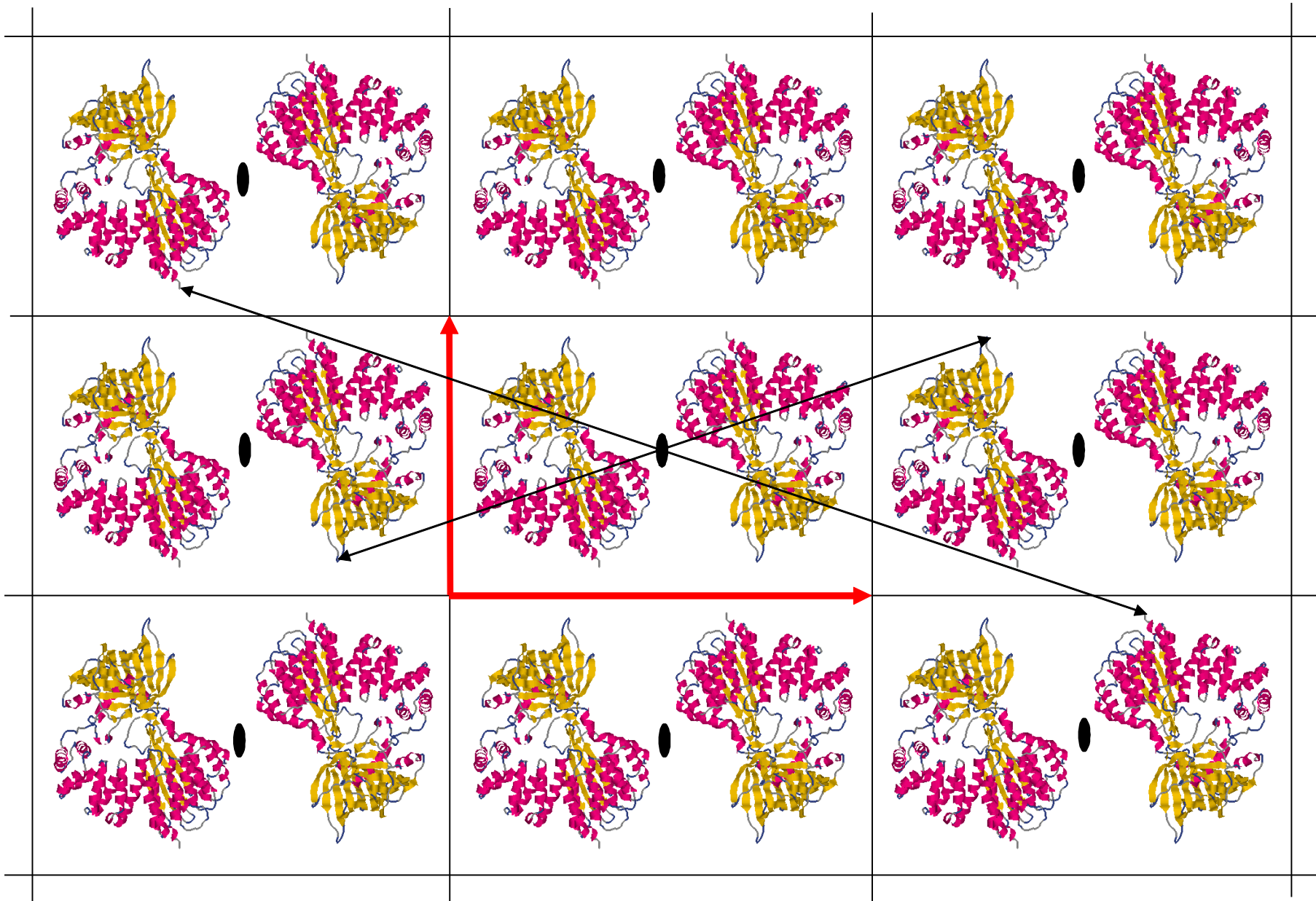
# Симметрия



# Элемент симметрии

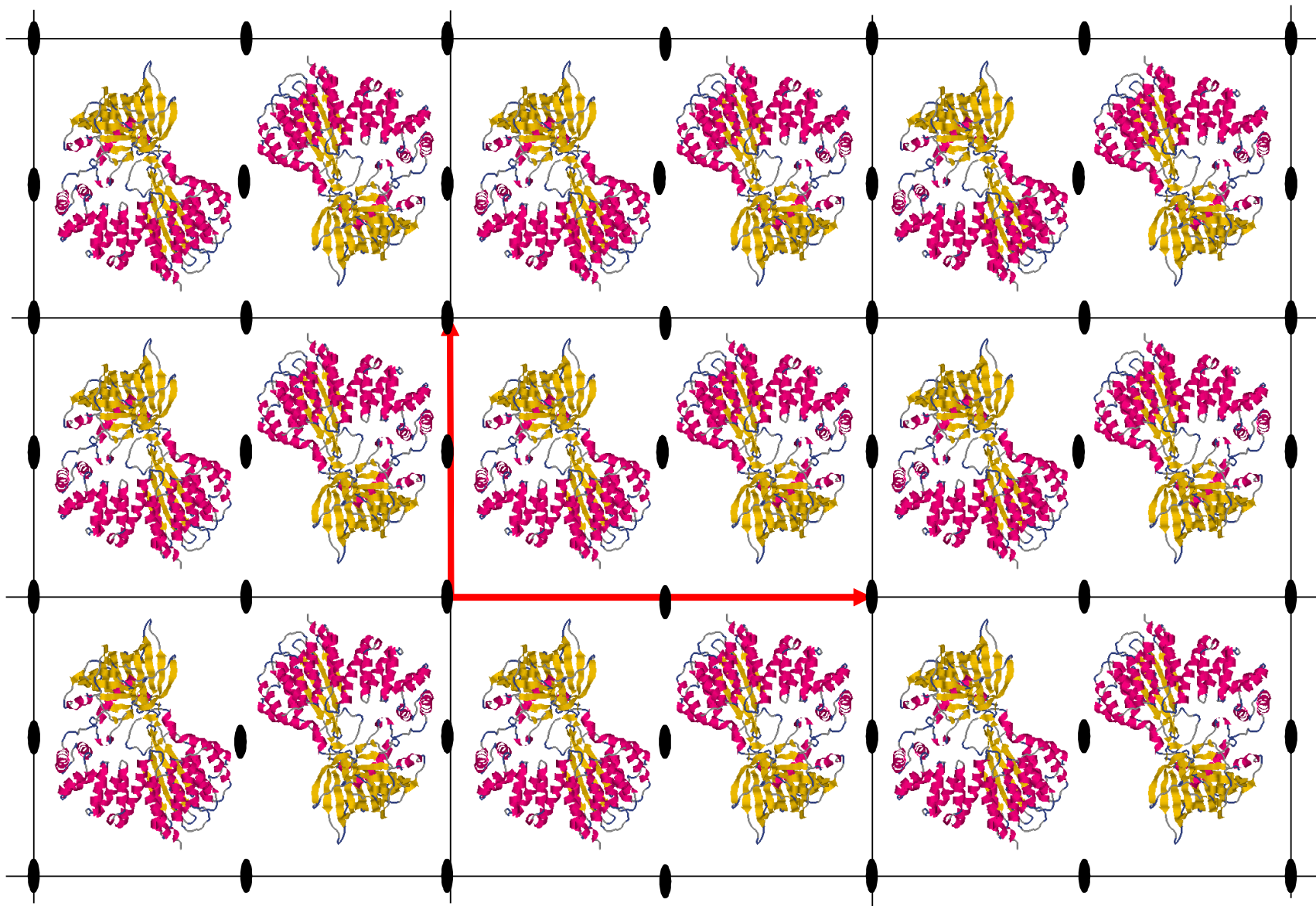


# Кристаллографическая симметрия



Кристаллографическая симметрия “действует” для всех точек кристалла

# Кристаллографическая симметрия



Периодичность порождает новые элементы симметрии

# 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.

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

Все симметрии конкретной кристаллической структуры образуют группу.

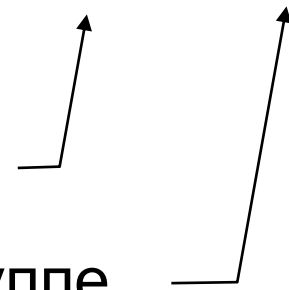
Существует конечное число (230) групп симметрии кристаллов.

Каждая группа имеет свое обозначение.

CRYST1 66.224 66.224 40.561 90.00 90.00 120.00 P 63 6

символ пространственной группы

число элементов симметрии в группе



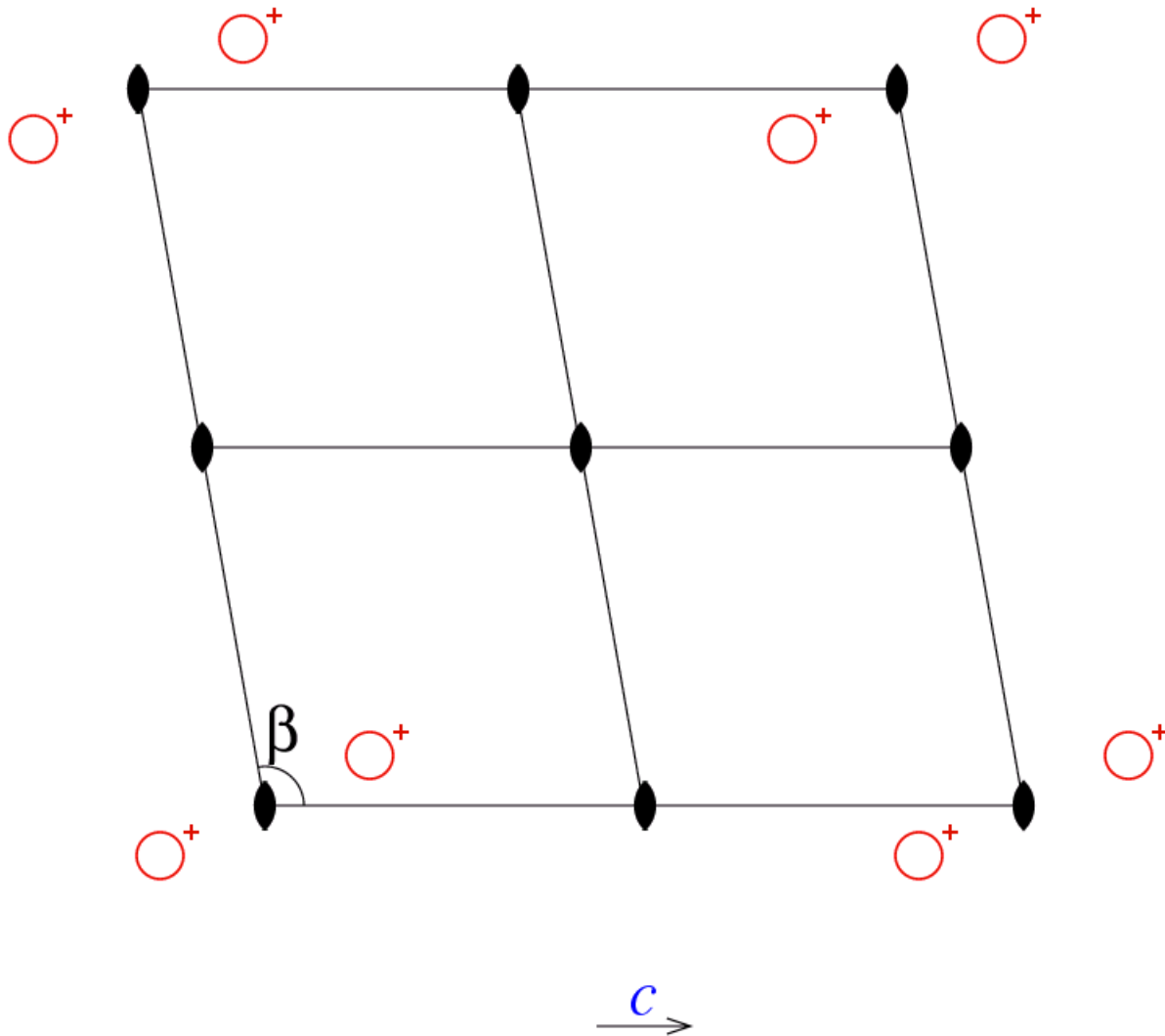


P2

P 1 2 1

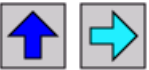
2

No. 3



1  $x, y, z$

2  $\bar{x}, y, \bar{z}$

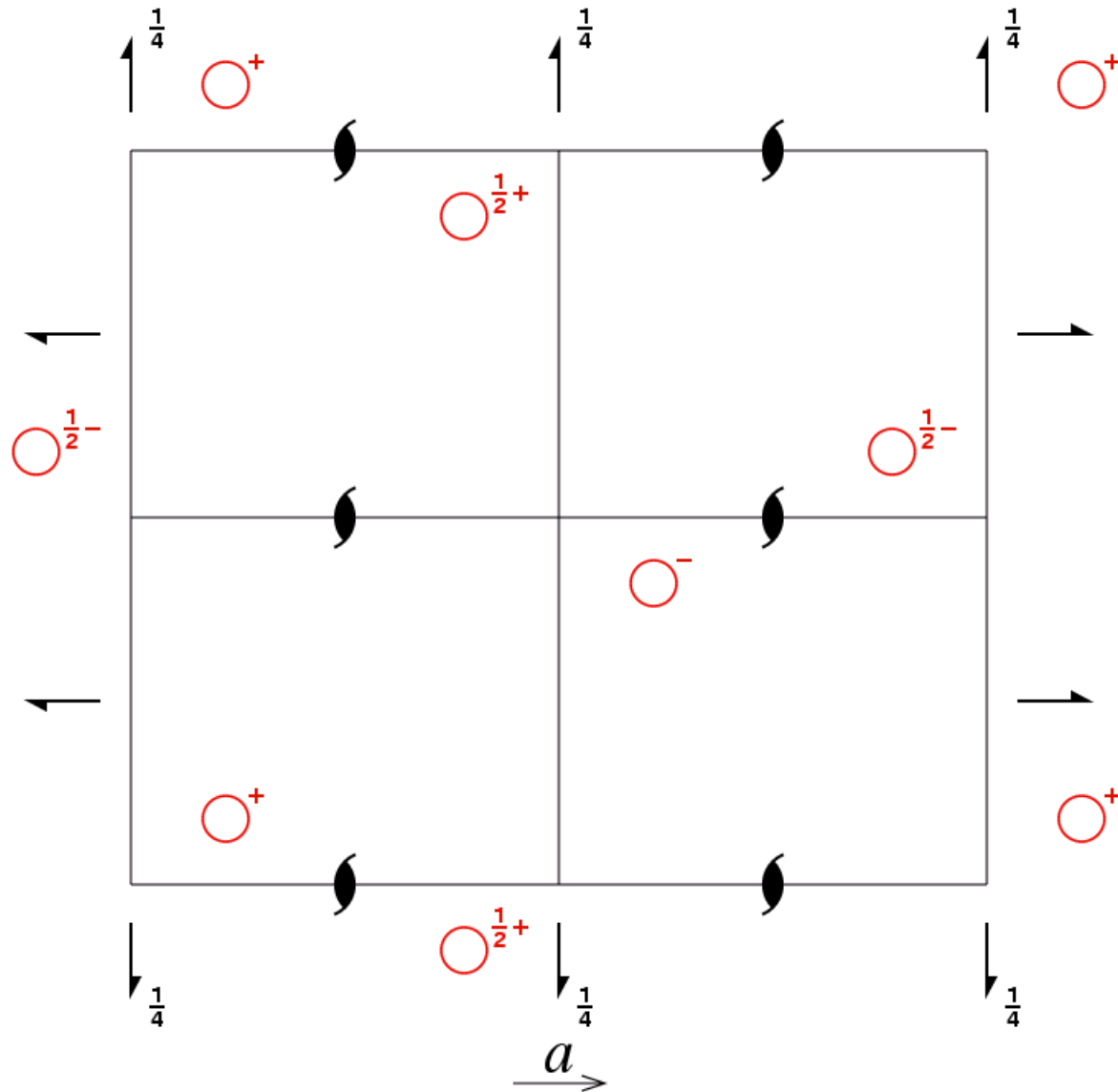


$P2_12_12_1$

$P2_12_12_1$

$222$

No. 19

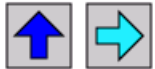


1  $x, y, z$

2  $\frac{1}{2} + x, \frac{1}{2} - y, \bar{z}$

3  $\bar{x}, \frac{1}{2} + y, \frac{1}{2} - z$

4  $\frac{1}{2} - x, \bar{y}, \frac{1}{2} + z$



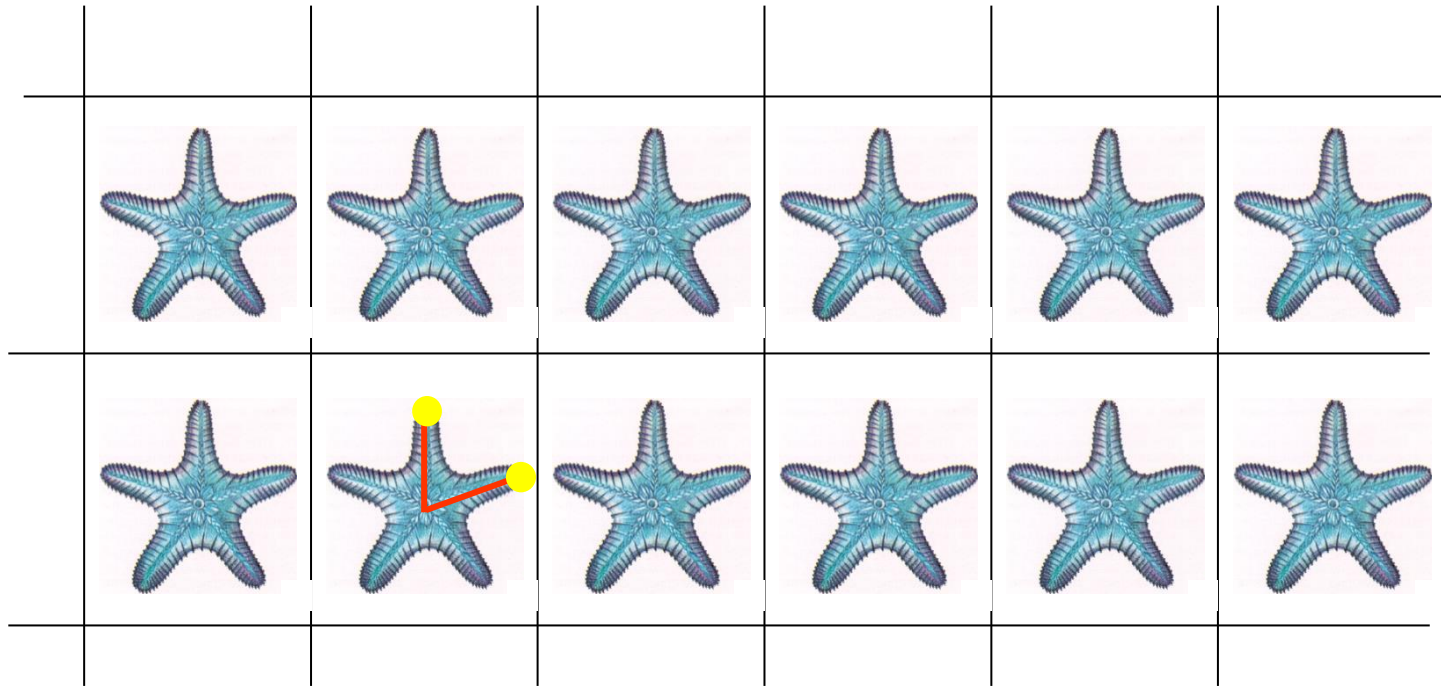
Что вносит кристаллографическая симметрия в рассеяние рентгеновских лучей?

$$P2 \quad \rho(-x, y, -z) = \rho(x, y, z)$$

$$F(-h, k, -l) = F(h, k, l)$$

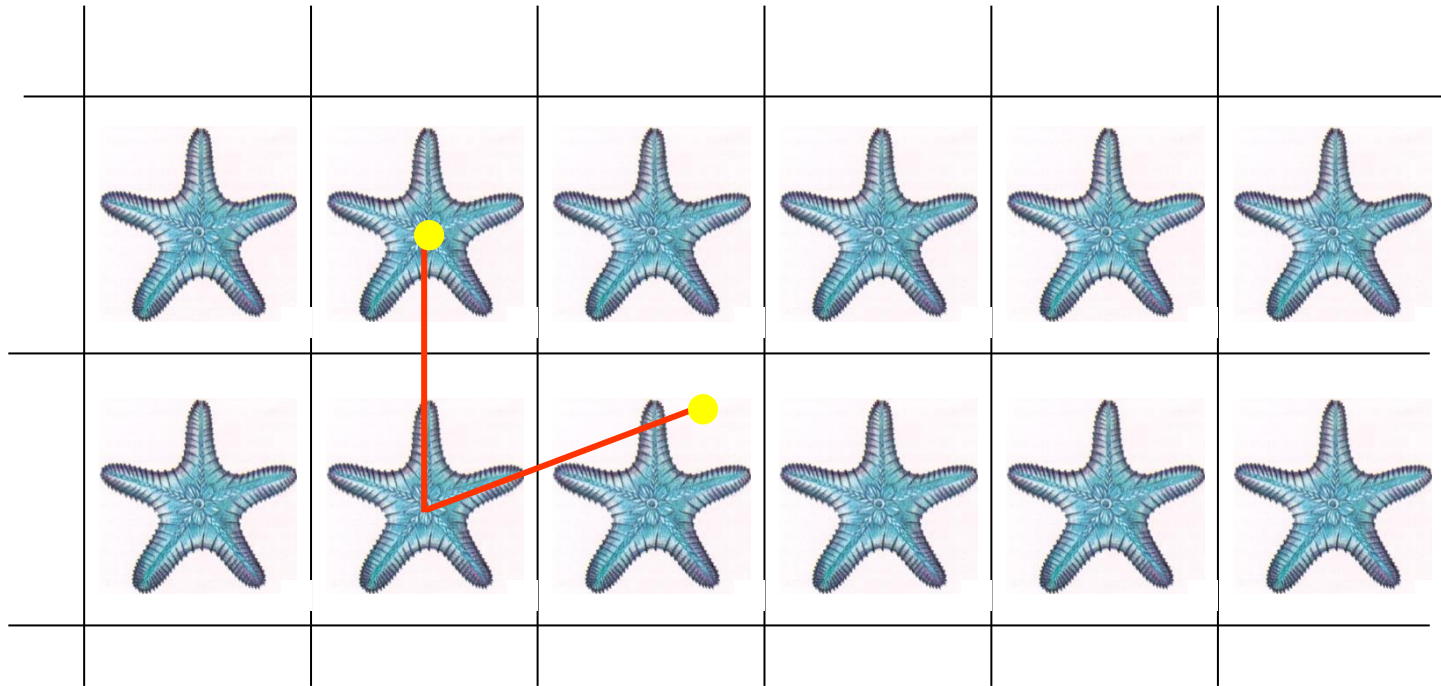
По симметрии картины распределения интенсивностей рефлексов можно делать выводы о симметрии изучаемого кристалла

# Некристаллографическая симметрия

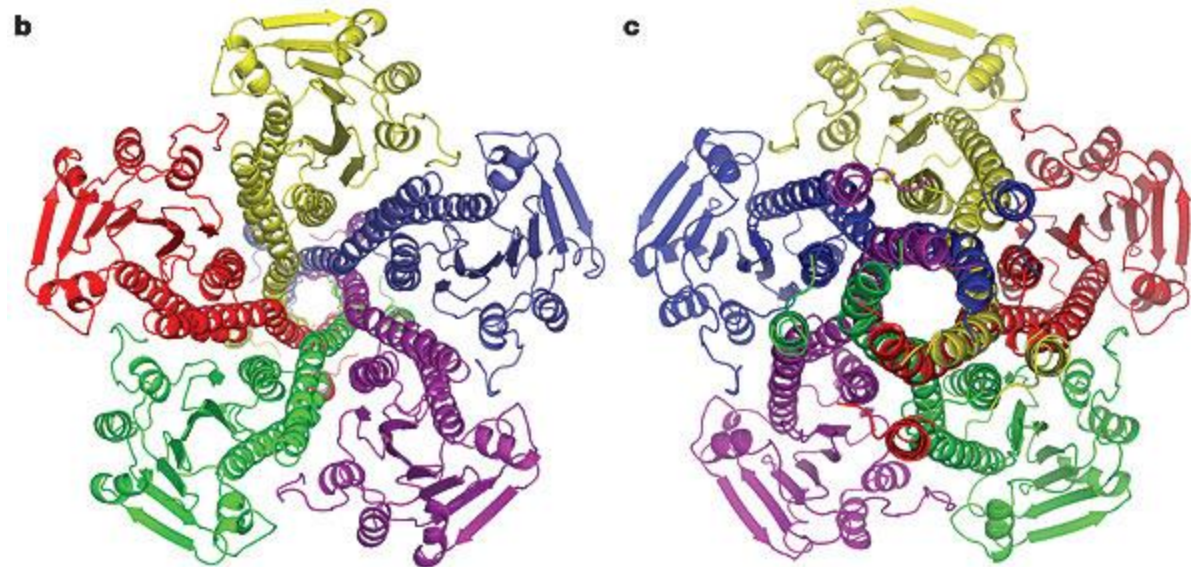
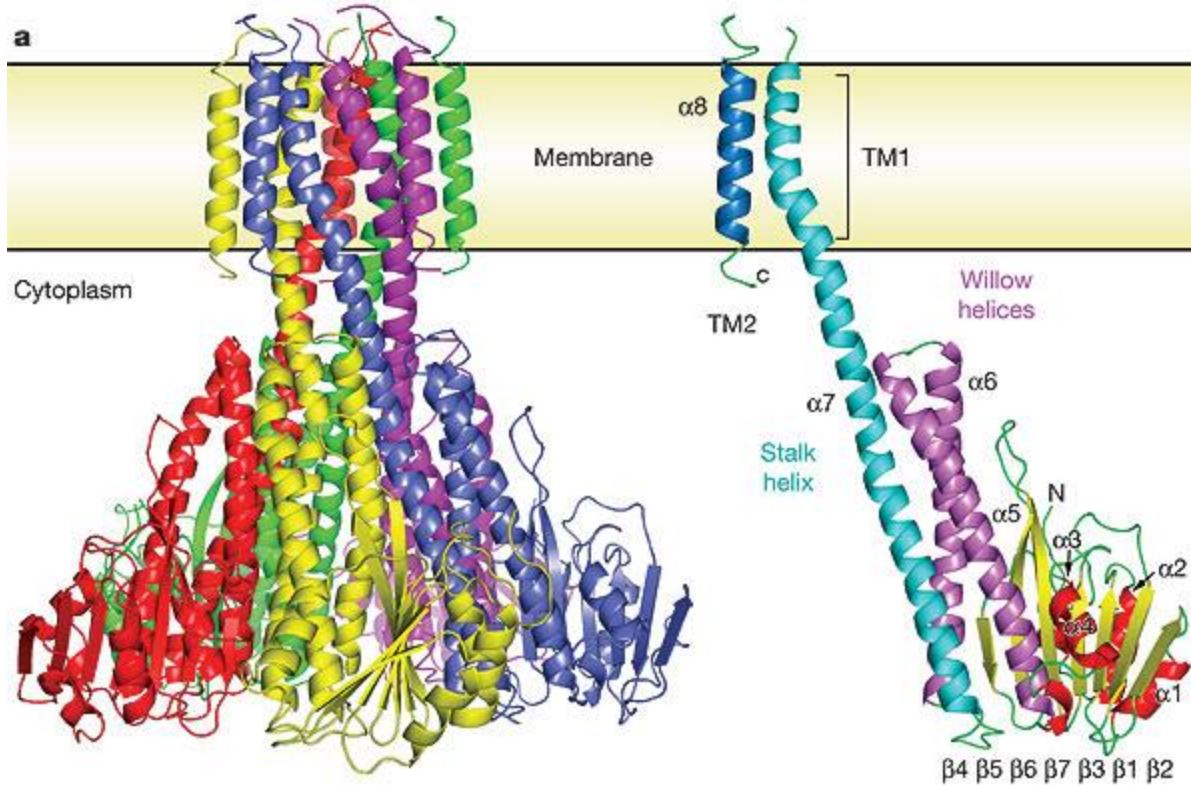


Некристаллографическая (локальная) симметрия имеет место только в ограниченной области пространства и не сохраняется для всего кристалла

# NON-CRYSTALLOGRAPHIC SYMMETRY



Некристаллографическая (локальная) симметрия имеет место только в ограниченной области пространства и не сохраняется для всего кристалла



Structure of the  
CorA Mg<sup>2+</sup> channel.

From the following  
article: [Crystal  
structure of the CorA  
Mg<sup>2+</sup> transporter](#)

*Nature* 440, 833-837  
(6 April 2006)

# Уточнение модели

- «Ручная» правка модели. (Комбинированные синтезы Фурье. Графические станции.)
- Автоматическое уточнение. (Программы уточнения. Уточнение параметров 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.*

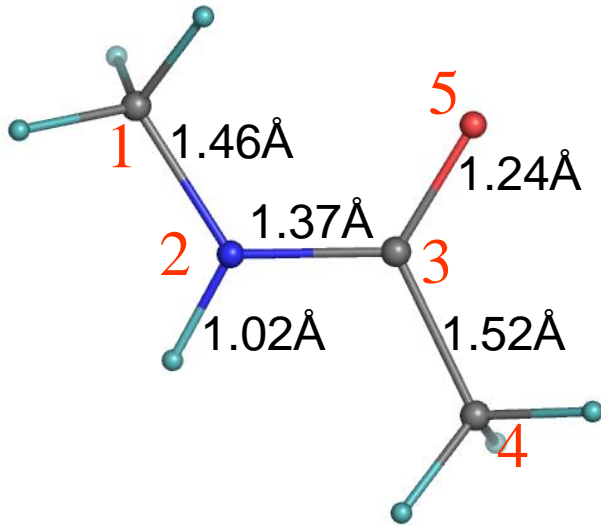
## Уточнение, как задача минимизации:

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

Проблемы:

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

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

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

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

Составной критерий

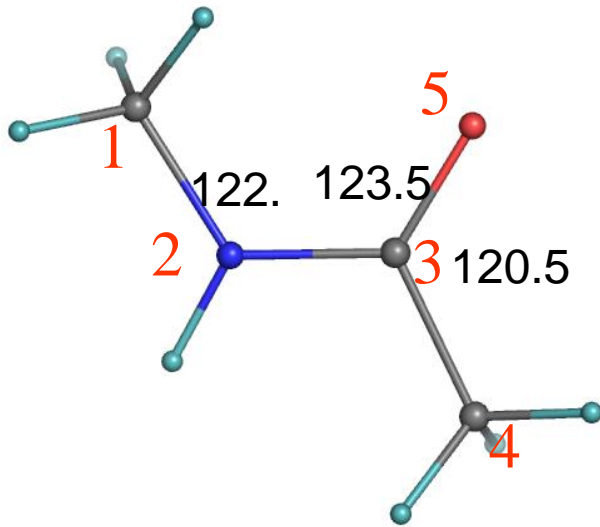
$$R = w_{X\text{-ray}} R_{X\text{-ray}} + w_{dist} R_{dist}$$

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



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

# Стереохимические ограничения



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

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

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

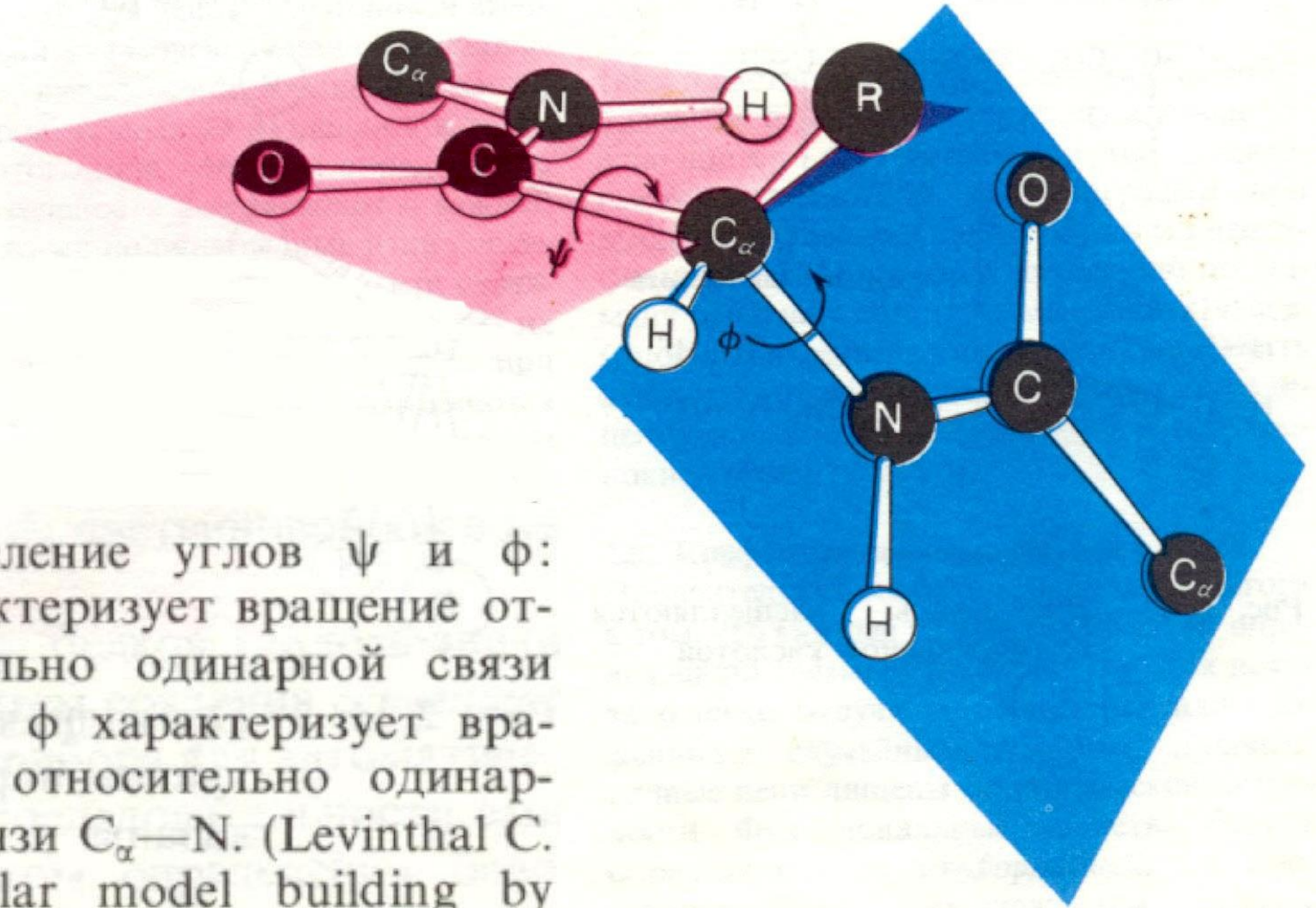
Хотим иметь

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

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

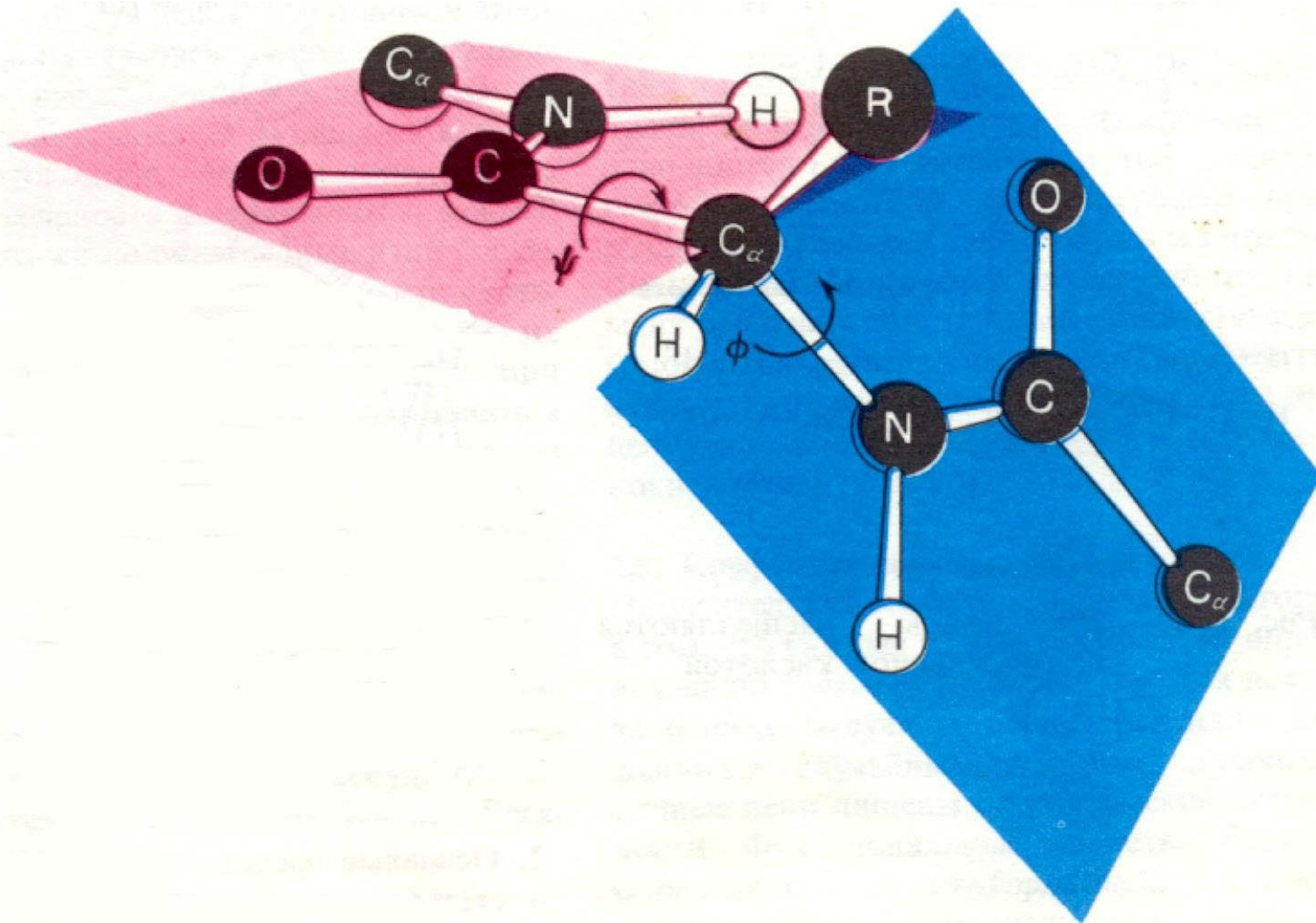
$$R = w_{X-ray} R_{X-ray} + w_{dist} R_{dist} + w_{angle} R_{angle}$$

## Двугранные углы

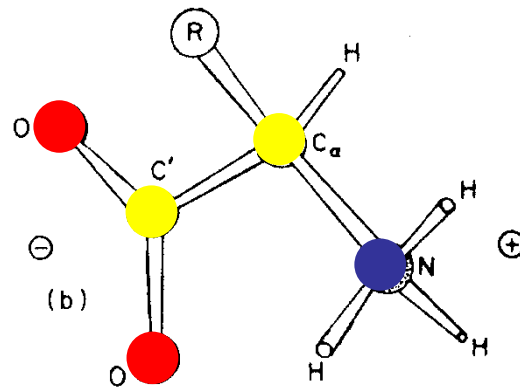
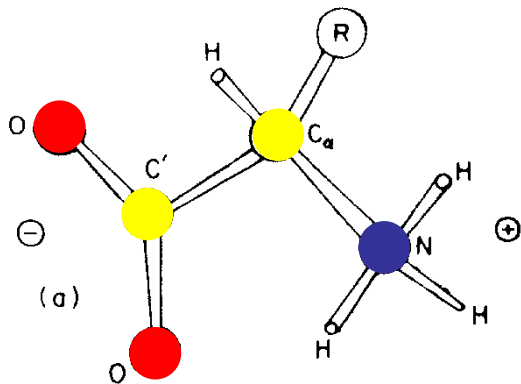


Определение углов  $\psi$  и  $\phi$ :  
 $\psi$  характеризует вращение относительно одинарной связи  $C_\alpha - C$ ;  $\phi$  характеризует вращение относительно одинарной связи  $C_\alpha - N$ . (Levinthal C. Molecular model building by computer, Scientific American, Inc., 1966.).

# Плоские группы



# Хиральность



# 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  $w_{X-ray} \sum_{hkl} (F_{hkl}^{calc} - F_{hkl}^{obs})^2 + w_{geom} R_{geom} \Rightarrow \min$

phenix.refine (P.Afonine *et al.*)

REFMAC (G.Murshudov *et al.*)

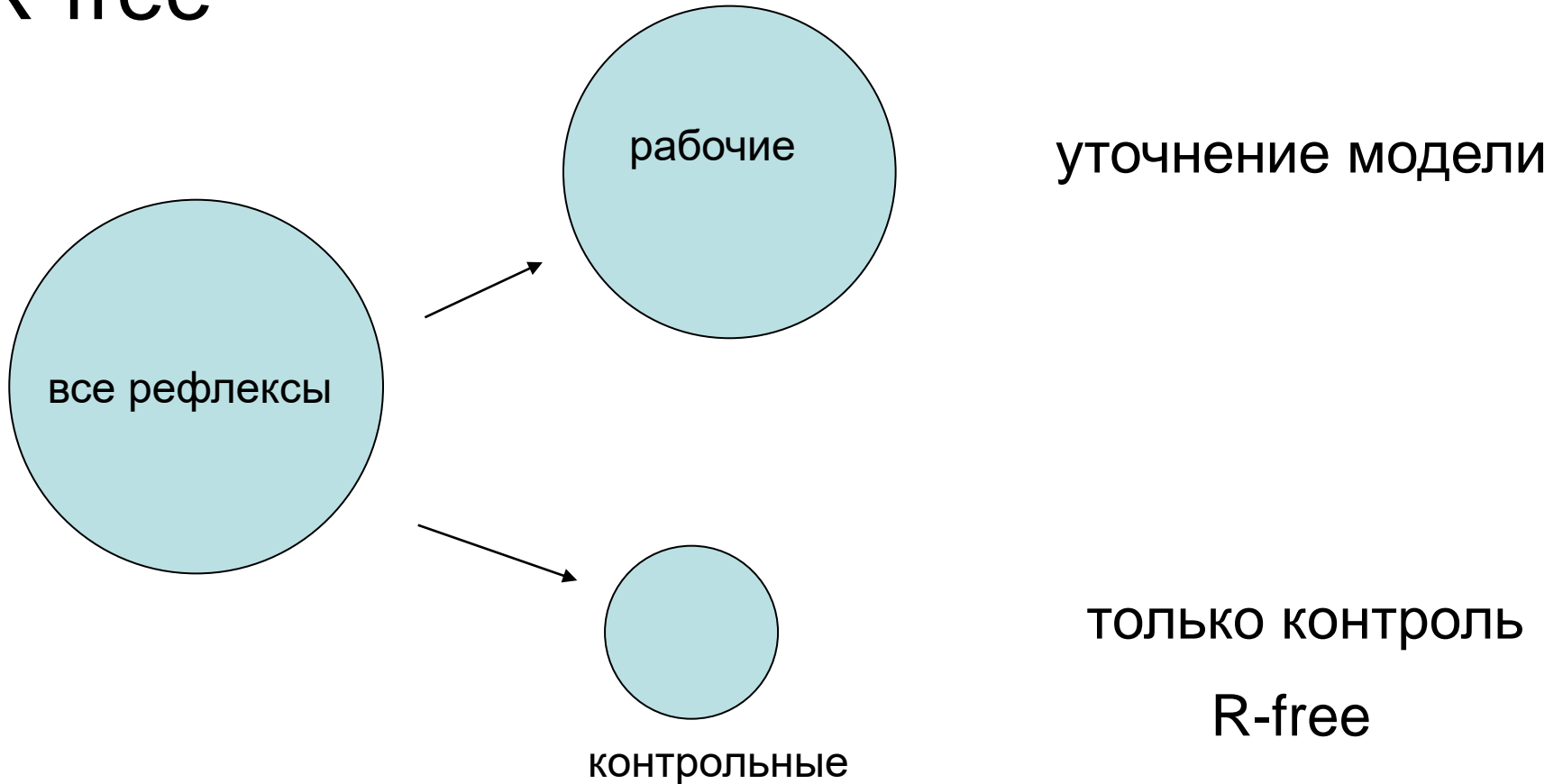
SHELX (G. Sheldrick)

BUSTER (G. Bricogne *et al.*)



*Снижение величины минимизируемого критерия может не сопровождаться улучшением параметров модели.*

R-free



```
REMARK      3
REMARK      3  FIT TO DATA USED IN REFINEMENT.
REMARK      3  CROSS-VALIDATION METHOD           : THROUGHOUT
REMARK      3  FREE R VALUE TEST SET SELECTION    : RANDOM
REMARK      3  R VALUE           (WORKING + TEST SET) : 0.15621
REMARK      3  R VALUE           (WORKING SET)       : 0.15185
REMARK      3  FREE R VALUE                               : 0.19471
REMARK      3  FREE R VALUE TEST SET SIZE      (%)    : 10.1
REMARK      3  FREE R VALUE TEST SET COUNT        : 5989
REMARK      3
```

# Улучшение моделирования структуры

*Для того, чтобы рассчитать теоретические значения структурных факторов, необходимо аккуратно рассчитать распределение электронной плотности в объекте. Координат атомов для этого не достаточно.*

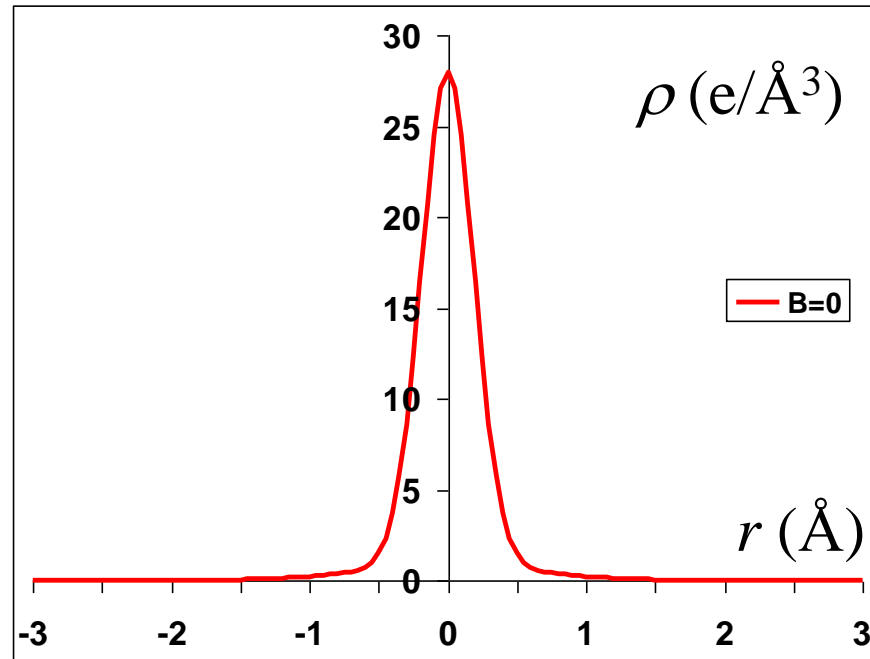
Urzhumtsev A.G., Lunin V.Y. (2019). Introduction to crystallographic refinement of macromolecular atomic models. *Crystallography Reviews*. 25:3, 164-262.

**Вода:** от 20 до 80% объема элементарной ячейки кристалла занято водой. Ее нельзя игнорировать при аккуратных расчетах.

- Связанная вода. (Аналогично атомам молекулы белка, но водородные связи вместо ковалентных).
- Неупорядоченный растворитель – bulk solvent. (Равномерное распределение плотности в области растворителя).

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

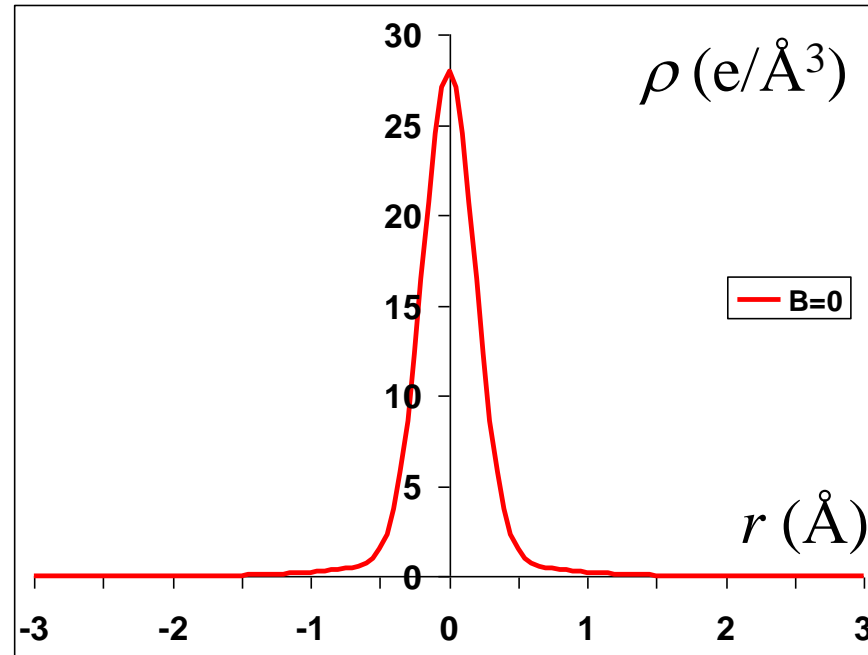
Динамическая и  
статическая  
неупорядоченность



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

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

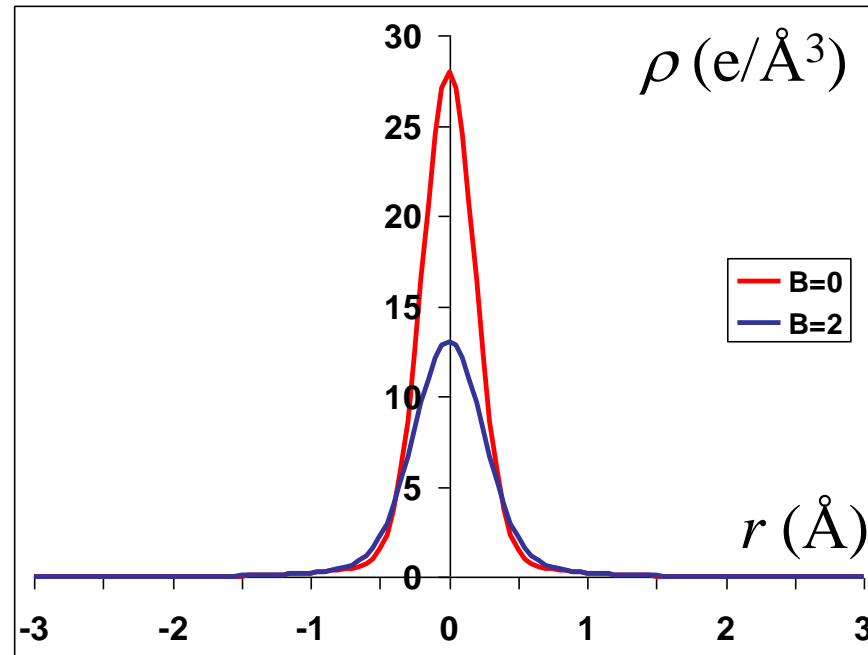
# Радиальное распределение электронной плотности в атоме углерода при разных значениях температурного фактора $B$ .



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

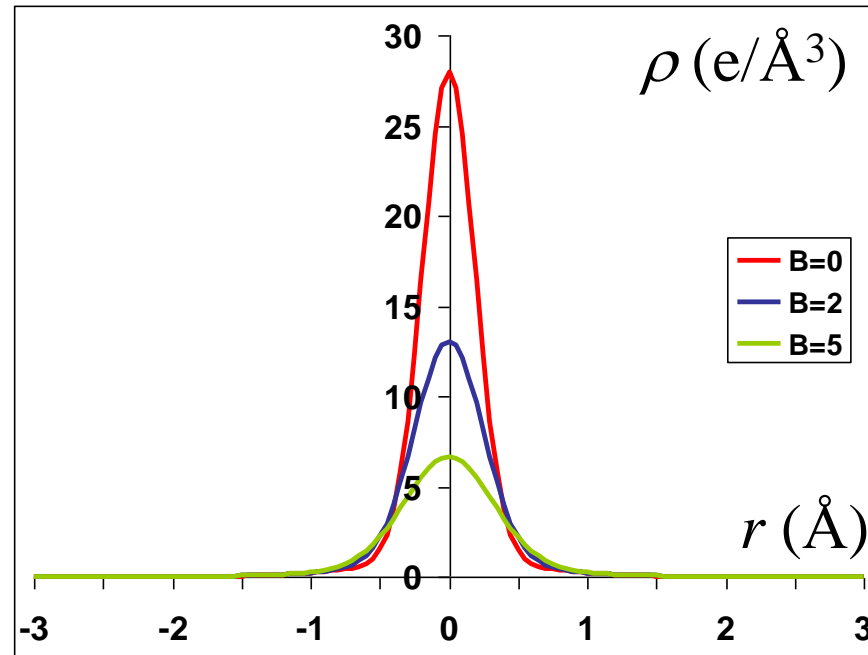
«Расплывание» плотности моделируется увеличением ширины гауссовых пиков.

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



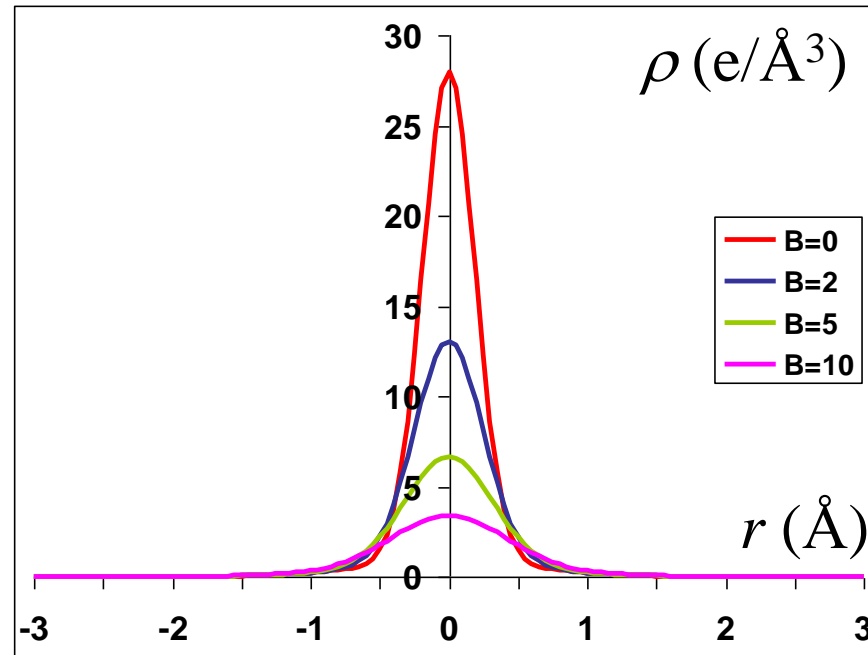
$$\rho(\mathbf{r}) = \sum_{j=1}^5 C_j \left( \frac{4\pi}{B_j + B} \right)^{3/2} \exp\left( -\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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



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

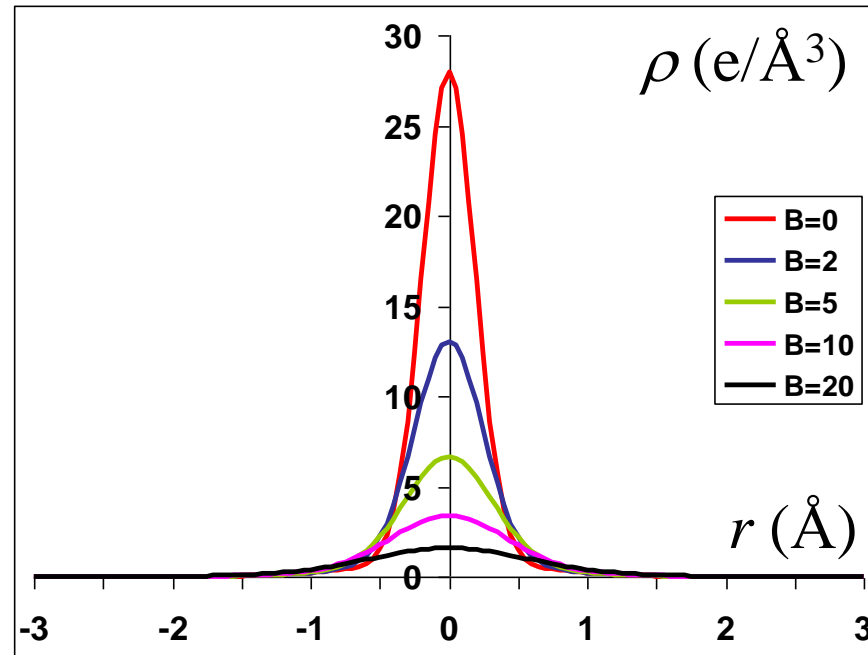
# Радиальное распределение электронной плотности в атоме углерода при разных значениях температурного фактора $B$ .



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



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



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

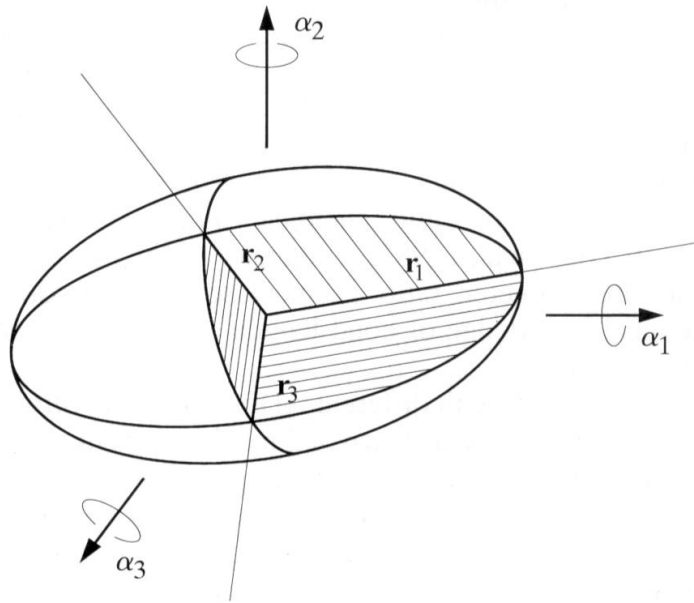
# Изотропный температурный фактор Atomic Displacement Parameter (ADP)

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

$$f(s) \Rightarrow f(s) \exp\left[ -B \frac{s^2}{4} \right]$$

ATOM	30	N	SER	2	13.117	9.840	39.210	1.000	<b>12.49</b>
------	----	---	-----	---	--------	-------	--------	-------	--------------

# 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  $\exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{12}hka^*b^* \cos \gamma^* + 2U_{13}hla^*c^* \cos \beta^* + 2U_{23}klb^*c^* \cos \alpha^*)]$ .

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

# TLS - уточнение

TLS-параметры описывают вибрацию единой группы атомов.

В модели выделяются группы атомов, движущиеся как жесткое тело.

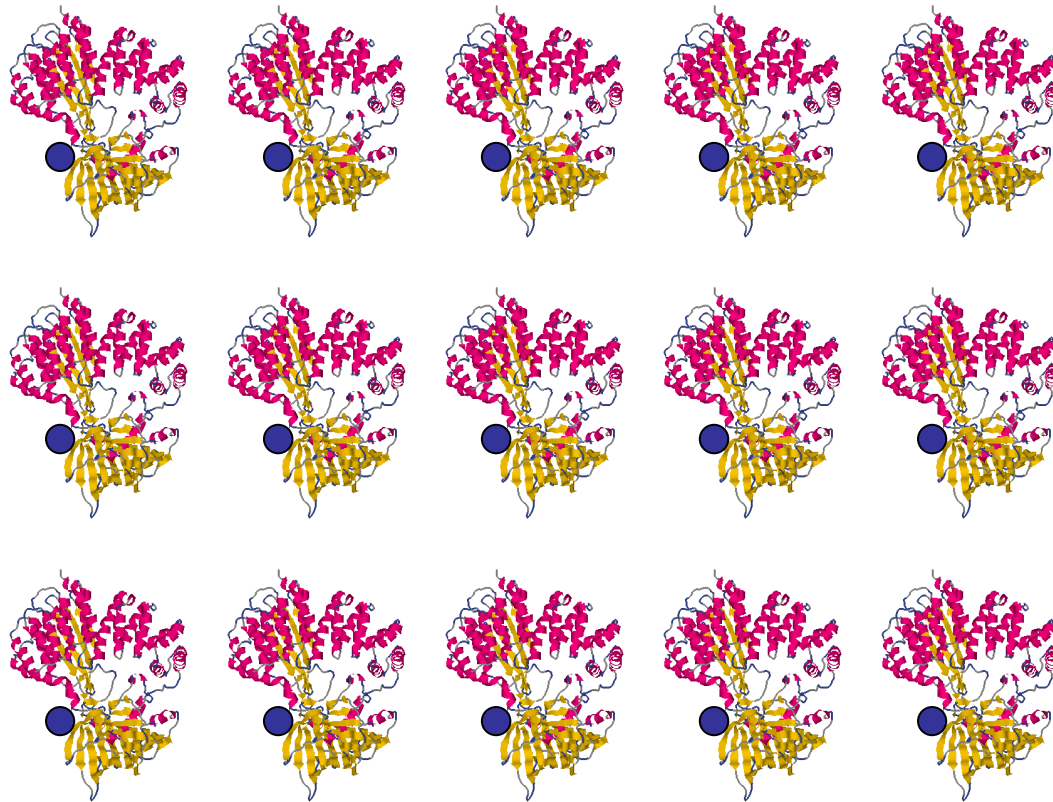
Каждая группа совершает возвратно-поступательные и вращательные колебания.

Для описания движения требуется 20 параметров на всю группу.

Если группы большие (больше 20 атомов), то такое описание требует меньше параметров, чем индивидуальные изотропные температурные факторы атомов.

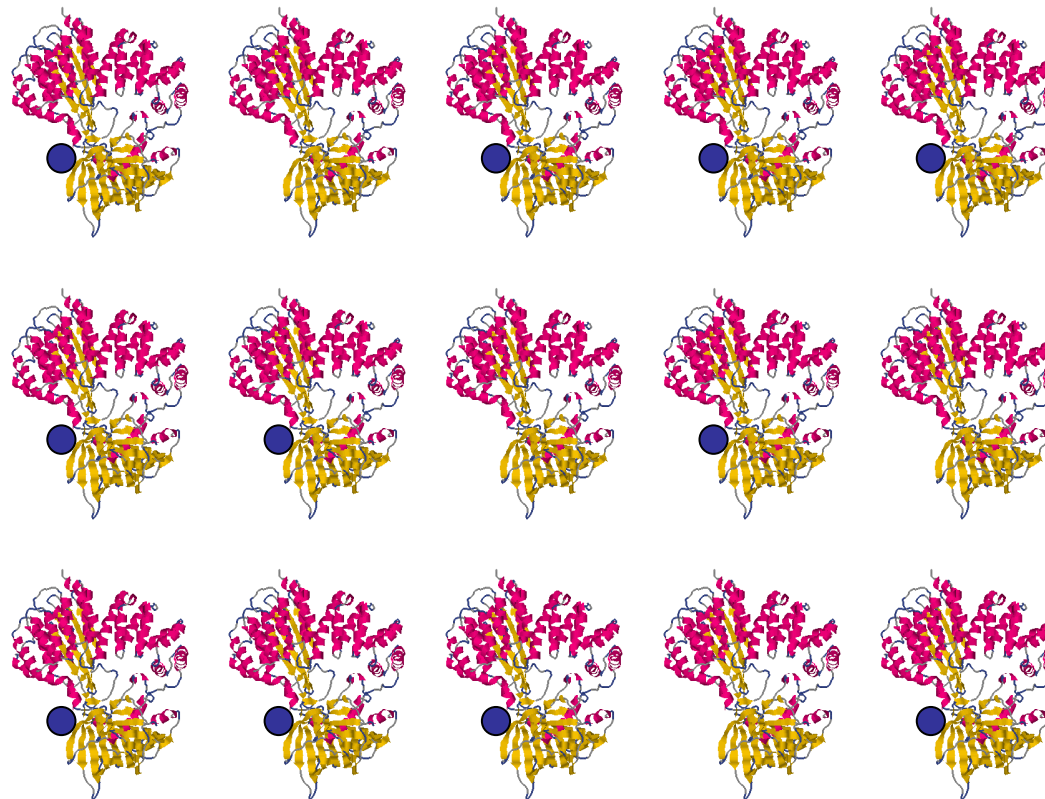
# Static disorder

*Коэффициент заполнения / заселенность / occupancy*



Идеальный кристалл - содержимое всех элементарных ячеек идентично.

# Коэффициент заполнения / заселенность / occupancy



"Реальный" кристалл - молекула "воды" присутствует только в 73% элементарных ячеек.

# Коэффициент заполнения / заселенность / occupancy

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

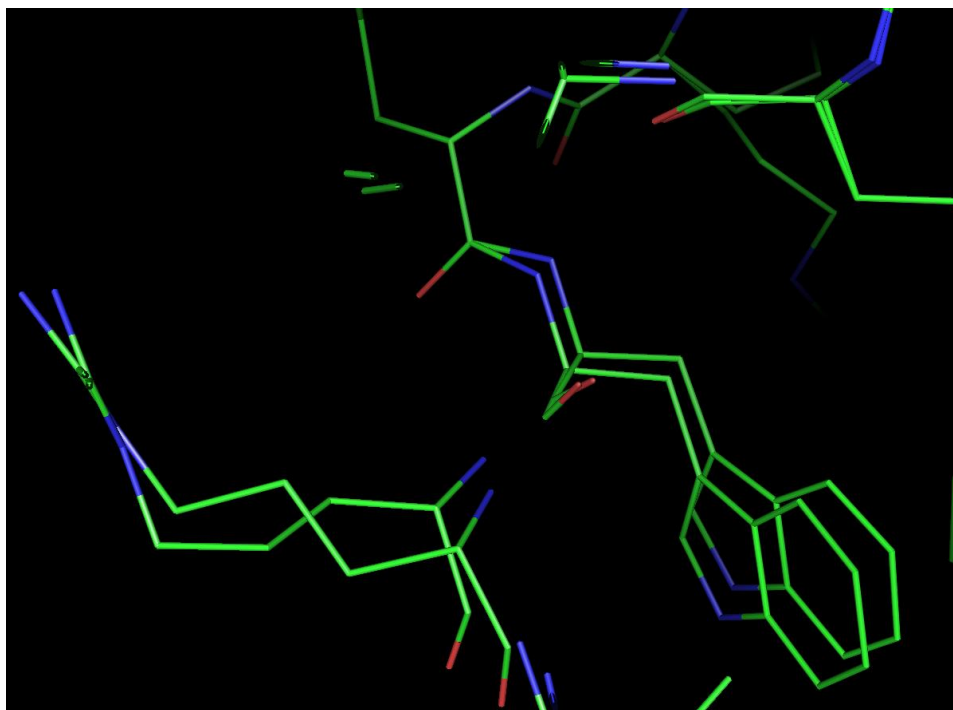
АТОМ	6626	О	НОН	4951	21.882	15.425	26.797	<b>0.500</b>	8.74
					<i>X</i>	<i>Y</i>	<i>Z</i>	<i>T</i>	<i>B</i>

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

$$f_j(s) \Rightarrow T_j f_j(s) \exp\left[ -B_j \frac{s^2}{4} \right]$$

# Коэффициент заполнения / заселенность / occupancy

## Альтернативные конформации



ATOM	5490	N	ATRP	295	29.848	2.643	16.199	<b>0.499</b>	7.13
ATOM	5514	N	BTRP	295	30.271	2.787	16.200	<b>0.501</b>	6.52



# Likelihood-based refinement

"Потерянные" атомы – не знаем координат, но знаем,  
что эти атомы “где-то” есть.

Правдоподобие частичной модели

Вероятность получить для всех рефлексов

$$L(\{\mathbf{r}_j\}) = \left| F_{hkl}^{part}(\{\mathbf{r}_j\}) + F_{hkl}^{lost}(\{\mathbf{u}_k\}) \right| = F_{hkl}^{obs}$$

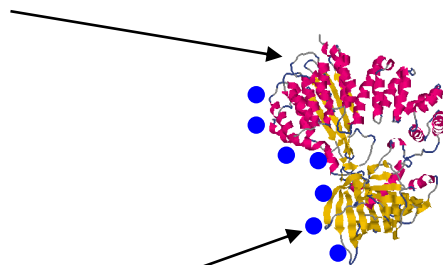
выбирая  $\{\mathbf{u}_k\}$  **случайным** образом.

*ML*-уточнение:  $L(\{\mathbf{r}_j\}) \Rightarrow \max$

По сути, мы ищем “заготовку”  $\{\mathbf{r}_j\}$ , которую легче всего превратить в правильную модель, добавив недостающие атомы.

Частичная модель

$$\left\{ \mathbf{r}_j \right\}_{j=1}^M \quad \vec{\mathbf{F}}_{hkl}^{part}$$



$$\vec{\mathbf{F}}_{hkl}^{full} = \vec{\mathbf{F}}_{hkl}^{part} + \vec{\mathbf{F}}_{hkl}^{lost}$$

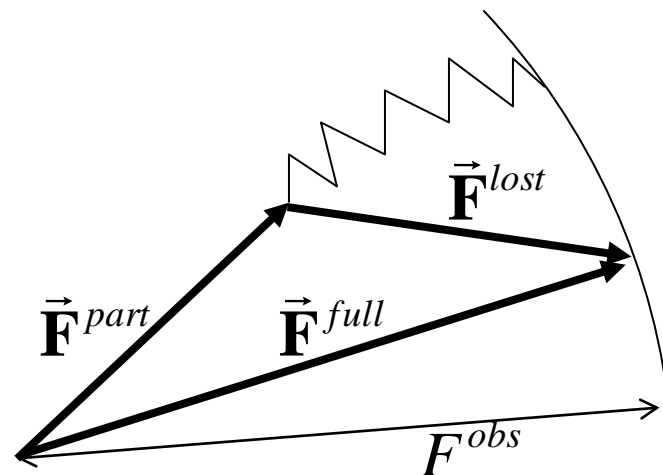
$$F_{hkl}^{full} = F_{hkl}^{obs}$$

"Потерянные" атомы

$$\left\{ \mathbf{u}_k \right\}_{k=M+1}^N \quad \vec{\mathbf{F}}_{hkl}^{lost}$$

Полная модель

$$\left\{ \mathbf{r}_j \right\}_{j=1}^M + \left\{ \mathbf{u}_k \right\}_{k=M+1}^N \quad \vec{\mathbf{F}}_{hkl}^{full}$$



“Обычное” уточнение

$$R_{X\text{-ray}} = \sum_{hkl} \left[ F_{hkl}^{part}(\{\mathbf{r}_j\}) - F_{hkl}^{obs} \right]^2 \Rightarrow \min$$

Подгонка  $F^{part}$  к  $F^{obs}$  может сдвигать координаты атомов с их правильных значений.

# Likelihood-based refinement

"Потерянные" атомы – не знаем координат, но знаем,  
что эти атомы “где-то” есть.

Правдоподобие частичной модели

Вероятность получить для всех рефлексов

$$L(\{\mathbf{r}_j\}) = \left| F_{hkl}^{part}(\{\mathbf{r}_j\}) + F_{hkl}^{lost}(\{\mathbf{u}_k\}) \right| = F_{hkl}^{obs}$$

выбирая  $\{\mathbf{u}_k\}$  **случайным** образом.

*ML*-уточнение:  $L(\{\mathbf{r}_j\}) \Rightarrow \max$

По сути, мы ищем “заготовку”  $\{\mathbf{r}_j\}$ , которую легче всего превратить в правильную модель, добавив недостающие атомы.

REMARK 3

REMARK 3 REFINEMENT.

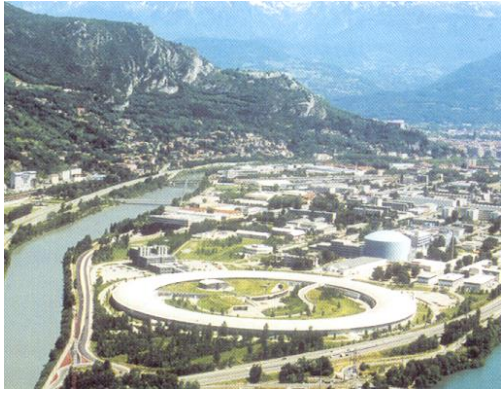
REMARK 3 PROGRAM : REFMAC

REMARK 3 AUTHORS : MURSHUDOV, VAGIN, DODSON

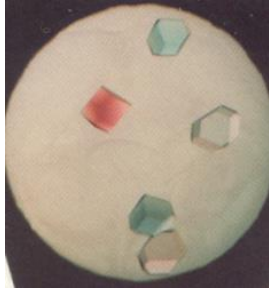
REMARK 3

REMARK 3 REFINEMENT TARGET : MAXIMUM LIKELIHOOD

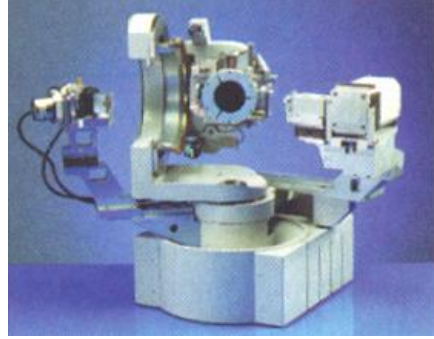
*X-ray structure analysis*



X-ray source

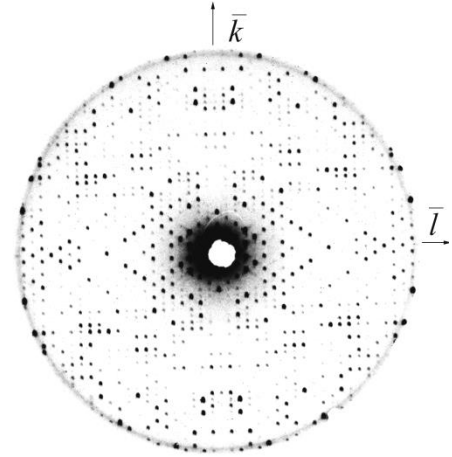


crystal

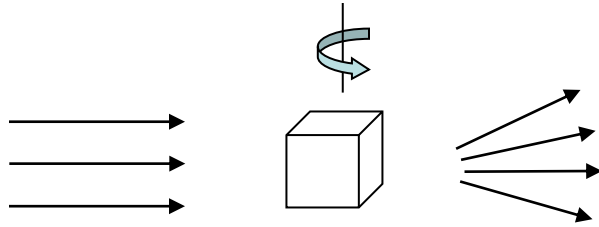


registration

Intensities  $I_s$  of diffracted beams



**X-ray experiment**



Kinematic theory of diffraction

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

$$F_s^2 \propto I_s$$

<b>h</b>	<b>k</b>	<b>l</b>	<b>F</b>	<b><math>\sigma</math></b>
0	0	6	46.09	2.74
0	0	8	212.95	5.00
0	0	20	98.75	3.15
0	1	6	188.33	5.06
0	1	7	14.88	8.00
0	1	8	226.02	7.9

# The solving of the structure

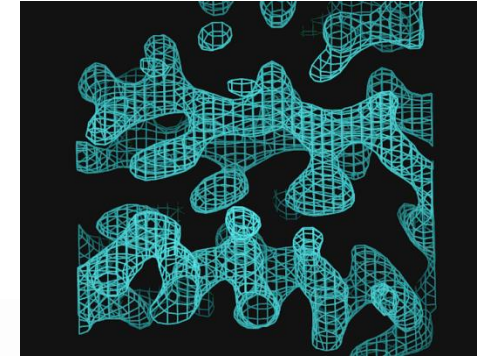
*X-ray structure analysis*

<b>h</b>	<b>k</b>	<b>l</b>	<b>F</b>	<b><math>\sigma</math></b>
0	0	6	46.09	2.74
0	0	8	212.95	5.00
0	0	20	98.75	3.15
0	1	6	188.33	5.06
0	1	7	14.88	8.00
0	1	8	226.02	7.9

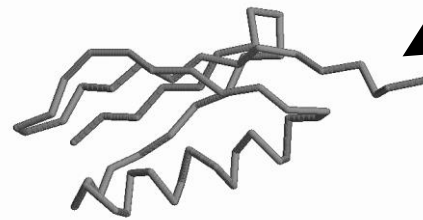
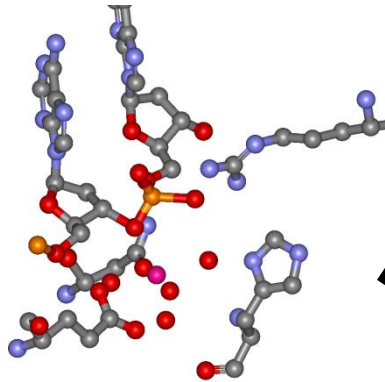


**The phase problem**

$$\Omega_{\kappa} = \{\mathbf{r} : \rho_S(\mathbf{r}) > \kappa\}$$



refined model



preliminary model

<b>Protein Data Bank</b>				<b>X</b>	<b>Y</b>	<b>Z</b>	<b>T</b>	<b>B</b>		
	5	CA	MET A	1	0	1.530	3.431	5.646	1.00	9.39
		C	MET A	1	0	1.452	4.960	5.500	1.00	7.10
		O	MET A	1	0	1.808	5.574	4.503	1.00	10.54
ATOM	9	CB	THR A	2	0	-0.430	7.045	7.578	1.00	23.54
ATOM	10	OG1	THR A	2	0	-1.549	7.435	6.701	1.00	27.09
ATOM	11	CG2	THR A	2	0	-0.265	7.733	8.906	1.00	21.71

# Рентгеновские лазеры (XFEL)

Ультракороткие мощные импульсы

“Обдирание электронов” – ионизация – кулоновский взрыв

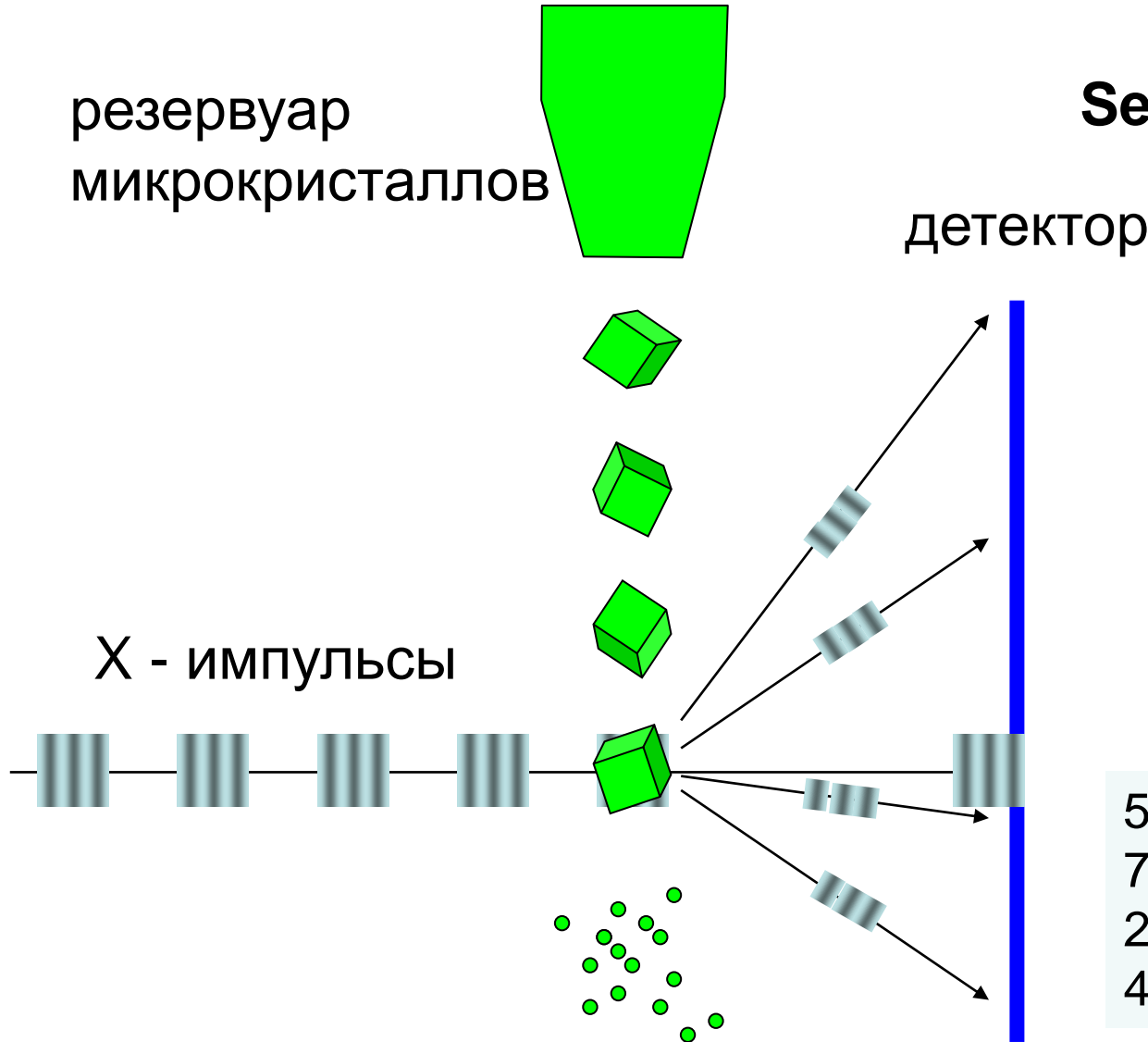
Регистрация до разрушения

Метод работает

Область использования - нанокристаллы

# «Проточная» рентгеновская кристаллография

## Serial crystallography



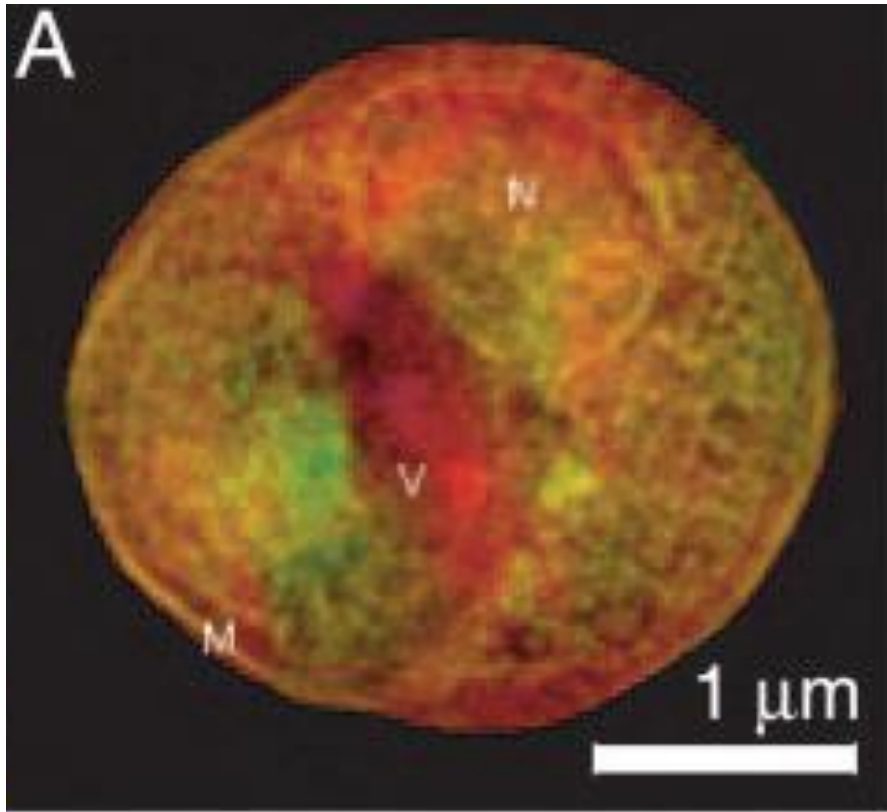
Одна  
рентгенограмма с  
одного кристалла

500 000 crystals  
762 000 frames  
20 000 contain a signal  
4 000 - indexed



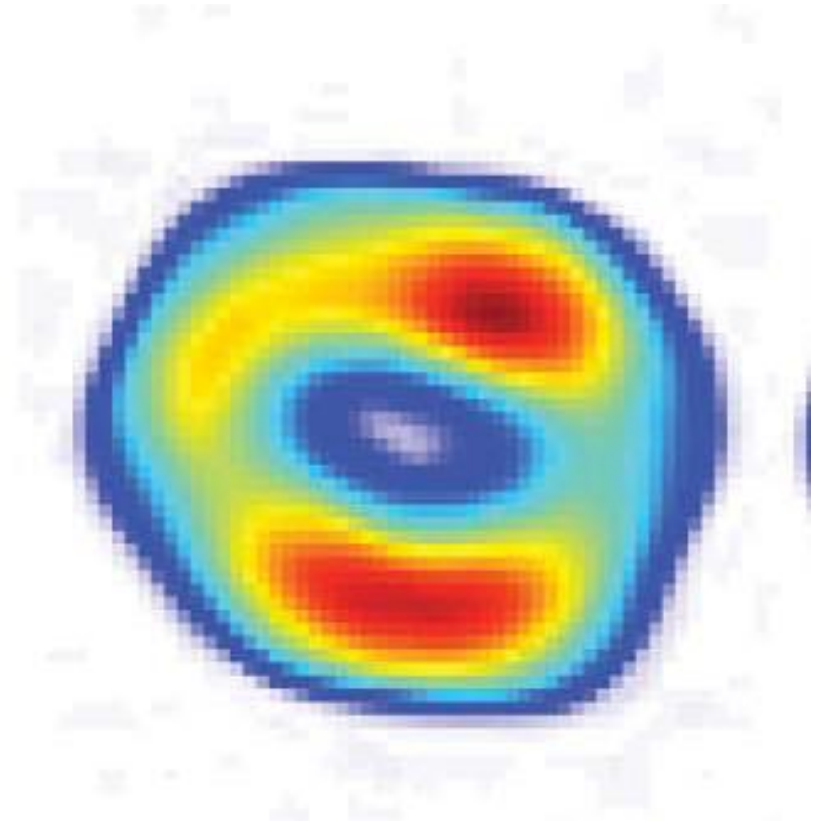
# Исследование отдельных частиц

## Дрожжевая клетка



**Shapiro, Thibault, Beetz, Elser,  
Howells, Jacobsen, Kirz, Lima, Miao,  
Neiman, and Sayre**  
PNAS, 2005 vol. 102 no. 43 15343–  
5346

## Мимивирус



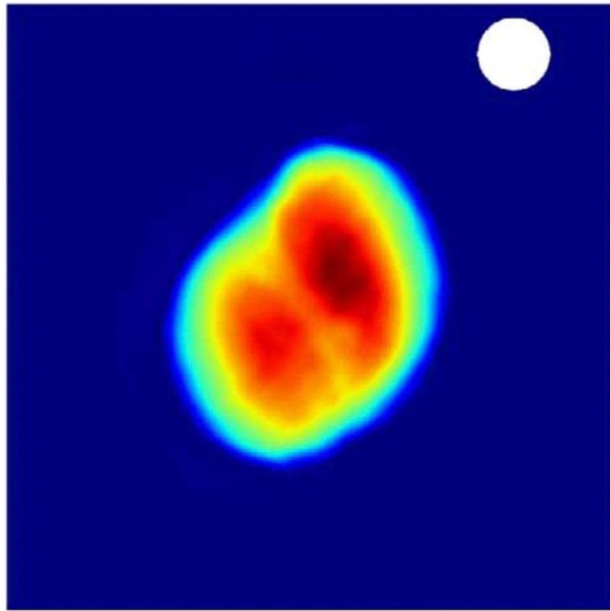
**Siebert et al., 2011, Nature, 470, 78-82**

# Imaging single cells in a beam of live cyanobacteria with an X-ray laser

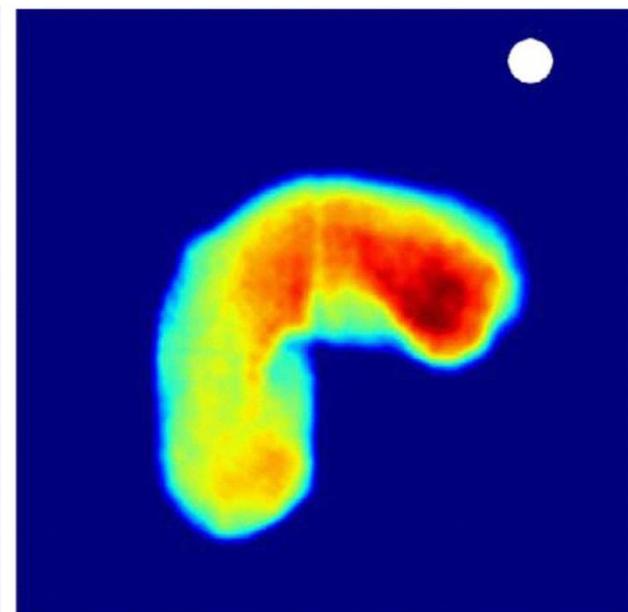
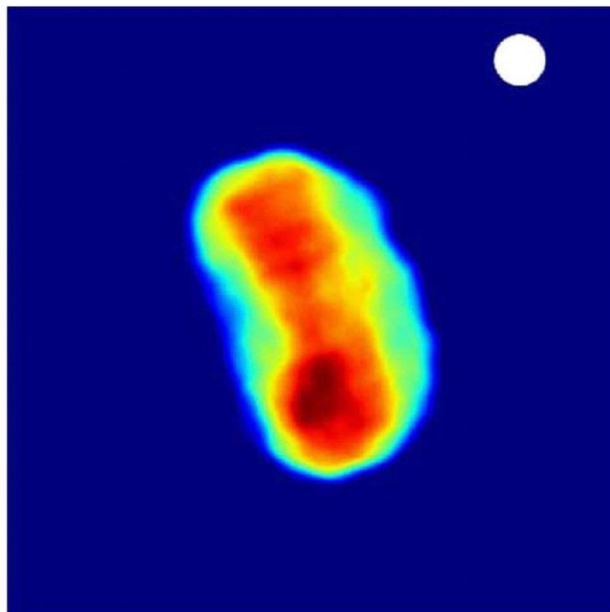
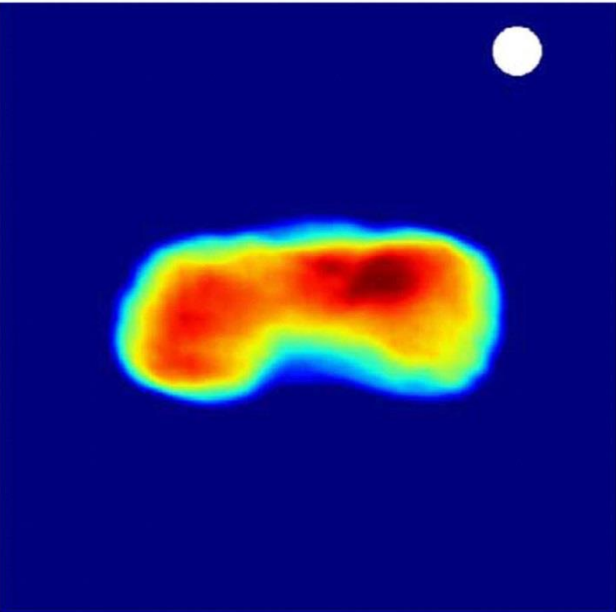
Schot *et al.*

Nature Communications 6, Article number: 5704

Published 11 February 2015



10,000 Å = 1 μm



Т.Бландел, Л.Джонсон. Кристаллография белка. "Мир", Москва, 1979

М.А.Порай-Кошиц. Основы структурного анализа химических соединений. Москва, "Высшая школа", 1989

Ч.Кантор, П.Шиммел. Биофизическая химия, том 2. Москва, "Мир", 1984

International Tables for Crystallography, vol. F. Crystallography of biological macromolecules. Second Edition. Wiley, 2011

И.Сердюк, Н.Закаи, Дж.Закаи. Методы в молекулярной биофизике. т.2. М.КДУ, 2010

Urzhumtsev A.G., Lunin V.Y. (2019). Introduction to crystallographic refinement of macromolecular atomic models. *Crystallography Reviews*. 25:3, 164-262.