

Введение в биологическую кристаллографию

Лекция 3, 2024 г.

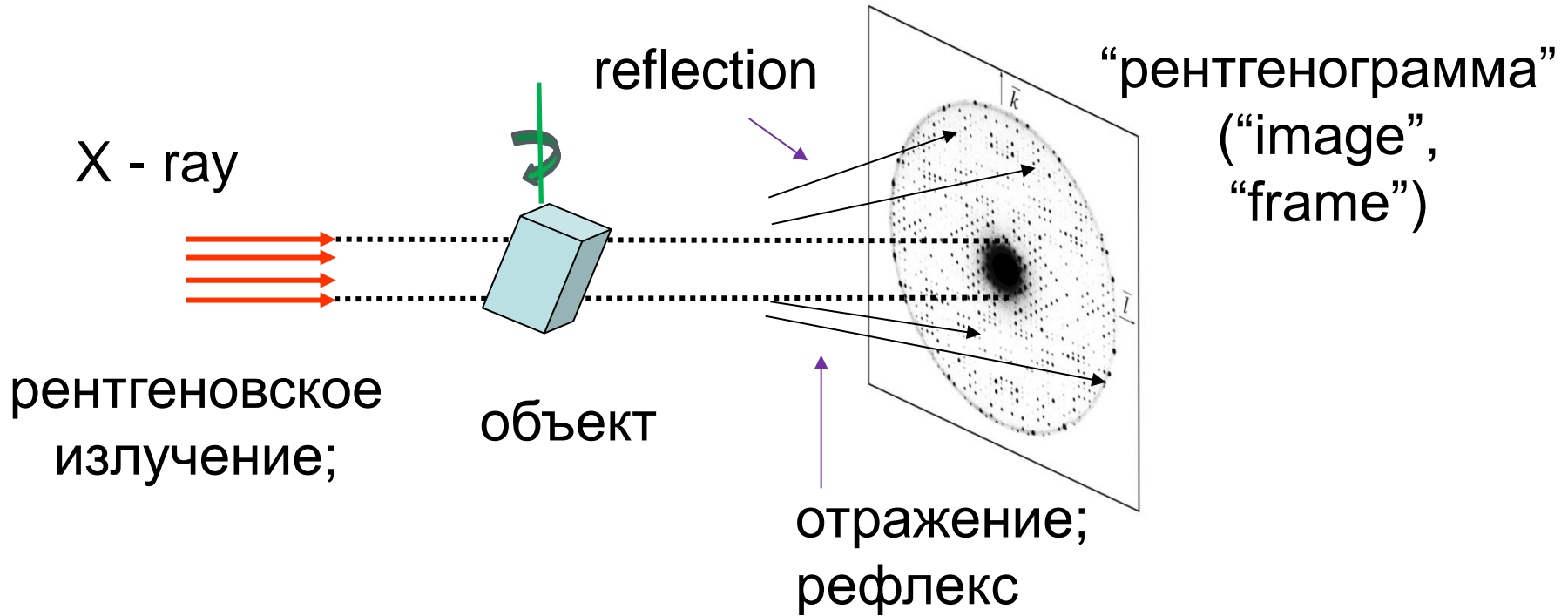
Владимир Юрьевич ЛУНИН

Институт Математических Проблем Биологии РАН

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

Пушино

Дифракционный эксперимент



- измеряем интенсивность (энергию) «отраженных» от объекта лучей;
- вращение объекта – трехмерный набор данных;
- разные объекты – разная картина рассеяния;
- ХОТИМ «восстановить» структуру объекта.

Первичный пучок

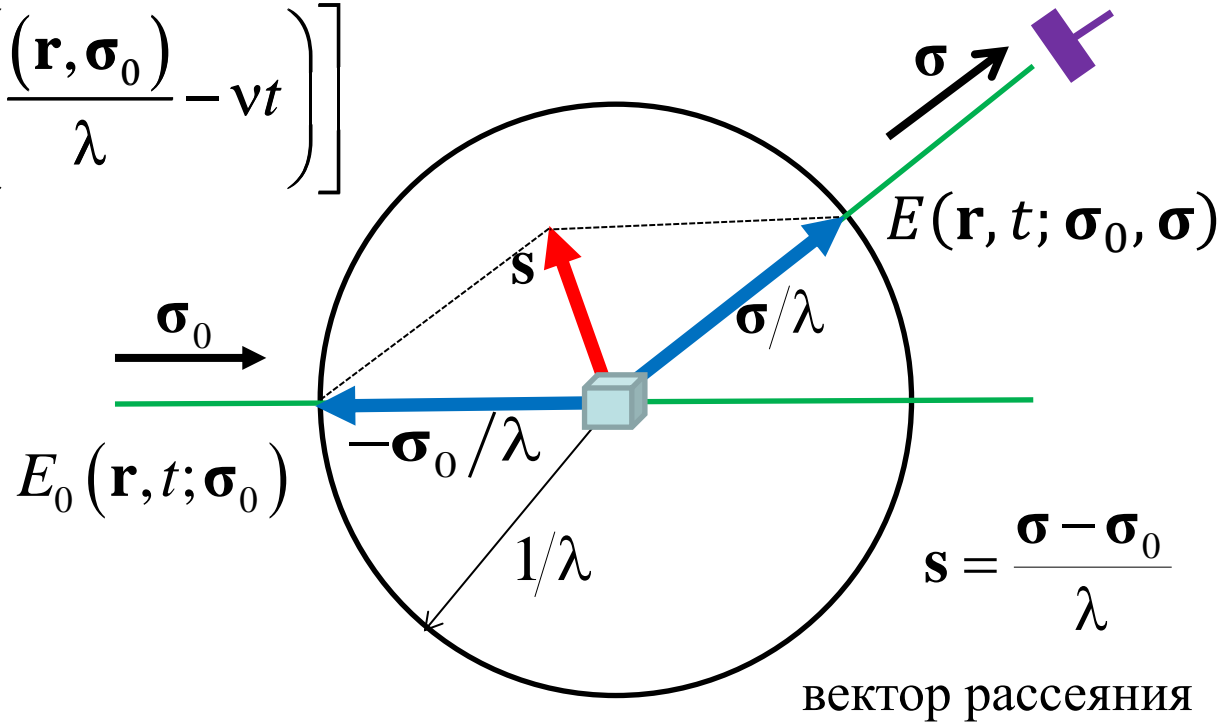
$$E_0(\mathbf{r}, t; \boldsymbol{\sigma}_0) = E_0 \sin \left[2\pi \left(\frac{(\mathbf{r}, \boldsymbol{\sigma}_0)}{\lambda} - \nu t \right) \right]$$

$$|\boldsymbol{\sigma}_0| = |\boldsymbol{\sigma}| = 1$$

λ - длина волны

R_{s-d} - расстояние
объект-детектор

$$\varepsilon \approx 10^{-12}$$



В окрестности пиксела детектора

$$E(\mathbf{r}, t; \boldsymbol{\sigma}_0, \boldsymbol{\sigma}) = \frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[2\pi \left(\frac{(\mathbf{r}, \boldsymbol{\sigma})}{\lambda} \right) - \varphi(\mathbf{s}) \right]$$

Вся информация об объекте в структурном факторе

$$F(\mathbf{s}), \varphi(\mathbf{s})$$

Волна на детекторе:

$$E(t; \boldsymbol{\sigma}_0, \boldsymbol{\sigma}) = E(t; \mathbf{s}) = \frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[2\pi \left(\frac{R_{s-d}}{\lambda} - \mathbf{v}t \right) - \varphi(\mathbf{s}) \right]$$

$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda}$ - вектор рассеяния, информация об эксперименте;

$F(\mathbf{s}), \varphi(\mathbf{s})$ - структурный фактор, информация о взаимном расположении электронов;

Измеряется интенсивность (энергия) рассеянных волн

$$I(\mathbf{s}) = \left(\frac{\varepsilon E_0}{R_{s-d}} \right)^2 F^2(\mathbf{s}) \Delta t \propto F^2(\mathbf{s})$$

Эксперимент дает набор величин модулей структурных факторов $F(\mathbf{s})$ для разных векторов рассеяния \mathbf{s} .

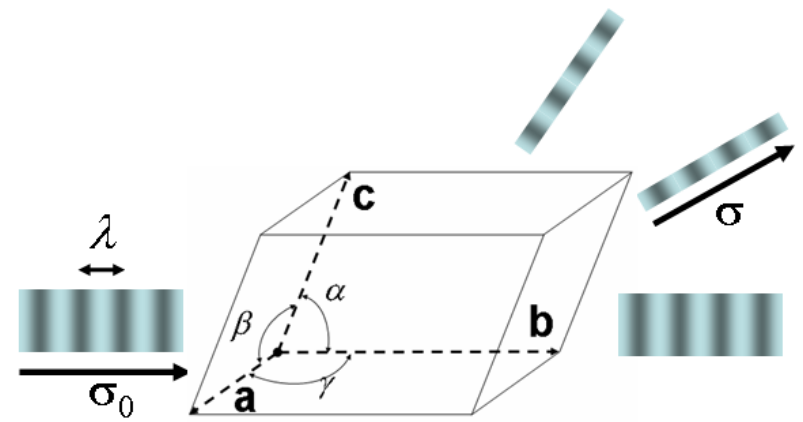
Информация о фазах $\varphi(\mathbf{s})$ в «стандартном» эксперименте теряется.

Рентгеновский эксперимент с монокристаллом позволяет измерить интенсивность волн, рассеянных в направлениях, определяемых условиями:

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

$$\mathbf{s}=(\boldsymbol{\sigma}_0-\boldsymbol{\sigma})/\lambda$$

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



h,k,l - целые

Амплитуда и фаза рассеянной волны зависят от распределения электронной плотности в элементарной ячейке и могут быть рассчитаны по формулам

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

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

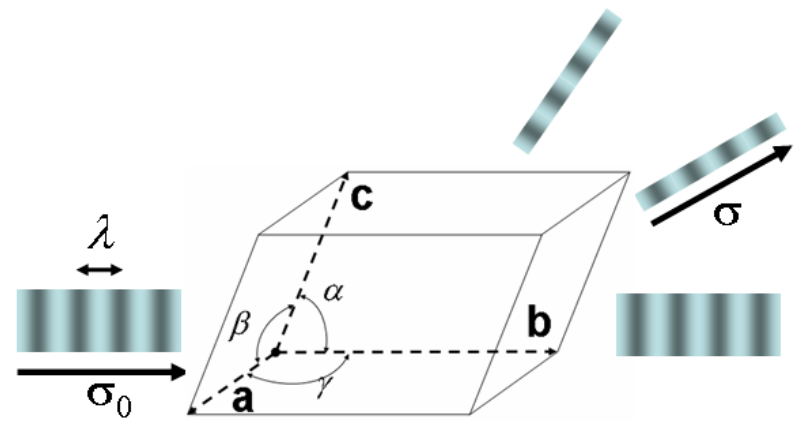
$$B_{hkl} = \int_V \rho(x, y, z) \sin[2\pi(hx + ky + lz)] dx dy dz$$

Рентгеновский эксперимент с монокристаллом позволяет измерить интенсивность волн, рассеянных в направлениях, определяемых условиями (Лауэ – Брэгга – Вульфа):

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

$$\mathbf{s}=(\boldsymbol{\sigma}_0-\boldsymbol{\sigma})/\lambda$$

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



h,k,l - целые

Амплитуда и фаза рассеянной волны зависят от распределения электронной плотности в элементарной ячейке и могут быть рассчитаны по формулам

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

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

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

Структурные факторы, отвечающие Брэгговским рефлексам совпадают с коэффициентами в разложении в ряд Фурье функции распределения электронной плотности.

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

Знание значений модулей и фаз структурных факторов дает возможность рассчитать "синтез Фурье электронной плотности"

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

Синтез Фурье рассчитывается по набору рефлексов S, для которых известны и модули и фазы структурных факторов.

Расчет значений фаз является центральной,

“Фазовой проблемой”

при расшифровке структуры методом РСА

Фазовая проблема



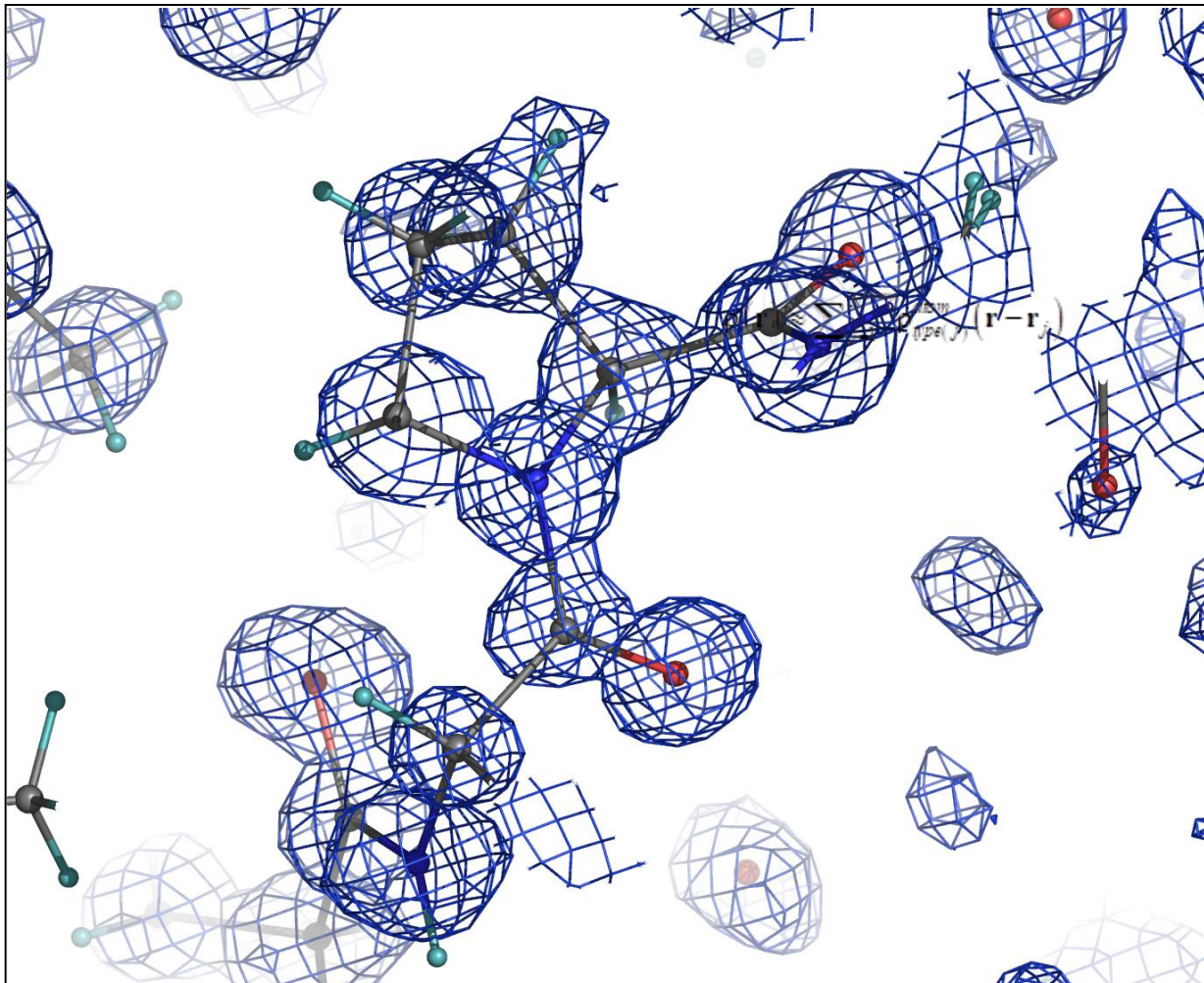
Молекулярное замещение
Molecular Replacement (MR)

Множественное изоморфное
замещение
*Multiple Isomorphous
Replacement (MIR)*

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

Интерпретация распределения электронной плотности

$$\rho(\mathbf{r}) \approx \sum_{j=1}^{N^{atoms}} \rho_{type(j)}^{atom}(\mathbf{r} - \mathbf{r}_j)$$

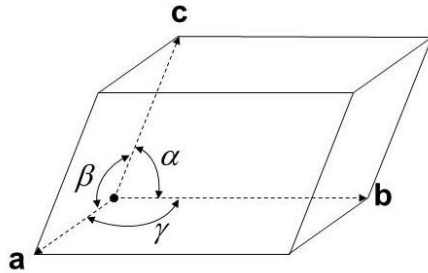


Уточнение
модели

PDB

«Прямое» и «обратное» пространства Real and reciprocal spaces

$\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$



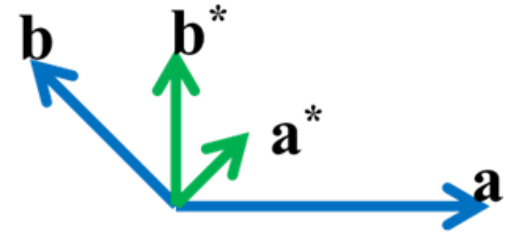
Базис обратного пространства

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

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

Для ортогонального базиса

$$\mathbf{a}^* = \frac{1}{a^2} \mathbf{a}, \quad \mathbf{b}^* = \frac{1}{b^2} \mathbf{b}, \quad \mathbf{c}^* = \frac{1}{c^2} \mathbf{c}$$



Координаты векторов

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

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

$$x = (\mathbf{r}, \mathbf{a}^*), \quad y = (\mathbf{r}, \mathbf{b}^*), \quad z = (\mathbf{r}, \mathbf{c}^*)$$

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

Скалярное произведение векторов

Условия дифракции Лауэ

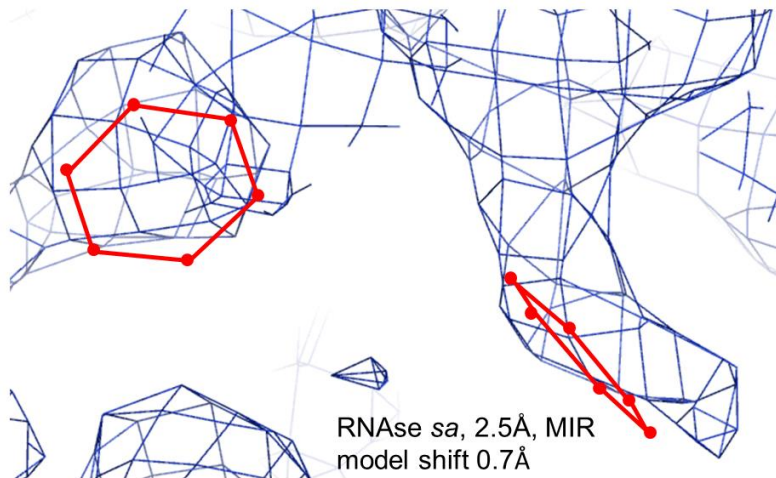
$$(\mathbf{r}, \mathbf{s}) = hx + ky + lz$$

$$\mathbf{s} \in \mathfrak{R}'$$

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.



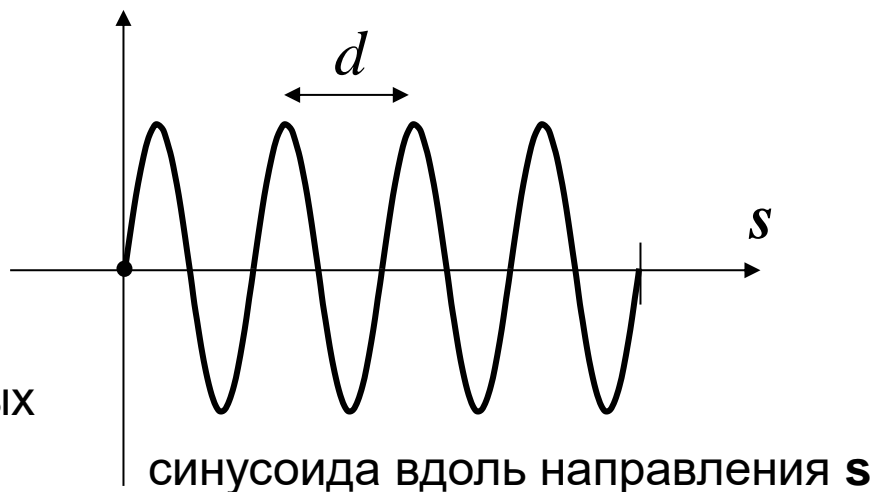
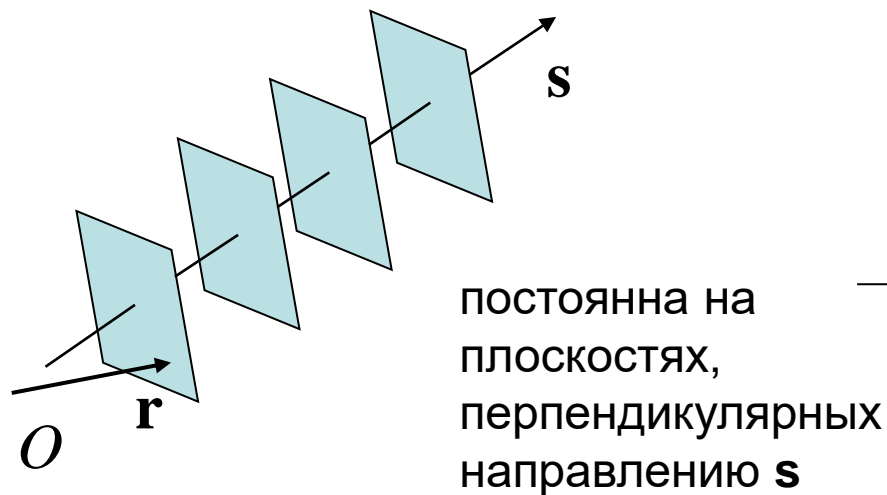
При работе с набором данных разрешения 2\AA , типичной оценкой точности определения координат атомов после уточнения является величина 0.02\AA .

Гармоника Фурье

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

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

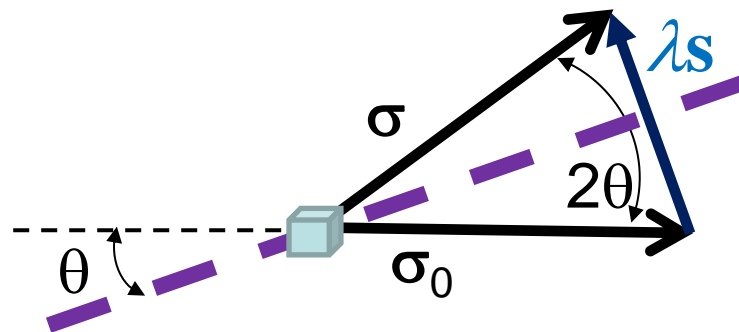
$$h = (\mathbf{s}, \mathbf{a}), k = (\mathbf{s}, \mathbf{b}), l = (\mathbf{s}, \mathbf{c})$$



Разрешение, соответствующее гармонике Фурье

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

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

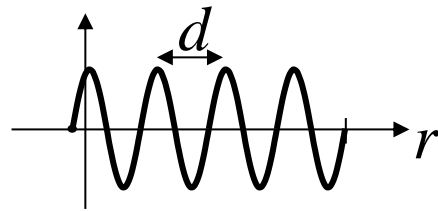
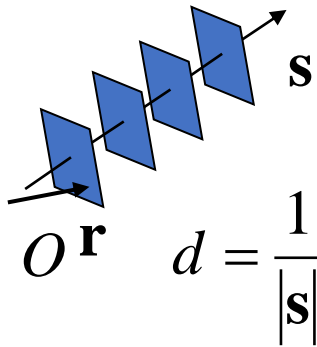


Синтез Фурье электронной плотности

$$\rho(\mathbf{r}) = \frac{1}{|V|} \sum_{\mathbf{s} \in S} F(\mathbf{s}) \cos(2\pi(\mathbf{s}, \mathbf{r}) - \varphi(\mathbf{s}))$$

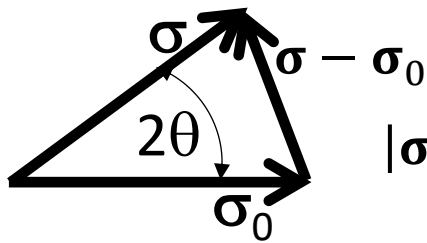
Разрешение, соответствующее рефлексу \mathbf{s}

$$H_{\mathbf{s}}(\mathbf{r}) = \cos 2\pi(\mathbf{s}, \mathbf{r})$$



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

2θ - угол рассеяния



$$|\sigma - \sigma_0| = 2 \sin \theta$$

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

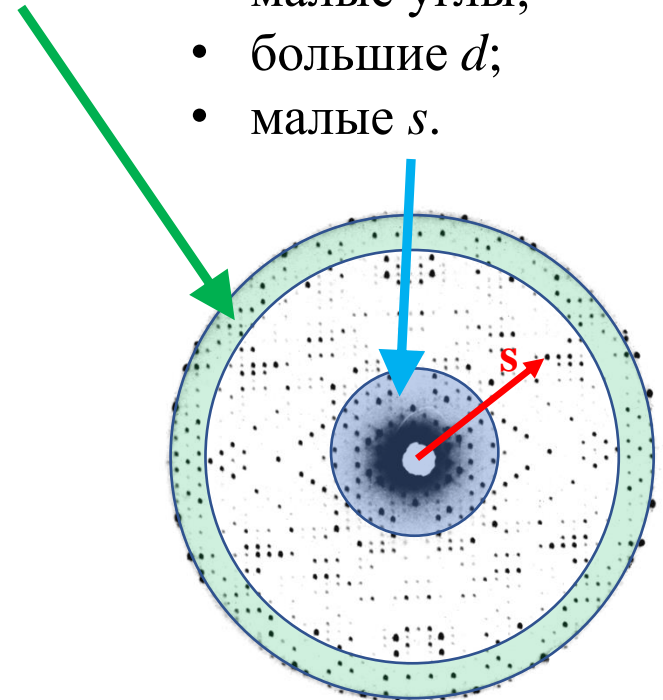
$$d = \frac{\lambda}{2 \sin \theta}$$

Высокое разрешение:

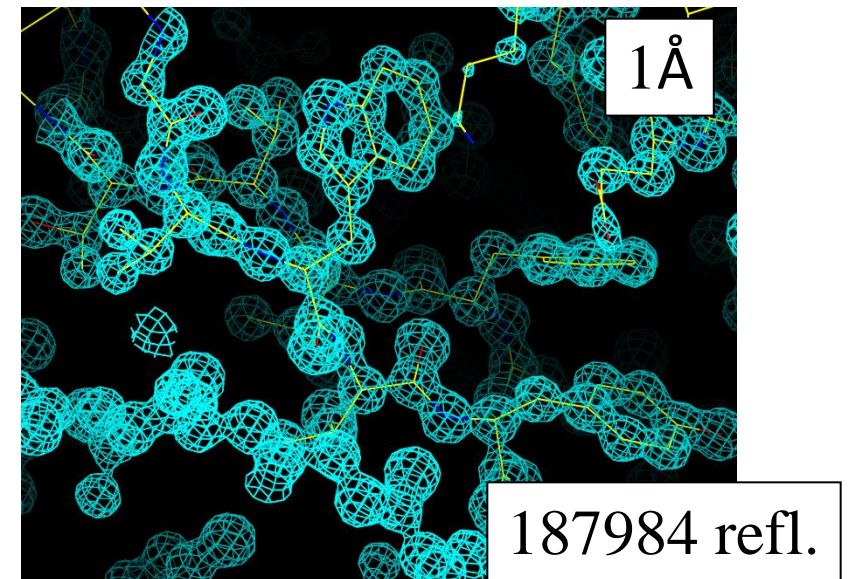
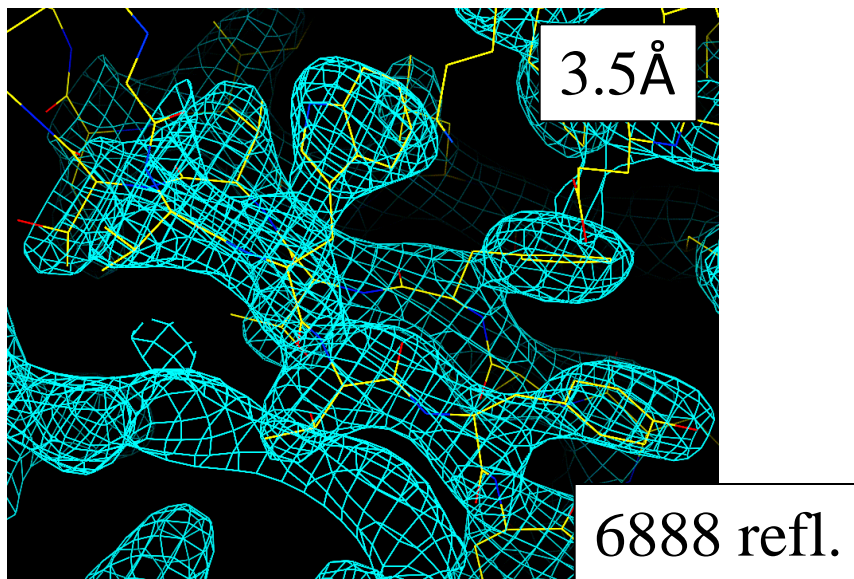
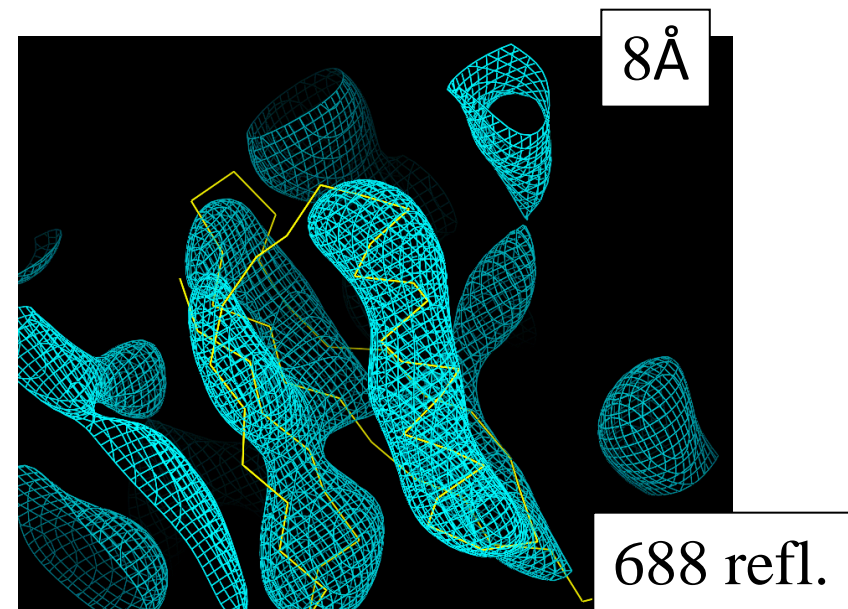
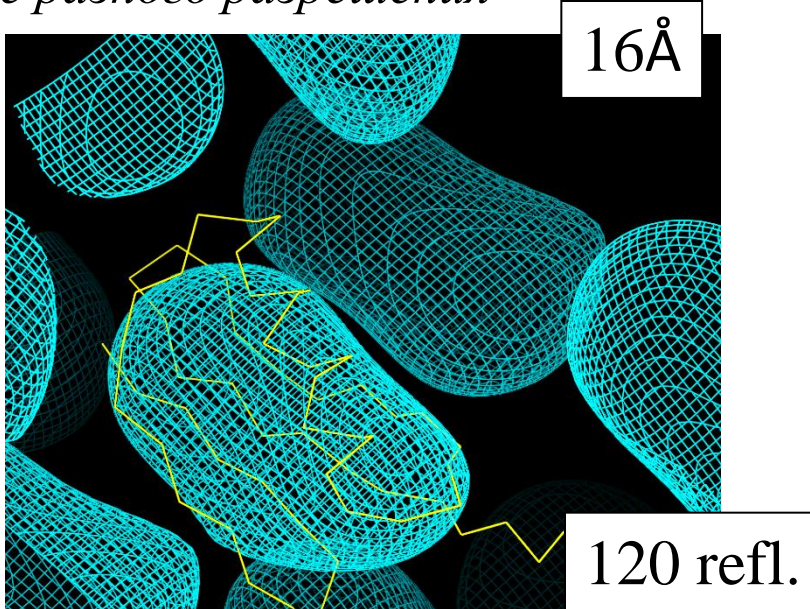
- большие углы;
- малые d ;
- большие s .

Низкое разрешение:

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



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



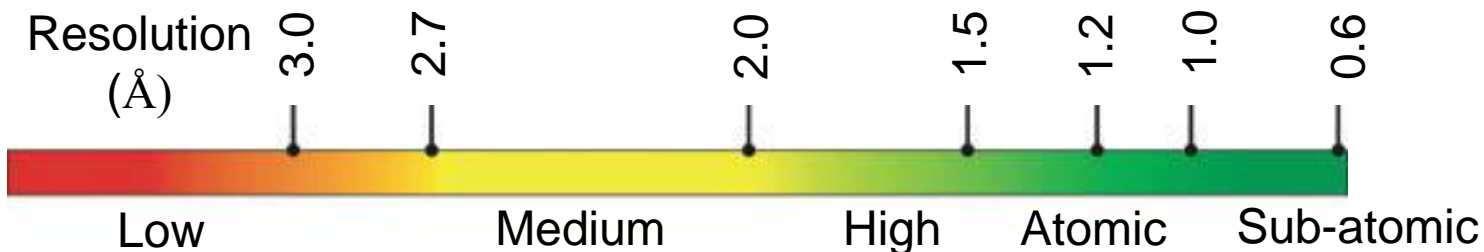
Разрешение набора структурных факторов, измеряемых в реальном эксперименте, определяется качеством кристалла.

crambin, 0.54 Å (Teeter et al., 1993); 0.48 Å (2011)

antifreeze protein, 0.62 Å (Ko et al., 2003);

aldose reductase, 0.66 Å (Howard et al., 2004)].

```
REMARK      3
REMARK      3  DATA USED IN REFINEMENT.
REMARK      3  RESOLUTION RANGE HIGH (ANGSTROMS)  :  1.4
REMARK      3  RESOLUTION RANGE LOW  (ANGSTROMS)  : 30.0
REMARK      3  DATA CUTOFF                    (SIGMA(F)) : NONE
REMARK      3  COMPLETENESS FOR RANGE              (%)      : 97.0
REMARK      3  NUMBER OF REFLECTIONS                : 53449
REMARK      3
```



Комбинированные синтезы Фурье

Фазовая проблема.

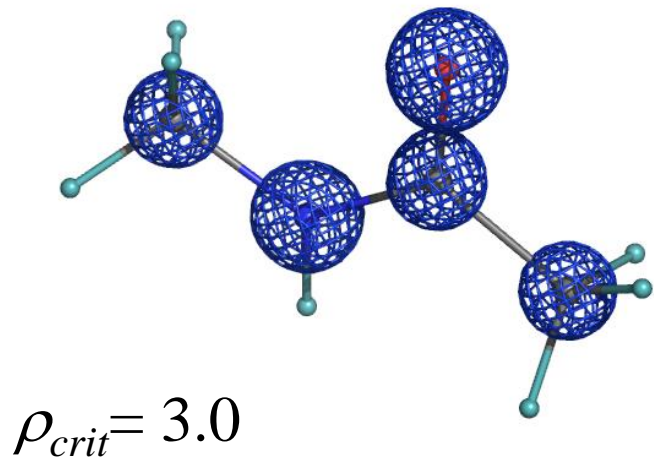
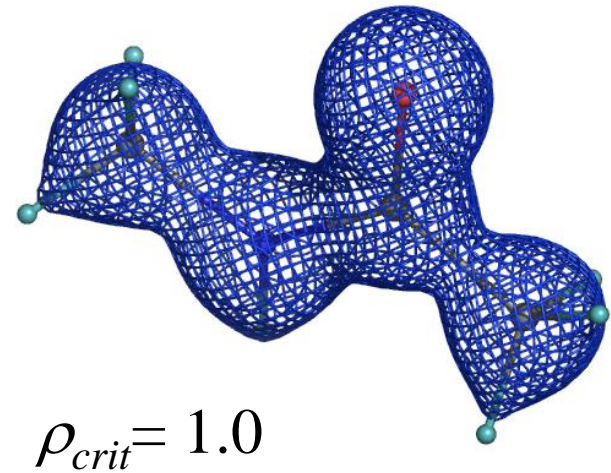
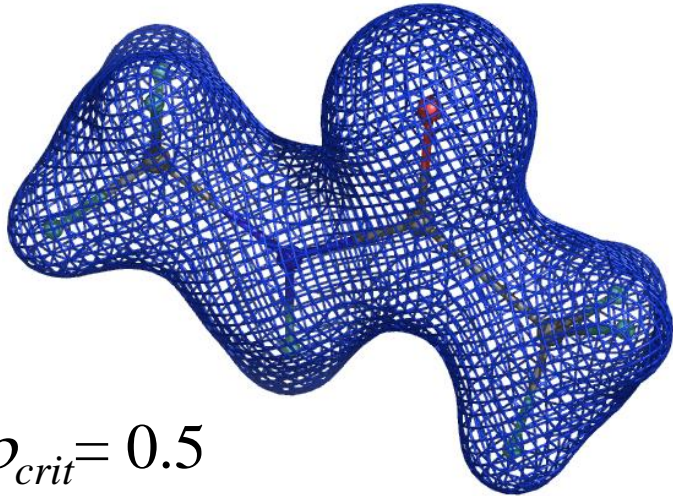
Метод молекулярного замещения

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^{obs} , φ^{PDB});
- **внести корректировки в модель в соответствии с рассчитанным синтезом.**

Как "показать" распределение электронной плотности?

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



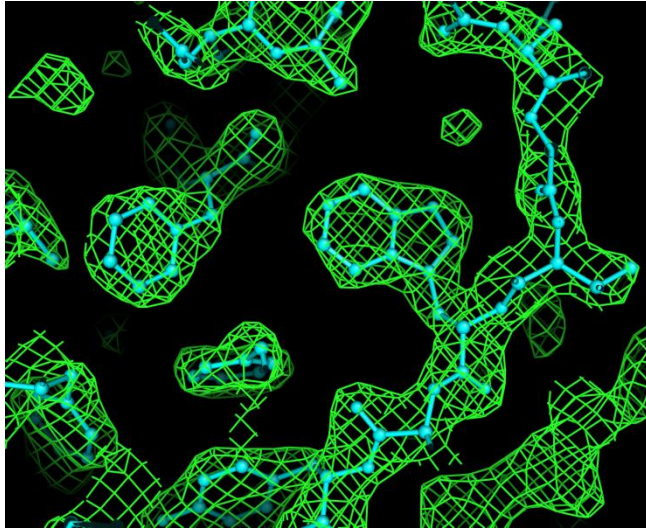
σ – шкала (σ -scale, z-score)

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

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

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

Комбинированные синтезы Фурье

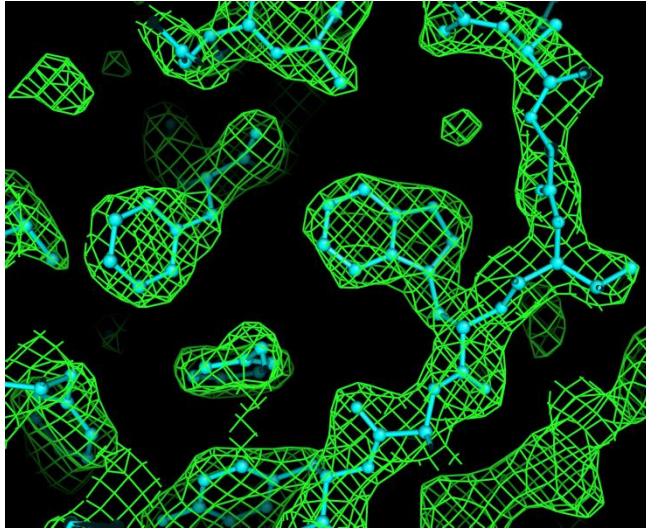


1.5σ

$$F_{full}, \varphi_{full}$$

$$F_{full} = F_{obs}$$

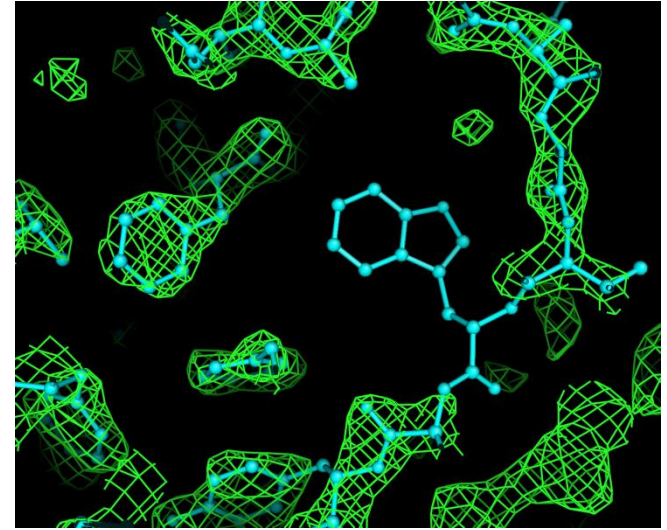
Комбинированные синтезы Фурье



$$F_{full}, \varphi_{full}$$

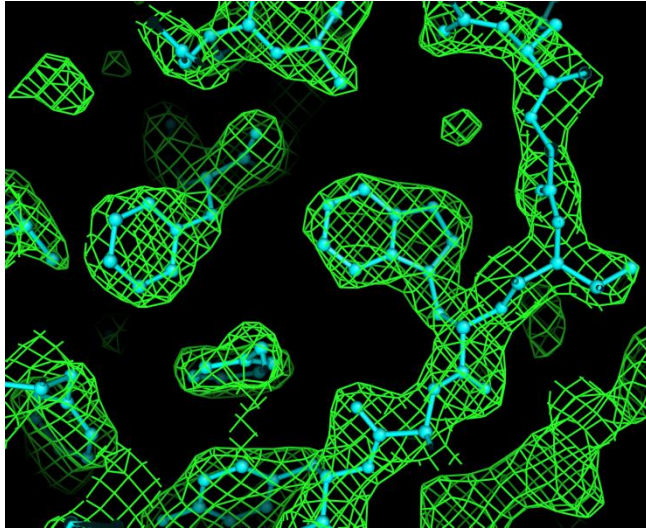
$$F_{full} = F_{obs}$$

1.5σ



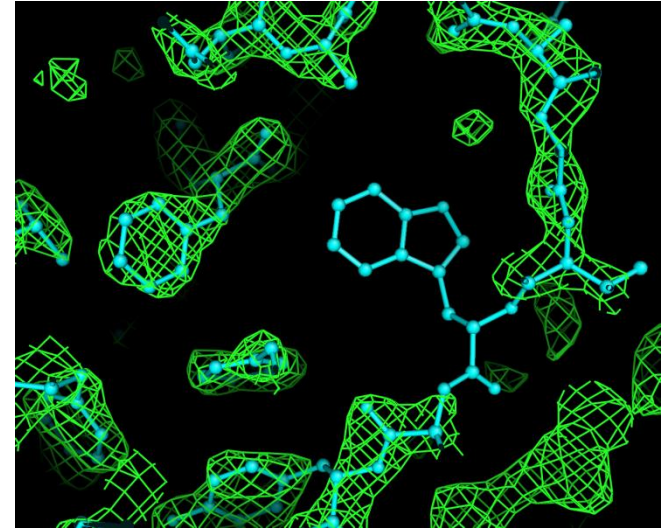
$$F_{part}, \varphi_{part}$$

Комбинированные синтезы Фурье



$$F_{full}, \varphi_{full}$$

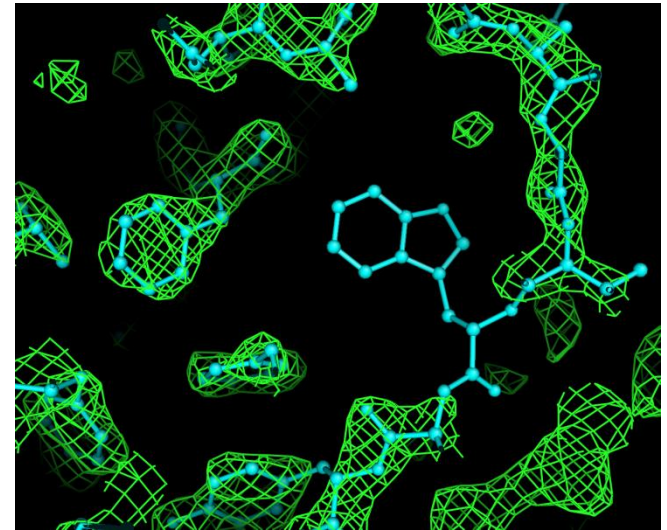
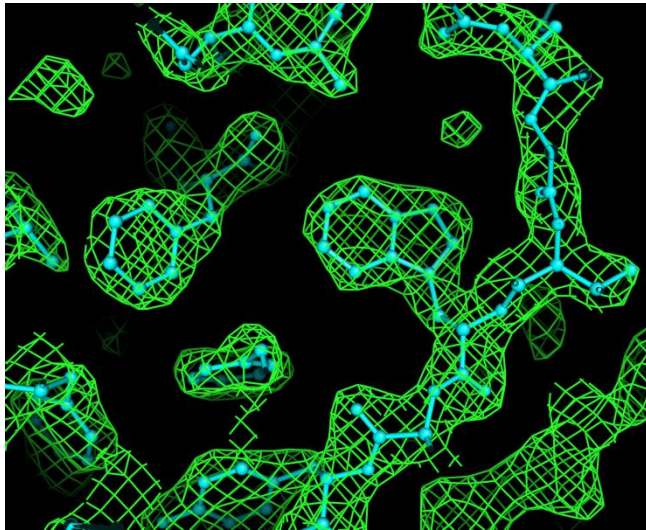
$$F_{full} = F_{obs}$$



$$F_{part}, \varphi_{part}$$

$$F_{full}, \varphi_{part}$$

Комбинированные синтезы Фурье

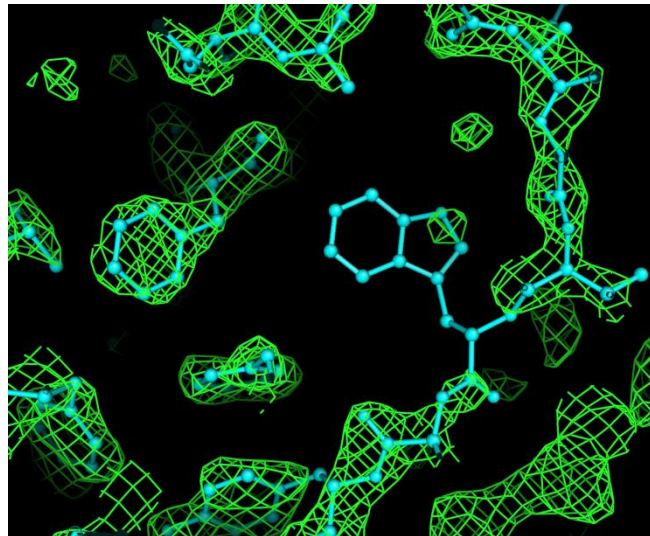


1.5σ

F_{full}, φ_{full}

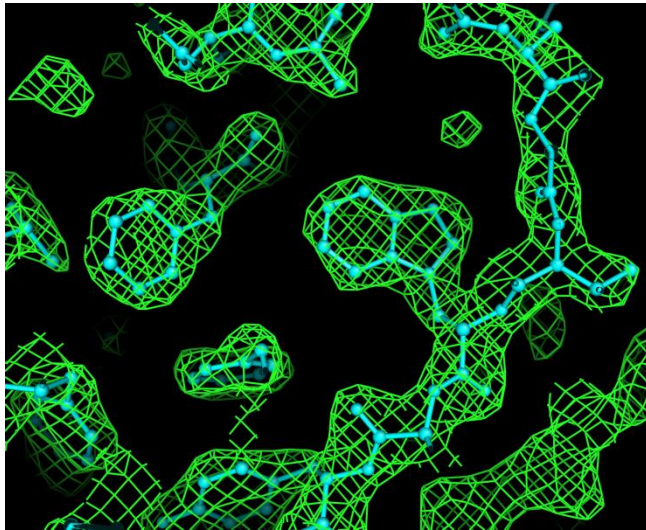
$F_{full} = F_{obs}$

F_{part}, φ_{part}

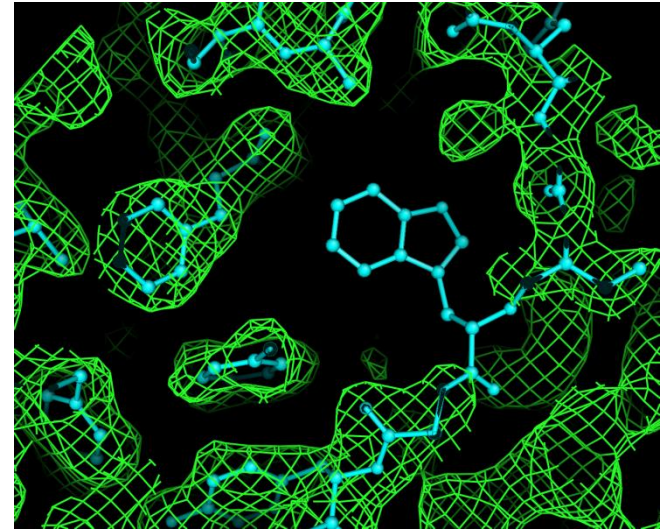


F_{full}, φ_{part}

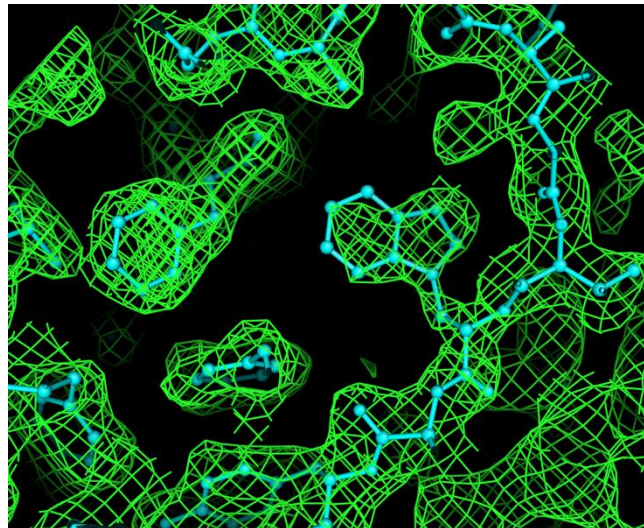
Комбинированные синтезы Фурье



F_{full}, φ_{full}
 1.5σ



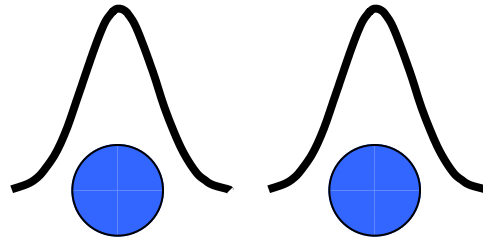
F_{part}, φ_{part}
 0.5σ



$F_{full}, \varphi_{part}, 0.5 \sigma$

«Комбинированные» синтезы Фурье. Корректировка атомной модели.

«точная»
структура



F^{obs}

φ^{true}

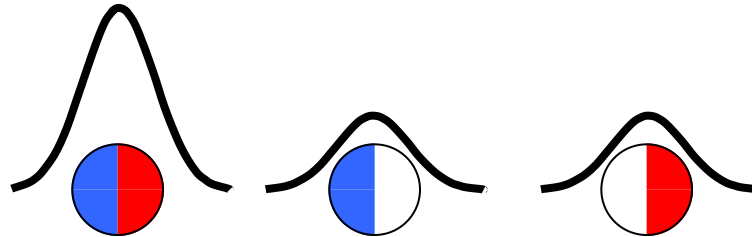
модель



F^{calc}

φ^{calc}

комбинированный
синтез Фурье



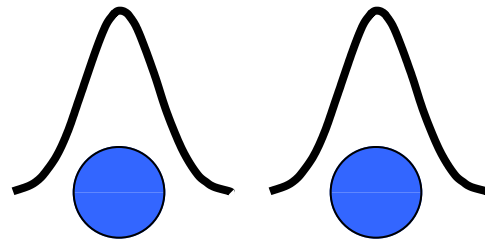
F^{obs}

φ^{calc}

- можем добавлять в модель «потерянные» атомы
- можем удалять из модели «неправильные» атомы

«Комбинированные» синтезы Фурье.

«точная»
структура



F^{obs}

φ^{true}

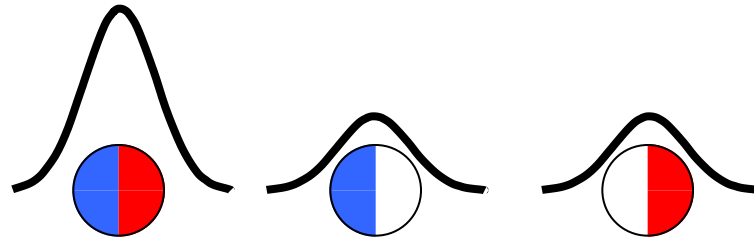
модель



F^{calc}

φ^{calc}

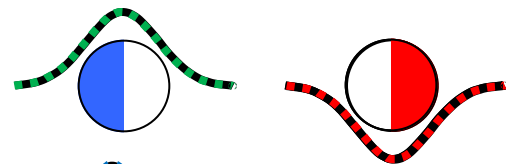
комбинированный
синтез Фурье



F^{obs}

φ^{calc}

разностный
синтез Фурье



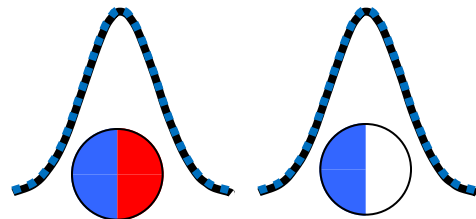
$F^{obs} - F^{calc}$

φ^{calc}

«2-1»

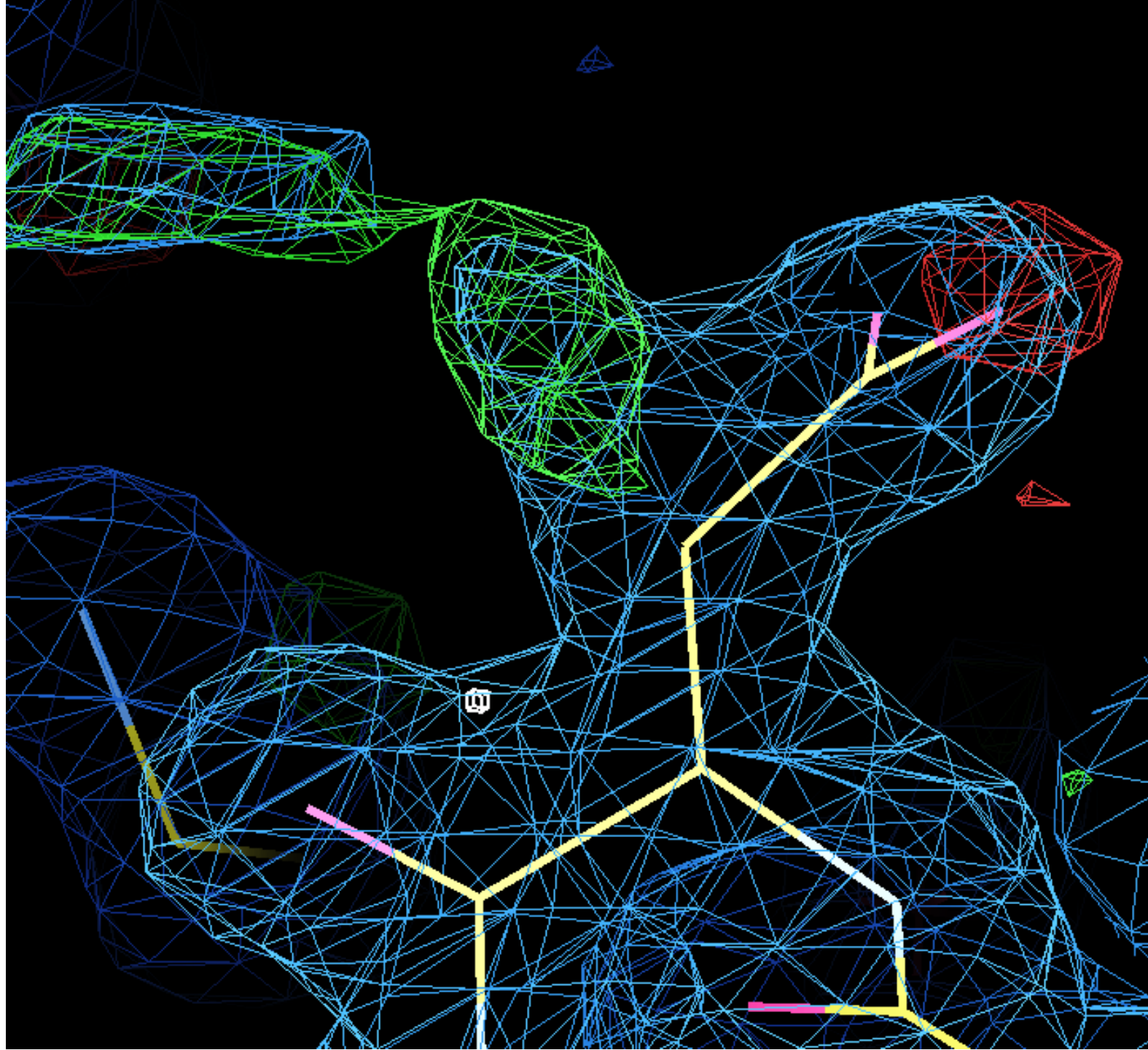
$2F^o - F^c$

синтез Фурье



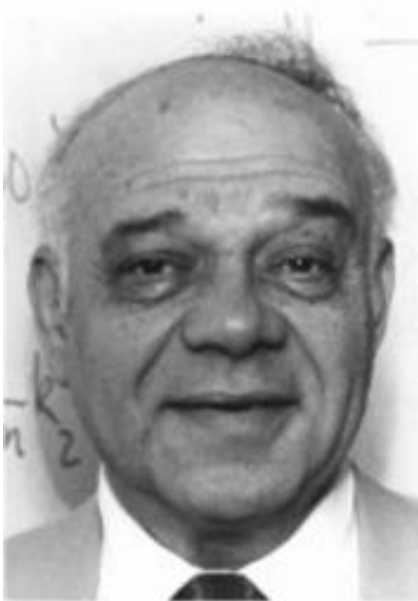
$2F^{obs} - F^{calc}$

φ^{calc}



Jerome Karle Gerbert Hauptman

F^{Karle} ,
 φ^{Karle}



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

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

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



F^{Hauptman} ,
 φ^{Karle}



Взвешенные синтезы Фурье

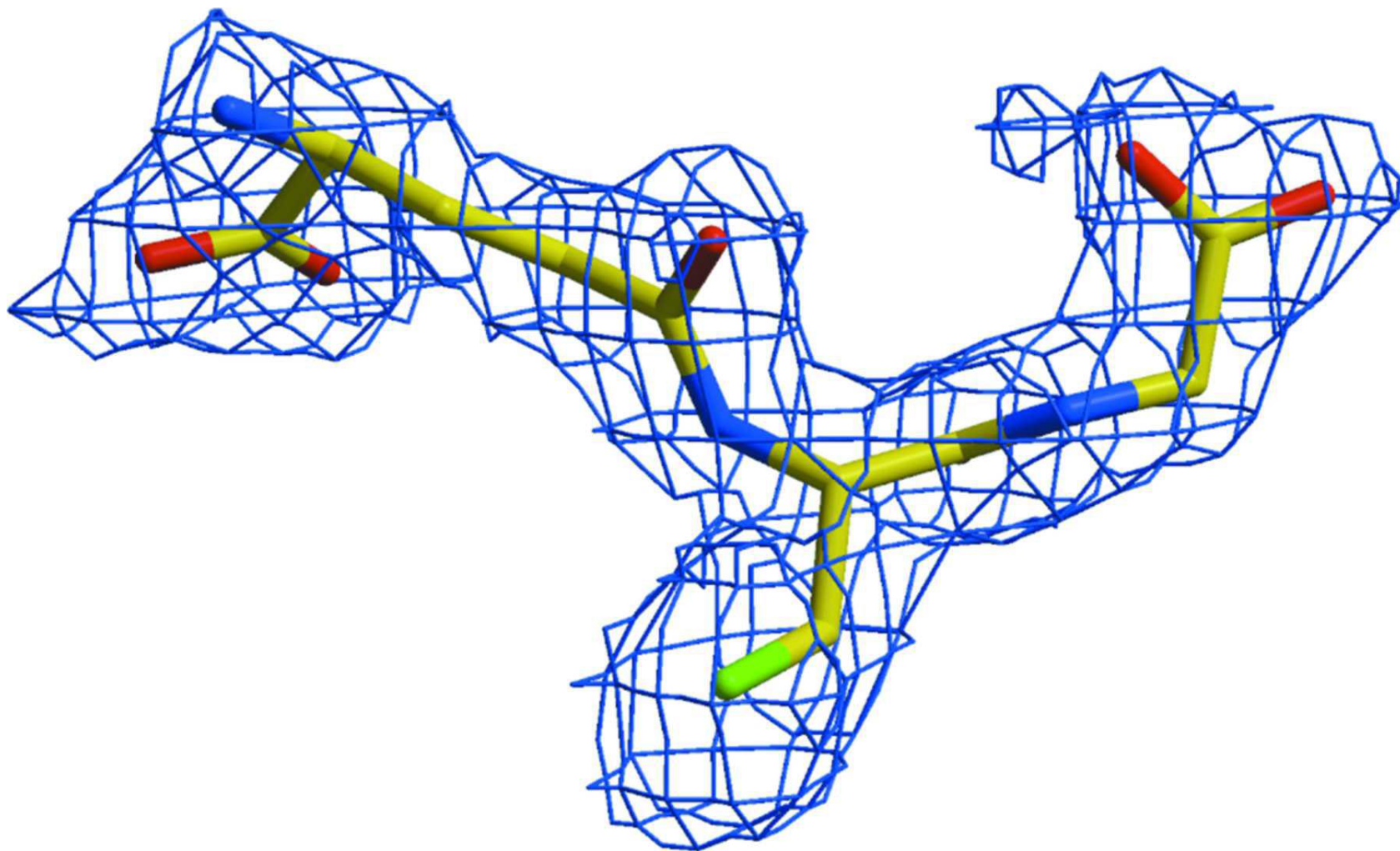
$F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - синтез Фурье

$w_{hkl} F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - 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)

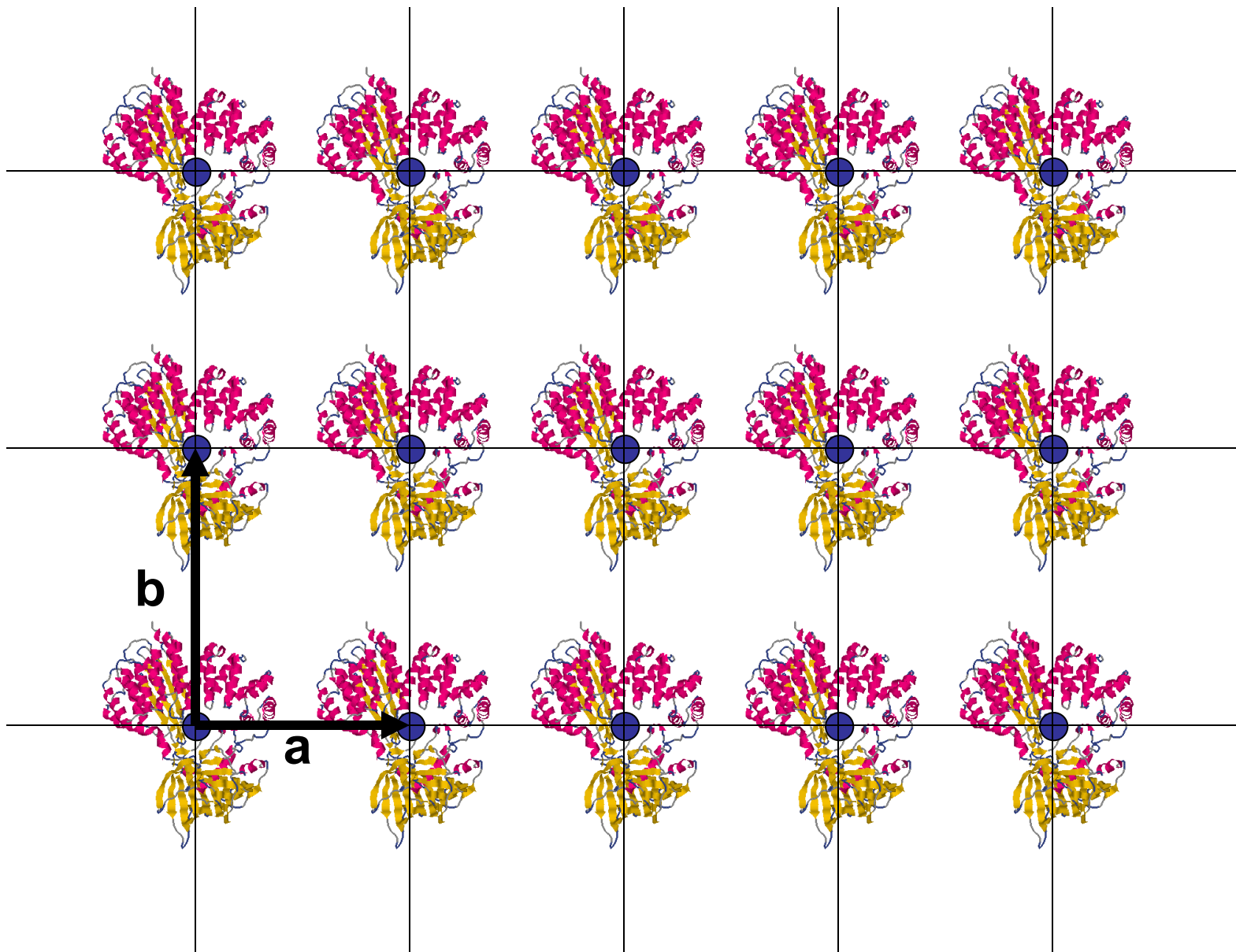
Final σ_A -weighted $2|F_o| - |F_c|$ electron density (Read, 1986) for the GSH molecule bound to subunit A of DmGSTE6 contoured at 1σ .

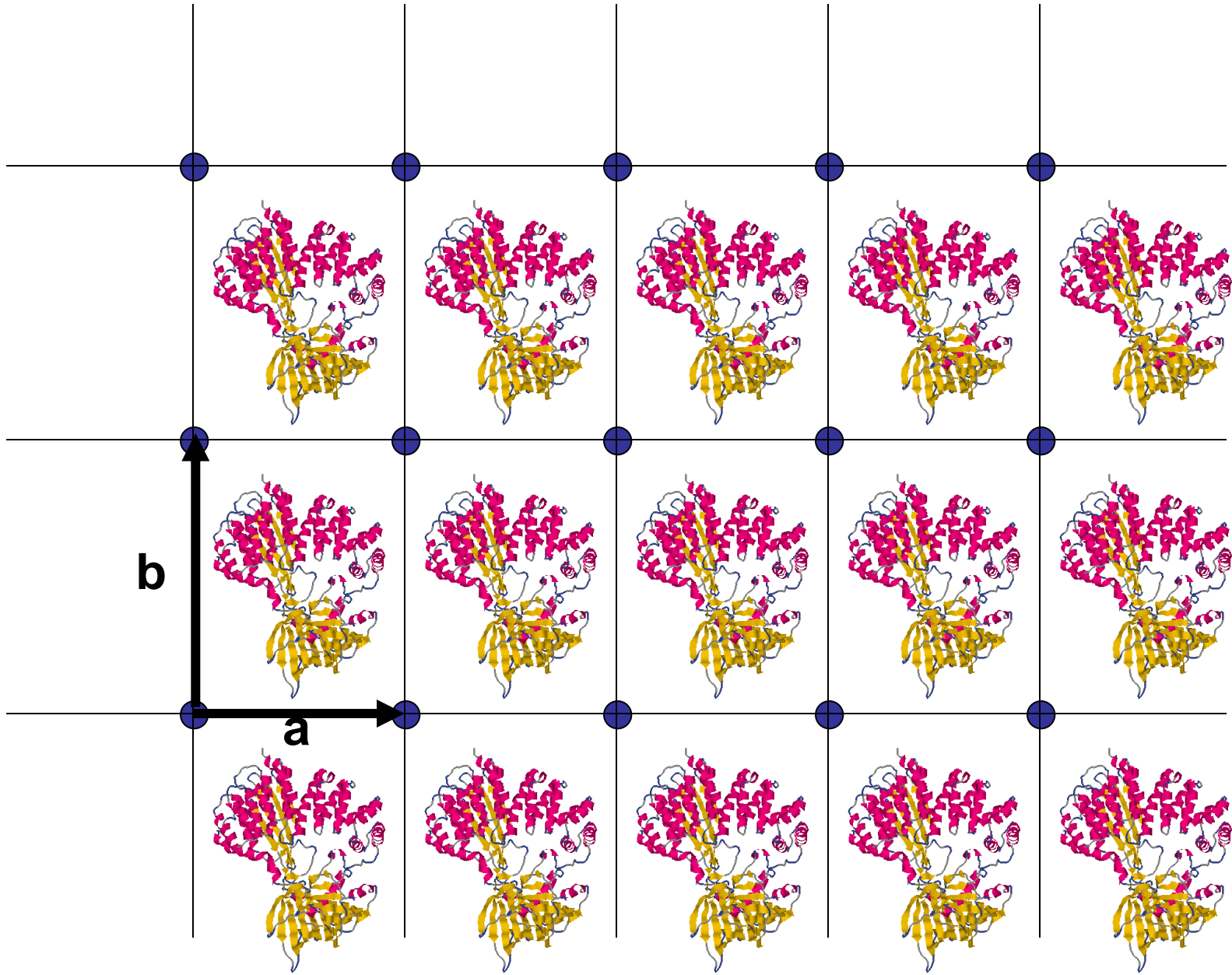


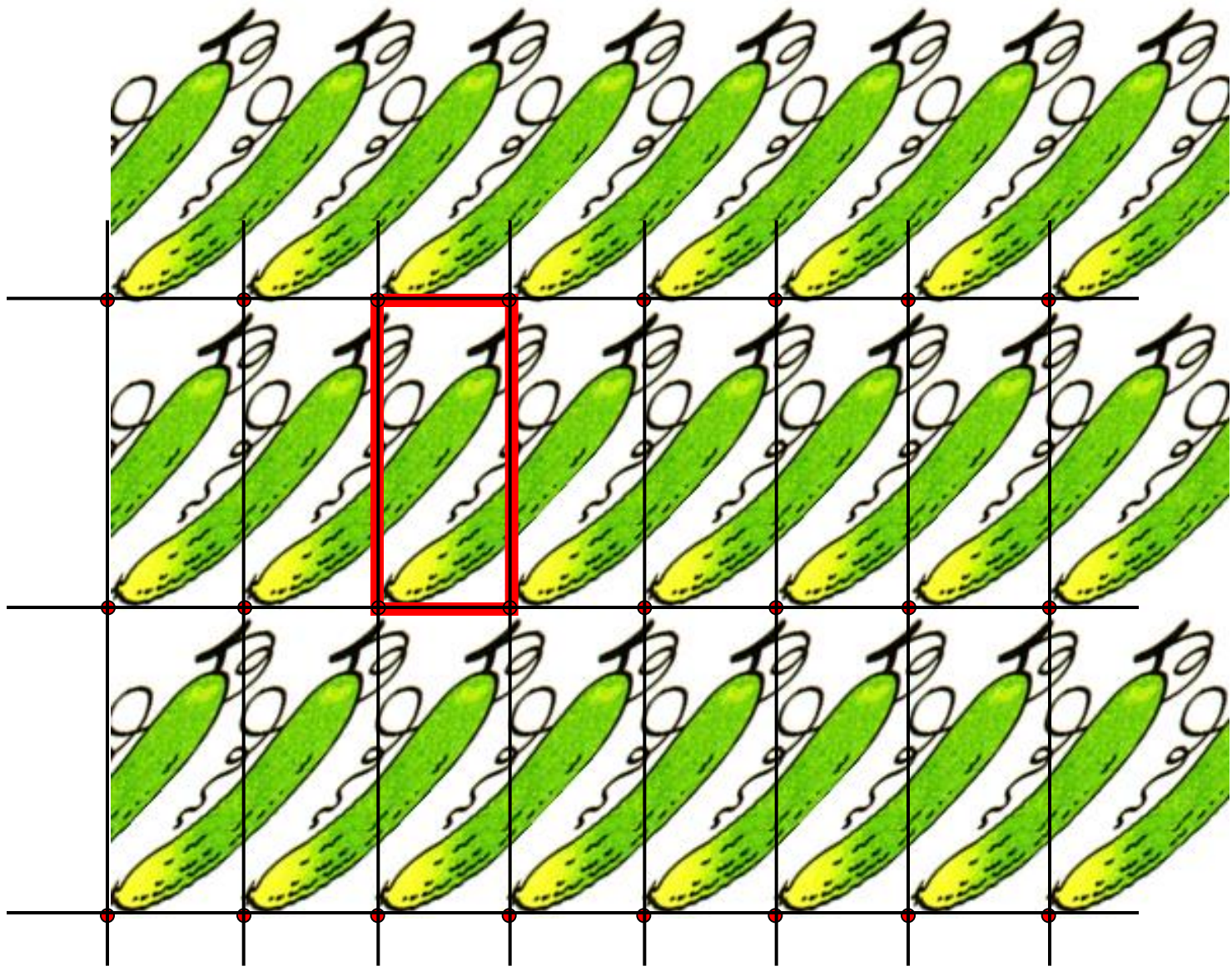
IUCr Journals

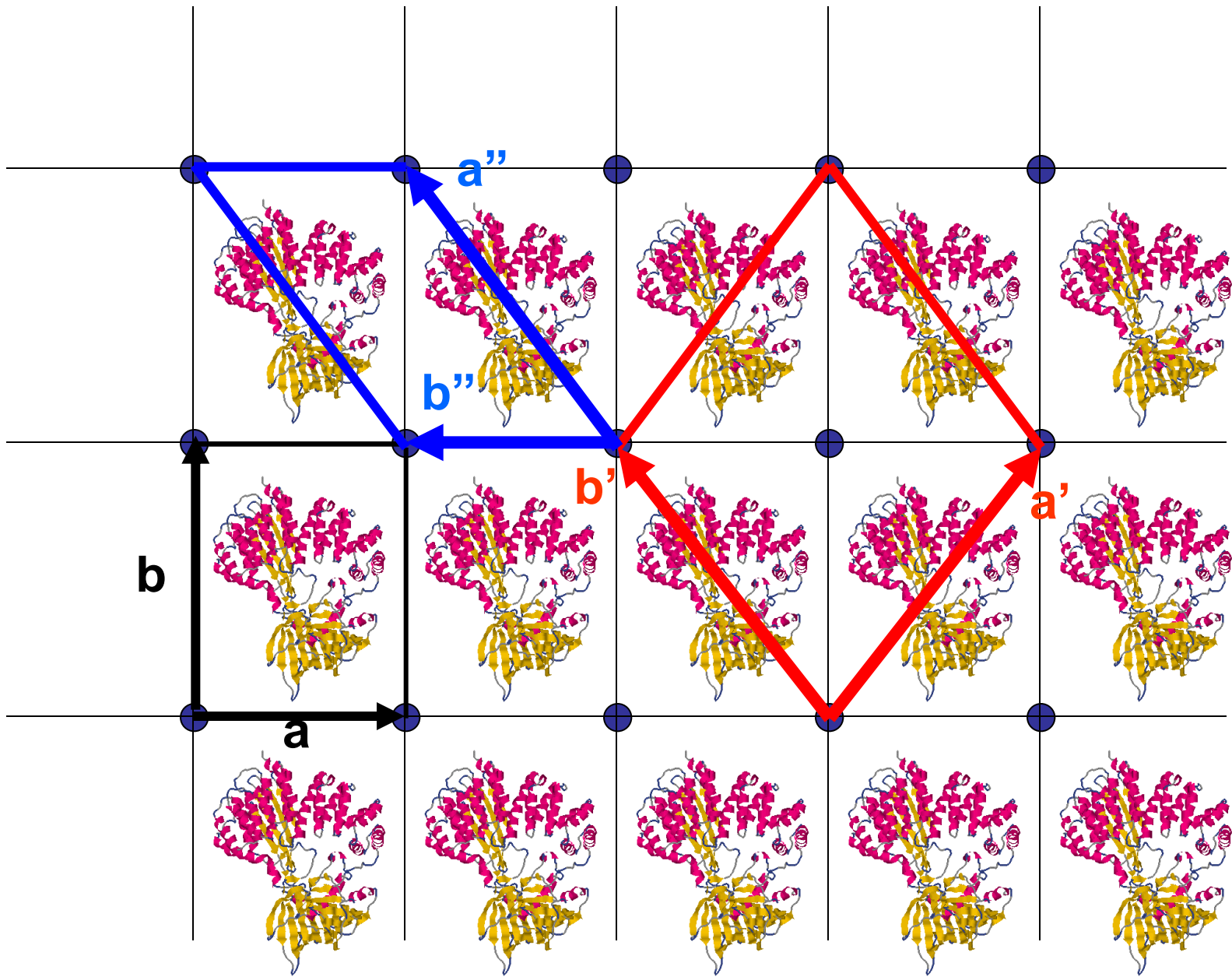
Использование комбинированного синтеза для поиска места расположения лиганда.

Выбор элементарной ячейки

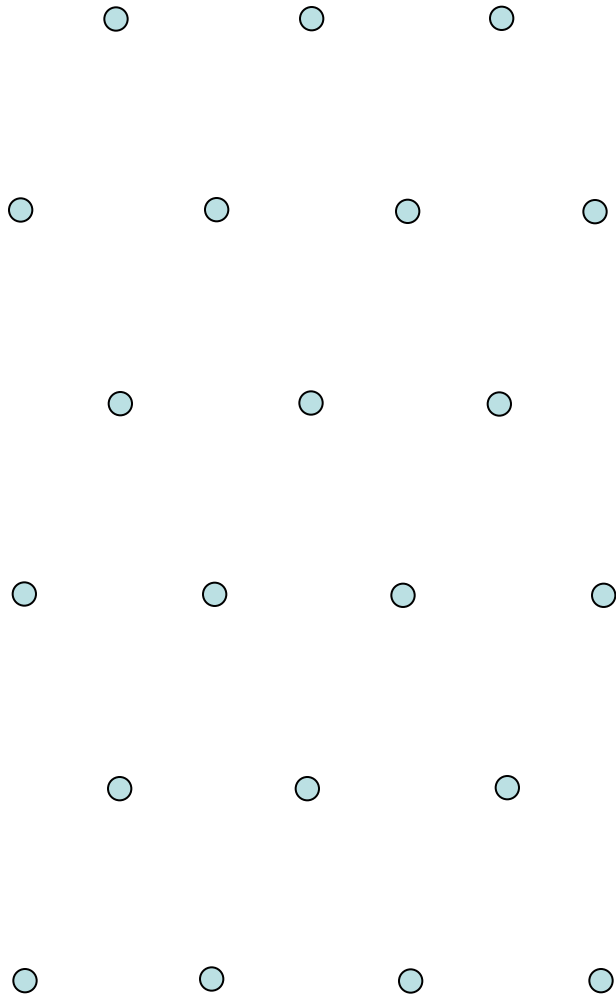




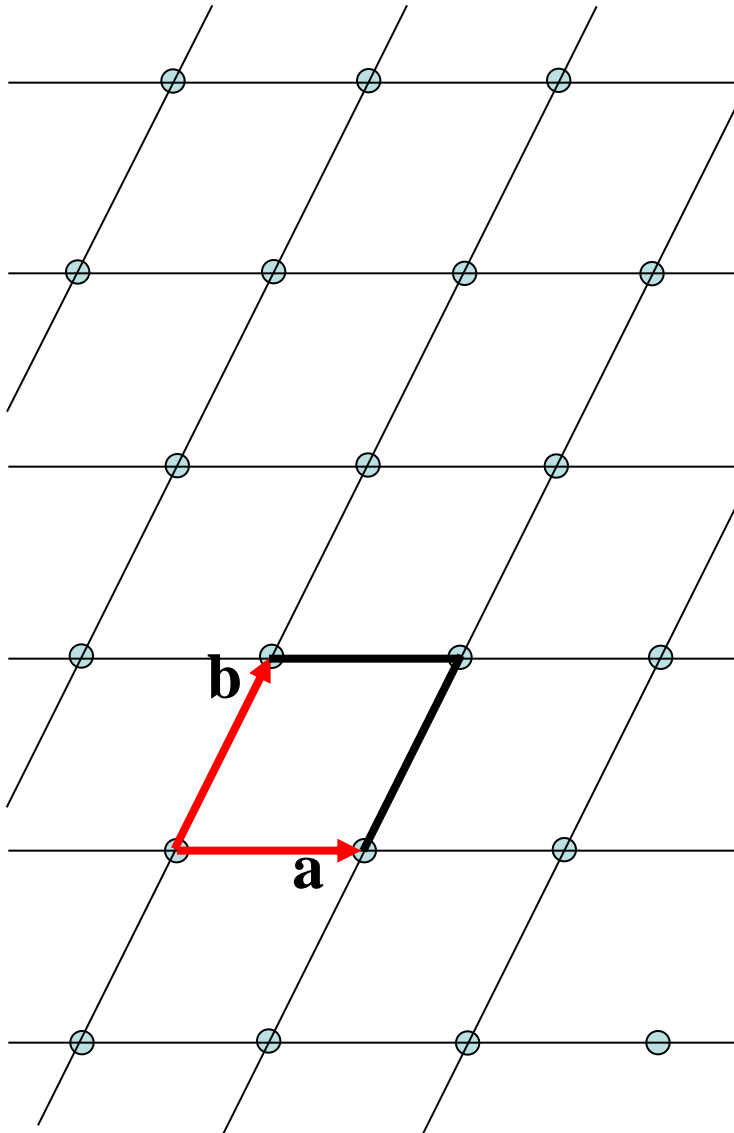




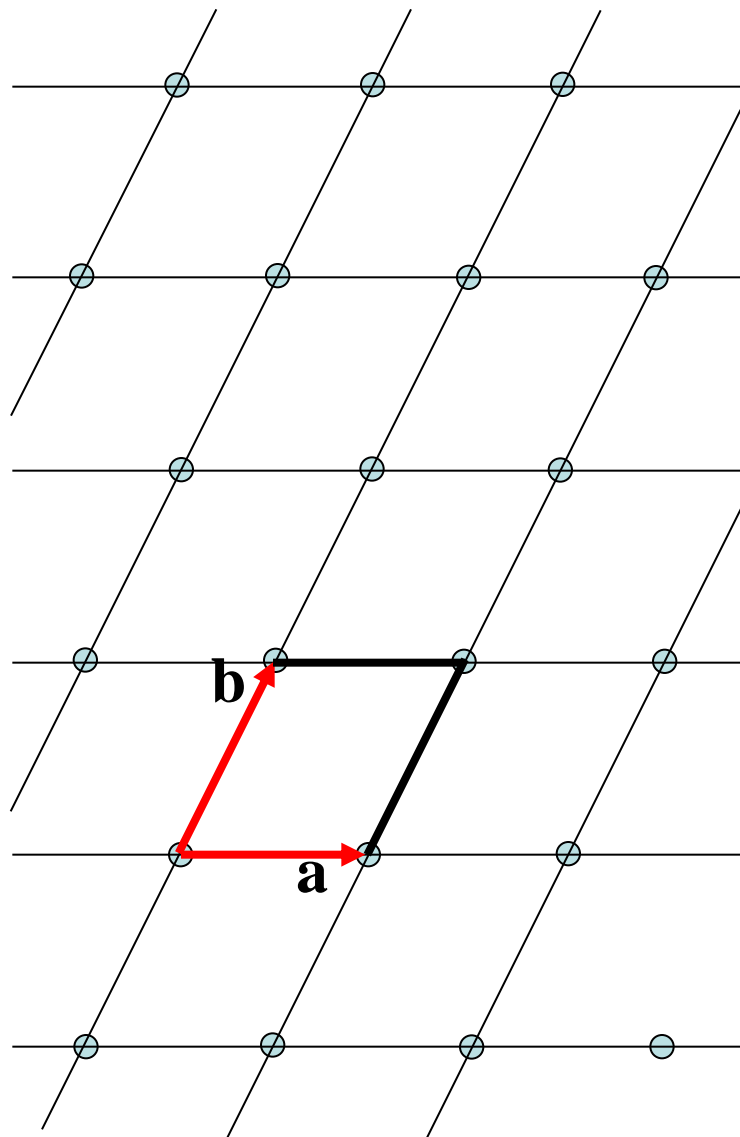
Центрированные ячейки



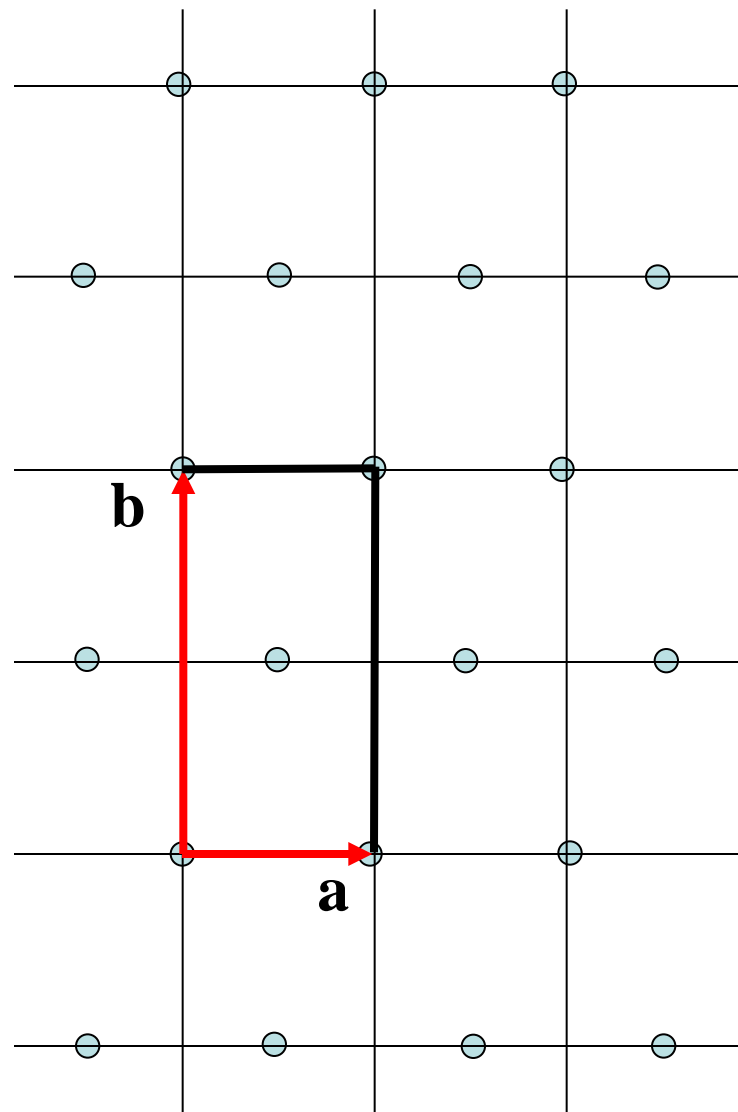
примитивная косоугольная ячейка



примитивная косоугольная ячейка



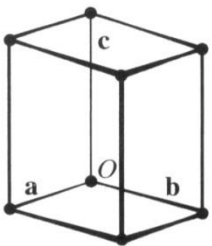
ортогональная ячейка;
дополнительная
трансляция $(1/2, 1/2)$



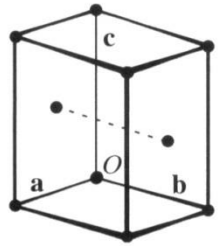
примитивная

базоцентрированная

P



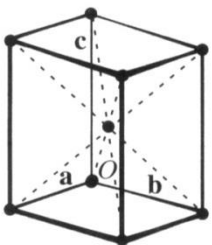
A primitive unit cell (*P*)



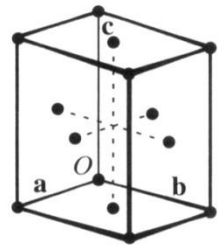
A unit cell centred in the (010) planes (*B*)

**A
B
C**

I



A body-centred unit cell (*I*)



A face-centred unit cell (*F*)

F

объемоцентрированная

гранецентрированная

дважды – объемоцентрированная (ромбоэдрическая)

R

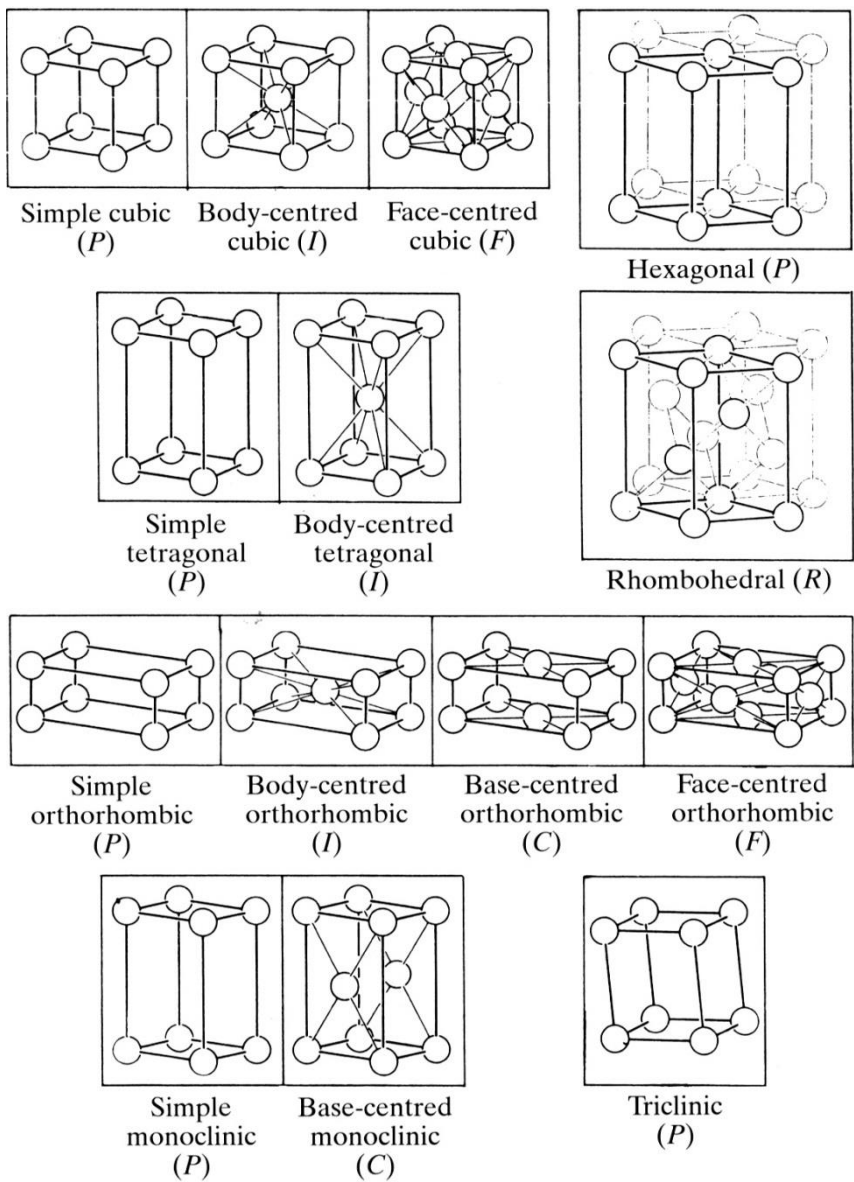


Fig. 2.1.3.3. The 14 Bravais lattices. Reproduced with permission from Burzlaff & Zimmermann (1995). Copyright (1995) International Union of Crystallography.

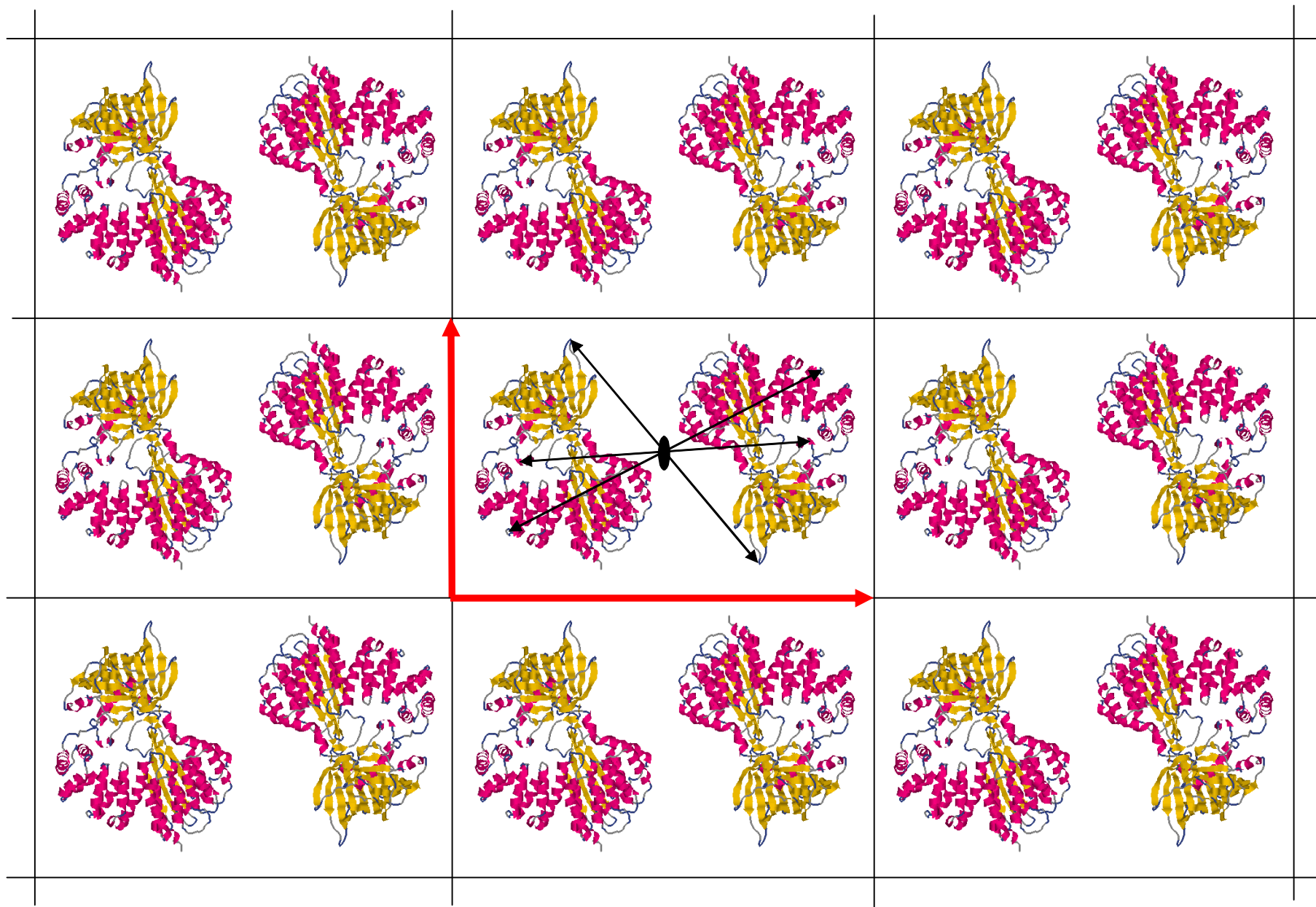
CRYST1 66.224 66.224 40.561 90.00 90.00 120.00 **P** 63 6

тип решетки

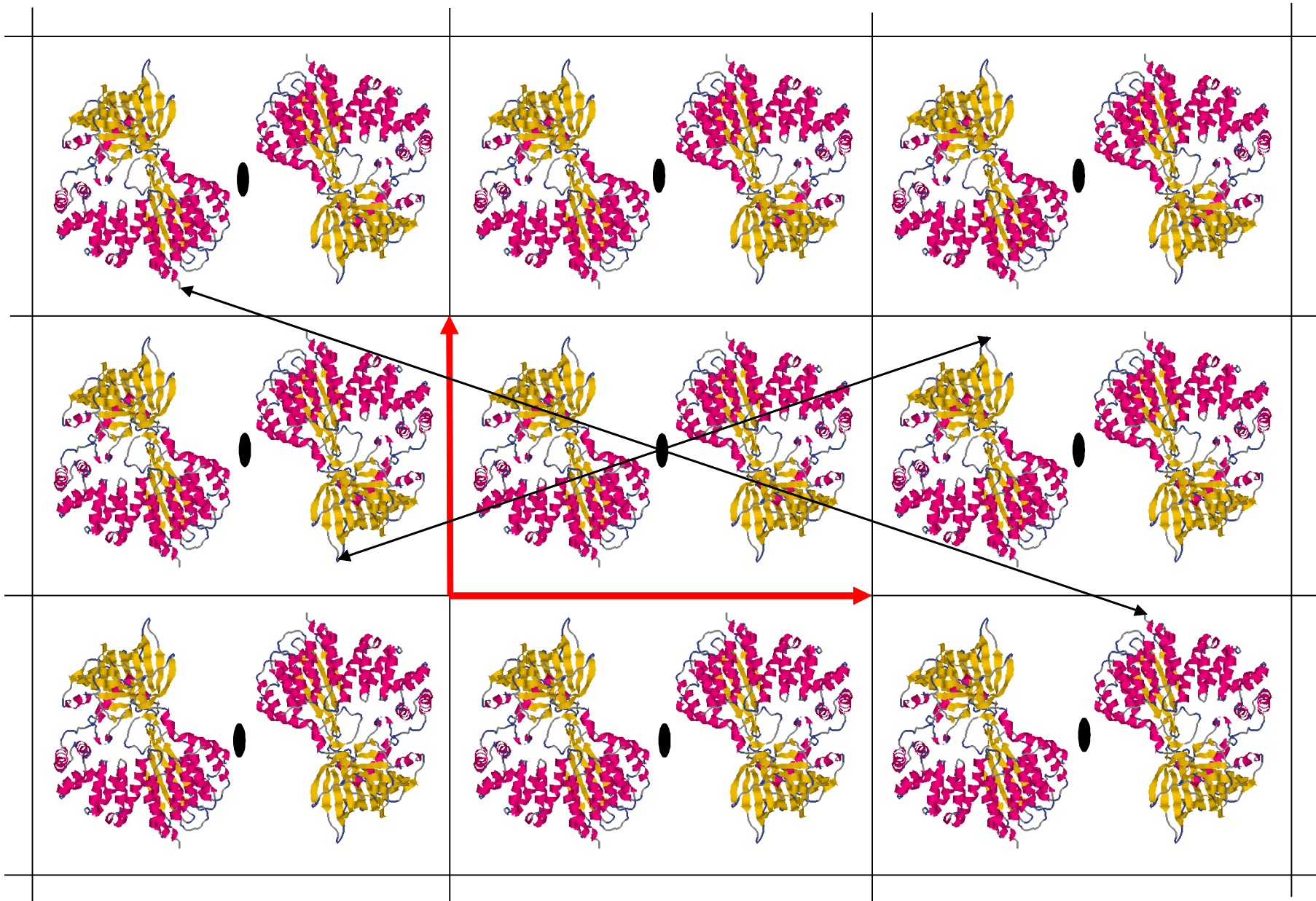


- Элементарная ячейка - математический объект. Она вводится для удобства работы.
- Выбор элементарной ячейки и начала координат в значительной мере произвольны.
- В начале координат может не находиться никакого атома.
- Молекула не всегда лежит в выбранной элементарной ячейке целиком.
- Выбор элементарной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

Элемент симметрии

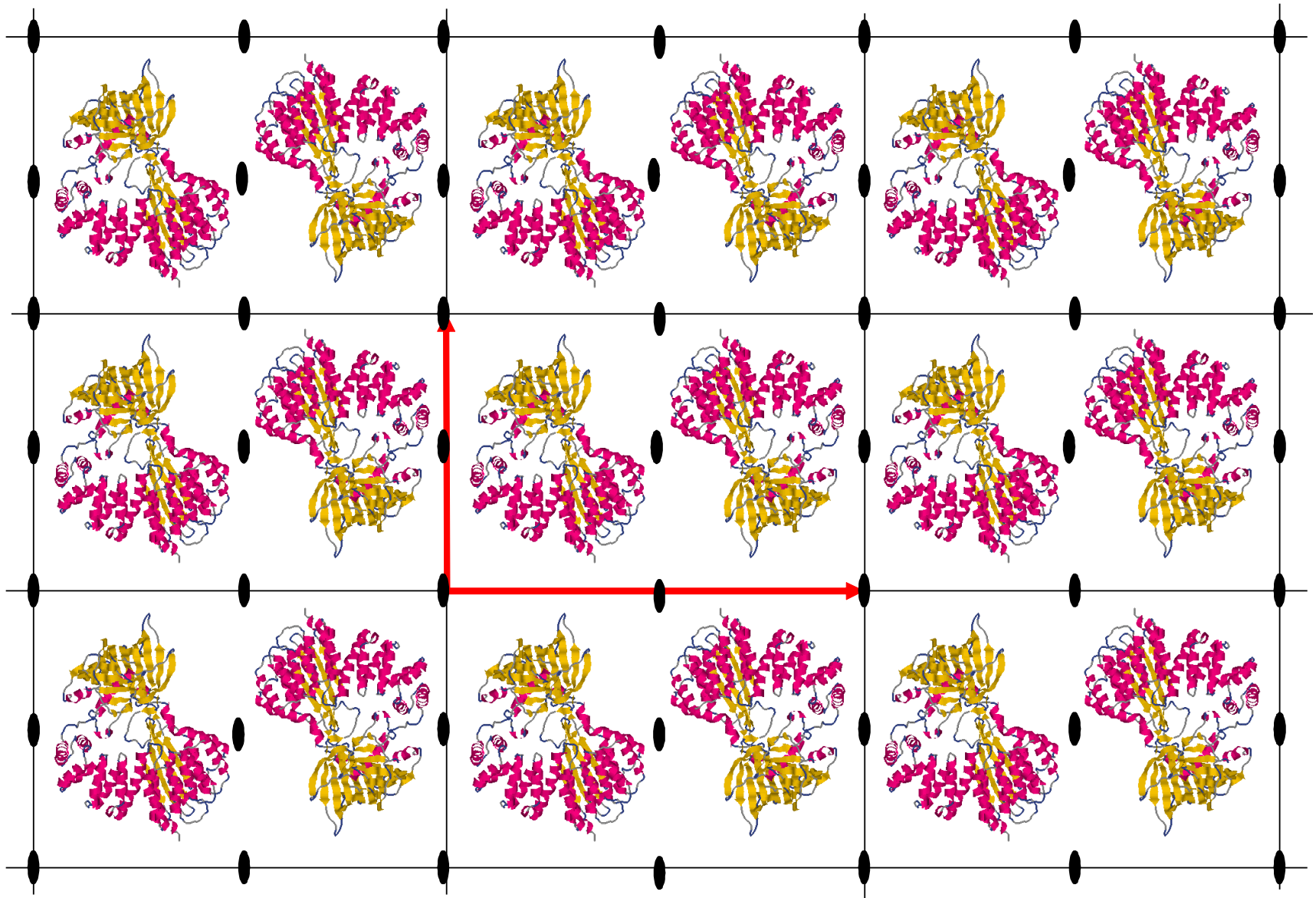


Кристаллографическая симметрия



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.

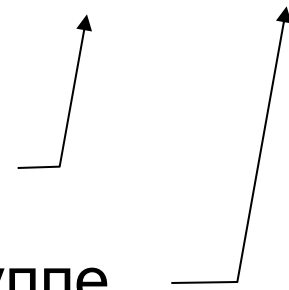
Существует конечное число (230) групп симметрии кристаллов.

Каждая группа имеет свое обозначение.

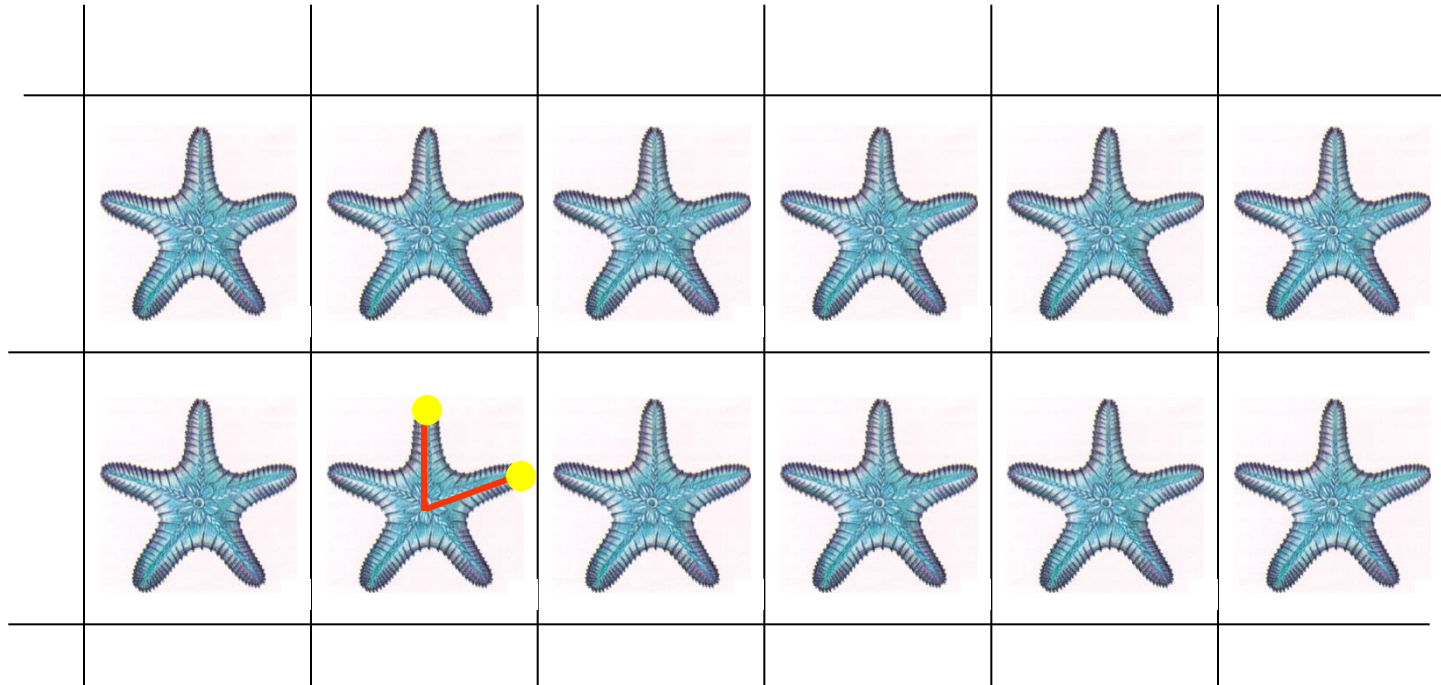
CRYST1 66.224 66.224 40.561 90.00 90.00 120.00 P 63 6

символ пространственной группы

число элементов симметрии в группе

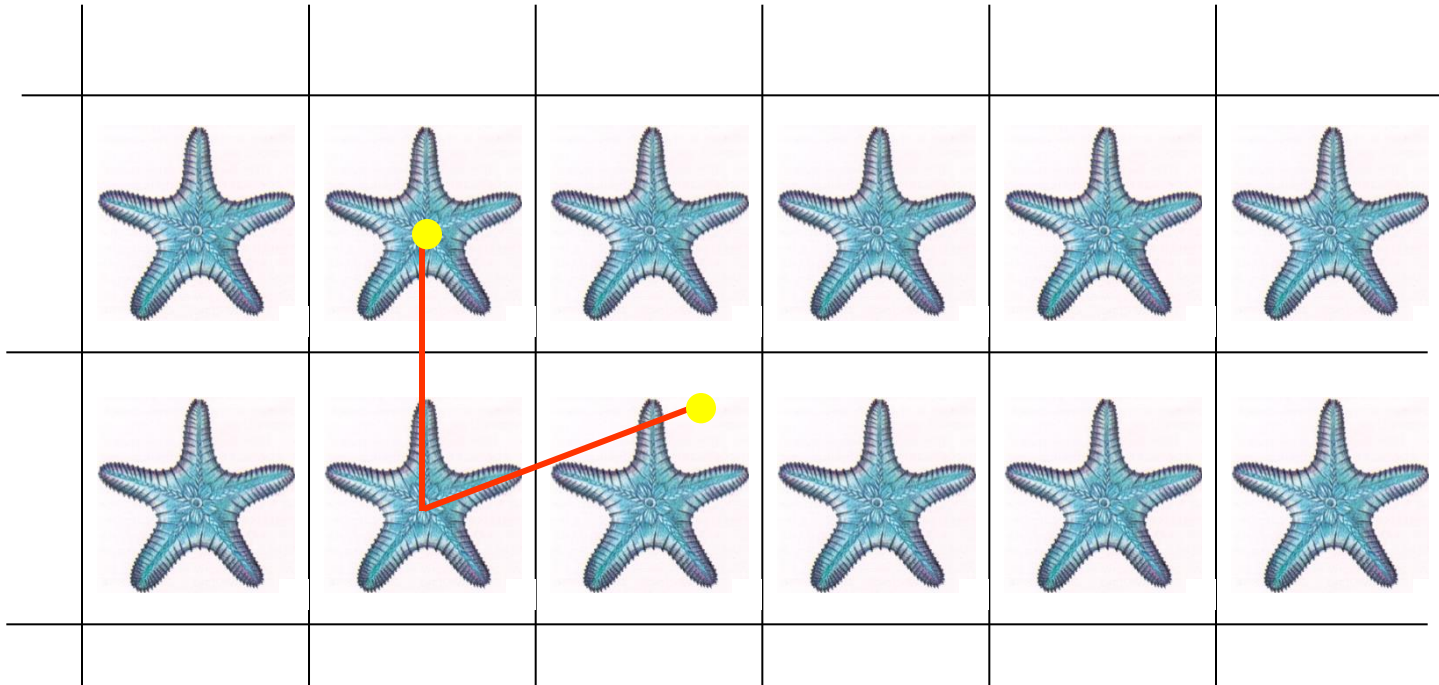


Некристаллографическая симметрия

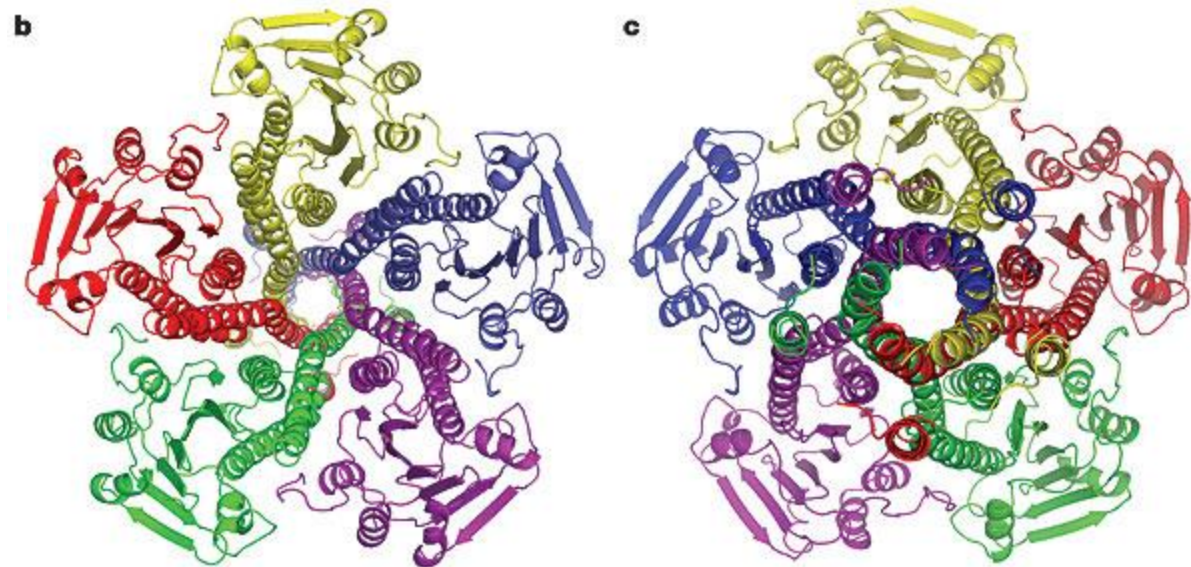
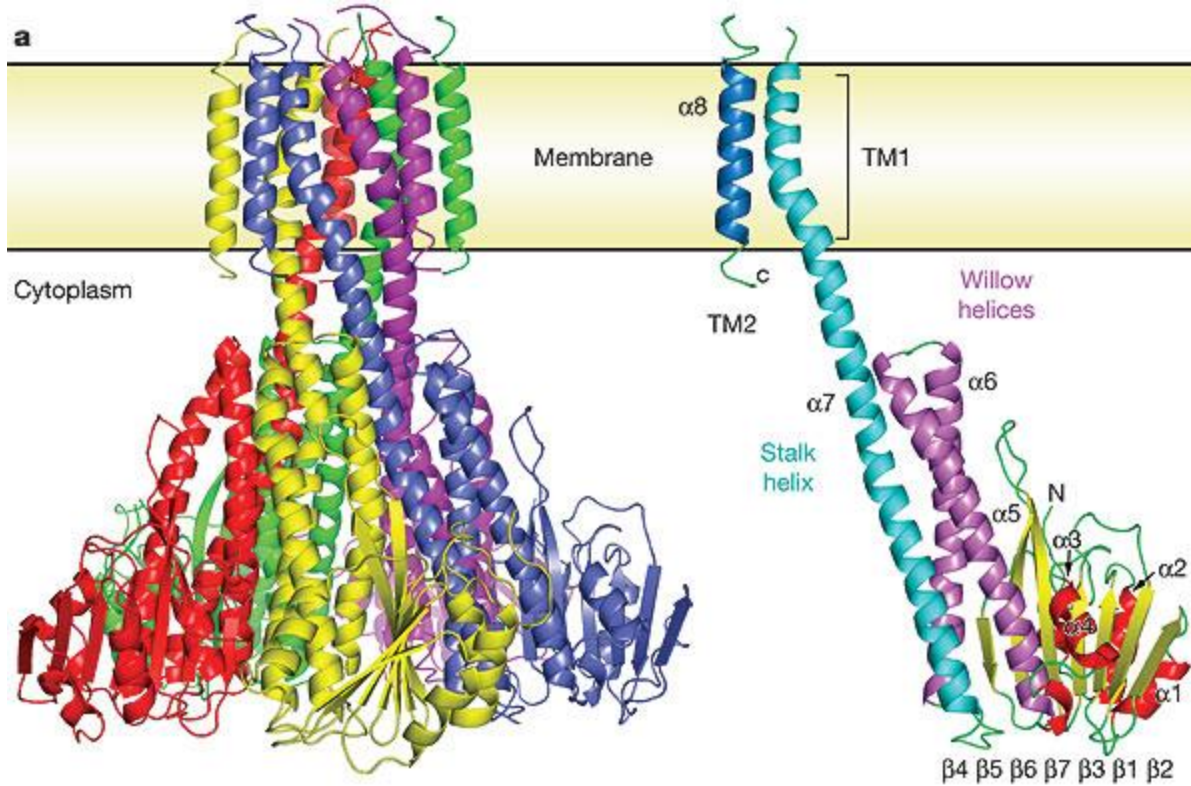


Некристаллографическая (локальная) симметрия имеет место только в ограниченной области пространства и не распространяется на весь кристалл

NON-CRYSTALLOGRAPHIC SYMMETRY



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



Structure of the
CorA Mg²⁺ channel.

From the following
article: [Crystal
structure of the CorA
Mg²⁺ transporter](#)

Nature 440, 833-837
(6 April 2006)

Уточнение модели

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

Относительные и абсолютные координаты атомов

Относительные координаты: x, y, z ; безразмерные;
оси координат не ортогональны;
 $r = xa + yb + zc$

Абсолютные координаты (PDB): Оси координат ортогональны;
единица длин - ангстрем

**Матрица перехода
от абсолютных
к относительным
координатам**

```
SCALE1 0.019231 0.000000 0.000000 0.000000
SCALE2 0.000000 0.017065 0.000000 0.000000
SCALE3 0.000000 0.000000 0.016155 0.000000
```

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

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

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.

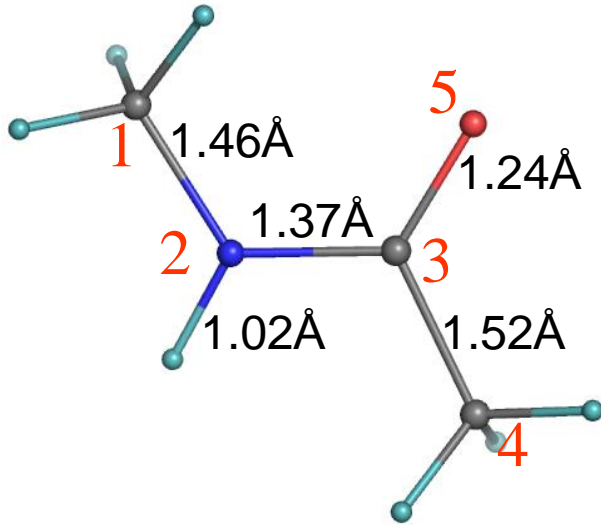
Уточнение, как задача минимизации:

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

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

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

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

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

Составной критерий

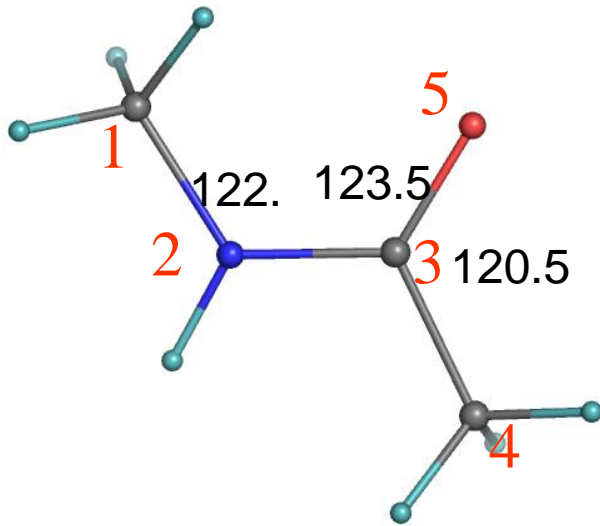
$$R = w_{X\text{-ray}} R_{X\text{-ray}} + w_{dist} R_{dist}$$

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



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

Стереохимические ограничения



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

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

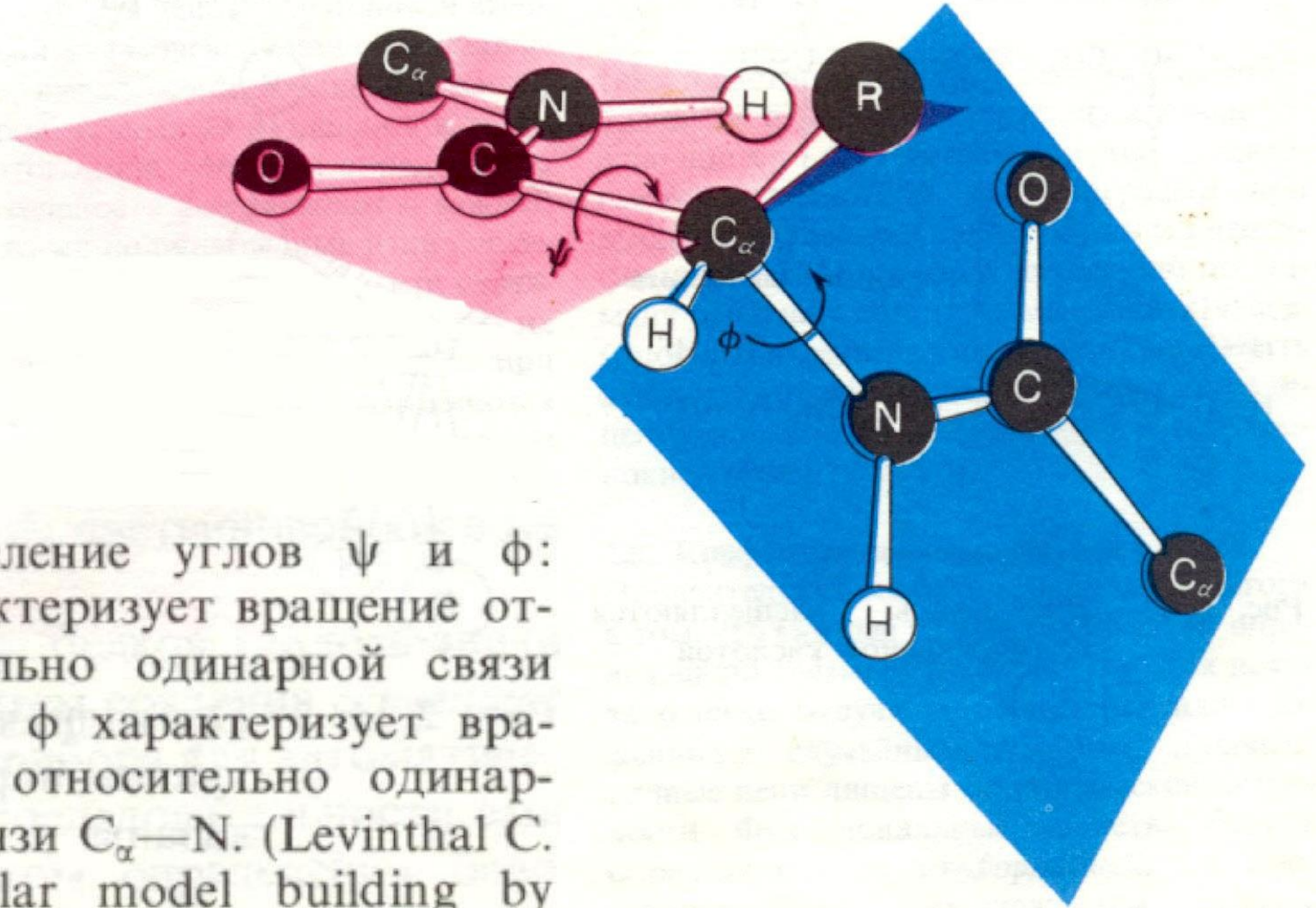
Хотим иметь

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

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

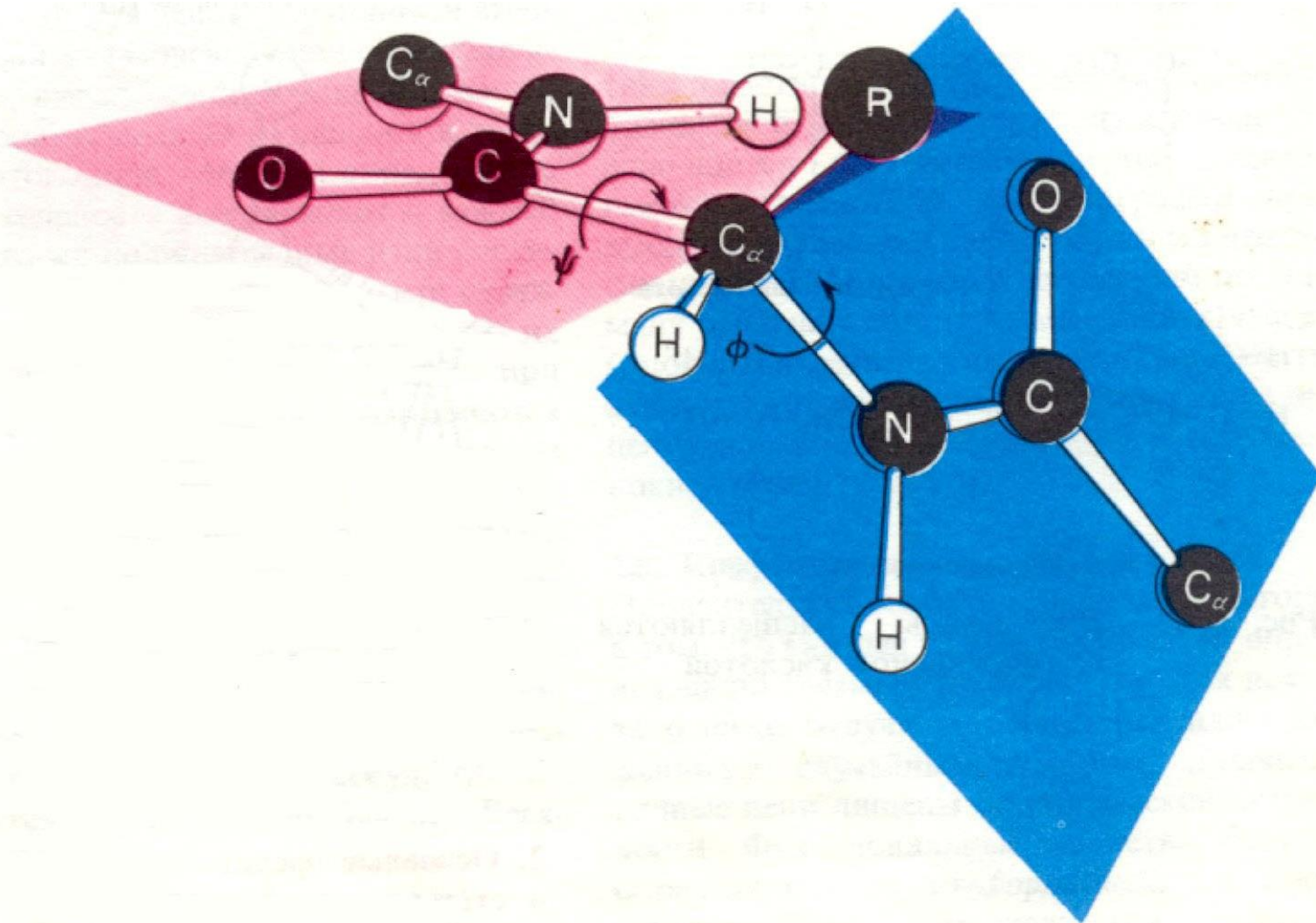
$$R = w_{X-ray} R_{X-ray} + w_{dist} R_{dist} + w_{angle} R_{angle}$$

Двугранные углы

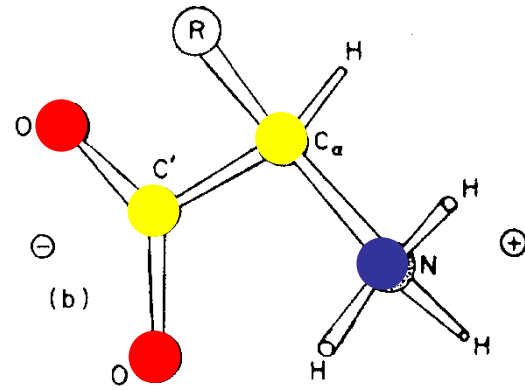
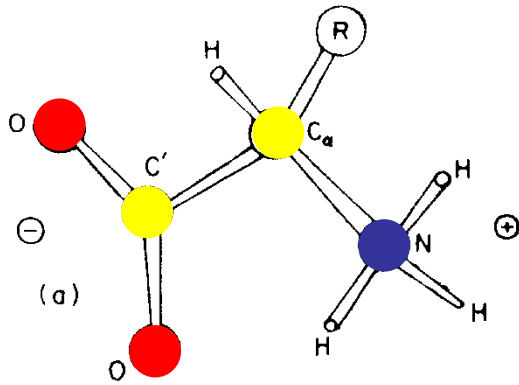


Определение углов ψ и ϕ :
 ψ характеризует вращение относительно одинарной связи $C_\alpha - C$; ϕ характеризует вращение относительно одинарной связи $C_\alpha - N$. (Levinthal C. Molecular model building by computer, Scientific American, Inc., 1966.).

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.*)

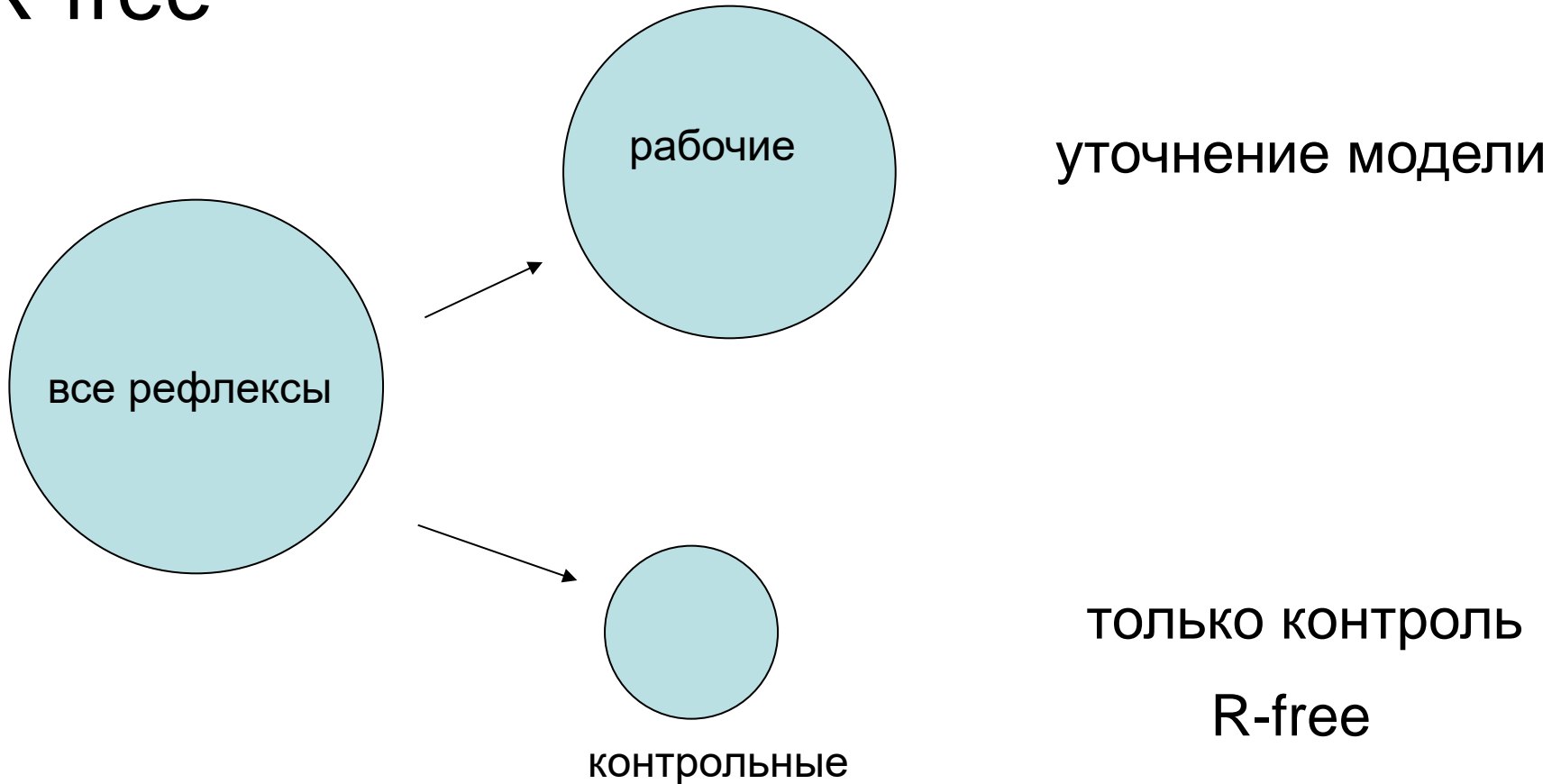
REFMAC (G.Murshudov *et al.*)

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 объекта

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 растворителем (водой). Его нельзя игнорировать при аккуратных расчетах.

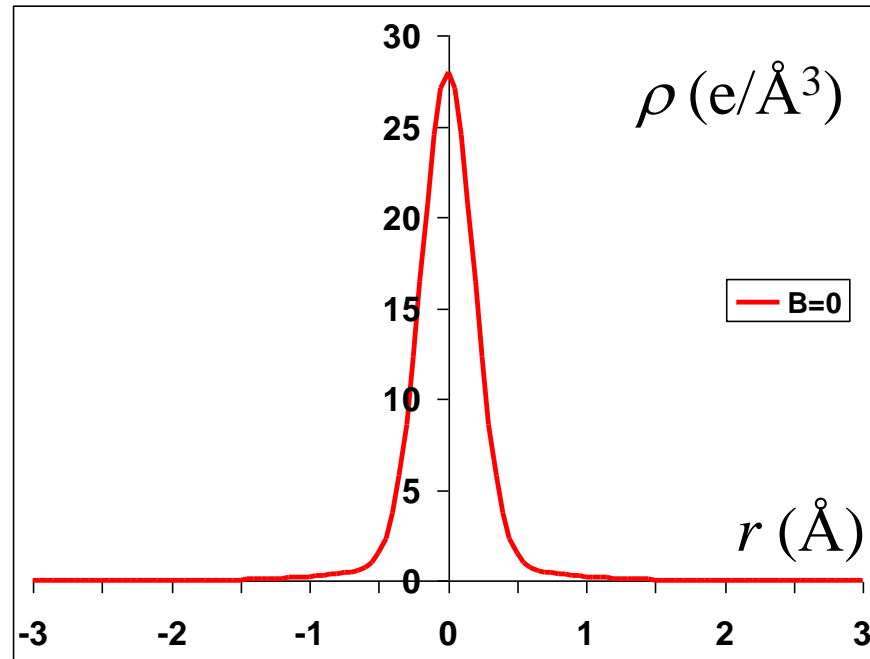
- Связанная вода. (Аналогично атомам молекулы белка, но водородные связи вместо ковалентных).
- Неупорядоченный растворитель – bulk solvent. (Равномерное распределение плотности в области растворителя).

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*)

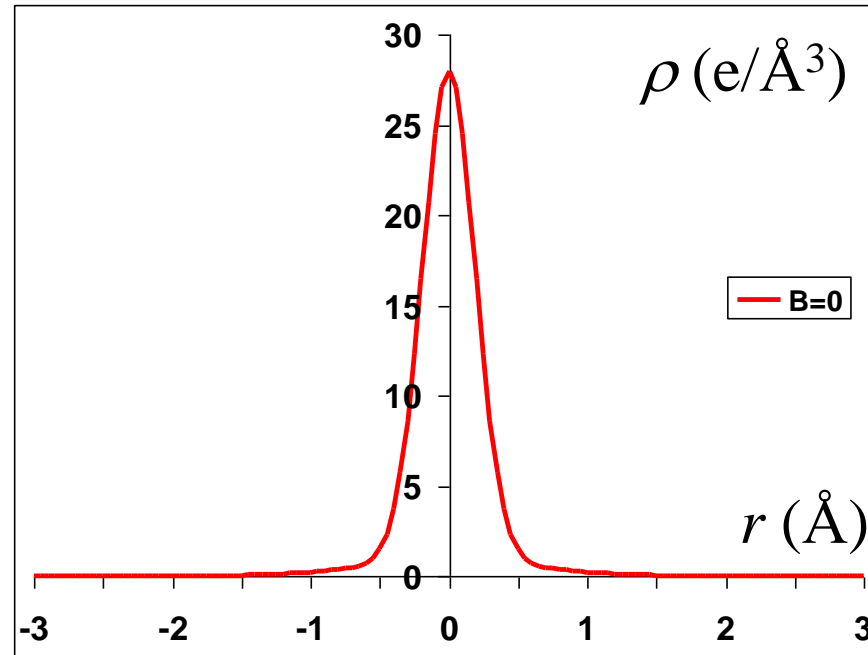
Динамическая и
статическая
неупорядоченность



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

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

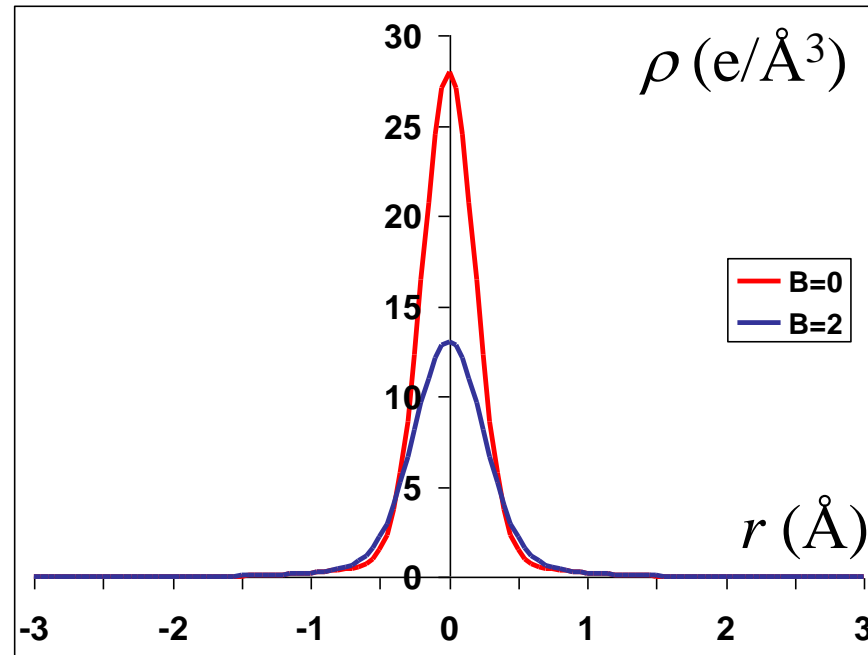
Радиальное распределение электронной плотности в атоме углерода при разных значениях температурного фактора B .



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

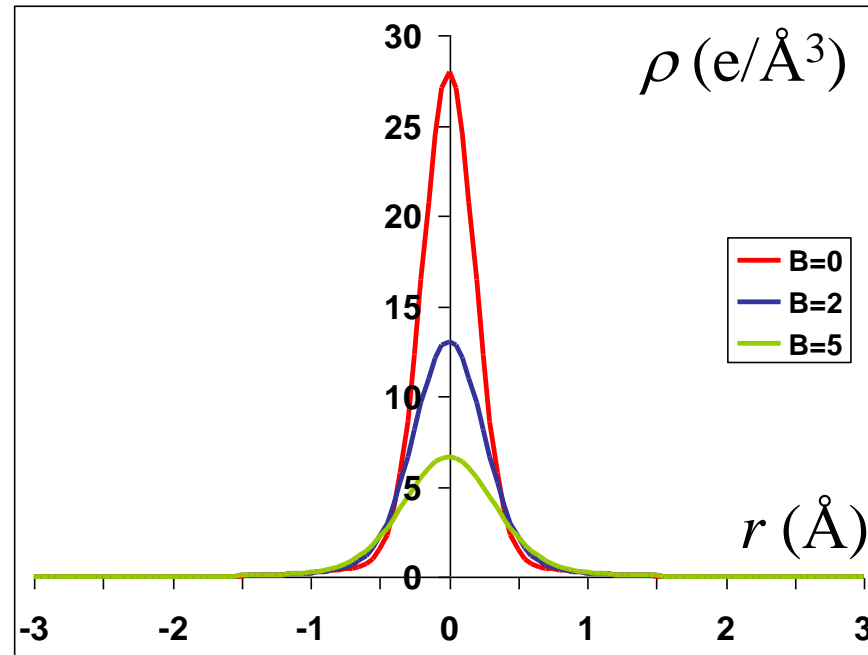
«Расплывание» плотности моделируется увеличением ширины гауссовых пиков.

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



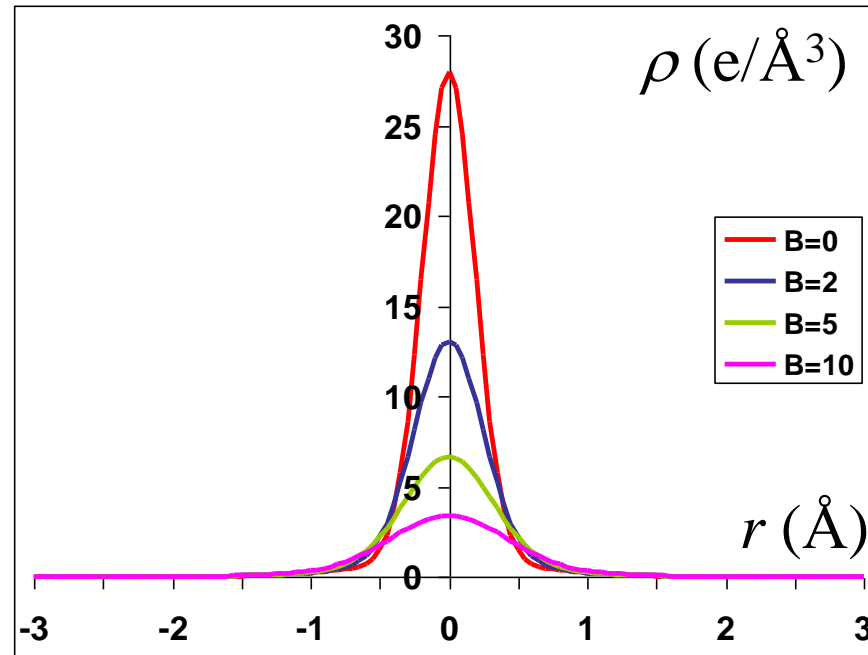
$$\rho(\mathbf{r}) = \sum_{j=1}^5 C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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



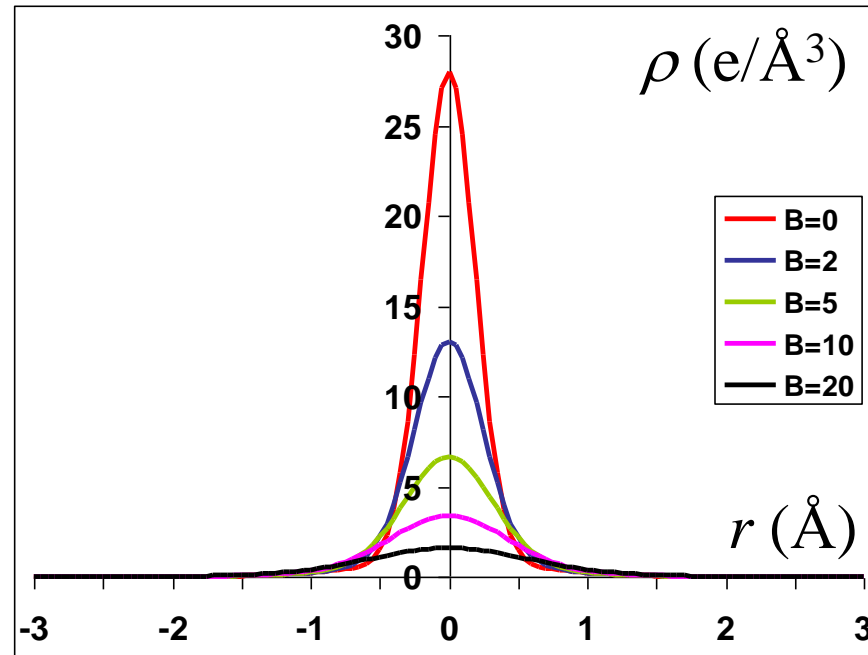
$$\rho(\mathbf{r}) = \sum_{j=1}^5 C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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)

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

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

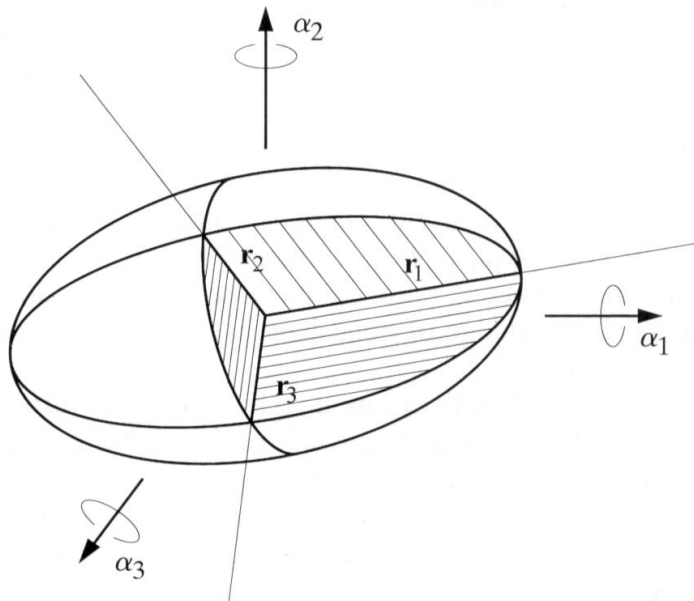
$$B = \frac{8\pi^2}{3} \langle |\Delta \mathbf{r}|^2 \rangle$$

АТОМ	30	N	SER	2	13.117	9.840	39.210	1.000	12.49
------	----	---	-----	---	--------	-------	--------	-------	--------------

ADP – не только подвижность атомов, но и гетерогенность молекул в кристалле.

Dynamic and static disorder.

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.

TLS - уточнение

(Translation/ Libration/ Screw)

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.

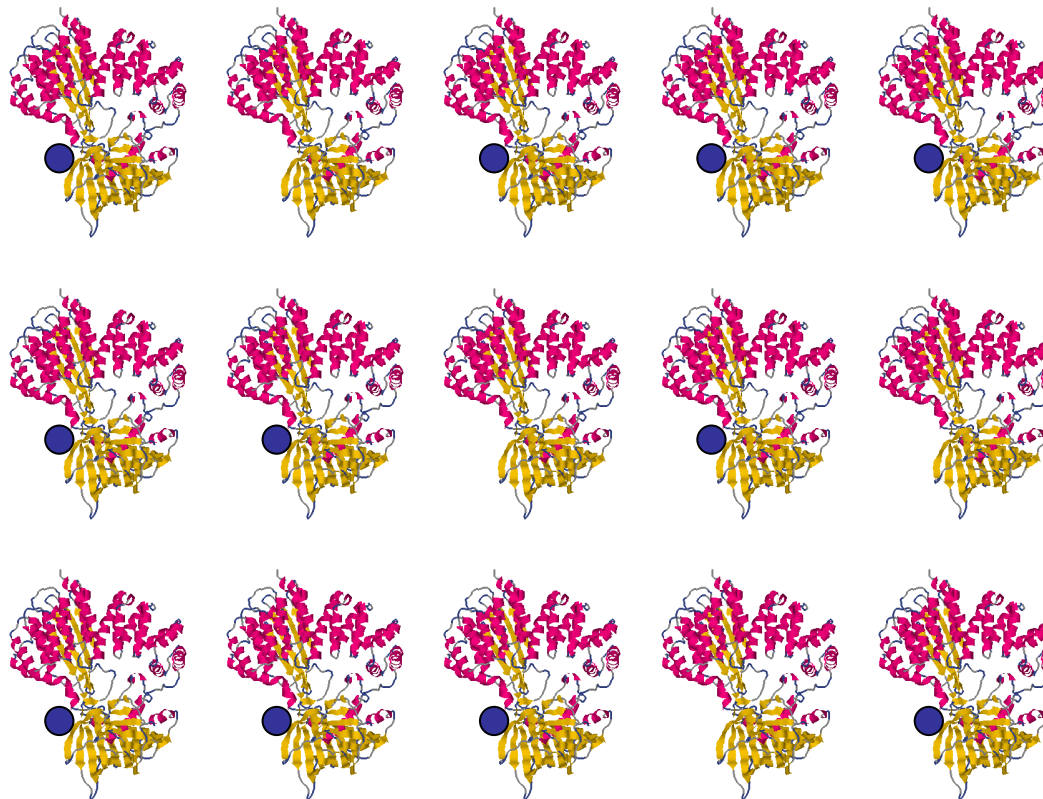
Static disorder

Коэффициент заполнения / заселенность / occupancy



Идеальный кристалл - содержимое всех элементарных ячеек идентично.

Коэффициент заполнения / заселенность / occupancy



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

Коэффициент заполнения / заселенность / occupancy

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

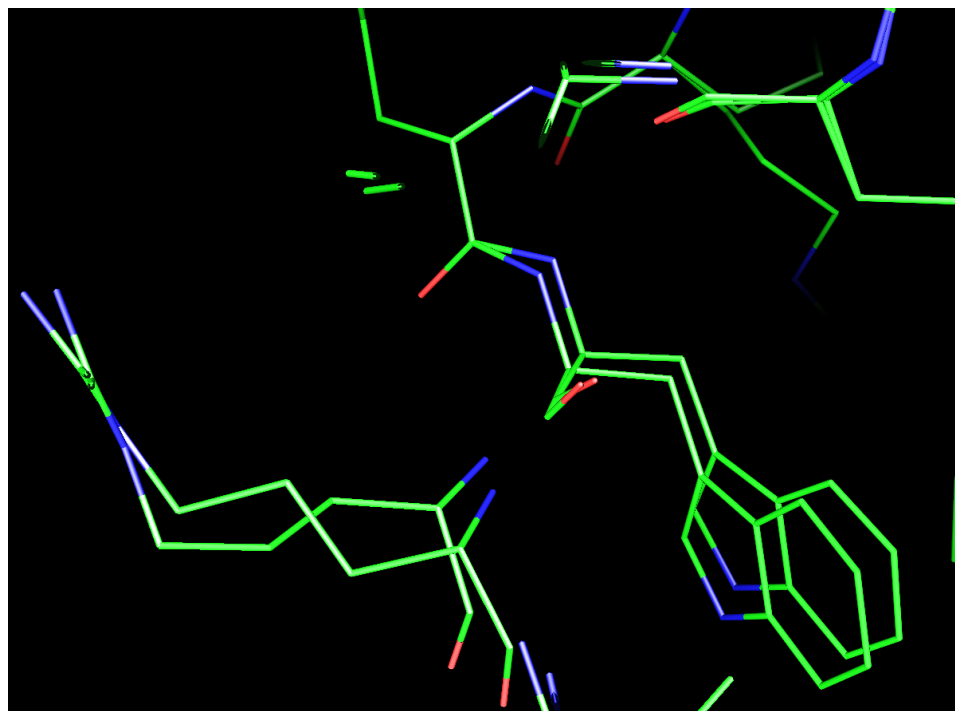
АТОМ	6626	0	НОН	4951	21.882	15.425	26.797	0.500	8.74
					<i>X</i>	<i>Y</i>	<i>Z</i>	<i>T</i>	<i>B</i>

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

$$f(s) = T \sum_{j=1}^5 C_j \exp \left(-\frac{B_j + B}{4} s^2 \right)$$

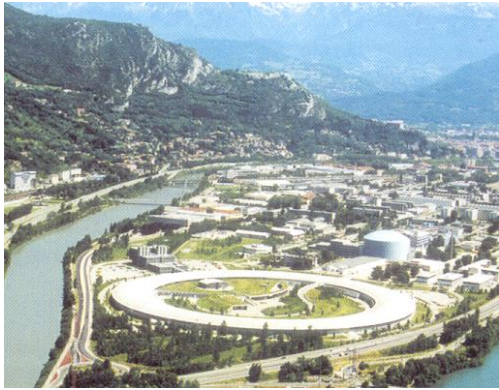
Коэффициент заполнения / заселенность / occupancy

Альтернативные конформации

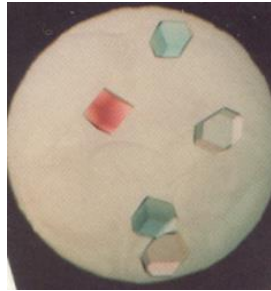


ATOM	5490	N	ATRP	295	29.848	2.643	16.199	0.499	7.13
ATOM	5514	N	BTRP	295	30.271	2.787	16.200	0.501	6.52

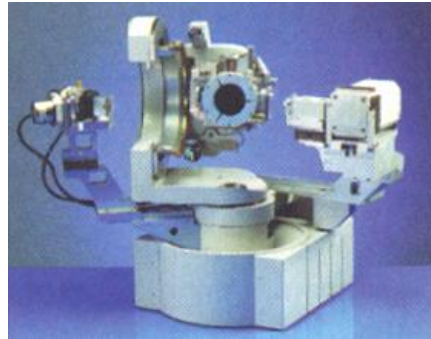
X-ray structure analysis



X-ray source

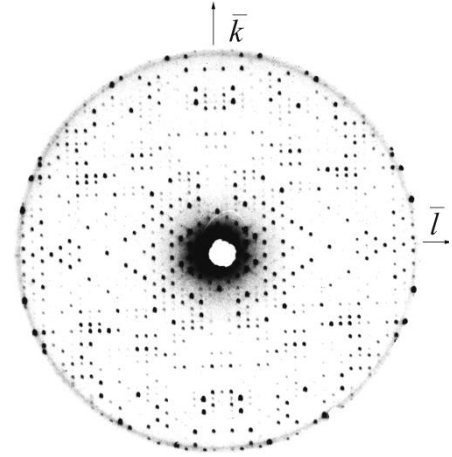


crystal

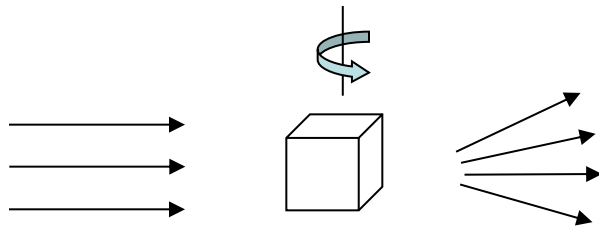


registration

Intensities I_s of diffracted beams



X-ray experiment

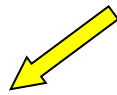


Kinematic theory of diffraction

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

$$F_s^2 \propto I_s$$

h	k	l	F	σ
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

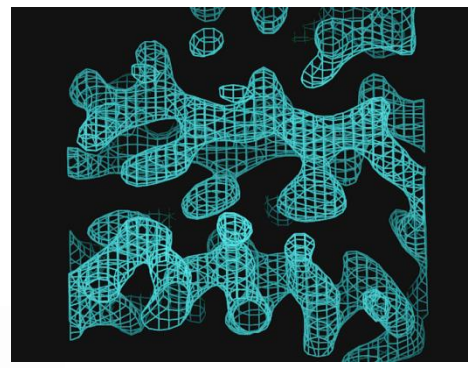
X-ray structure analysis



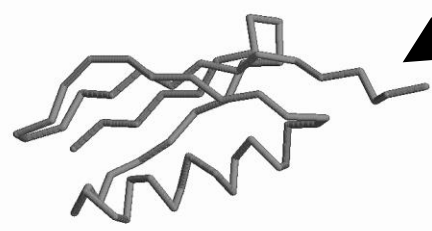
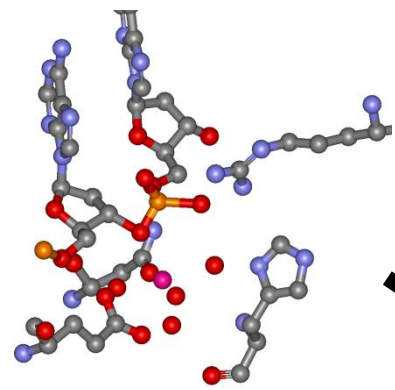
$$\Omega_{\kappa} = \{\mathbf{r} : \rho_S(\mathbf{r}) > \kappa\}$$

h	k	l	F	σ
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



refined model



preliminary model

Protein Data Bank					X	Y	Z	T	B		
		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. (Diffraction before distraction)

Метод работает

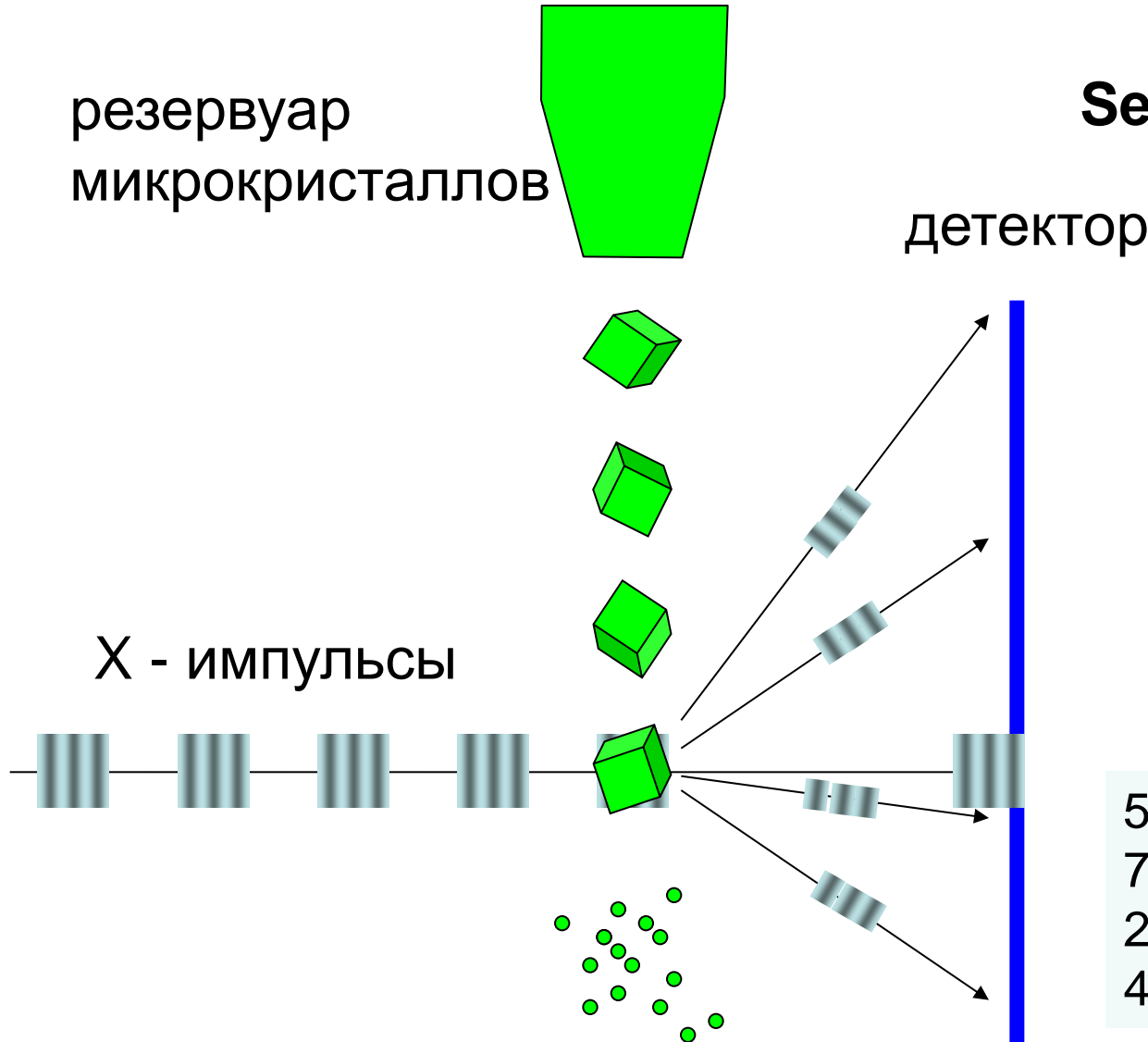
Область использования – нанокристаллы, отдельные молекулы



«Проточная» кристаллография (serial crystallography).

«Проточная» рентгеновская кристаллография

Serial crystallography

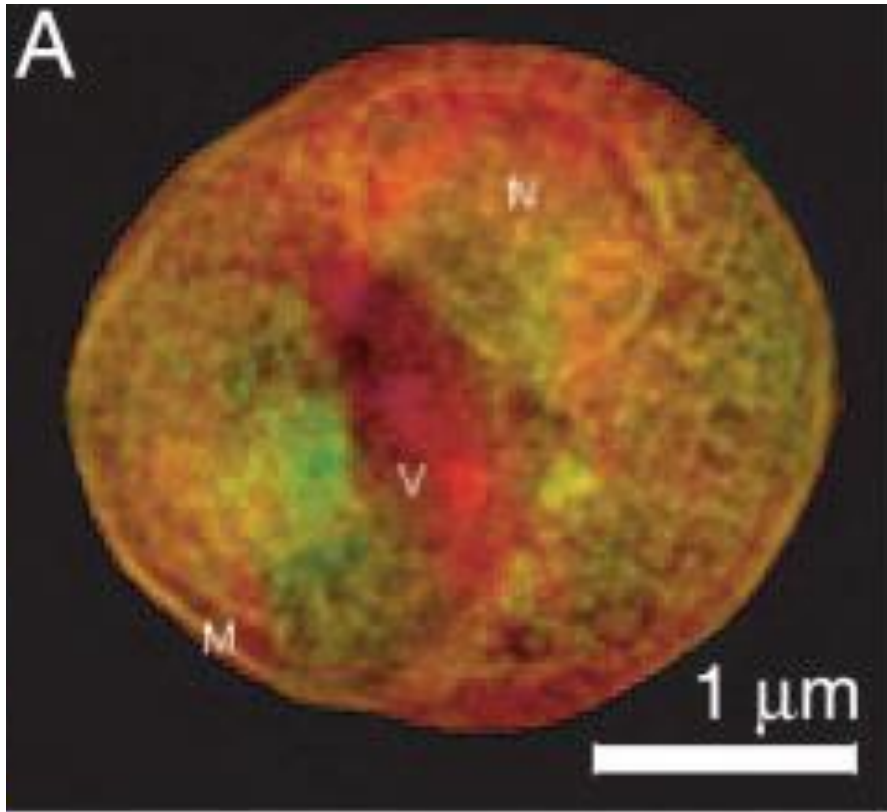


Одна
рентгенограмма с
одного кристалла

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

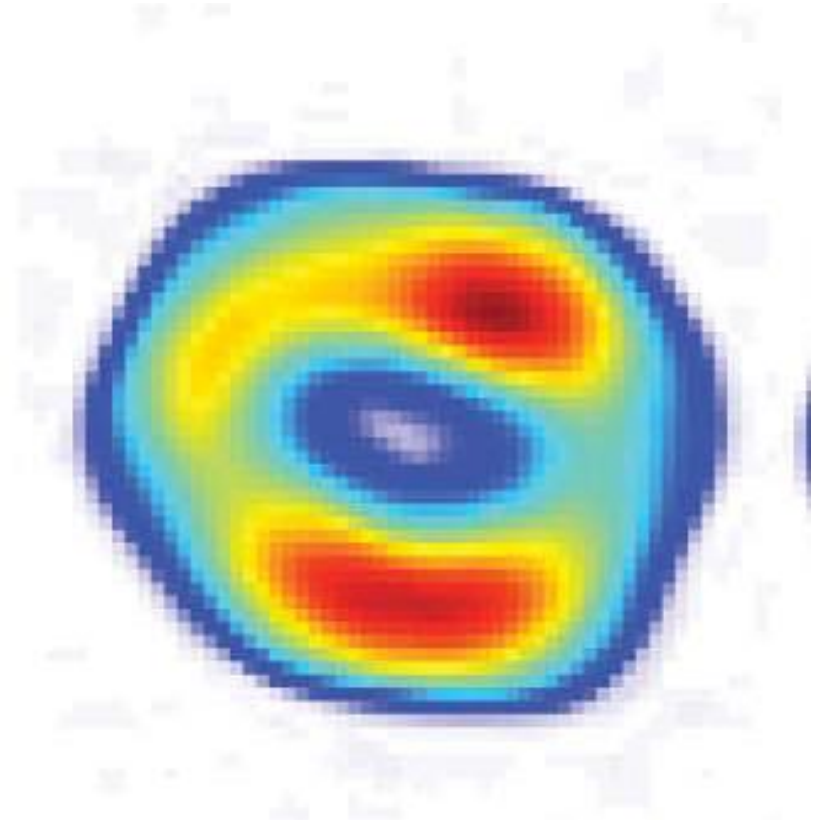
Исследование отдельных частиц

Дрожжевая клетка



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

Мимивирус



200 nm

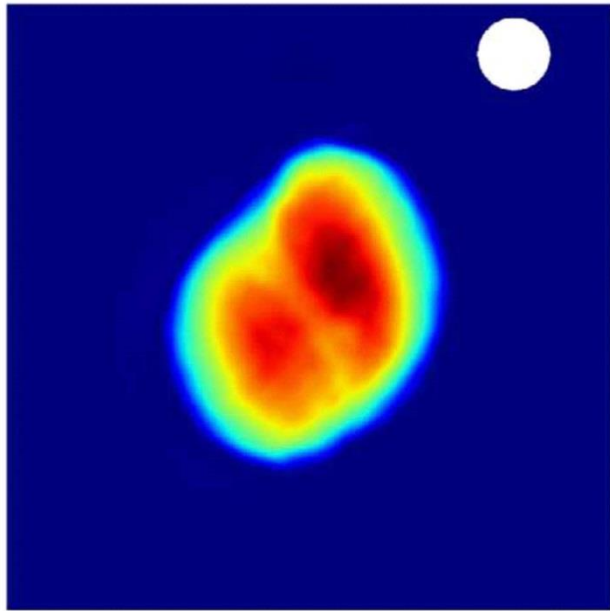
Siebert et al., 2011, Nature, 470, 78-82

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



$10,000 \text{ \AA} = 1 \text{ \mu m}$

