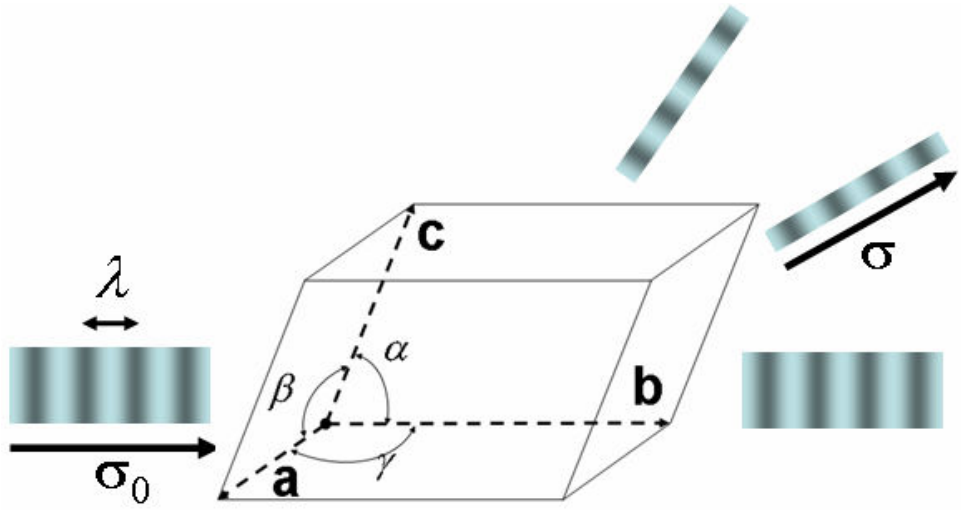
Введение в кристаллографию макромолекул

Лекция 3

Владимир Юрьевич ЛУНИН

Институт Математических Проблем Биологии РАН
Пущино

http://www.impb.ru/lmc

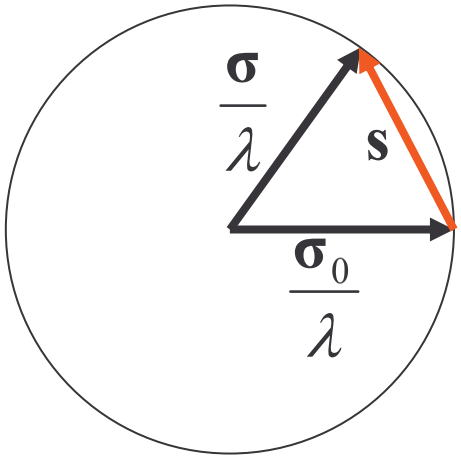


Рентгеновский эксперимент с монокристаллом позволяет практически измерить интенсивность волн, рассеянных в направлениях, определяемых условиями

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

$$\left(\frac{\boldsymbol{\sigma}-\boldsymbol{\sigma}_0}{\lambda},\mathbf{a}\right) = h, \left(\frac{\boldsymbol{\sigma}-\boldsymbol{\sigma}_0}{\lambda},\mathbf{b}\right) = k,$$

$$\left(\frac{\boldsymbol{\sigma}-\boldsymbol{\sigma}_0}{\lambda},\mathbf{c}\right) = l$$



h,k,l - целые числа (индексы рефлекса)

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

Рассеянная волна отличается от первичной множителем F в амплитуде и сдвигом φ фазы

Вклады в амплитуду и фазу рассеянной волны зависят от распределения электронной плотности в элементарной ячейке и для известного распределения электронной плотности могут быть рассчитаны по формулам

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

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

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

F - модуль структурного фактора, φ - фаза структурного фактора

Измеряемая интенсивность рассеянной волны пропорциональна квадрату модуля структурного фактора

Ряды Фурье для функций нескольких переменных

Любая функция может быть разложена, при этом единственным образом, в ряд Фурье.

$$\rho(x, y, z) \qquad 0 \le x, y, z \le 1$$

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

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

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

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

Формулы для расчета коэффициентов Фурье совпадают с формулами для структурных факторов

Знание структурных факторов дает возможность рассчитать «синтез Фурье» электронной плотности

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

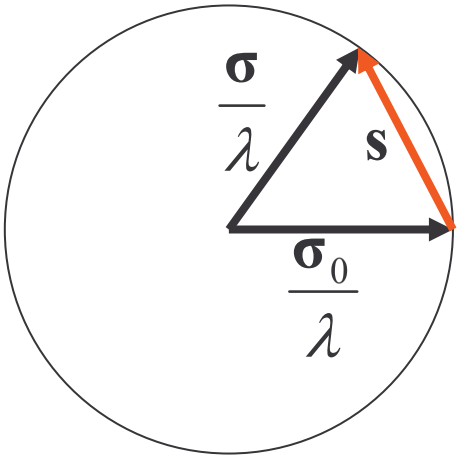
Отличие синтеза Фурье от «настоящего» распределения электронной плотности - конечный набор слагаемых при расчете синтеза.

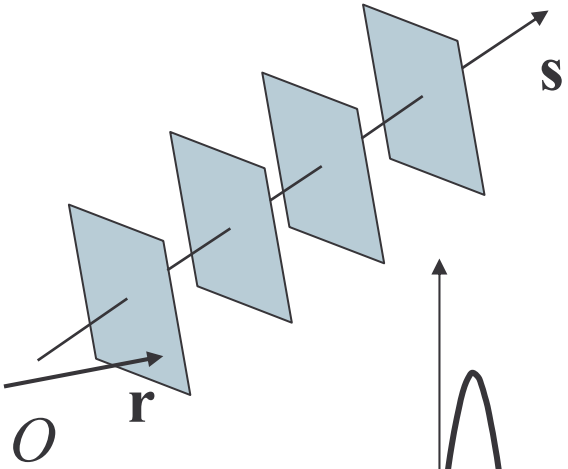
Гармоника Фурье

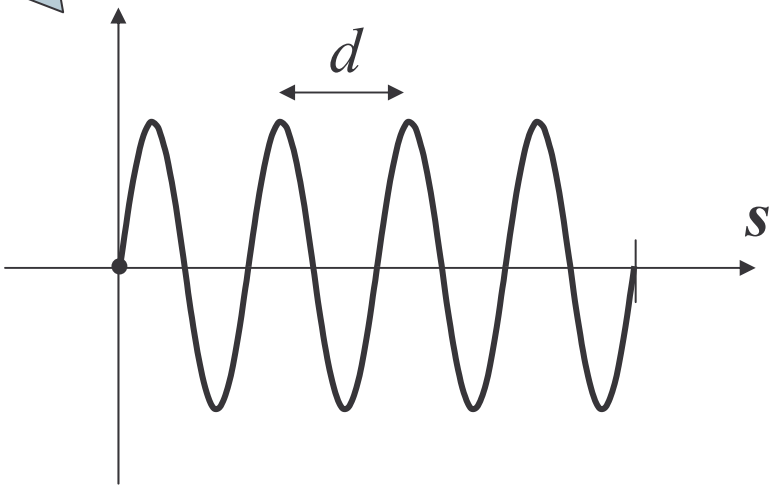
$$F_{hkl}\cos\left[2\pi\left(hx+ky+lz\right)-\varphi_{hkl}\right]=F(\mathbf{s})\cos\left[2\pi(\mathbf{s},\mathbf{r})-\varphi(\mathbf{s})\right]$$

Разрешение, соответствующее гармонике Фурье

$d=\dfrac{1}{|\mathbf{s}|}$ - расстояние между соседними максимумами в направлении **s**; минимальный размер «видимых» деталей



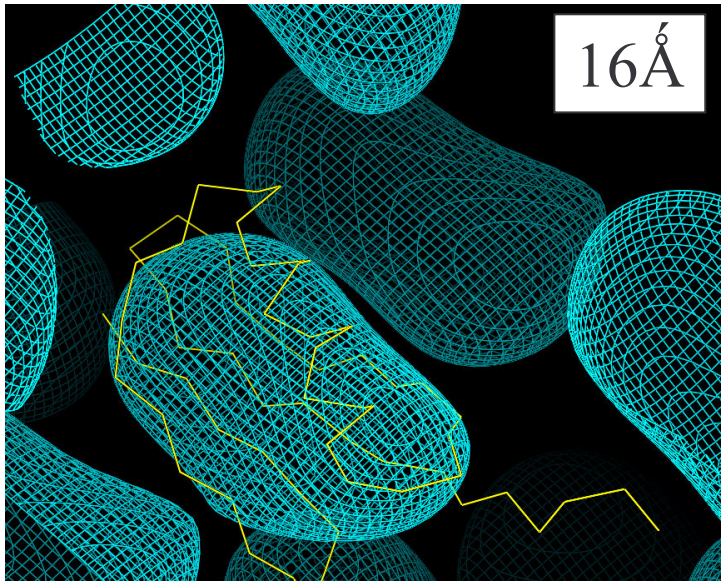


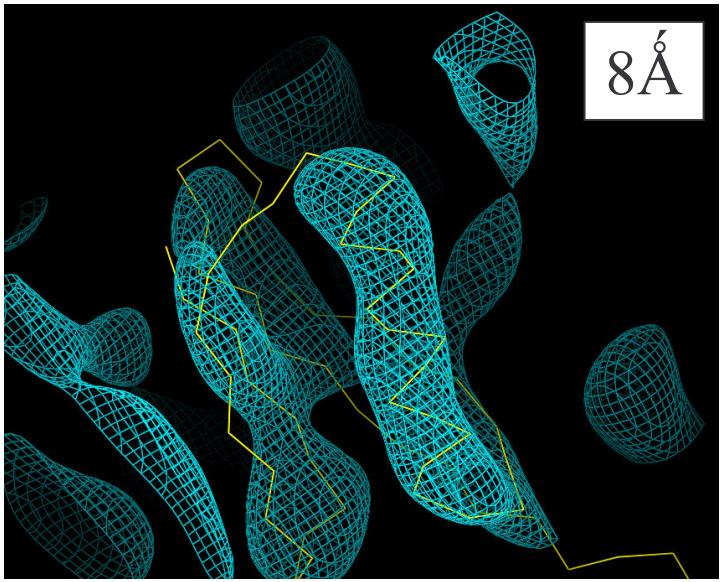


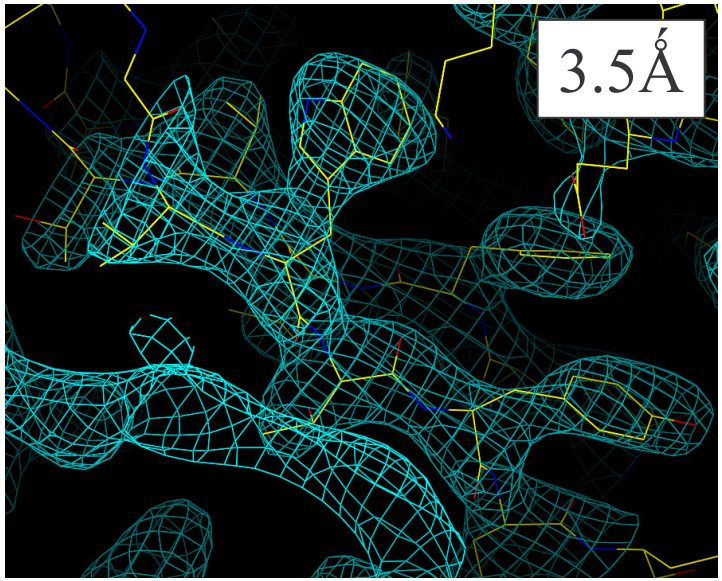
$\mathbf{s}=\dfrac{\boldsymbol{\sigma}-\boldsymbol{\sigma}_0}{\lambda}$ - вектор рассеяния

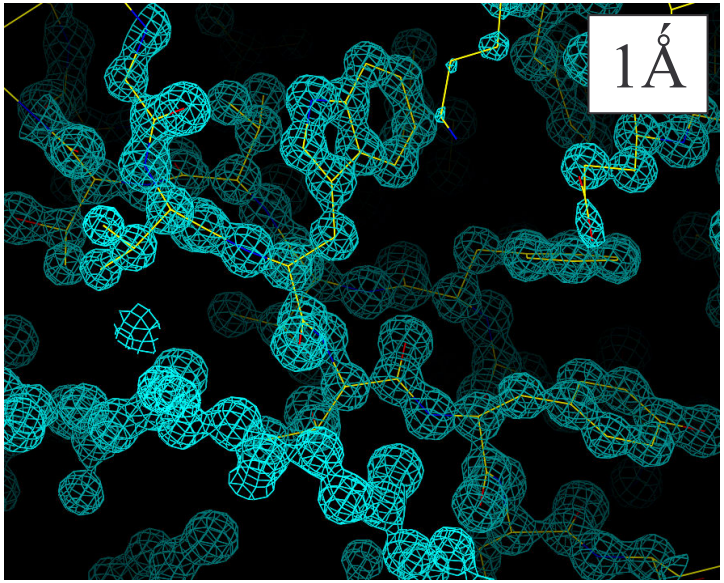
Views of the high-density region in Protein G constructed on the base of Fourier syntheses of different resolutions.

X-ray structure analysis

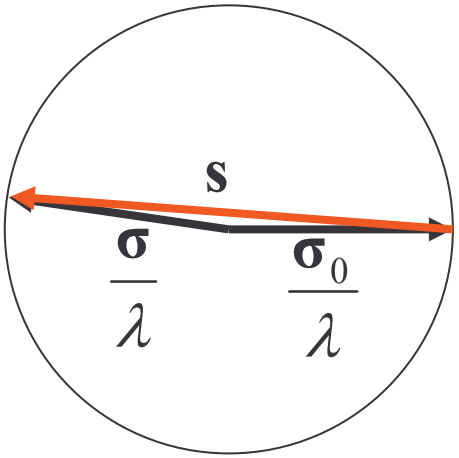








$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \qquad d = \frac{1}{|\mathbf{s}|}$$



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

$$|\mathbf{s}|_{\max} = \frac{2}{\lambda} \qquad d_{\min}^{theoretical} = \frac{\lambda}{2}$$

Картина рассеяния рентгеновских лучей (модули и фазы структурных факторов), теоретически, содержит информацию о синтезе Фурье электронной плотности разрешения $\lambda/2$.

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

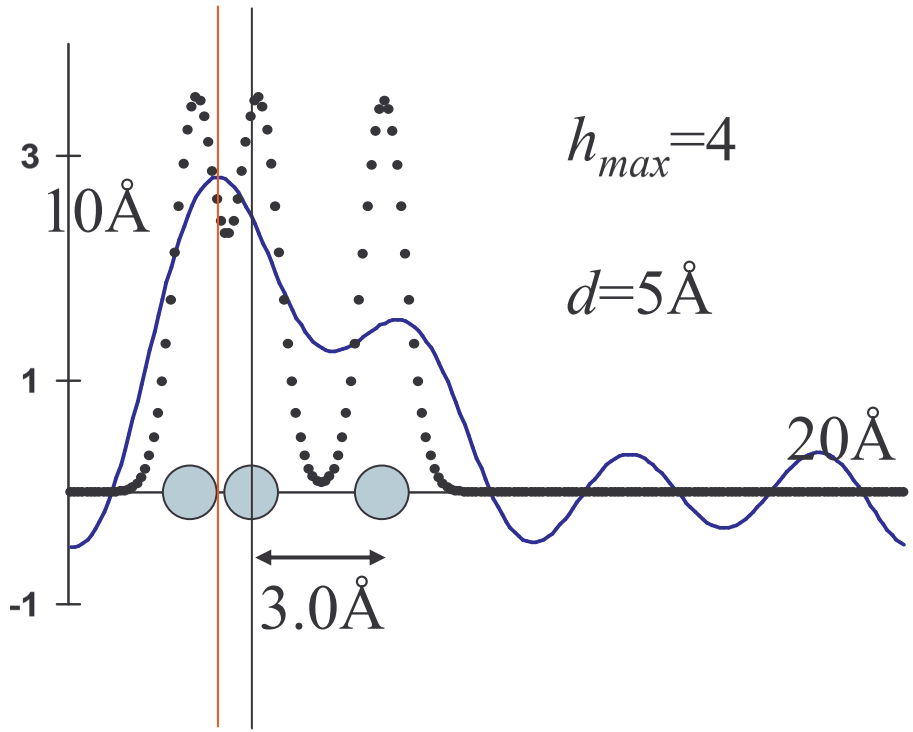
crambin, 0.54 A (Teeter et al., 1993);

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

aldose reductase, 0.66 A (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
```

Разрешение набора структурных факторов характеризует размер деталей «видимых» на картах распределения электронной плотности. Оно не связано прямо с точностью определения координат атомов методом РСА!



1. Рентгеновский дифракционный эксперимент с монокристаллом позволяет измерить модули коэффициентов в разложении функции распределения электронной плотности в ряд Фурье.

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

2. Наличие модели структуры дает возможность рассчитать гипотетические значения модулей и фаз и сравнить модули с экспериментальными значениями

3. Знание значений модулей и фаз дает возможность рассчитать "синтез Фурье электронной плотности"

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

Функция Паттерсона

$$\rho(x) \approx F_0 + 2\sum_{h=1}^{h_{\max}} F_h \cos[2\pi(hx) - \varphi_h]$$ - требует знания фаз

Функция Паттерсона

$$\rho(x) \approx F_0 + 2\sum_{h=1}^{h_{\max}} F_h \cos[2\pi(hx) - \varphi_h]$$ - требует знания фаз

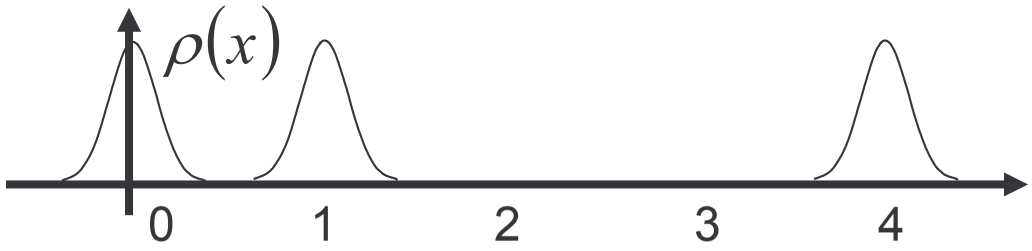
$$P(x) \approx (F_0)^2 + 2\sum_{h=1}^{h_{\max}} (F_h)^2 \cos[2\pi(hx)]$$ - может быть рассчитана по экспериментальным данным

Функция Паттерсона

$$\rho(x) \approx F_0 + 2\sum_{h=1}^{h_{\max}} F_h \cos[2\pi(hx) - \varphi_h]$$ - требует знания фаз

$$P(x) \approx (F_0)^2 + 2\sum_{h=1}^{h_{\max}} (F_h)^2 \cos[2\pi(hx)]$$ - может быть рассчитана по экспериментальным данным

$$\rho(x) \approx \sum_{j=1}^{N} \rho_0(x - r_j)$$

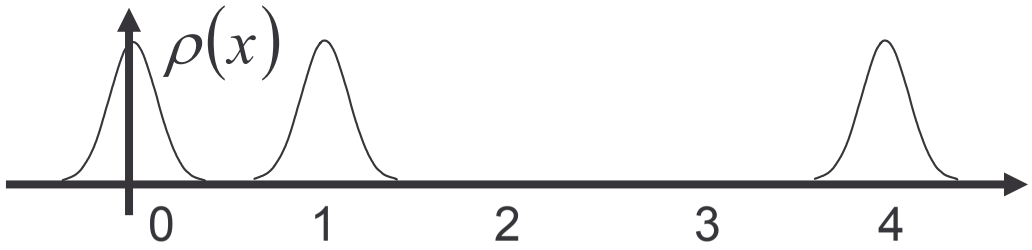


Функция Паттерсона

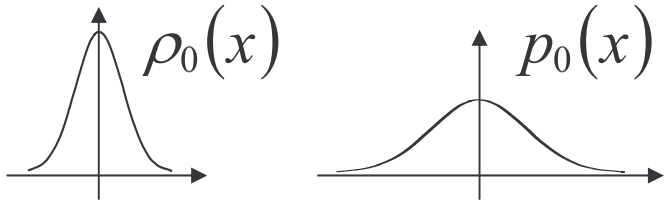
$$\rho(x) \approx F_0 + 2\sum_{h=1}^{h_{\max}} F_h \cos[2\pi(hx) - \varphi_h]$$ - требует знания фаз

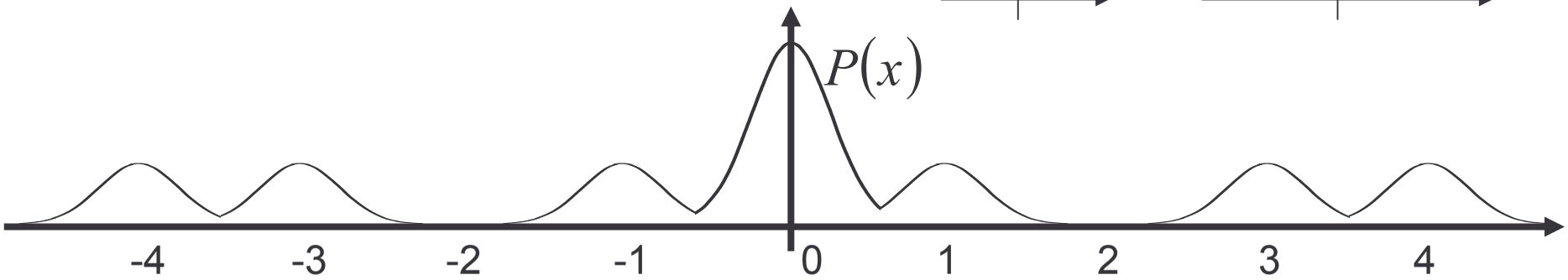
$$P(x) \approx (F_0)^2 + 2\sum_{h=1}^{h_{\max}} (F_h)^2 \cos[2\pi(hx)]$$ - может быть рассчитана по экспериментальным данным

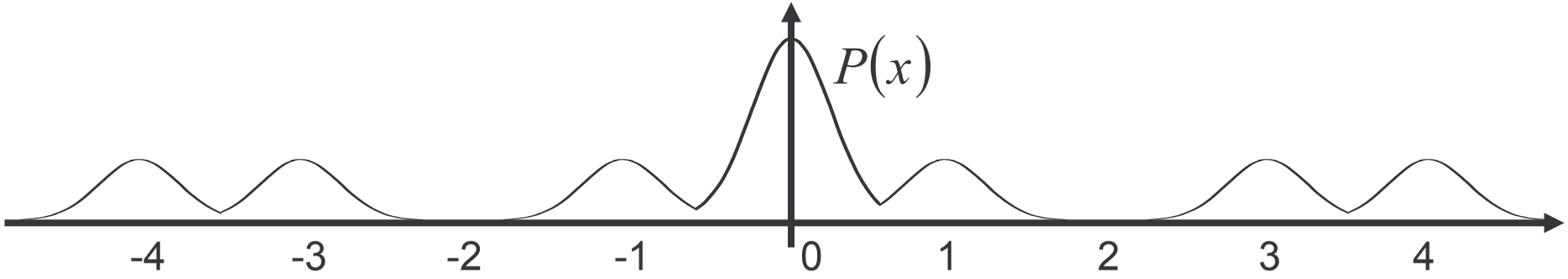
$$\rho(x) \approx \sum_{j=1}^{N} \rho_0(x - r_j)$$



$$P(x) \approx \sum_{j,k=1}^{N} p_0(x - (r_j - r_k))$$







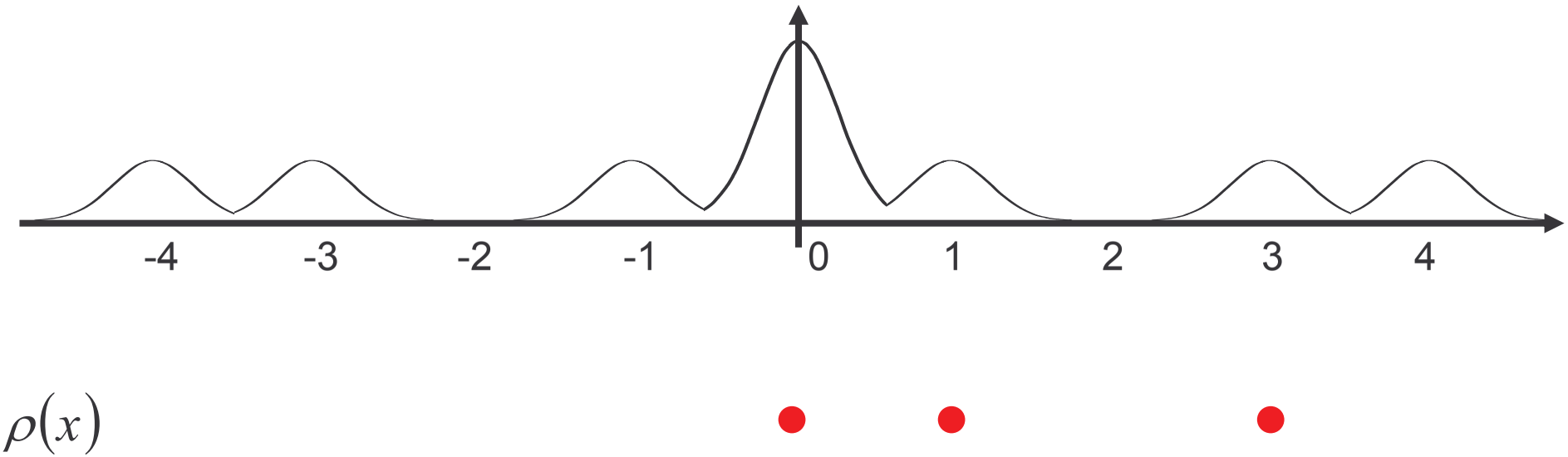
$\rho(x)$ $\quad \mathbf{0}, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_N$

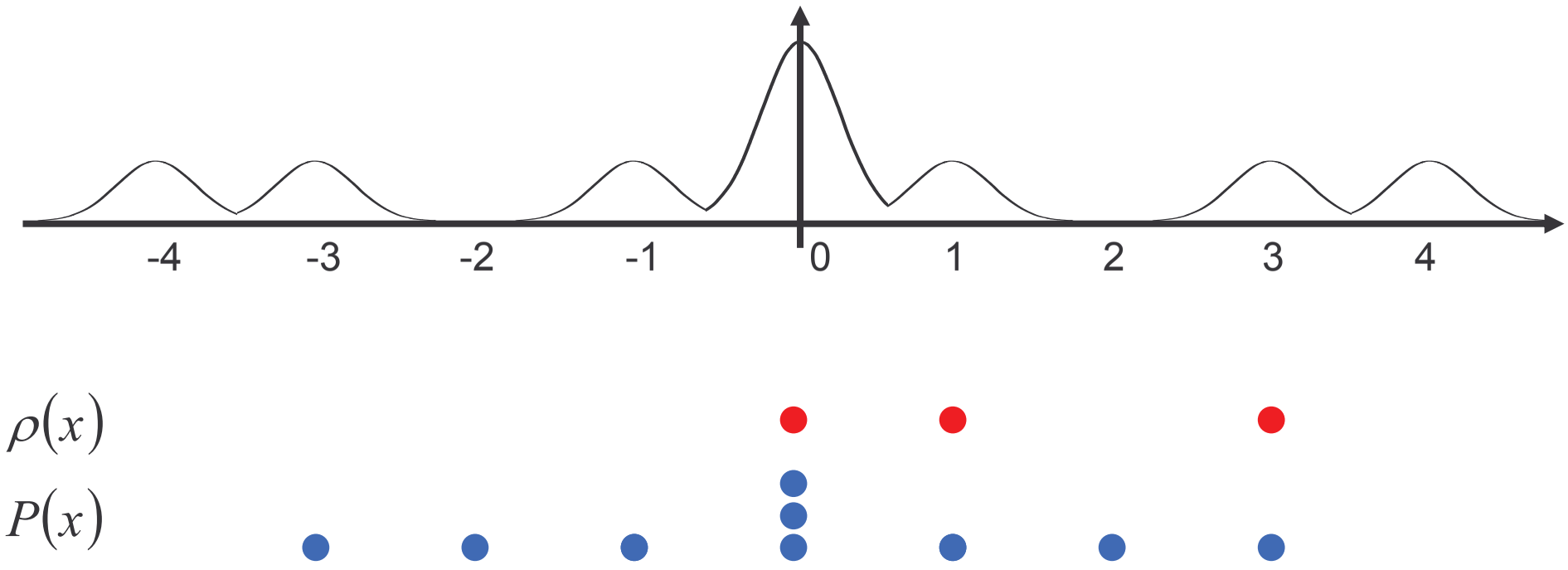
$$P(x) \approx \sum_{j,k=1}^{N} p_0\left(x - \left(r_j - r_k\right)\right)$$

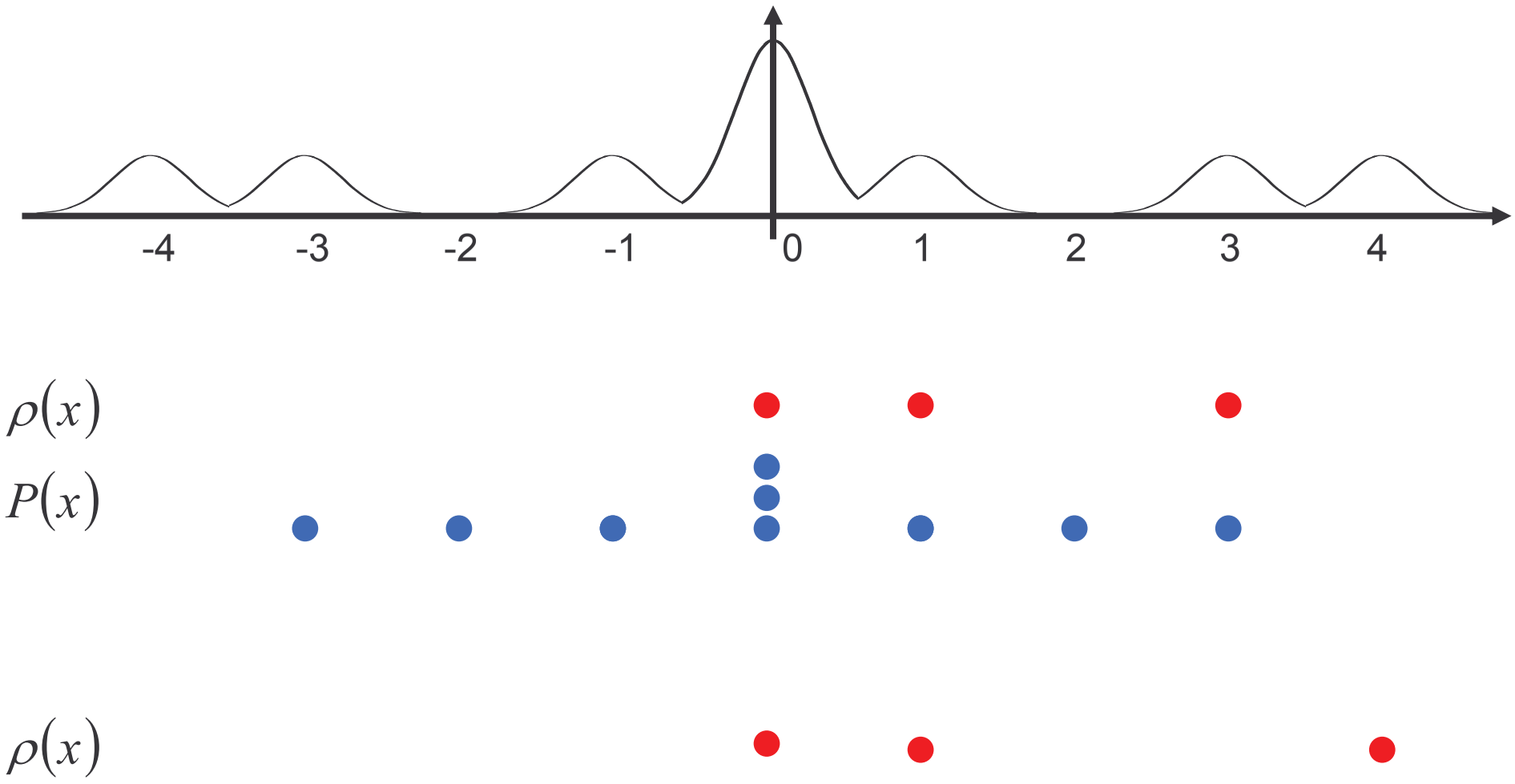
$P(x)$

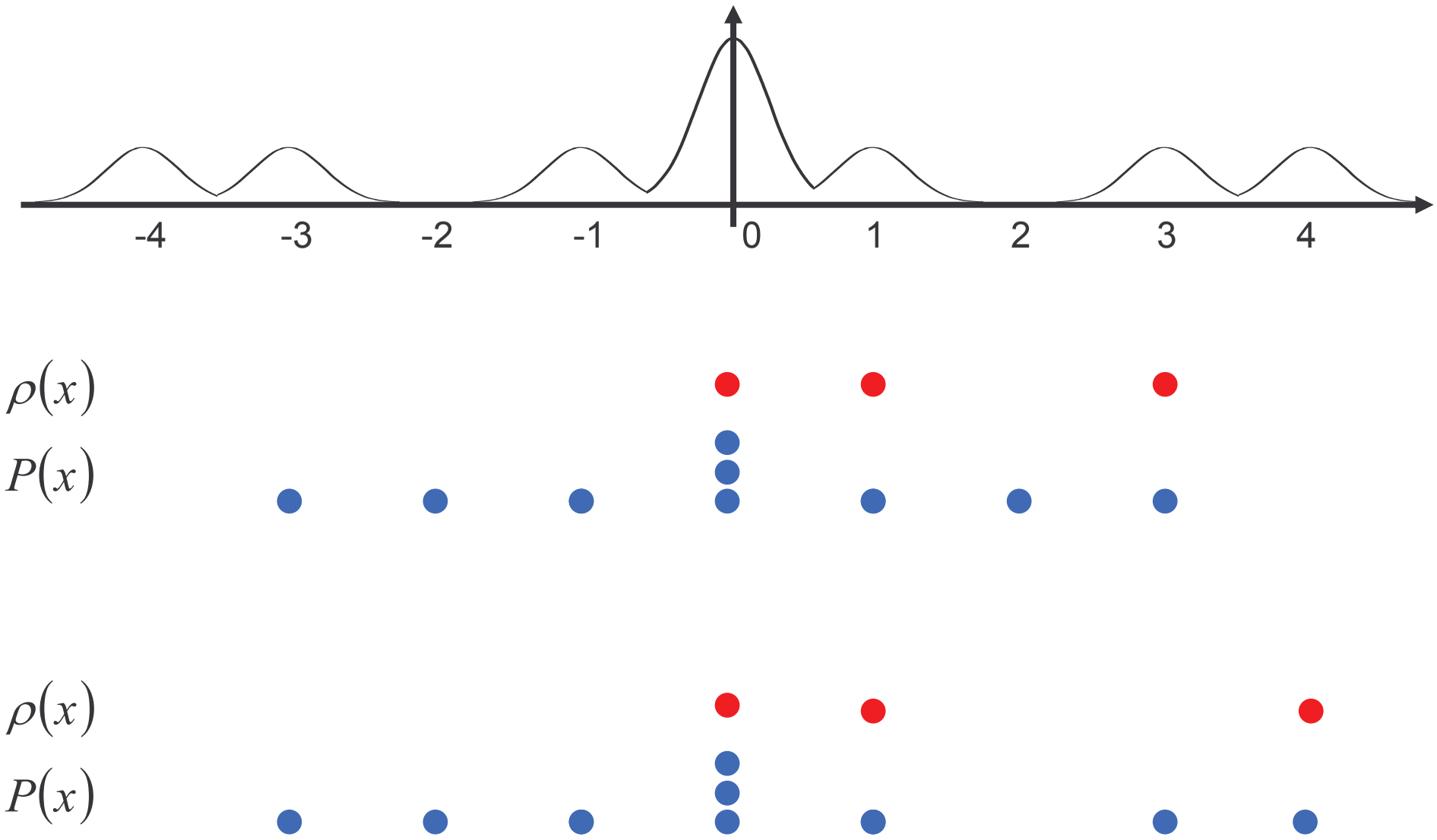
0	$\mathbf{r}_2$	$\mathbf{r}_3$	. . .	$\mathbf{r}_N$
$-\mathbf{r}_2$	**0**	$\mathbf{r}_3-\mathbf{r}_2$	. . .	$\mathbf{r}_N-\mathbf{r}_2$
$-\mathbf{r}_3$	$\mathbf{r}_2-\mathbf{r}_3$	**0**	. . .	$\mathbf{r}_N-\mathbf{r}_3$
. . .	. . .	. . .	. . .	. . .
$-\mathbf{r}_N$	$\mathbf{r}_2-\mathbf{r}_N$	$\mathbf{r}_3-\mathbf{r}_N$	. . .	**0**

Исходная структура содержится в пиках функции Паттерсона


-4
-3
-2
-1
0
1
2
3
4
$\rho(x)$


-4
-3
-2
-1
0
1
2
3
4
$\rho(x)$
$P(x)$


-4
-3
-2
-1
0
1
2
3
4
$\rho(x)$
$P(x)$
$\rho(x)$


-4
-3
-2
-1
0
1
2
3
4
$\rho(x)$
$P(x)$
$\rho(x)$
$P(x)$

Паттерсоновские методы.

Надо идентифицировать N^2 пиков.

Пики - более широкие, нежели "атомные" пики.

Не все коэффициенты можно измерить.

⇓

Используются для решения структур, состоящих из небольшого числа атомов. Применяются для определения субструктур («тяжелых» атомов) в белках.

Подходы к решению фазовой проблемы

SIR - Single Isomorphous Replacement
Однократное изоморфное замещение;

MIR - Multiple Isomorthous Replacement
Множественное изоморфное замещение;

AD - Anomalous Dispersion
Аномальное рассеяние;

MAD - Multiwavelength Anomalous Diffraction
Многоволновое аномальное рассеяние;

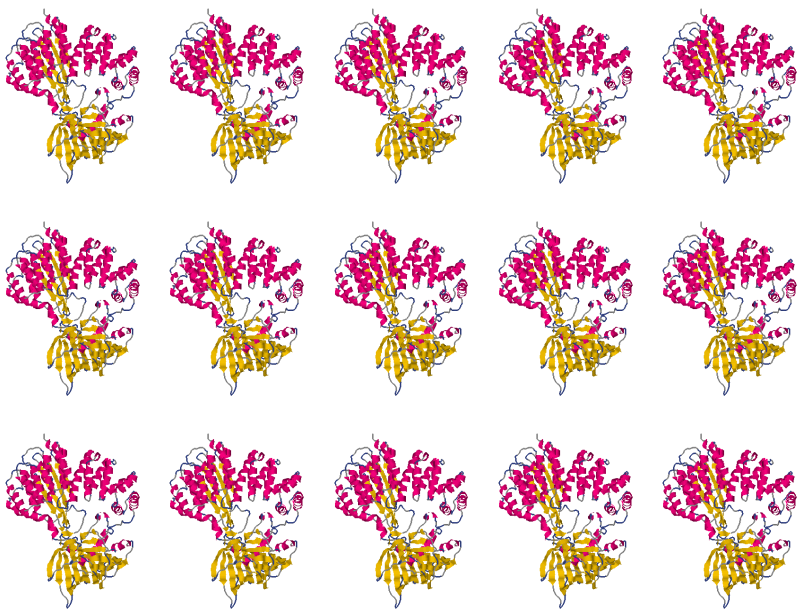
MR - Molecular Replacement
Молекулярное замещение;

Direct Methods (*ab initio* methods)
"Прямые" методы.

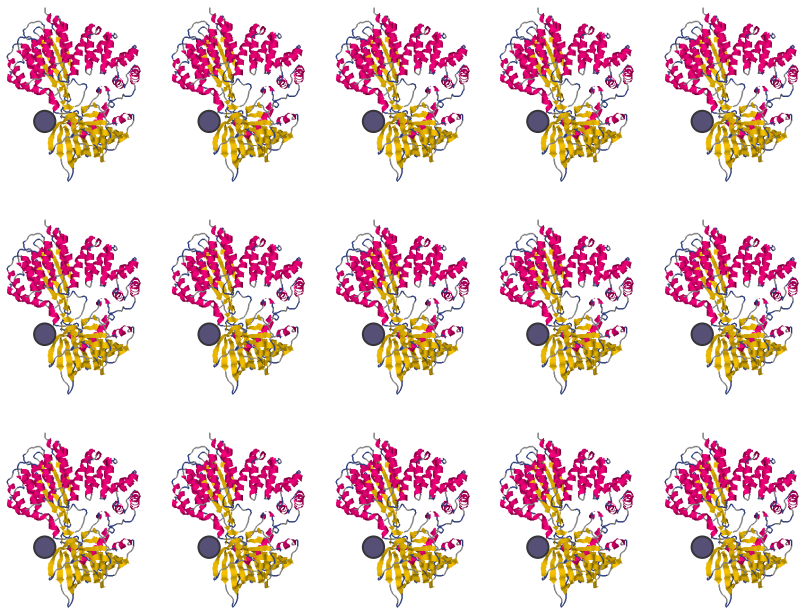
Подходы к решению фазовой проблемы

SIR - Single Isomorphous Replacement
Однократное изоморфное замещение;

"нативный" белок



изоморфное производное



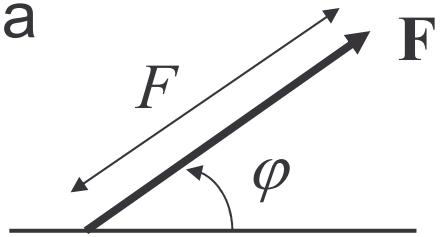
$$\{F^P\} \qquad \{F^{PH}\}$$

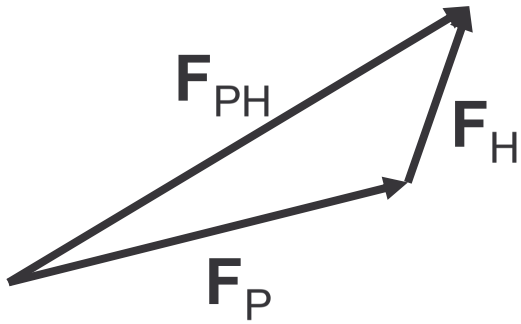
$$\{\varphi^P\}$$

1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \quad \Longrightarrow \quad \mathbf{r}_H \quad \Longrightarrow \quad F_H, \varphi_H$$

2. Фазовая диаграмма

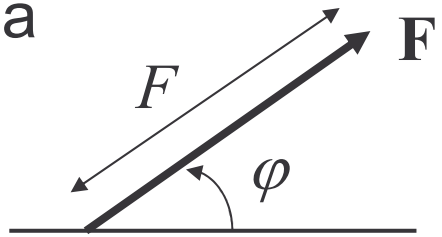


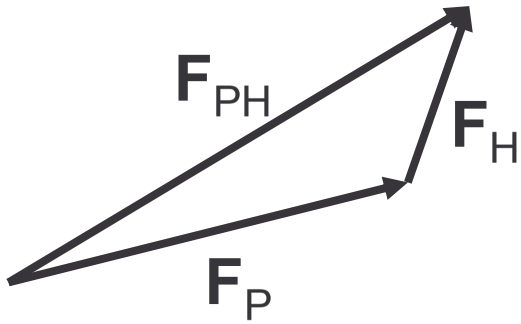


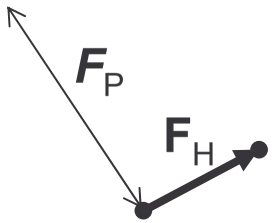
1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \Longrightarrow \mathbf{r}_H \Longrightarrow F_H, \varphi_H$$

2. Фазовая диаграмма



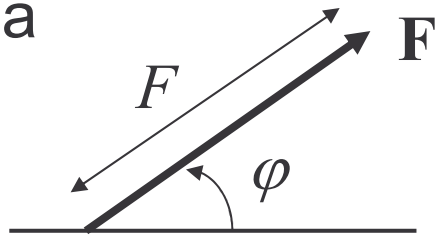


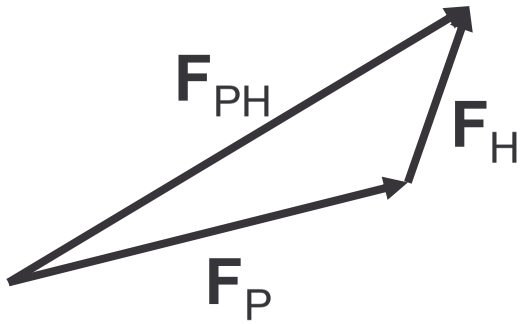


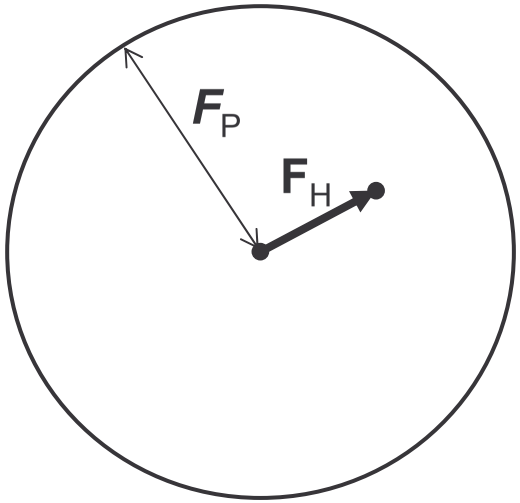
1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \quad \Longrightarrow \quad \mathbf{r}_H \quad \Longrightarrow \quad F_H, \varphi_H$$

2. Фазовая диаграмма



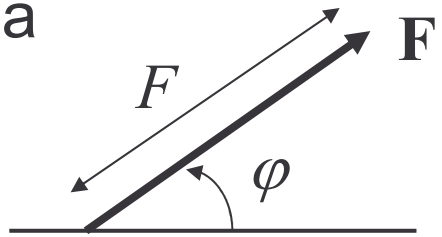


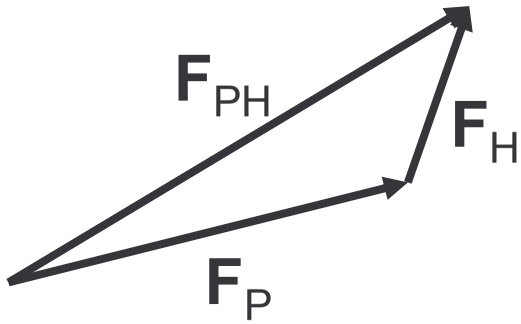


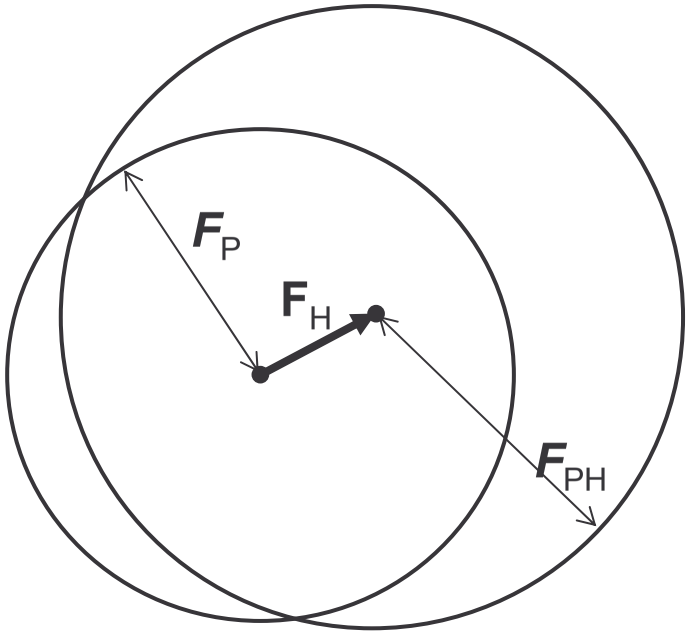
1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \Longrightarrow \mathbf{r}_H \Longrightarrow F_H, \varphi_H$$

2. Фазовая диаграмма



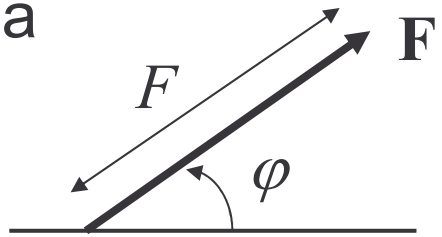


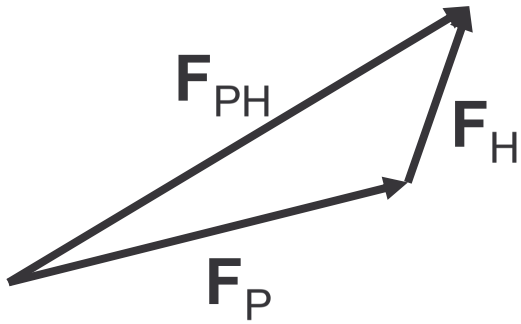


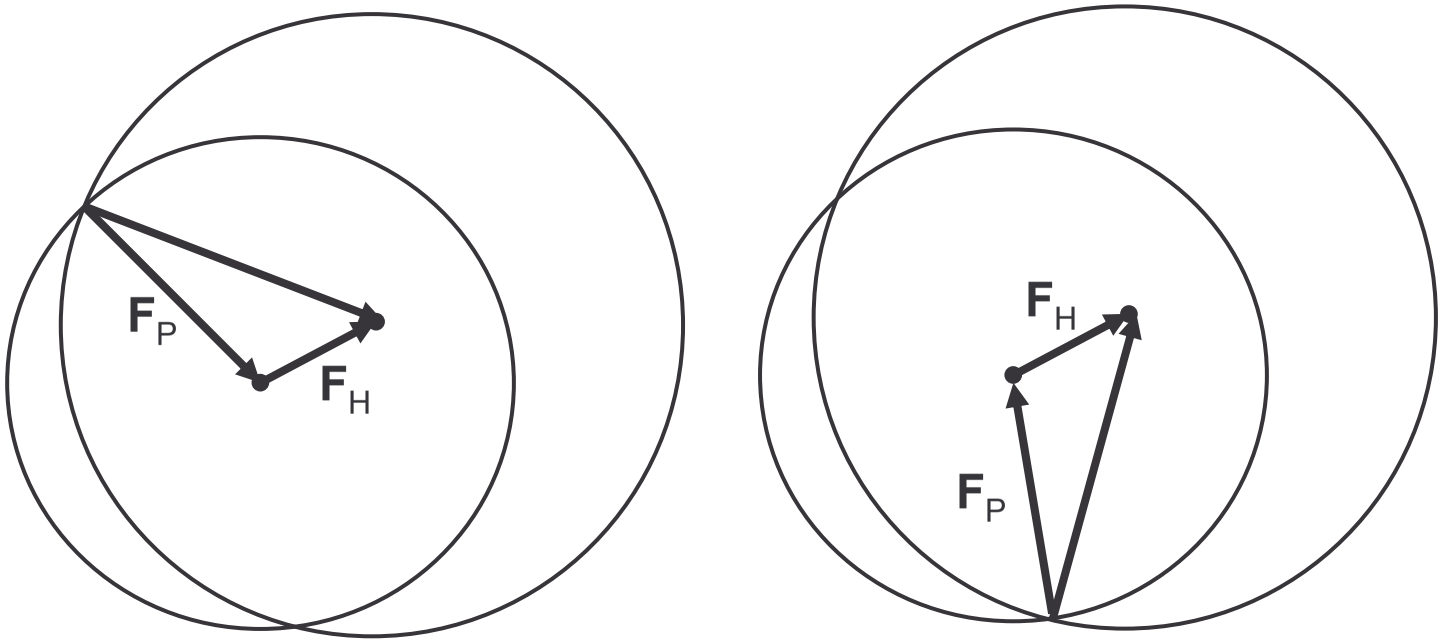
1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \Longrightarrow \mathbf{r}_H \Longrightarrow F_H, \varphi_H$$

2. Фазовая диаграмма



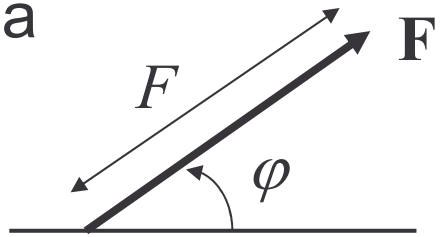


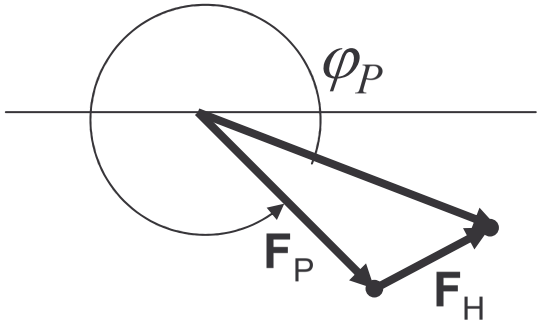


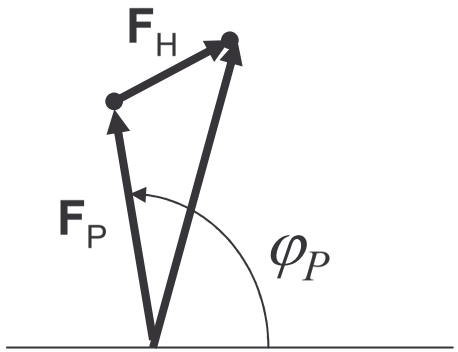
1. Определение положений тяжелых атомов.
(Паттерсоновские методы)

$$F_H \approx \left| F_{PH} - F_P \right| \Longrightarrow \mathbf{r}_H \Longrightarrow F_H, \varphi_H$$

2. Фазовая диаграмма







два возможных
значения фазы

- 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 Isomorthous Replacement
Множественное изоморфное замещение;

AD - Anomalous Dispersion
Аномальное рассеяние

Из кинематической теории рассеяния следует, что

$$F(-\mathbf{s}) = F(\mathbf{s})$$

("закон Фриделя").

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

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

$$F_{hkl} = \sqrt{A_{hkl}^2 + B_{hkl}^2}$$

При наличии аномально рассеивающих атомов закон нарушается. Измерения $F(-\mathbf{s})$ и $F(\mathbf{s})$ становятся независимыми. Их сравнение позволяет решать фазовую проблему.

AD - Anomalous Dispersion
Аномальное рассеяние

Из кинематической теории рассеяния следует, что

$$F(-\mathbf{s}) = F(\mathbf{s})$$

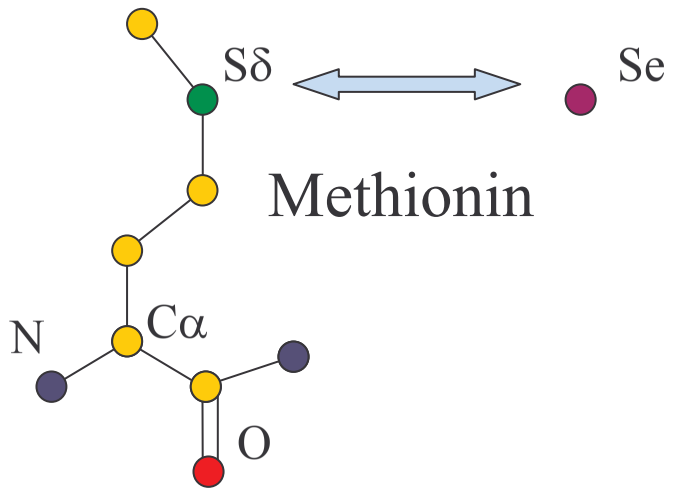
(закон Фриделя). При наличии аномально рассеивающих атомов (а.р.а.) закон нарушается. Измерения $F(-\mathbf{s})$ и $F(\mathbf{s})$ становятся независимыми. Их сравнение позволяет решать фазовую проблему.

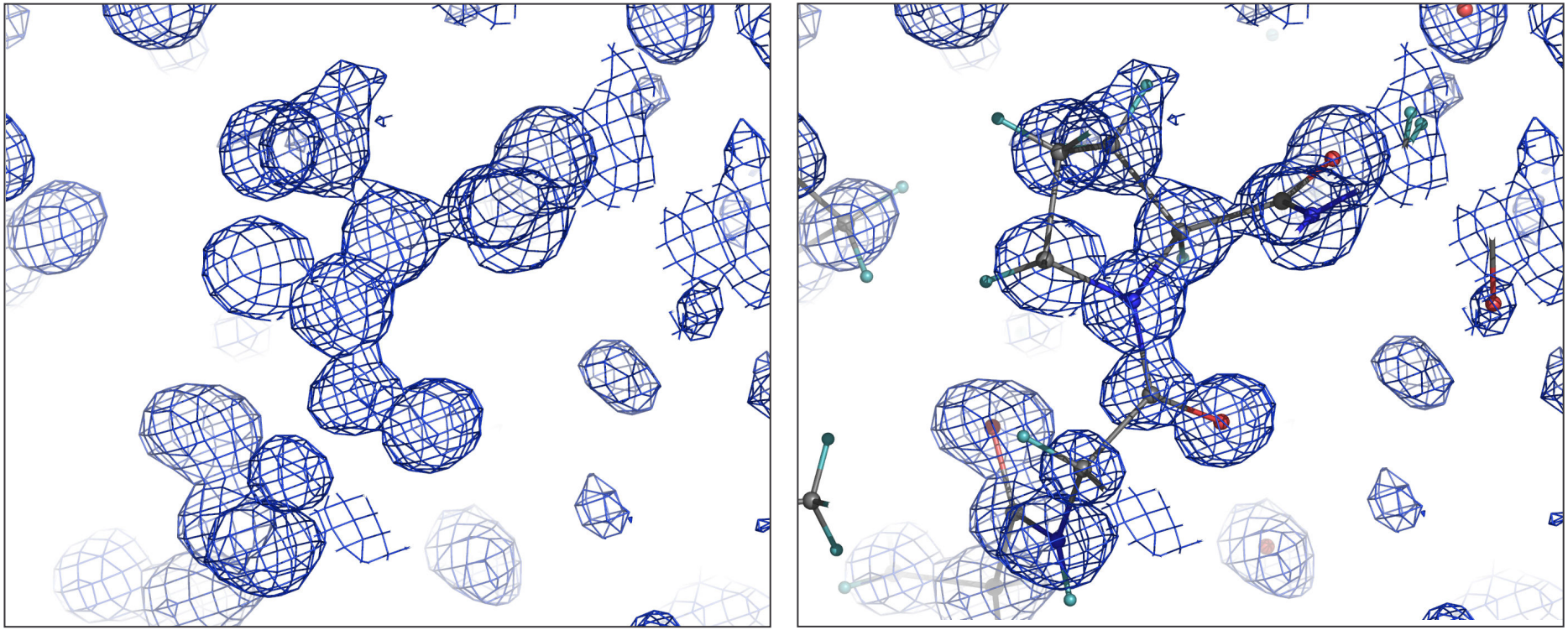
☺ • метод позволяет решать фазовую проблему;
☺ • в белках встречаются а.р.а.; можно использовать производные с тяжелыми атомами;
☹ • проблема определения мест нахождения аномально рассеивающих атомов;
☹ • фазы определяются неоднозначно;

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.



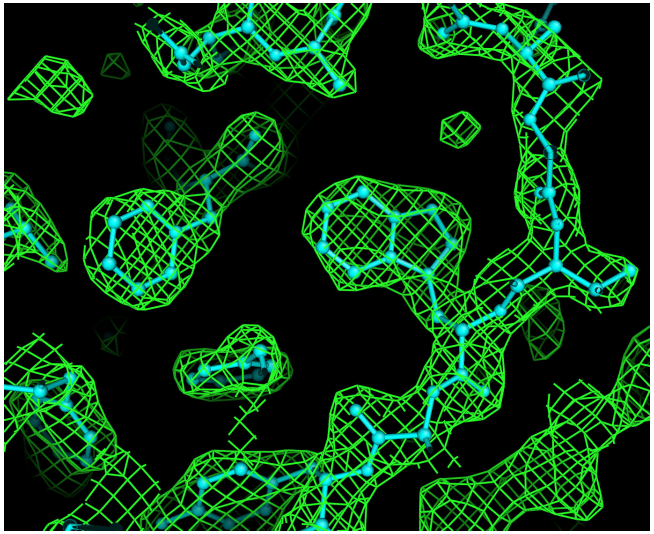


Синтез Фурье электронной плотности для белка альдоз-редуктазы. Разрешение 0.9Å. Фазы определены методом MAD.

MR - Molecular Replacement
Молекулярное замещение;

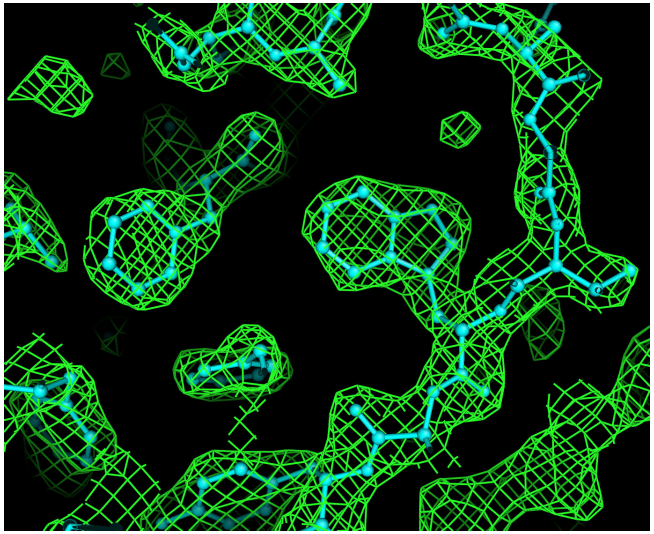
Комбинированные синтезы Фурье

Комбинированные синтезы Фурье

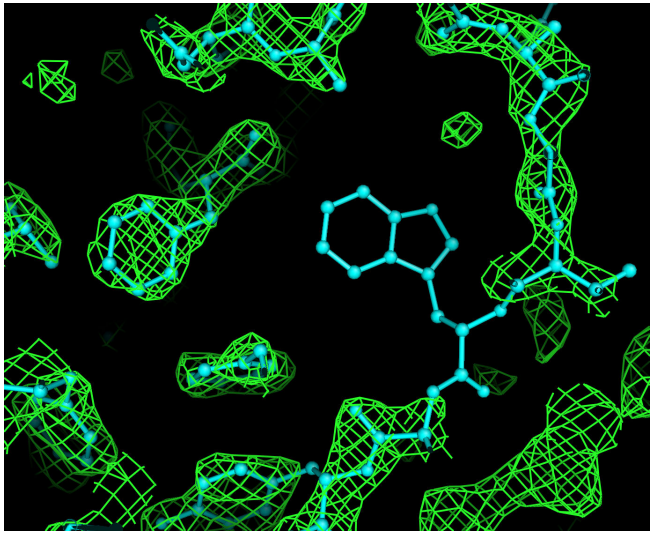


F_{full}, φ_{full}

Комбинированные синтезы Фурье

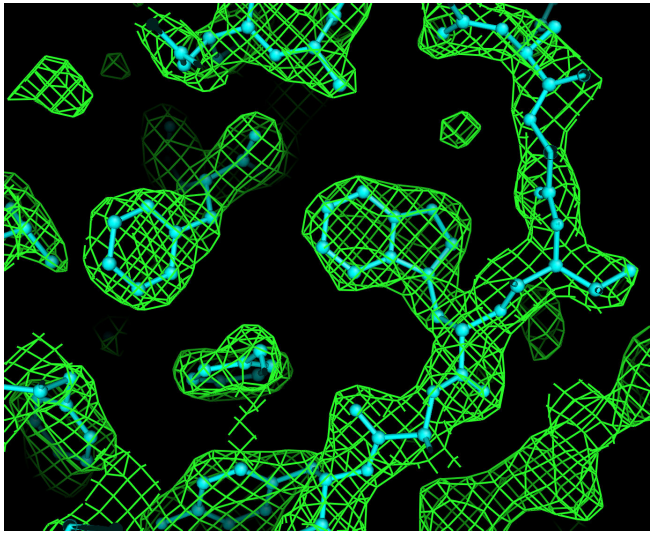


F_{full}, φ_{full}

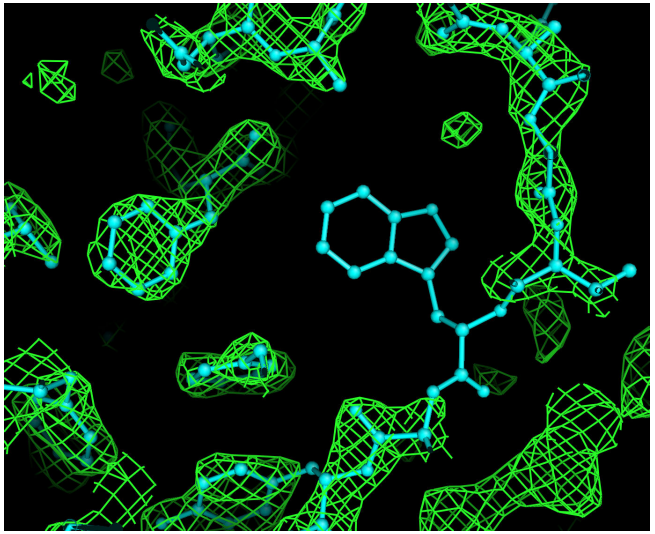


F_{part}, φ_{part}

Комбинированные синтезы Фурье



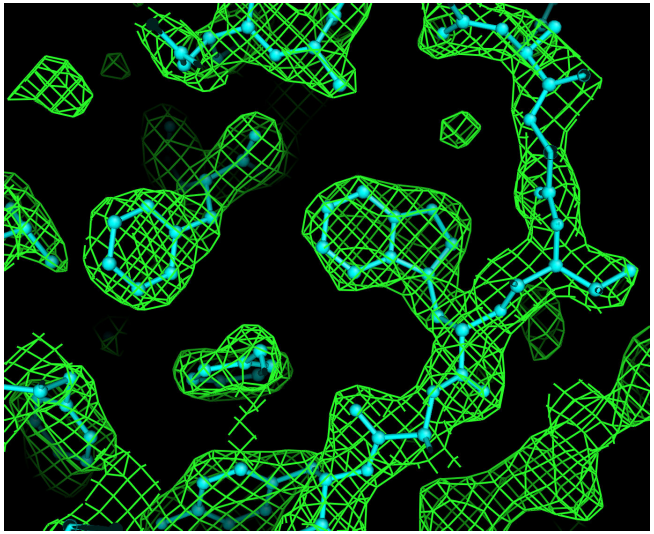
F_{full}, φ_{full}



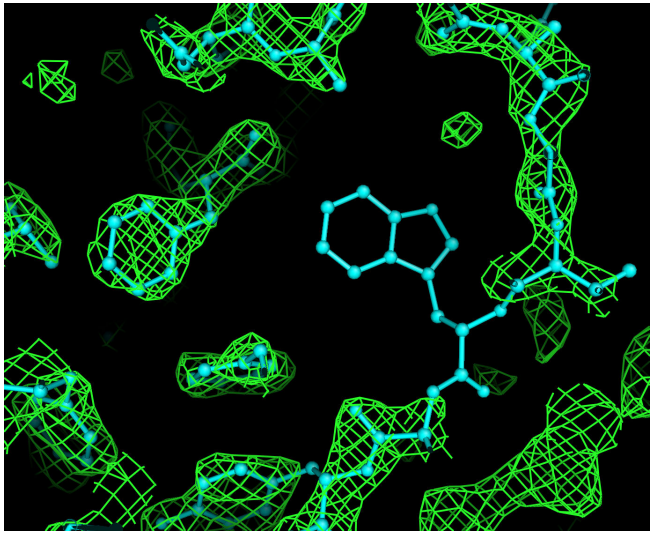
F_{part}, φ_{part}

F_{full}, φ_{part}

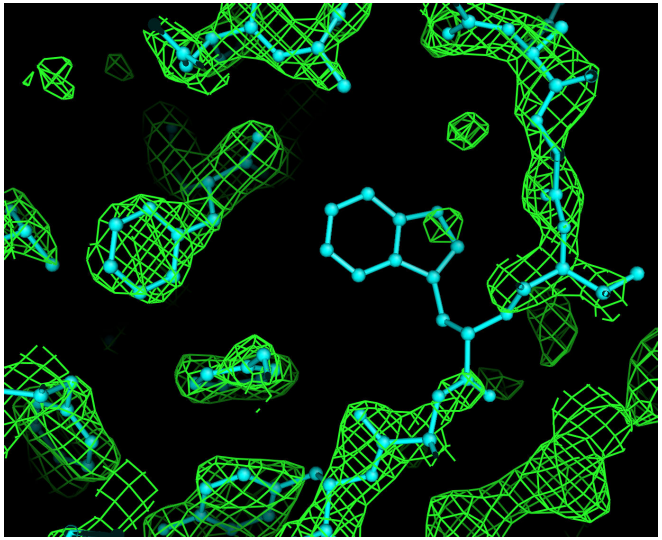
Комбинированные синтезы Фурье



F_{full}, φ_{full}

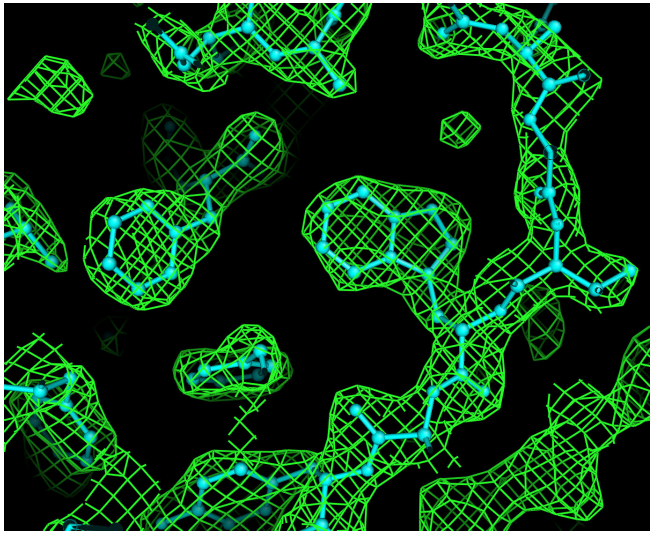


F_{part}, φ_{part}

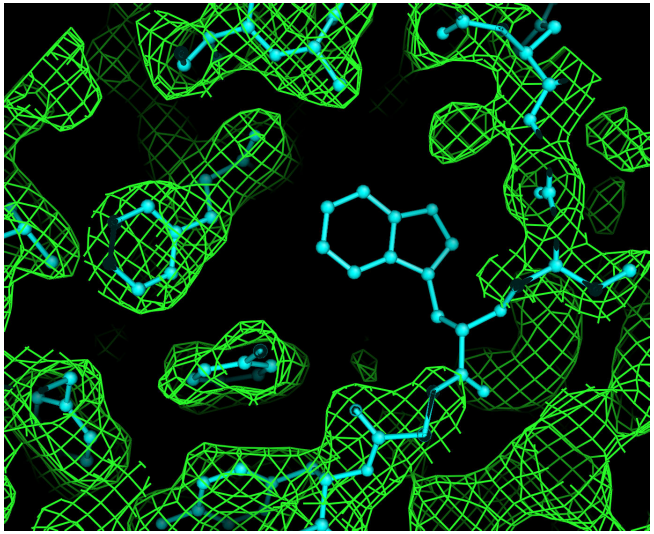


F_{full}, φ_{part}

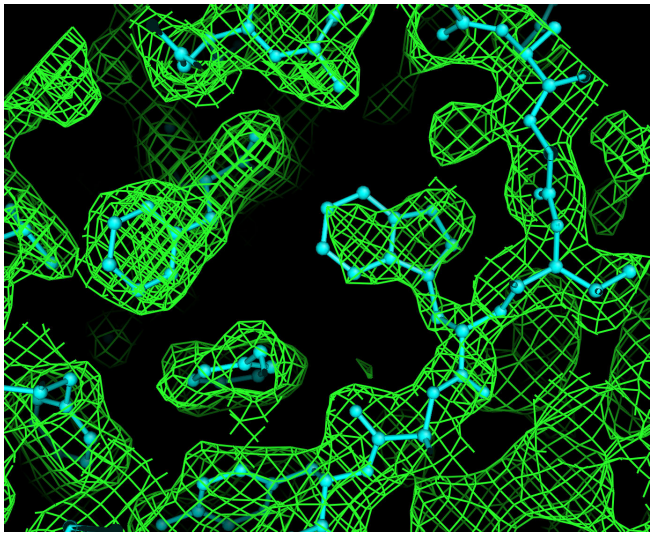
Комбинированные синтезы Фурье



F_{full}, φ_{full}
1.5 σ



F_{part}, φ_{part}
0.5 σ



F_{full}, $\varphi_{part,}$ 0.5 σ

Комбинированные синтезы Фурье

модель	структурные факторы
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M, \ldots \mathbf{r}_N$	$F_{hkl}^{full}, \varphi_{hkl}^{full}$
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$F_{hkl}^{part}, \varphi_{hkl}^{part}$

Синтез Фурье	$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$\mathbf{r}_M, \ldots \mathbf{r}_N$
$F_{hkl}^{full}, \varphi_{hkl}^{full}$	+	+
$F_{hkl}^{part}, \varphi_{hkl}^{part}$	+	-
$F_{hkl}^{full}, \varphi_{hkl}^{part}$	?	?

Комбинированные синтезы Фурье

модель	структурные факторы
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M, \ldots \mathbf{r}_N$	$F_{hkl}^{full}, \varphi_{hkl}^{full}$
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$F_{hkl}^{part}, \varphi_{hkl}^{part}$

Синтез Фурье	$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$\mathbf{r}_M, \ldots \mathbf{r}_N$
$F_{hkl}^{full}, \varphi_{hkl}^{full}$	+	+
$F_{hkl}^{part}, \varphi_{hkl}^{part}$	+	-
$F_{hkl}^{full}, \varphi_{hkl}^{part}$	+	1/2

Комбинированные синтезы Фурье

модель	структурные факторы
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M, \ldots \mathbf{r}_N$	$F_{hkl}^{full}, \varphi_{hkl}^{full}$
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M \quad \mathbf{u}_1, \ldots \mathbf{u}_L$	$F_{hkl}^{part}, \varphi_{hkl}^{part}$

Синтез Фурье	$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$\mathbf{r}_{M+1}, \ldots \mathbf{r}_N$	$\mathbf{u}_1, \ldots \mathbf{u}_L$
$F_{hkl}^{full}, \varphi_{hkl}^{full}$	+	+	-
$F_{hkl}^{part}, \varphi_{hkl}^{part}$	+	-	+
$F_{hkl}^{full}, \varphi_{hkl}^{part}$	+	1/2	1/2

Комбинированные синтезы Фурье

модель	структурные факторы
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M, \ldots \mathbf{r}_N$	$F_{hkl}^{full}, \varphi_{hkl}^{full}$
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M \quad \mathbf{u}_1, \ldots \mathbf{u}_L$	$F_{hkl}^{part}, \varphi_{hkl}^{part}$

	Синтез Фурье	$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$\mathbf{r}_M, \ldots \mathbf{r}_N$	$\mathbf{u}_1, \ldots \mathbf{u}_L$
	$F_{hkl}^{full}, \varphi_{hkl}^{full}$	1	1	0
-1	$F_{hkl}^{part}, \varphi_{hkl}^{part}$	1	0	1
2	$F_{hkl}^{full}, \varphi_{hkl}^{part}$	1	1/2	1/2
	$2F_{hkl}^{full} - F_{hkl}^{part}, \varphi_{hkl}^{part}$	1	1	0

Комбинированные синтезы Фурье

модель	структурные факторы
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M, \ldots \mathbf{r}_N$	$F_{hkl}^{full}, \varphi_{hkl}^{full}$
$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M \quad \mathbf{u}_1, \ldots \mathbf{u}_L$	$F_{hkl}^{part}, \varphi_{hkl}^{part}$

	Синтез Фурье	$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$	$\mathbf{r}_M, \ldots \mathbf{r}_N$	$\mathbf{u}_1, \ldots \mathbf{u}_L$
	$F_{hkl}^{full}, \varphi_{hkl}^{full}$	1	1	0
-1	$F_{hkl}^{part}, \varphi_{hkl}^{part}$	1	0	1
2	$F_{hkl}^{full}, \varphi_{hkl}^{part}$	1	1/2	1/2
	$\boxed{\left(2F^{obs} - F^{calc}, \varphi^{calc}\right)}$	1	1	0

MR - Molecular Replacement
Молекулярное замещение;

План действий:

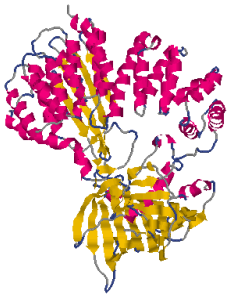
А. Найти в банке белок с известными координатами (пробную модель), чья структура ожидается быть похожей на структуру исследуемого белка.

Б. Разместить пробную модель в элементарной ячейке так, чтобы получить максимальное соответствие этой модели исследуемой структуре.

В. Рассчитать по оптимально размещенной пробной модели значения фаз структурных факторов.

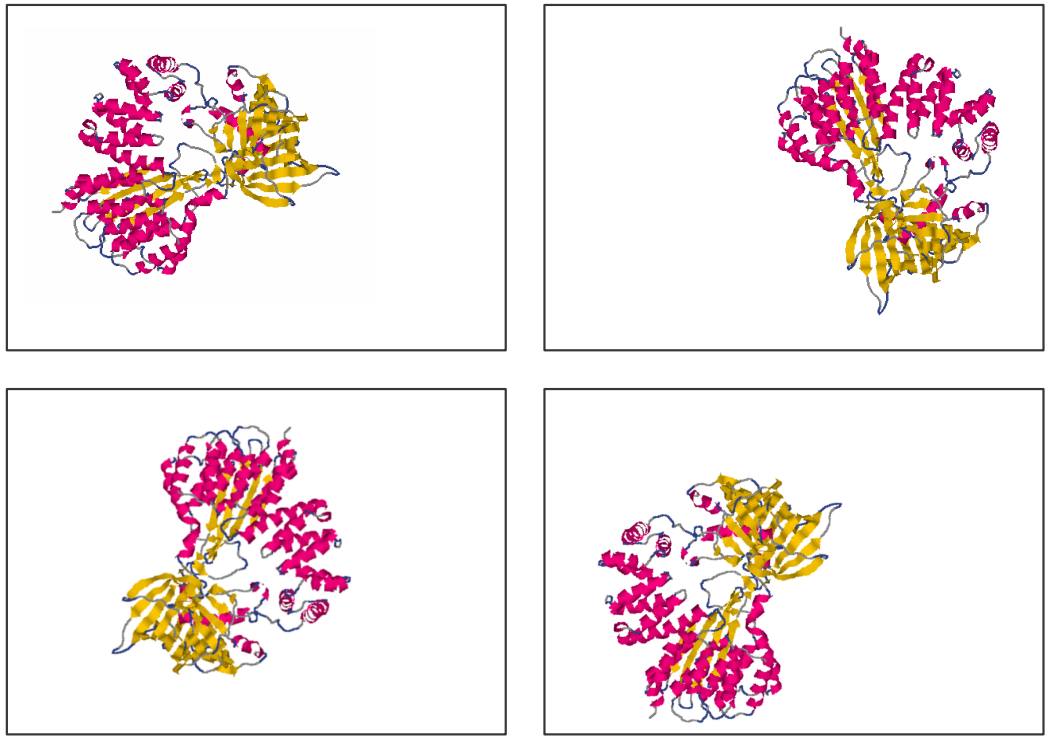
Г. Использовать комбинированные синтезы Фурье для корректировки модели.

MR - Molecular Replacement
Молекулярное замещение;



пробная модель

?



В общем случае положение твердого тела определяется шестью параметрами $(\alpha, \beta, \gamma, t_x, t_y, t_z)$.

углы вращения

вектор трансляции

MR - Molecular Replacement
Молекулярное замещение;

Оптимальное размещение пробной модели

Для каждого допустимого набора параметров (α, β, γ, t_x, t_y, t_z) можно рассчитать соответствующие значения модулей структурных факторов $F_{hkl}^{calc}(\alpha, \beta, \gamma, t_x, t_y, t_z)$

Хотим иметь наилучшее совпадение с экспериментально определенными модулями F_{hkl}^{obs}

$$\sum_{hkl}\left(F_{hkl}^{calc}(\alpha, \beta, \gamma, t_x, t_y, t_z) - F_{hkl}^{obs}\right)^2 \quad \Rightarrow \quad \min$$

Задачу можно разделить на два этапа: поиск углов вращения ("функция вращения") и поиск трансляции ("функция трансляции").

MR - Molecular Replacement
Молекулярное замещение;

С нахождением оптимальной ориентации и положения пробной модели работа не кончается!

Надо откорректировать модель.

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 гомологии между пробной моделью и исследуемой структурой.

MR - Molecular Replacement
Молекулярное замещение;

Где взять гомологичную пробную модель?

- среди структур, определенных РСА;
- среди структур, определенных ЯМР;
- электронная микроскопия;
- теоретическое предсказание модели.

"Прямые" методы

Основаны на информации общего типа:

- атомность;
- неотрицательность функции распределения электронной плотности;
- разделенность атомов в пространстве;
- связность областей высокой электронной плотности;
- гистограммы синтезов Фурье.

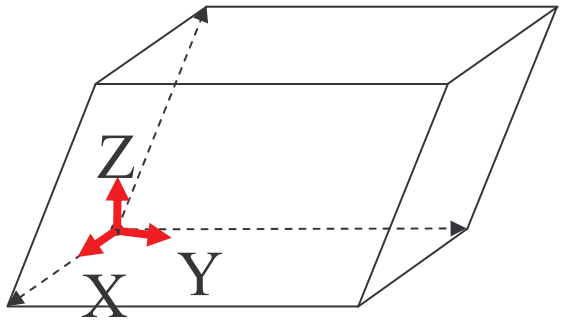
Низкомолекулярные структуры -
рутинное использование;

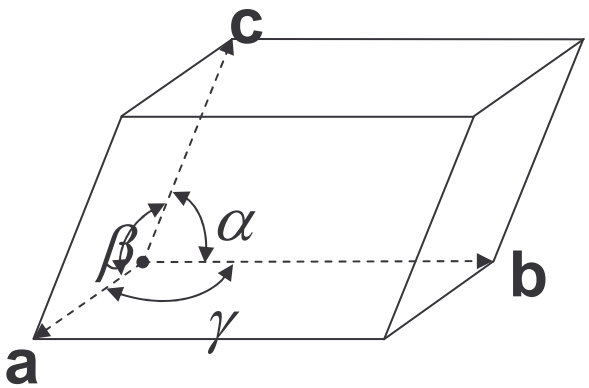
Небольшие белки, высокое разрешение -
ряд успешных попыток;

Большие комплексы, среднее разрешение -
в процессе разработки, отдельные успешные попытки.

Модель структуры

```
ATOM      30  N   SER    2      13.117   9.840  39.210 1.000 12.49
```





абсолютные координаты
(X,Y,Z)
Ортогональная система координат,
единица измерения Å.

относительные координаты
(x,y,z):
$\mathbf{r}=x\mathbf{a}+y\mathbf{b}+z\mathbf{c}$

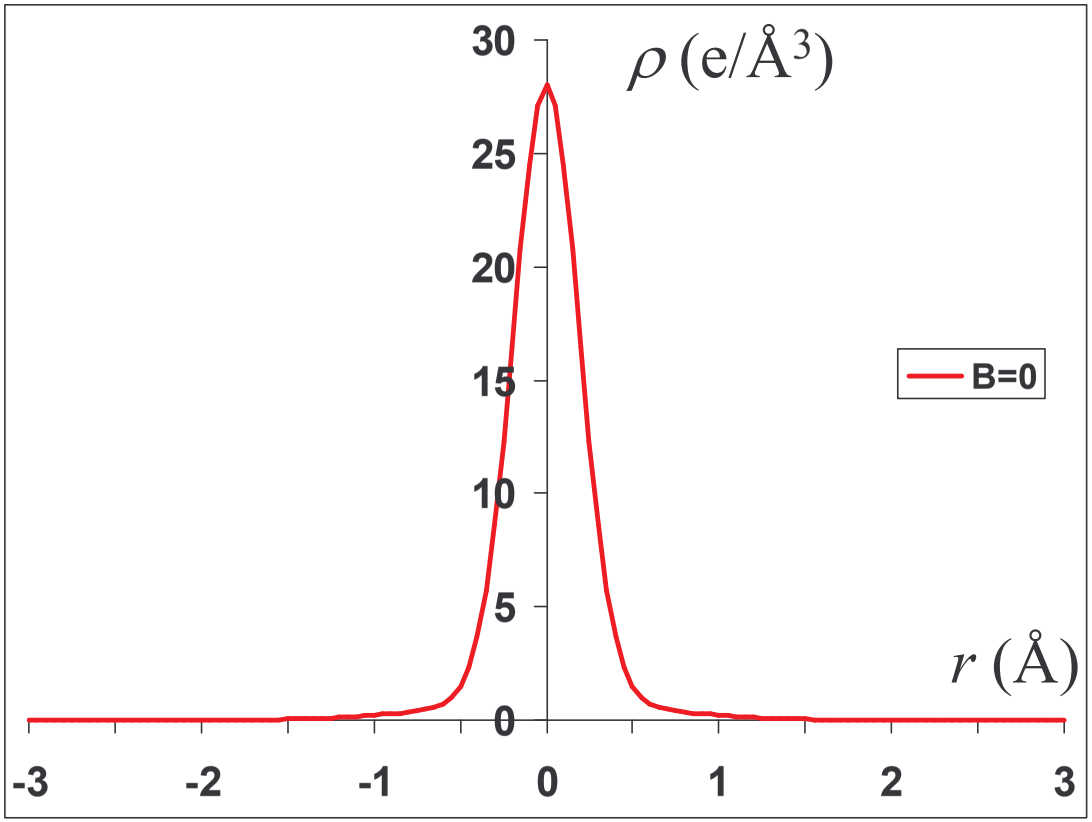
```
CRYST1    49.430    66.790    47.400   90.00   92.40   90.00
SCALE1     0.020231  0.000000  0.000848          0.000000
SCALE2     0.000000  0.014972  0.000000          0.000000
SCALE3     0.000000  0.000000  0.021116          0.000000
```

U

S

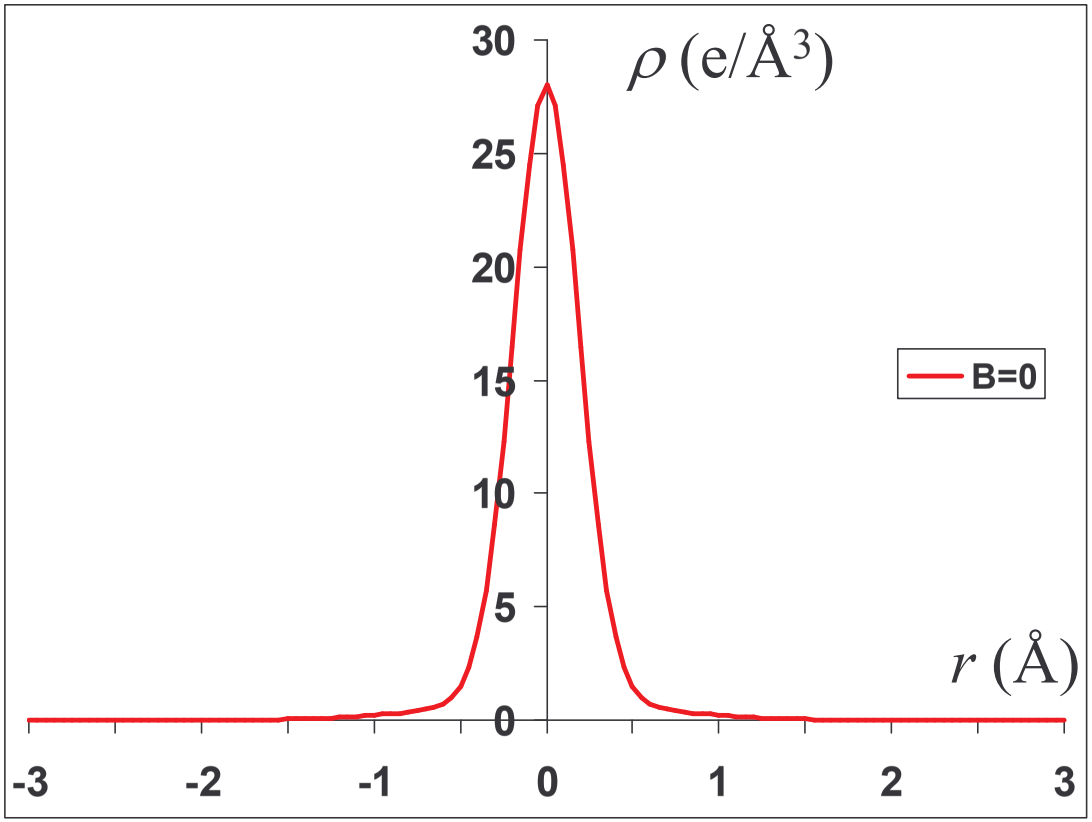
```
xfrac = S11X + S12Y + S13Z + U1
yfrac = S21X + S22Y + S23Z + U2
zfrac = S31X + S32Y + S33Z + U3
```

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



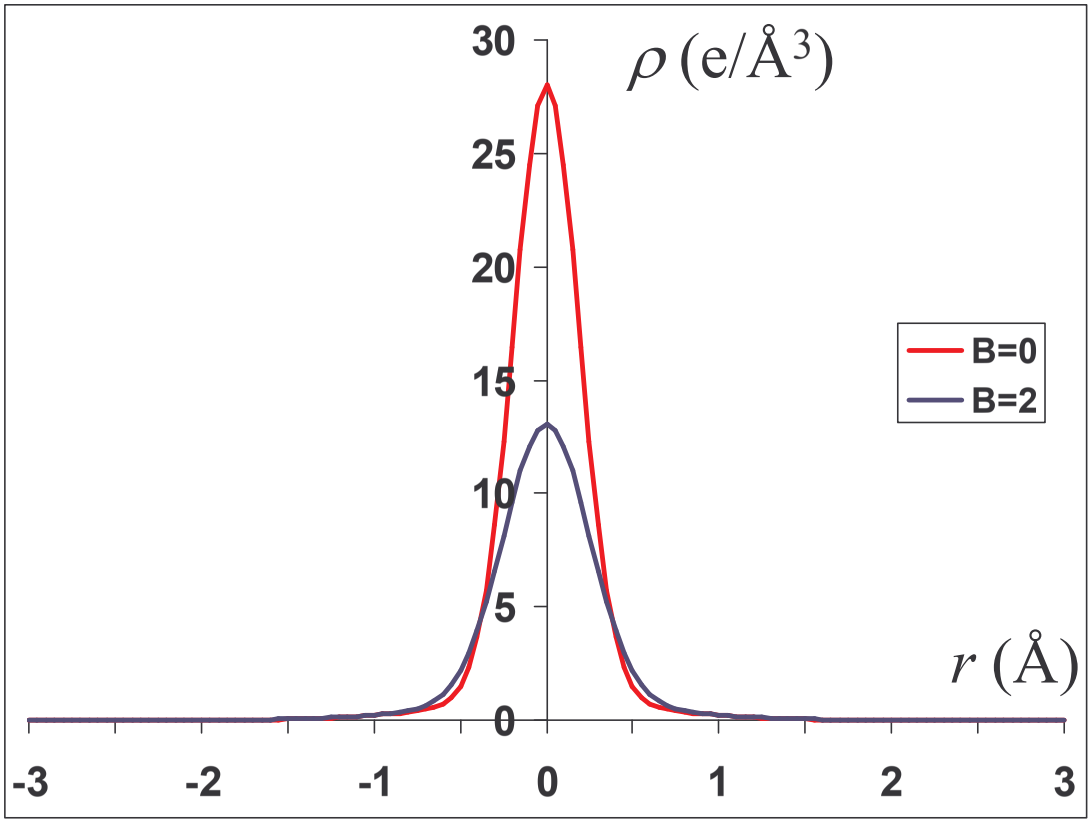
$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j} \right)$$

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



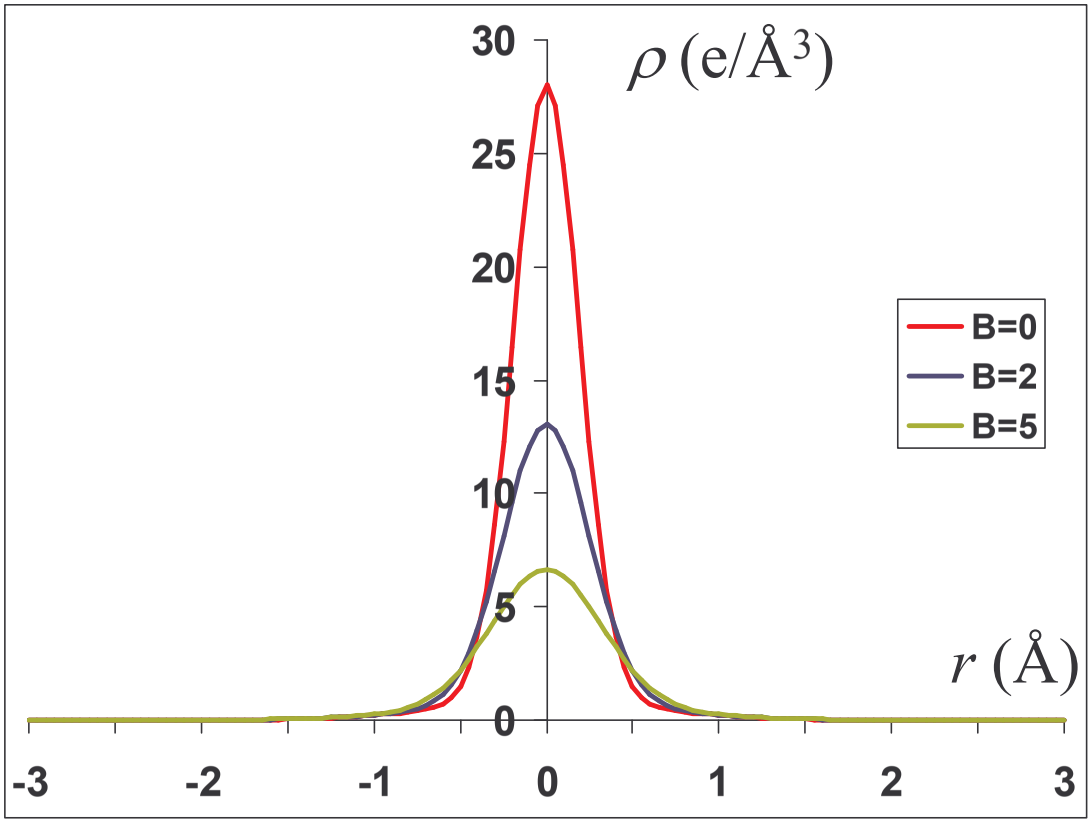
$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j + \boxed{B}} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + \boxed{B}} \right)$$

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



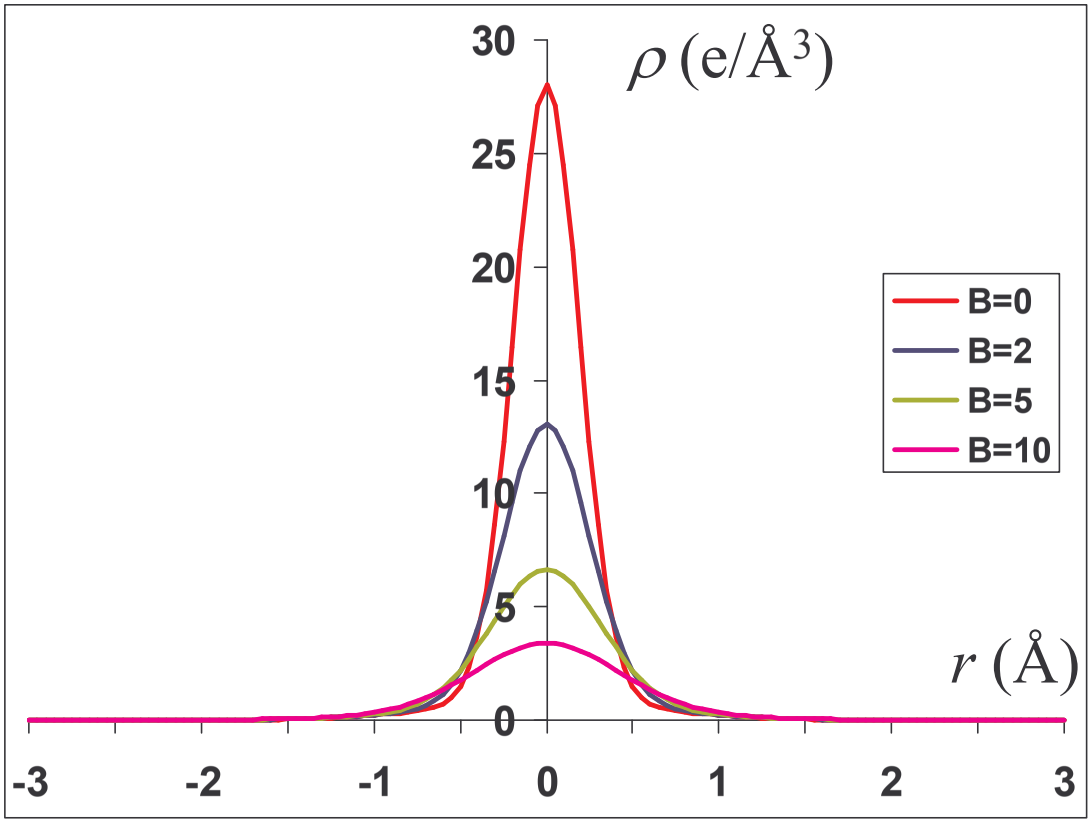
$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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



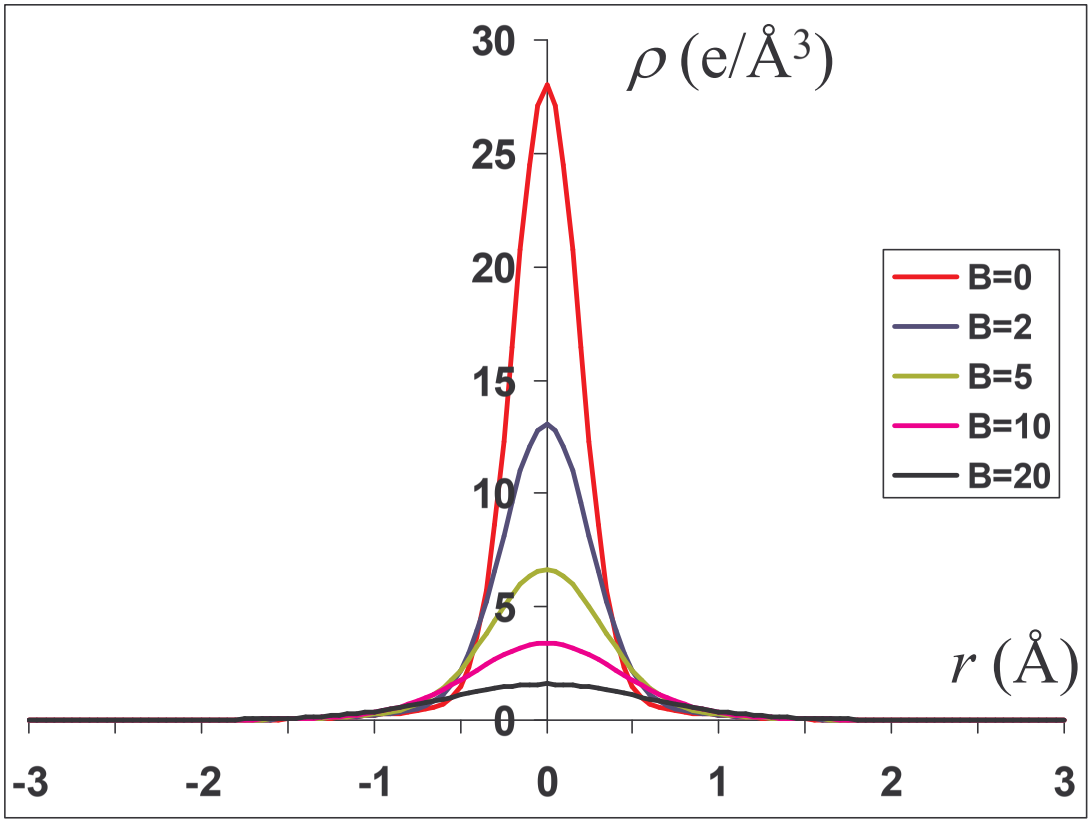
$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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



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

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

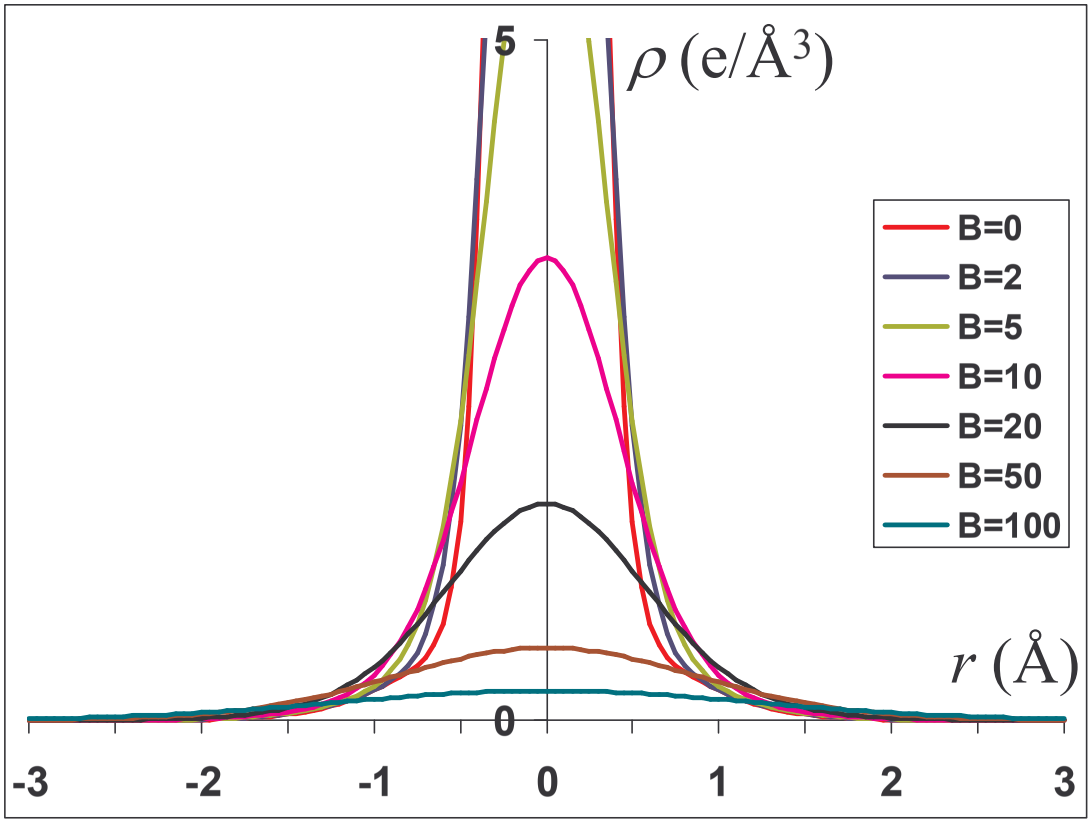


$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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

$$B = 8\pi^2 \overline{u^2}$$

u - смещение атома



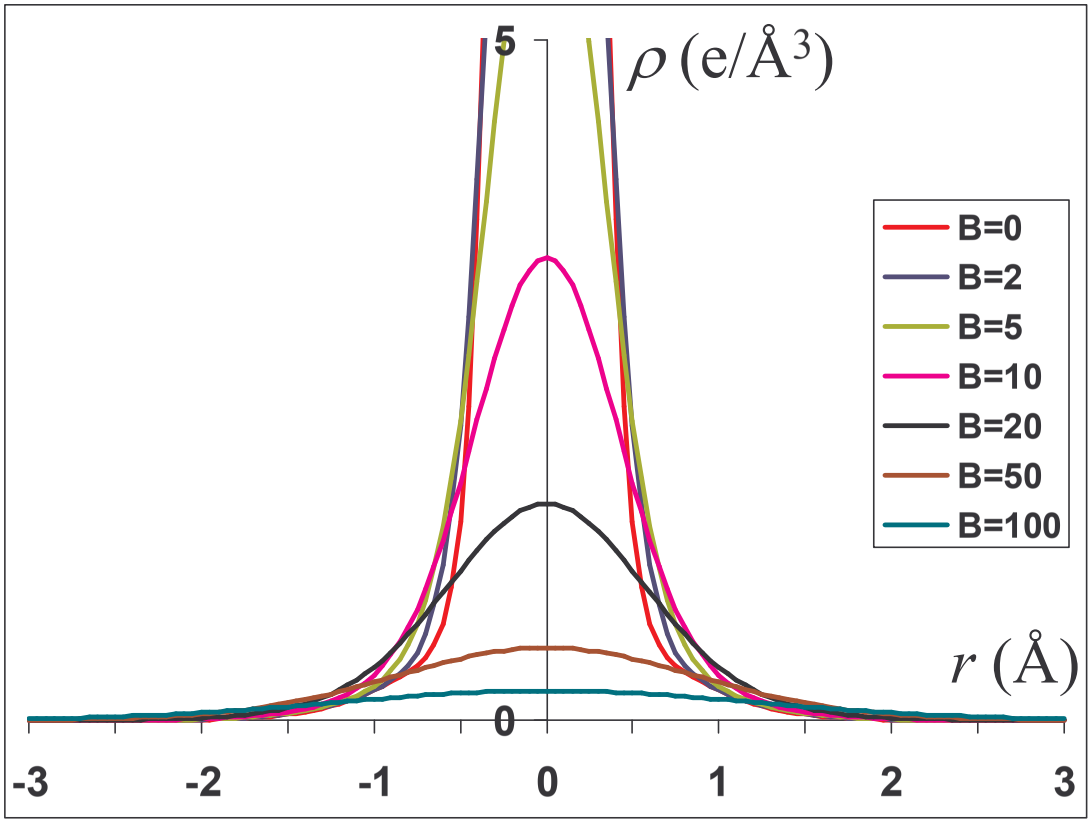
$$\rho(\mathbf{r}) = \sum_{j=1}^{5} C_j \left(\frac{4\pi}{B_j + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r}|^2}{B_j + B} \right)$$

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

$$B = 8\pi^2 \overline{u^2}$$

u - смещение атома

B	$\sqrt{\overline{u^2}}$
2	0.16
5	0.25
10	0.36
20	0.50
50	0.80
100	1.13



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

Изотропный температурный фактор

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

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

```
ATOM      30  N   SER    2      13.117   9.840  39.210 1.000 12.49
```

Анизотропный температурный фактор

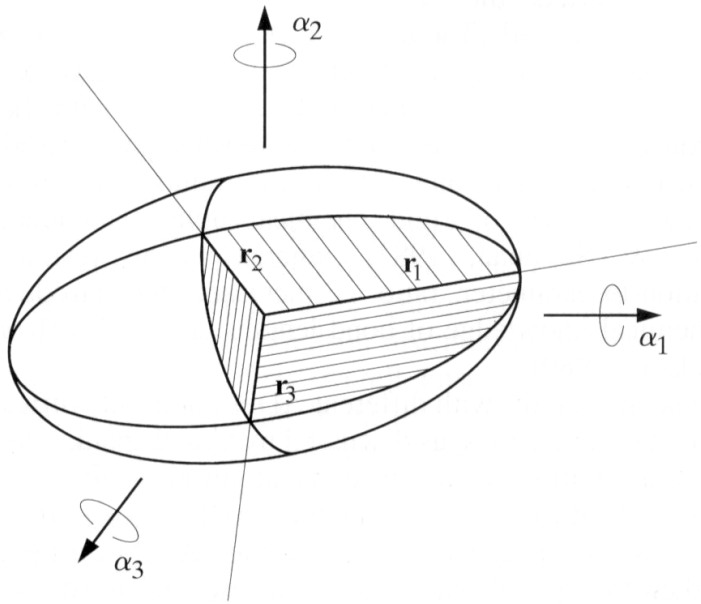


Fig. 18.4.1.1. The thermal-ellipsoid model used to represent anisotropic atomic displacement, with major axes indicated. The ellipsoid is drawn with a specified probability of finding an atom inside its contour. Six parameters are necessary to describe the ellipsoid: three represent the dimensions of 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    1510   2105   1130    447   -393  -1019
```

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.

```
ATOM   6626  O   HOH  4951      21.882  15.425  26.797 0.500  8.74
```

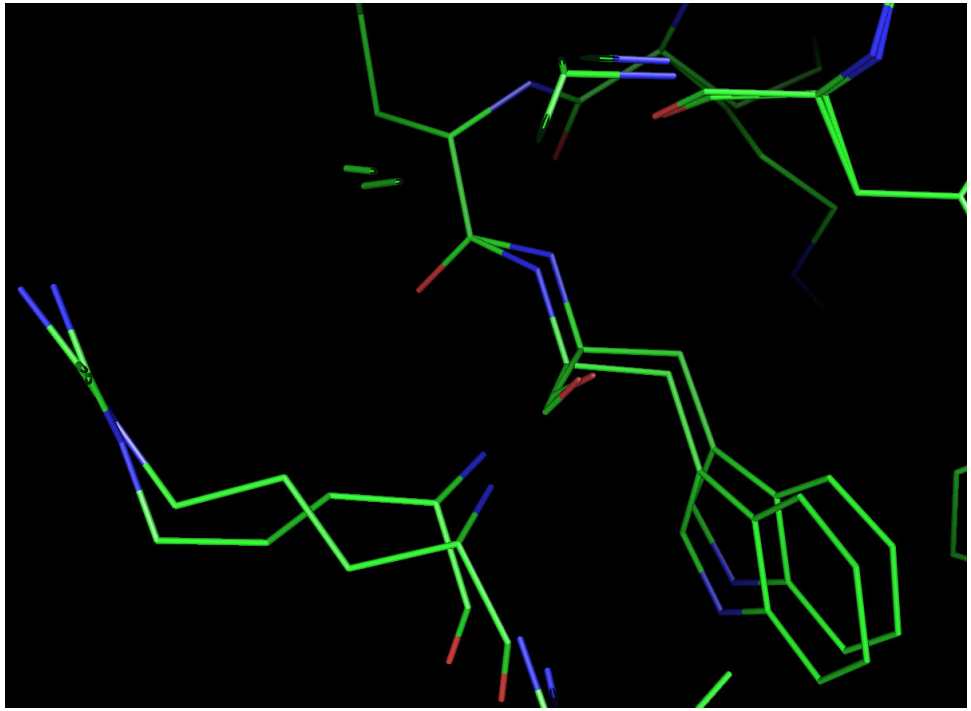
X *Y* *Z* *T* *B*

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

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

Коэффициент заполнения / заселенность / occupancy

Альтернативные конформации



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

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}^{calc} - экспериментально определенные модули структурных факторов

Хотим иметь

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

Получили задачу на минимизацию (нелинейный "метод наименьших квадратов" / least squares / LSQ).

Стандартный фактор достоверности / R-factor

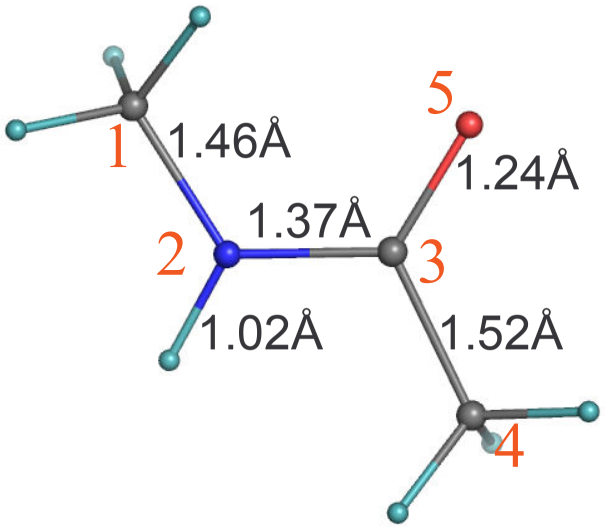
$$R = \frac{\sum_{hkl} \left| F_{hkl}^{calc} - F_{hkl}^{obs} \right|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Типичная задача:	
endonuclease *Sm*	
Число неводородных атомов	3 694
Число независимых параметров	36 940
Число независимых ограничений (1.1Å set)	108 000

Проблемы:

- много локальных максимумов; возможно только локальное уточнение; радиус сходимости ~ 0.7Å;
- модель "рассыпается".

Стереохимические ограничения



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

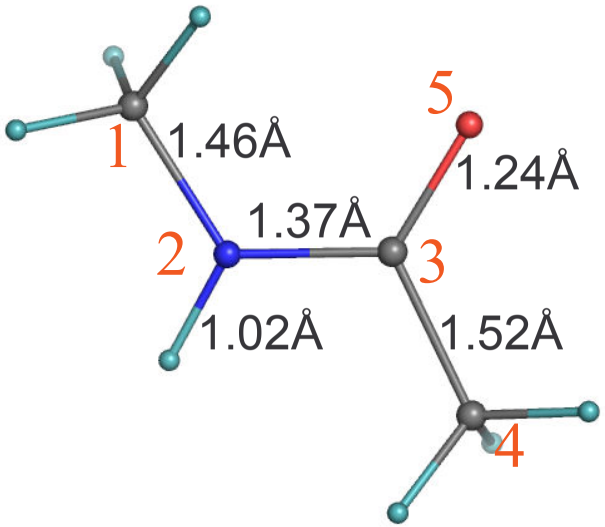
d_{ij}^{calc} - расстояние между i-ым и j-ым атомами в модели

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

Хотим иметь

$$\left(d_{12}^{calc} - 1.46\right)^2 + \left(d_{23}^{calc} - 1.37\right)^2 + \left(d_{34}^{calc} - 1.52\right)^2 + \left(d_{35}^{calc} - 1.24\right)^2 \Rightarrow \min$$

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

$$\left(d_{12}^{calc} - 1.46\right)^2 + \left(d_{23}^{calc} - 1.37\right)^2 + \left(d_{34}^{calc} - 1.52\right)^2 + \left(d_{35}^{calc} - 1.24\right)^2 \Rightarrow \min$$

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

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

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

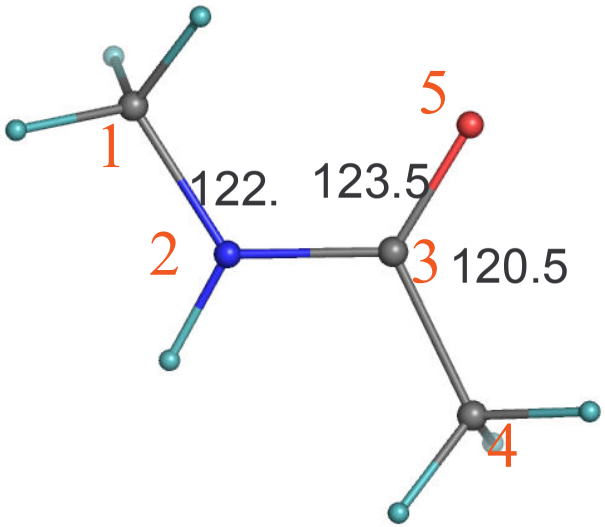
Составной критерий

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

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

☹ 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

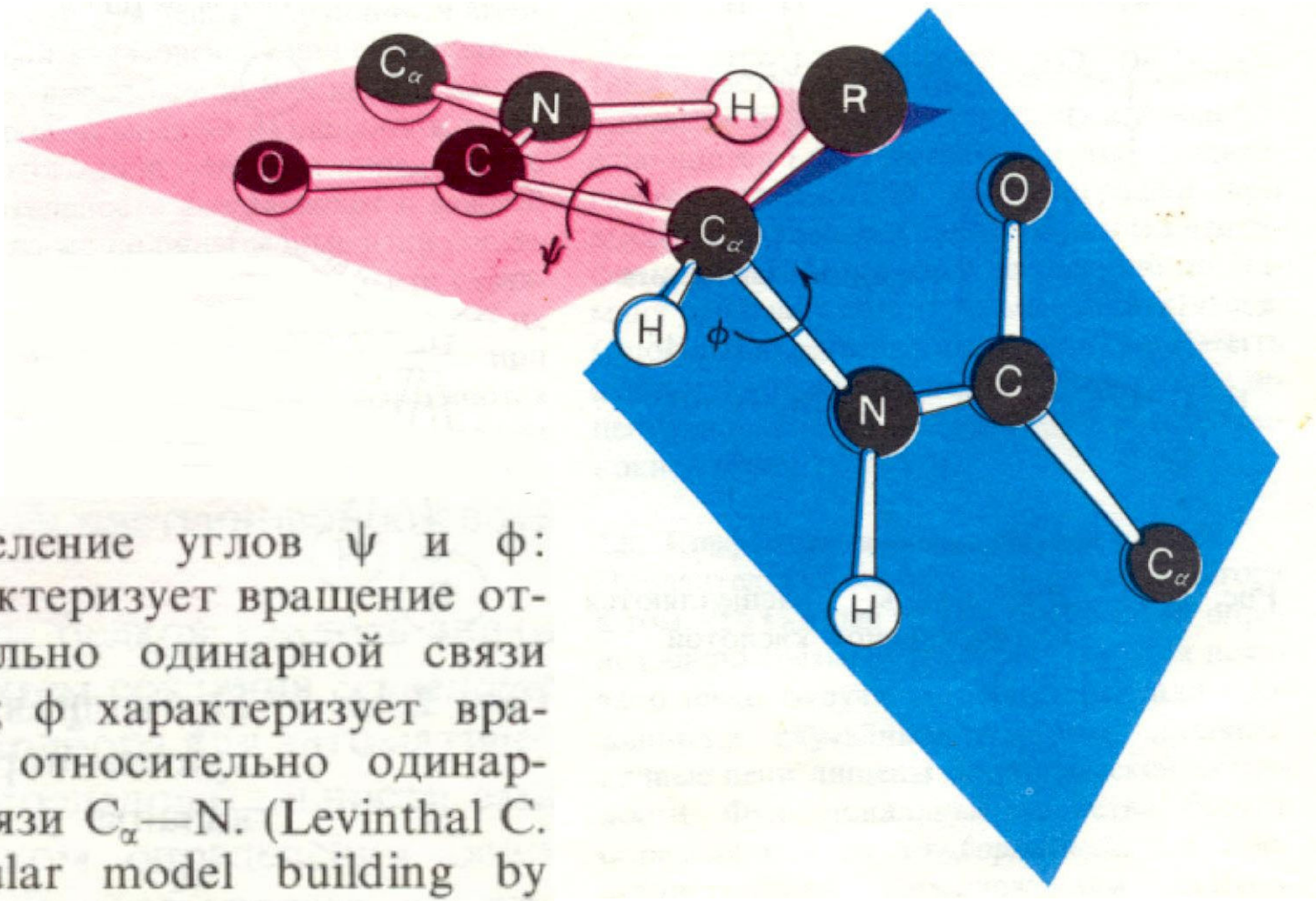
Хотим иметь

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

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

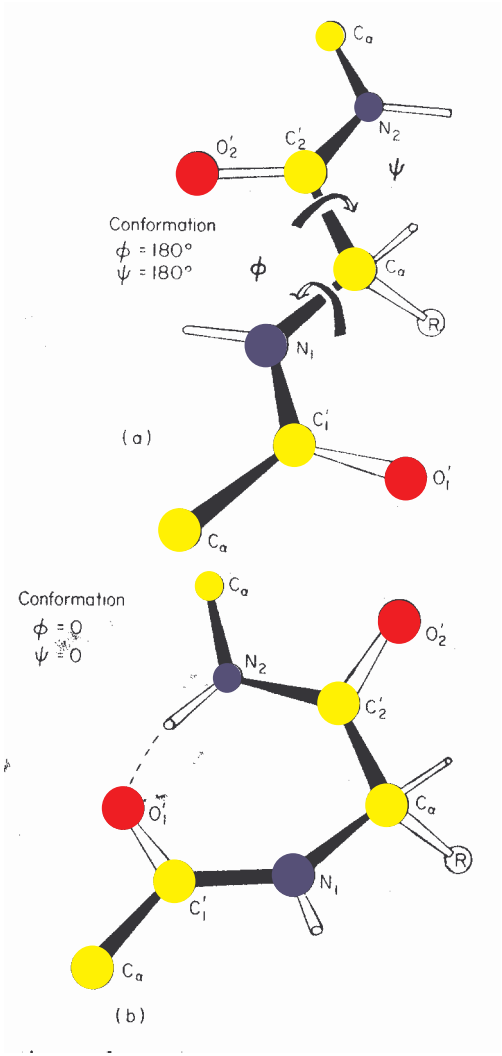
$$R = w_{X-ray} R_{X-ray} + w_{dist} R_{dist} + w_{angle} R_{angle}$$

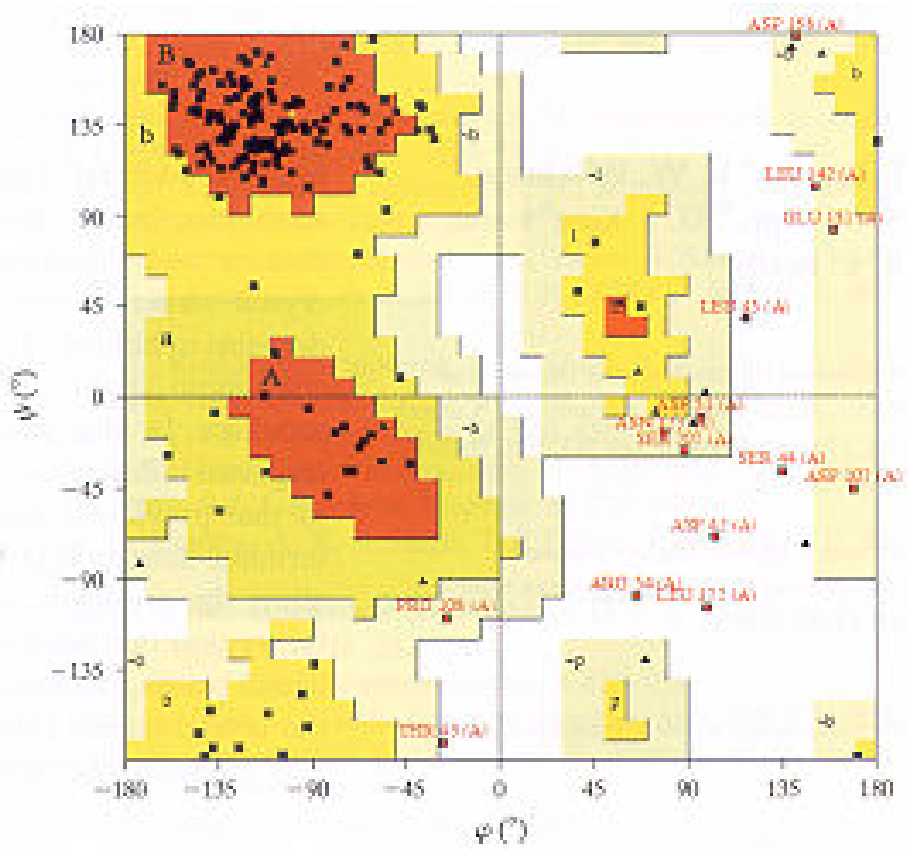
Двугранные углы



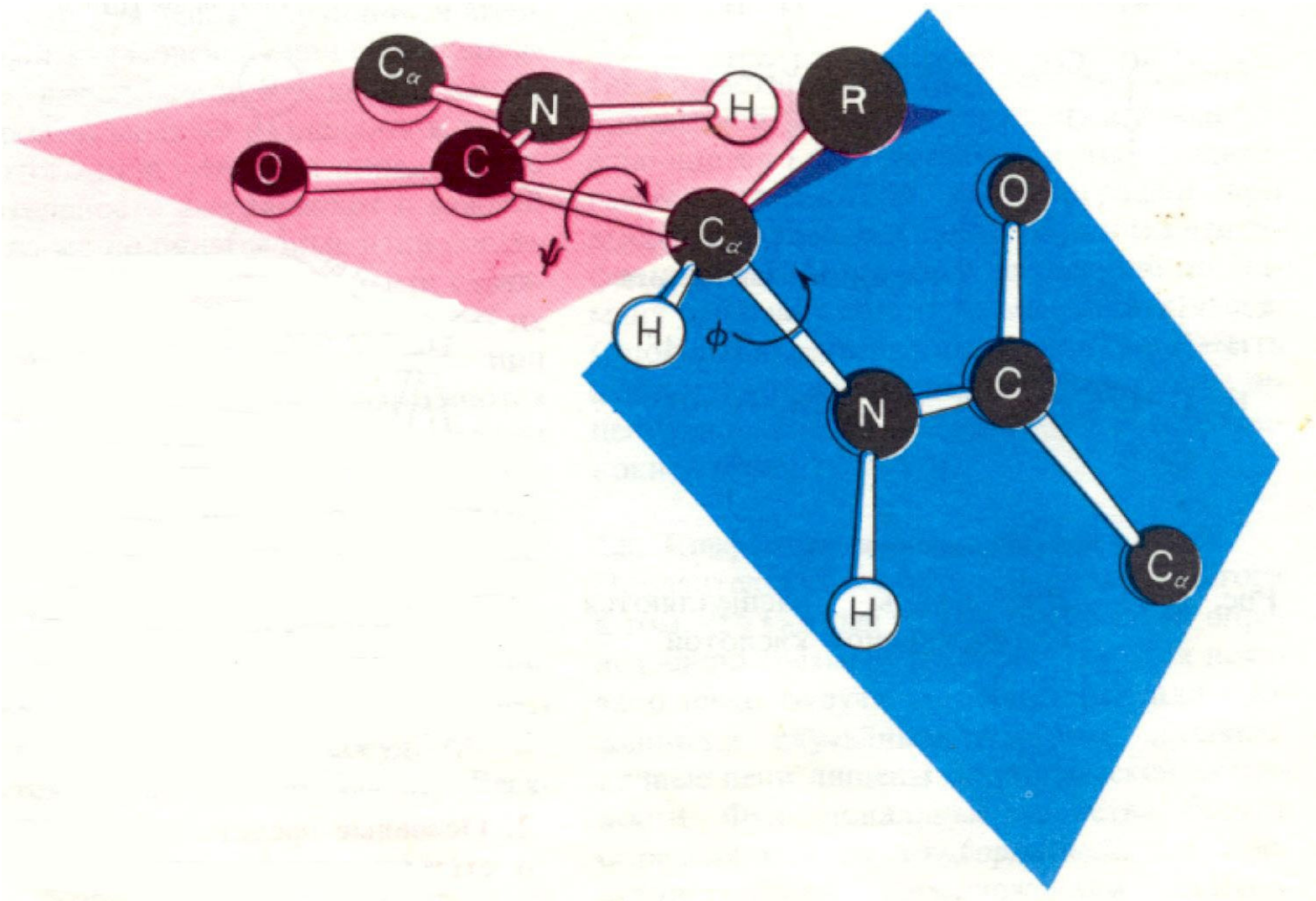
Определение углов ψ и ϕ: ψ характеризует вращение относительно одинарной связи $C_\alpha - C$; ϕ характеризует вращение относительно одинарной связи C_α—N. (Levinthal C. Molecular model building by computer, Scientific American, Inc., 1966.).

Двугранные углы

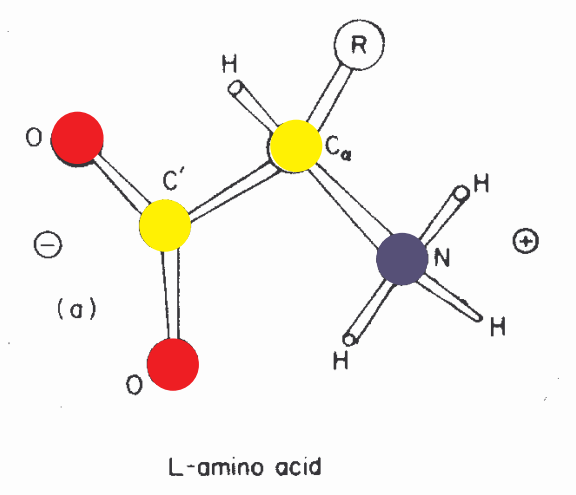




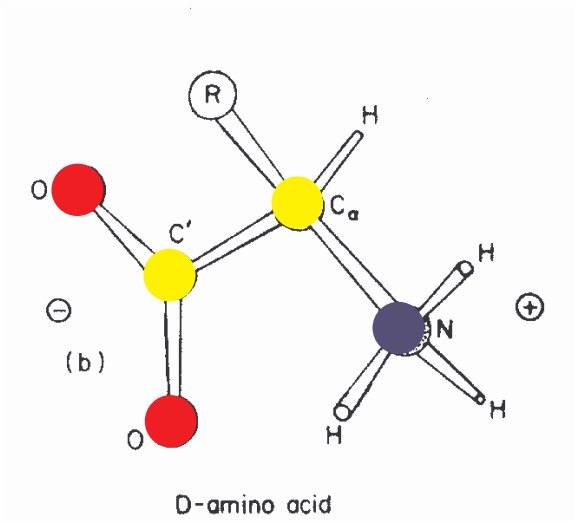
Плоские группы



Хиральность

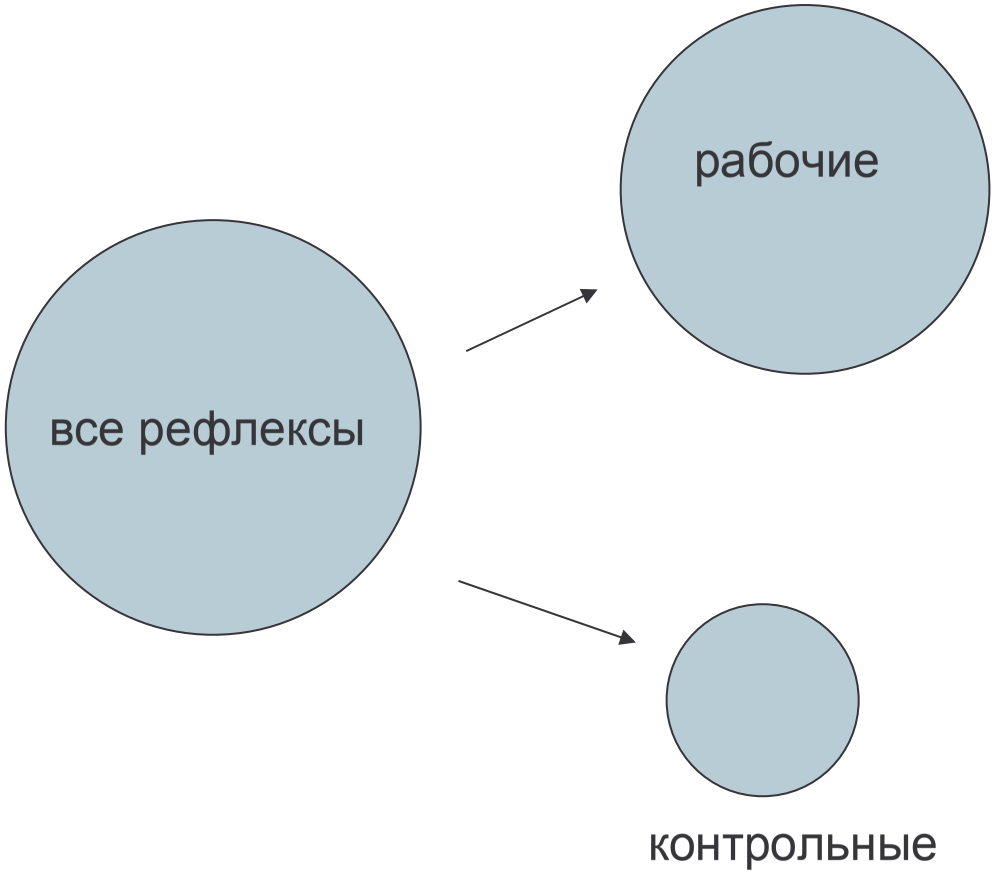


L-amino acid



D-amino acid

R-free



уточнение модели

только контроль

R-free

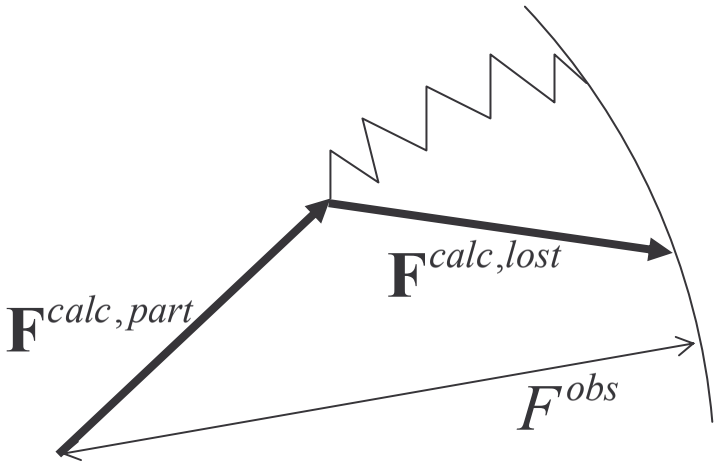
```
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   CROSS-VALIDATION METHOD          : THROUGHOUT
REMARK   3   FREE R VALUE TEST SET SELECTION  : RANDOM
REMARK   3   R VALUE     (WORKING + TEST SET) : 0.15621
REMARK   3   R VALUE            (WORKING SET) : 0.15185
REMARK   3   FREE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

Обычное уточнение

Частичная модель "Потерянные" атомы

$$\left\{\mathbf{r}_j\right\}_{j=1}^{M} \qquad \left\{\mathbf{u}_k\right\}_{k=M+1}^{N}$$

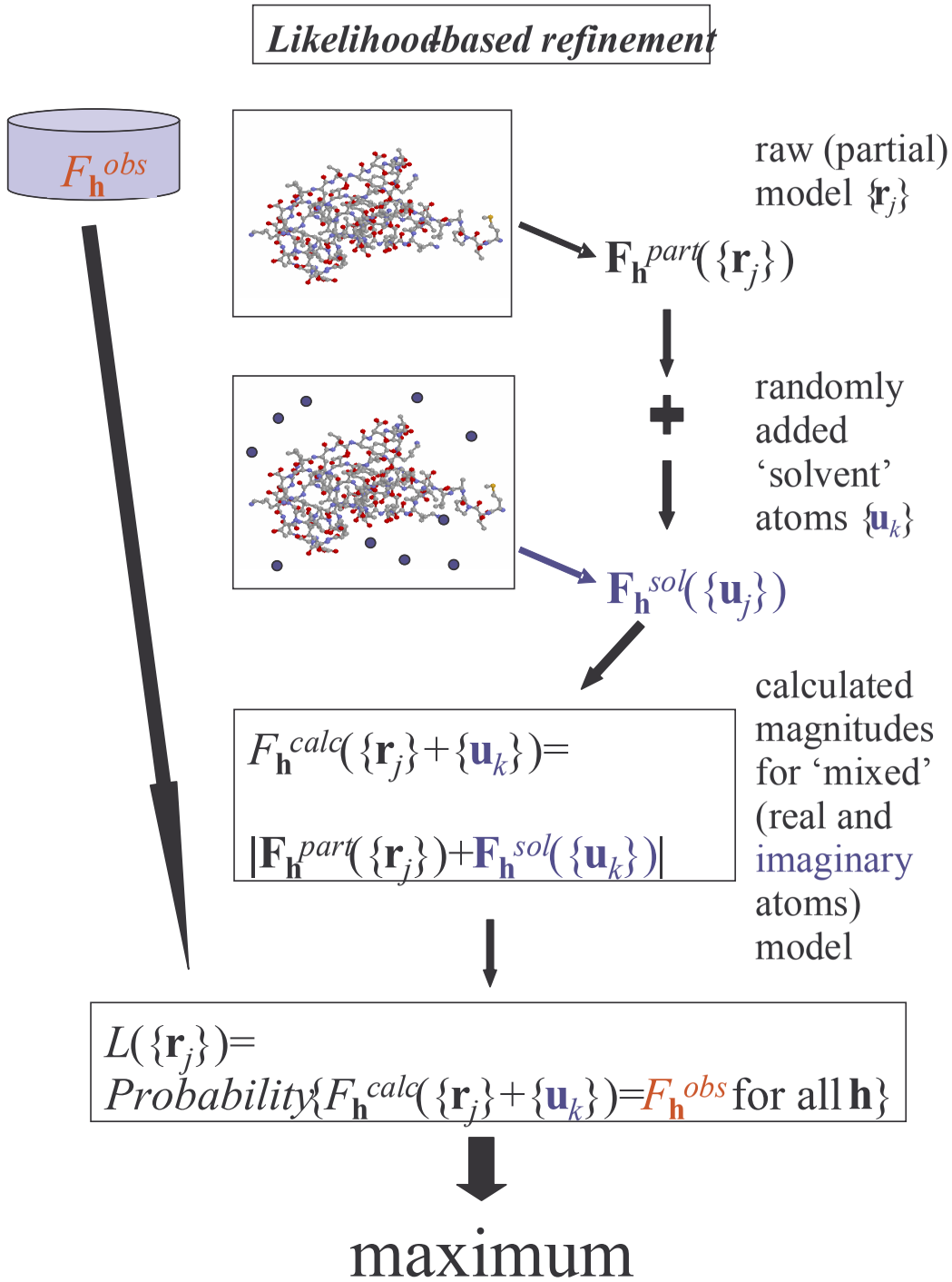
$$F_{hkl}^{full,true} = \left|\mathbf{F}_{hkl}^{calc,part}\left(\left\{\mathbf{r}_j^{true}\right\}\right) + \mathbf{F}_{hkl}^{calc,lost}\left(\left\{\mathbf{u}_k^{true}\right\}\right)\right| \approx F_{hkl}^{obs}$$



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

Максимизация правдоподобия

Likelihood-based refinement



```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : REFMAC
REMARK   3   AUTHORS     : MURSHUDOV,VAGIN,DODSON
REMARK   3
REMARK   3    REFINEMENT TARGET : MAXIMUM LIKELIHOOD
```

T.Terwilliger

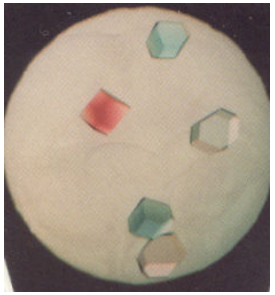
Pyrobaculum aerophilum
A pilot project in structural genomics
A test case to find the bottlenecks: summary of results from all laboratories in the project

Step	*Number of genes*
Cloned	*274*
Expressed in E. coli	*168*
Purified	*43*
Crystallized	*24*
X-ray structures	*10* ☺ ? ☹
NMR structures	*1*

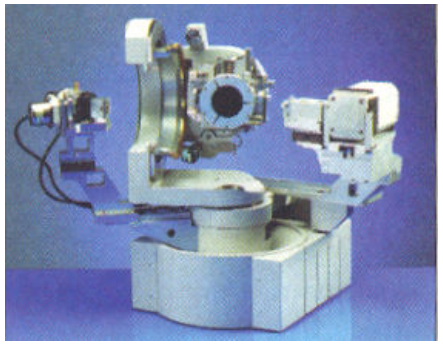
Next: how to increase the success rate?

X-ray structure analysis

X-ray source

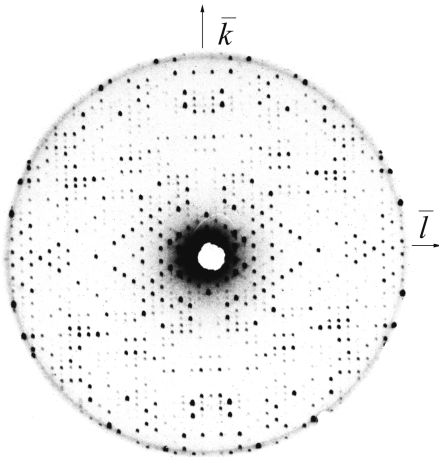


crystal

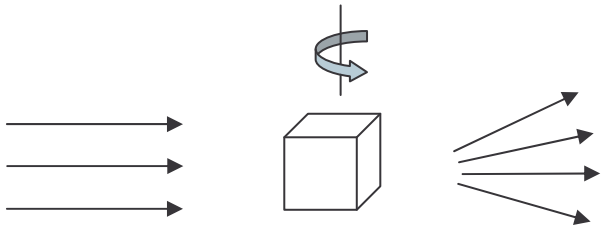


registration

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

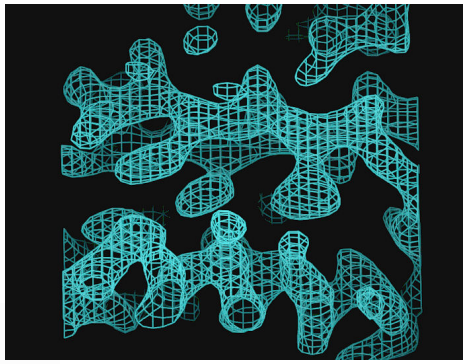
The solving of the structure

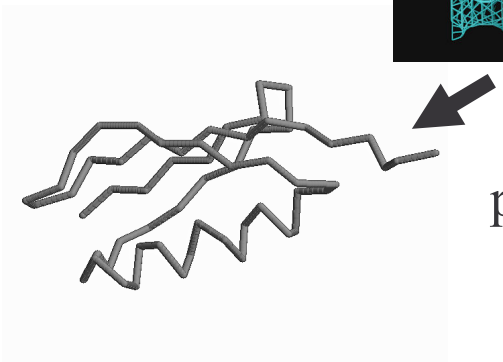
X-ray structure analysis

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

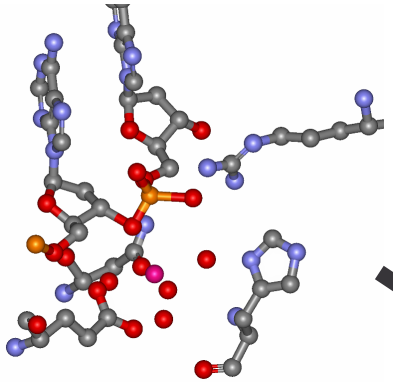
$\Omega_\kappa = \{\mathbf{r}: \rho_S(\mathbf{r}) > \kappa\}$





preliminary model

refined 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

1. Дж.Гласкер, К.Трублад. Анализ кристаллической структуры. "Мир", Москва, 1974
2. М.А.Порай-Кошиц. Основы структурного анализа химических соединений. Москва, "Высшая школа", 1989
3. Т.Бландел, Л.Джонсон. Кристаллография белка. "Мир", Москва, 1979
4. Ч.Кантор, П.Шиммел. Биофизическая химия, том 2. Москва, "Мир", 1984
5. International Tables for Crystallography, vol. F. Crystallography of biological macromolecules. by Kluwer Academic Publishers, Dordrecht/Boston/London, 2001