

**Определение формы белковых  
макромолекул в разбавленных системах  
по данным малоуглового рентгеновского  
и нейтронного рассеяния".**

**В.В. Волков** [vvo@ns.crys.ras.ru](mailto:vvo@ns.crys.ras.ru)  
[volkicras@mail.ru](mailto:volkicras@mail.ru)

**Институт кристаллографии имени А.В. Шубникова  
ФНИЦ «Кристаллография и фотоника» РАН,  
Москва**

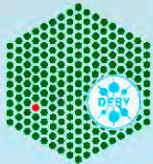
# Acknowledgements to the participants:

**BioSAXS group  
(EMBL-Hamburg)  
D.I. Svergun  
M.W. Roessle  
A.G. Kikhney  
E. Mylonas  
D. Franke  
and many others**

**Институт кристаллографии РАН**



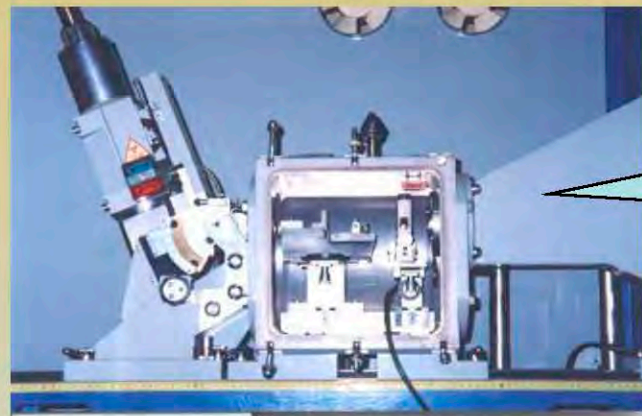
**П.В. Конарев  
С.В. Амарантов  
М.В. Петухов  
А.Е. Крюкова  
Э.В. Штыкова  
Л.А. Дадинова**



**ANSTO (Sydney,  
Australia)  
A.V. Sokolova**

- **Малоугловое рассеяние** – когерентное диффузное рассеяние монохроматических рентгеновских лучей и нейтронов вблизи первичного луча на **апериодических** флуктуациях электронной плотности в материалах (напр., при наличии микропор в твердом теле).
- Картина малоуглового рассеяния, как и дифракционная картина, является **результатом интерференции лучей**, когерентно рассеянных на образце. При типичных длинах волн излучения в диапазоне от 0,05 до 0,5 нм малоугловое рассеяние позволяет исследовать структуры размерами от **единиц до нескольких сотен нанометров**. Интерференционная картина рассеяния формируется сложением множества вторичных когерентно рассеянных волн, которые отличаются друг от друга по фазе. Фазовые отличия и амплитуды слагаемых зависят от пространственного распределения электронной плотности, то есть от структуры объекта, и определяют форму экспериментальной кривой рассеяния  $I(s)$ , анализ которой позволяет определить электронный радиус инерции и максимальный размер наночастиц в монодисперсных системах и их распределение по размерам в полидисперсных образцах.

# Схема малоуглового эксперимента (лаборатория)



Установка рентгеновского малоуглового рассеяния АМУР-К

Детектор  
Статическая картина интенсивности рассеяния

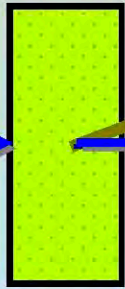
Образец на просвет

Монохроматор (пиролитический графит)

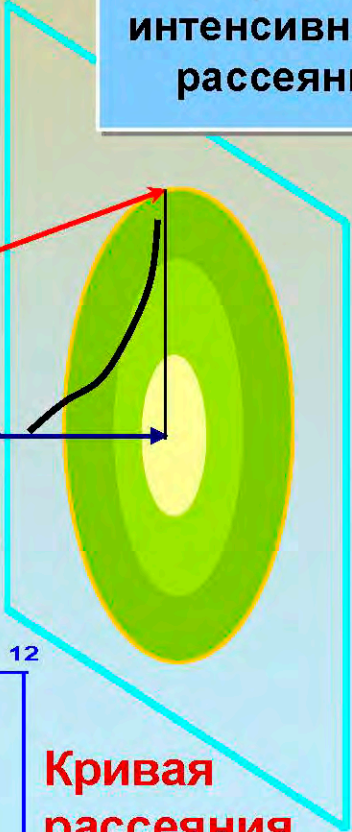
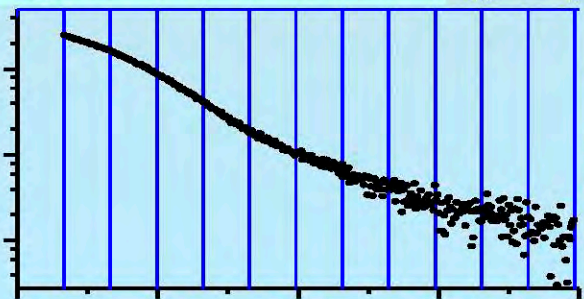
Рентгеновская трубка

Длина волны излучения 0.154 нм

Блочный коллиматор по схеме Кратки



log интенсивности



$k_1$   
 $k_0$   
 $s$   
 $2\theta$

Кривая рассеяния

S

# Схема малоуглового эксперимента (синхротрон)

Волновой вектор  
 $k_0 = 2\pi / \lambda$

Вектор рассеяния  $s = k_1 - k_0$   
 $s = 4\pi \sin \theta / \lambda$

Белый пучок  
РИ

Монохроматор  
(монокристалл,  
зеркало)

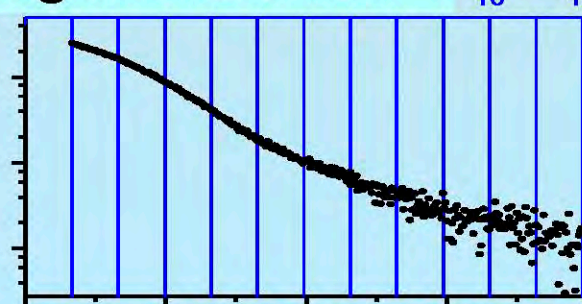
Образец,  
на просвет

Детектор  
Статическая  
картина  
интенсивности  
рассеяния

Источник монохроматического  
излучения:

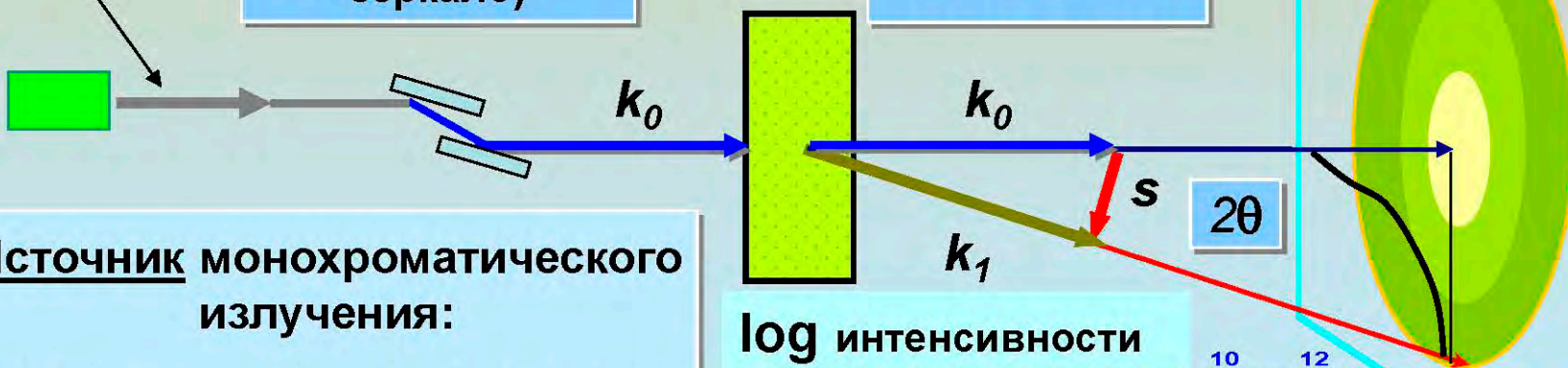
Трубка ( $\lambda = 0.07 - 0.2 \text{ nm}$ )  
Синхротрон ( $\lambda = 0.03 - 1 \text{ nm}$ )  
Тепловые  
нейтроны ( $\lambda = 0.1 - 5 \text{ nm}$ )

log интенсивности

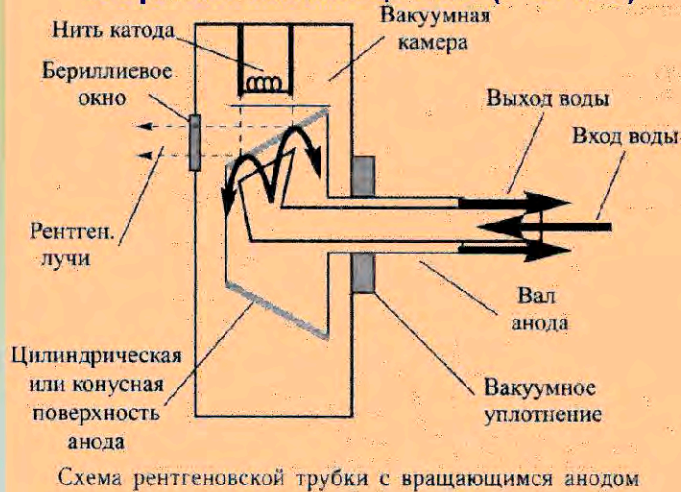


Кривая  
рассеяния

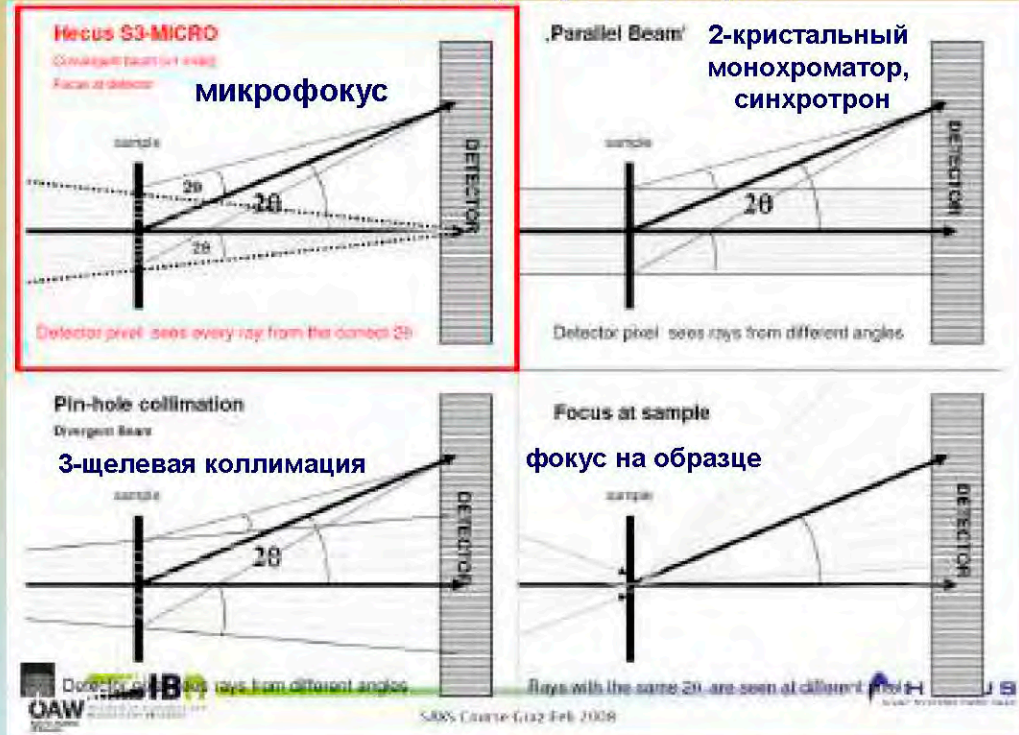
S



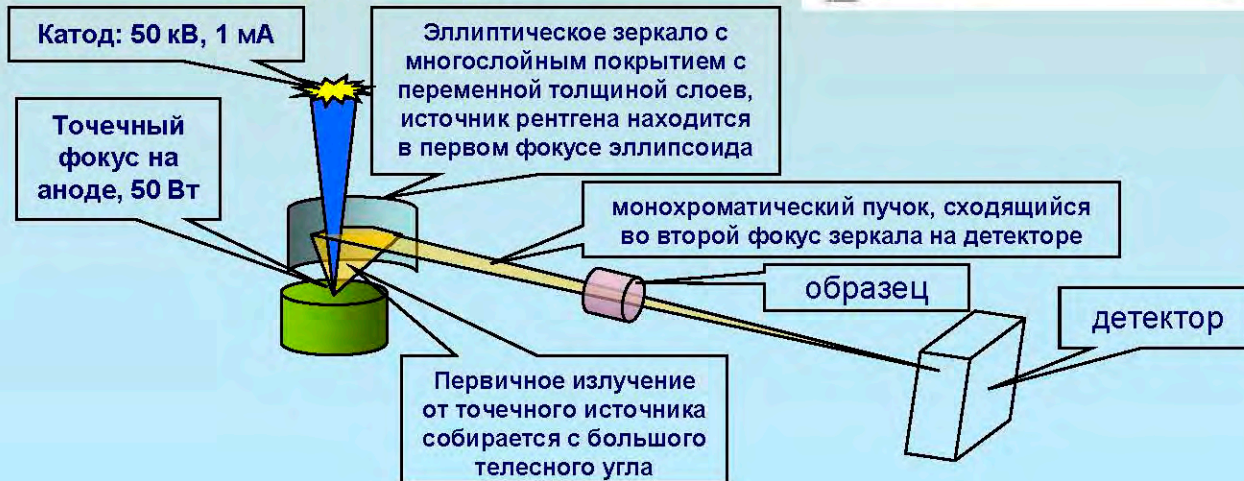
**Источник с вращающимся анодом:**  
 высокая яркость, но очень большая  
 потребляемая мощность (5-50 кВт)



**Коллимационные искажения сигнала рассеяния при конесном размере пучка на образце**



## Принцип микрофокусного дифрактометра



**Преимущества микрофокусного источника:**

1. Высокая яркость (200 млн. квантов/сек) – даже больше, чем у вращающегося анода!
2. Малое сечение пучка
3. Сходящийся пучок – отсутствие коллимационных искажений
4. Малая потребляемая мощность (30-60 Вт)

# Лабораторные установки МУР

## Специализированные установки



АМУР-К  
(ИК РАН)



Anton Paar  
SAXSess mc<sup>2</sup>



## Многоцелевые установки

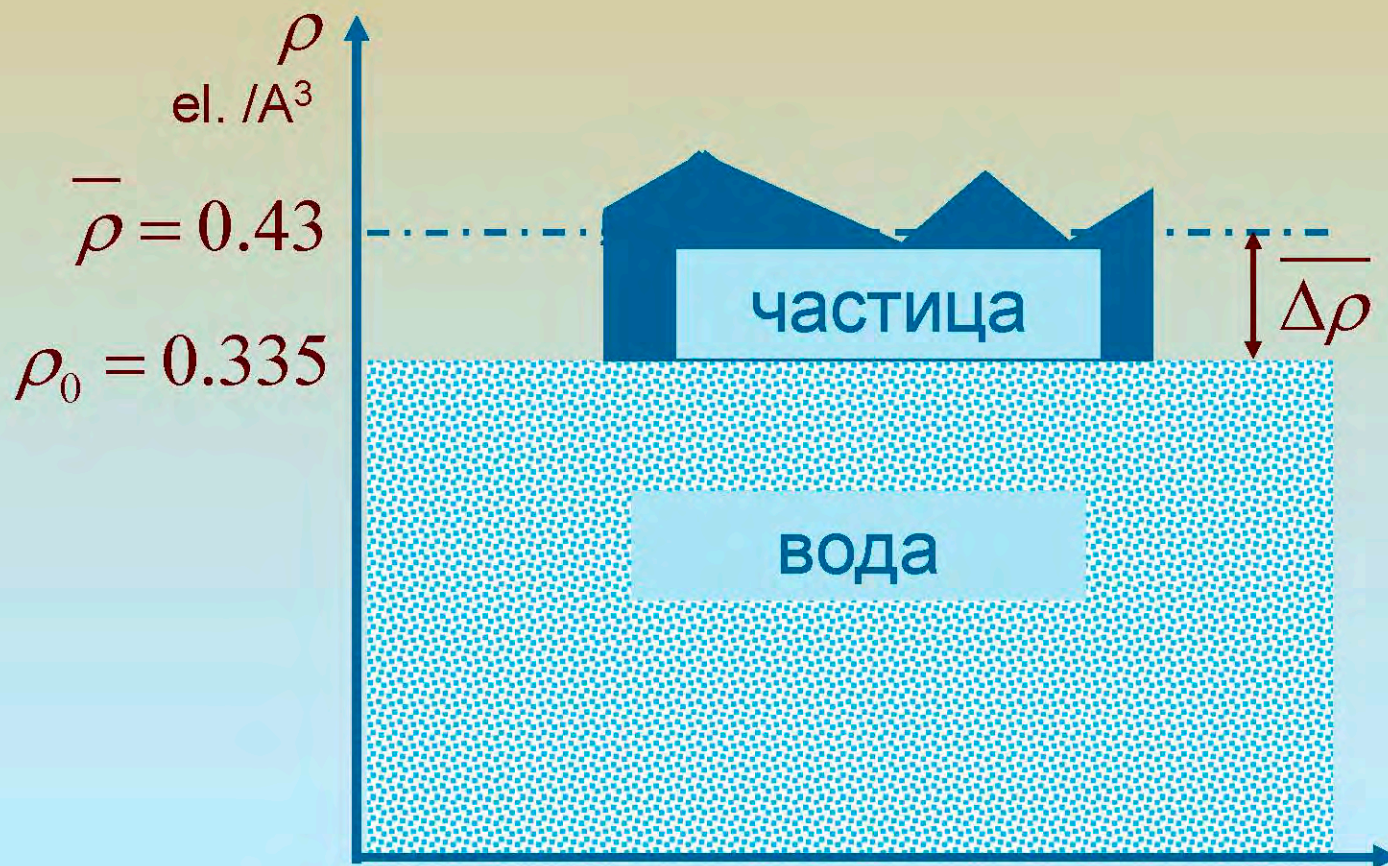


Rigaku  
S-MAX3000



BRUKER NANOSTAR

# Рассеивающий контраст для биоорганических соединений в воде





# Пакет обработки данных МУР: PRIMUS

**primus** File Tools BinAsc Gnom Peak Help

**Graphics Window**

Subtracted data **PLOT**

log I

(1) d27000.dat  
(2) d28000.dat  
(3) Subtr06.dat

20-Feb-2002 14:09:29 E:\PROTEINS\Conti\_01\Conti.c01

**Command Window**

```
Working directory: E:\PROTEINS\Conti_01\Conti.c01\
File to be opened: d20000.dat
Plot: view experimental data
Plot: view experimental data
Subtract: manipulation with data
```

**Help Window**

Program PRIMUS performs the following operations on data files such as: average, subtraction, and curve fitting and evaluation of parameters such as radius of gyration, zero intensity and molecular weight.

The program PRIMUS allows the user to perform these operations which is convenient for the user.

The program is written by: ...

Version 1.0 20.01.02

**GNOM Parameters**

Input1: Subtr06.dat Select nBeg: 1

Input2: none Select nEnd: 701

Output: gnom.out iUnits: 1

Job: 0 Rmin: 0.0 Rmax: 0.0

P(Rmin)=0  P(Rmax)=0

Display Search  Compute Errors

Reset Run Cancel

**Dialog**

Active	File Name	Range	Units	nBeg	Sync	nEnd	Conc	Multiplier
<input checked="" type="checkbox"/>	#1 d27000.dat	Select	1	1		701	1.000	1.000
<input checked="" type="checkbox"/>	#2 d28000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#3 d29000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#4 d30000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#5 d31000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#6 d32000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#7 d33000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#8 d34000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#9 d35000.dat	Select	1	1		701	1.000	1.000
<input type="checkbox"/>	#10 d36000.dat	Select	1	1		701	1.000	1.000

Plot: Average Subtract Divide ZerConw Adjust Guinier Flat Clear

Sasplot Merge Subcst Divcst ZerConc Adjmul Porod Rod Finish

OUT Subtr06.dat Select 1 1 701 1.000 1.000

Running | Input pending in Graphics Window

# Связь между прямым и обратным пространствами для одного неориентированного объекта

По определению Порода (1948) и Вилсона (1949),  $\chi(r)$  - это функция распределения **усредненного по ориентациям** сдвинутого объема в зависимости от величины сдвига (**корреляционная функция**). Это определение вытекает из определения функции Паттерсона, как самосвертки функции рассеивающей плотности.

$$I(s) = \langle I(\mathbf{s}) \rangle_{\Omega} = 4\pi \int_{r=0}^{\infty} r^2 \chi(r) \cdot \frac{\sin(sr)}{sr} dr$$

$$s = 4\pi \sin \theta / \lambda$$

$$\chi(r) = \langle P(\mathbf{r}) \rangle_{\omega} = \frac{1}{2\pi^2} \int_{s=0}^{\infty} s^2 I(s) \cdot \frac{\sin(sr)}{sr} ds$$

Здесь  $r^2 \chi(r) = p(r)$  - функция распределения по расстояниям. Полученные формулы замечательны еще и тем, что в них  $r$  и  $s$  скаляры, а не векторы. Функции  $\chi(r)$  можно придать вероятностные свойства, отнормировав ее.

# Малоугловое рентгеновское рассеяние от разбавленных растворов белков

$$I_i(s) = \langle A^2(s) \rangle_{\Omega} = 4\pi \int_{r=0}^{\infty} r^2 \langle \rho(\mathbf{r}) \otimes \rho(\mathbf{r}) \rangle_{\omega} \frac{\sin(sr)}{sr} dr$$

Интенсивность рассеяния от одной молекулы

Корреляционная функция (усредненная самосвертка функции электронной плотности)

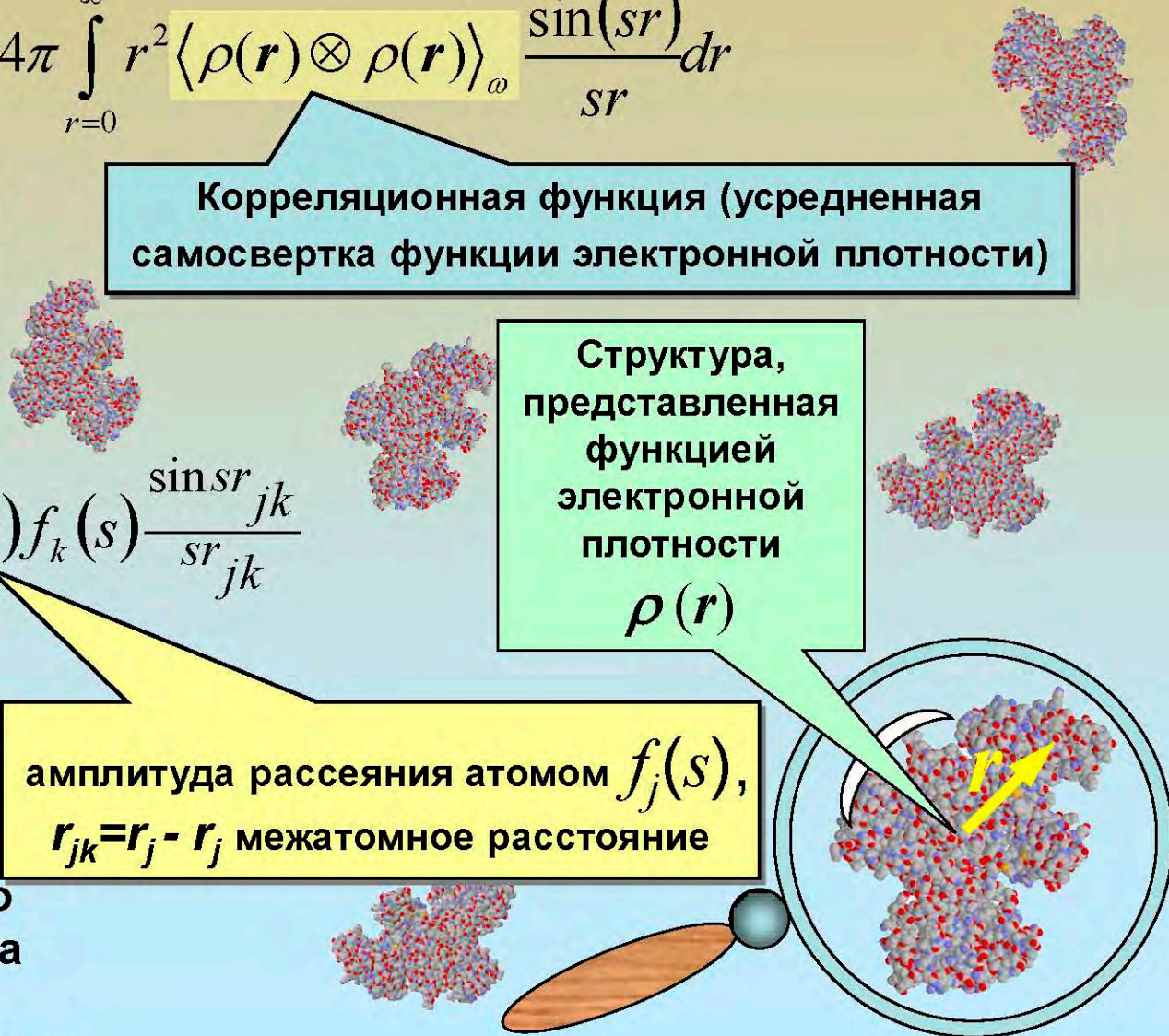
$$I_i(s) = \sum_j^N \sum_k^N f_j(s) f_k(s) \frac{\sin sr_{jk}}{sr_{jk}}$$

Структура, представленная функцией электронной плотности  $\rho(\mathbf{r})$

$$I_{total}(s) = \sum_{k=1}^M I_i(s)$$

амплитуда рассеяния атомом  $f_j(s)$ ,  
 $r_{jk} = r_j - r_j$  межатомное расстояние

Рассеяние от всего образца есть сумма интенсивностей рассеяния от M молекул



# Общий случай рассеяния от системы из $N$ одинаковых рассеивающих объектов

$$\begin{aligned} \langle I(\mathbf{s}) \rangle &= \langle N \rangle \langle f^2(\mathbf{s}) \rangle + \\ &+ \langle f(\mathbf{s}) \rangle^2 \iint_V 1 \cdot \frac{\sin(sr_{ij})}{sr_{ij}} \cdot \frac{dv_i}{v} \frac{dv_j}{v} + \\ &+ \langle f(\mathbf{s}) \rangle^2 \iint_V [W_{ij} - 1] \frac{\sin(sr_{ij})}{sr_{ij}} \frac{dv_i}{v} \frac{dv_j}{v} \end{aligned}$$

**Соотношение Цернике и Принса:**

$$I(s) = N \cdot f^2(s) \cdot \left\{ 1 + \frac{1}{v} \int_0^\infty 4\pi r^2 [W(r) - 1] \frac{\sin(sr)}{sr} \cdot dr \right\}$$

$W(r)$  – Фурье образ радиальной статистической функции размещения объектов в пространстве

# Сводка интегральных параметров наночастиц, определяемых по данным МУР от монодисперсных разбавленных систем

$$I(s) = \langle I(s) \rangle_{\Omega} = 4\pi \int_{r=0}^{\infty} p(r) \cdot \frac{\sin(sr)}{sr} dr$$

$$I(s) = I(0) \cdot \exp(-R_g^2 \cdot s^2 / 3) \text{ (формула Гинье),}$$

радиус инерции

$$\text{где } I(0) \sim N \cdot (V \cdot \rho)^2 \sim N \cdot M^2;$$

объем и плотность

$$R_g^2 = \int \rho(r) \cdot r^2 dr / \int \rho(r) dr = \int r^4 \cdot \gamma(r) dr / 2 \int r^2 \gamma(r) dr$$

интегральный  
радиус инерции

$$V = 4 \cdot \pi \int r^2 \cdot \gamma(r) dr = 4 \cdot \pi \int p(r) dr;$$

объем

$$S = \lim \{4 V [(1 - \gamma_o(r))/r]\}$$

площадь поверхности

$$I_m = 2 \int \gamma_o(r) dr;$$

средний размер

$$D_{\max}$$

максимальный размер

# График Гинье: определение радиуса инерции наночастиц

Для самых малых углов справедливо приближение:

$$\frac{\sin(sr)}{sr} = 1 - \frac{s^2 r^2}{3!} + \frac{s^4 r^4}{5!} - \dots$$

$$\ln[I(s)] \cong \ln[I(0)] - \frac{4\pi^2}{3} R_g^2 s^2$$

Линейная аппроксимация позволяет найти:

$I(0)$  - интенсивность рассеяния в нулевой угол;

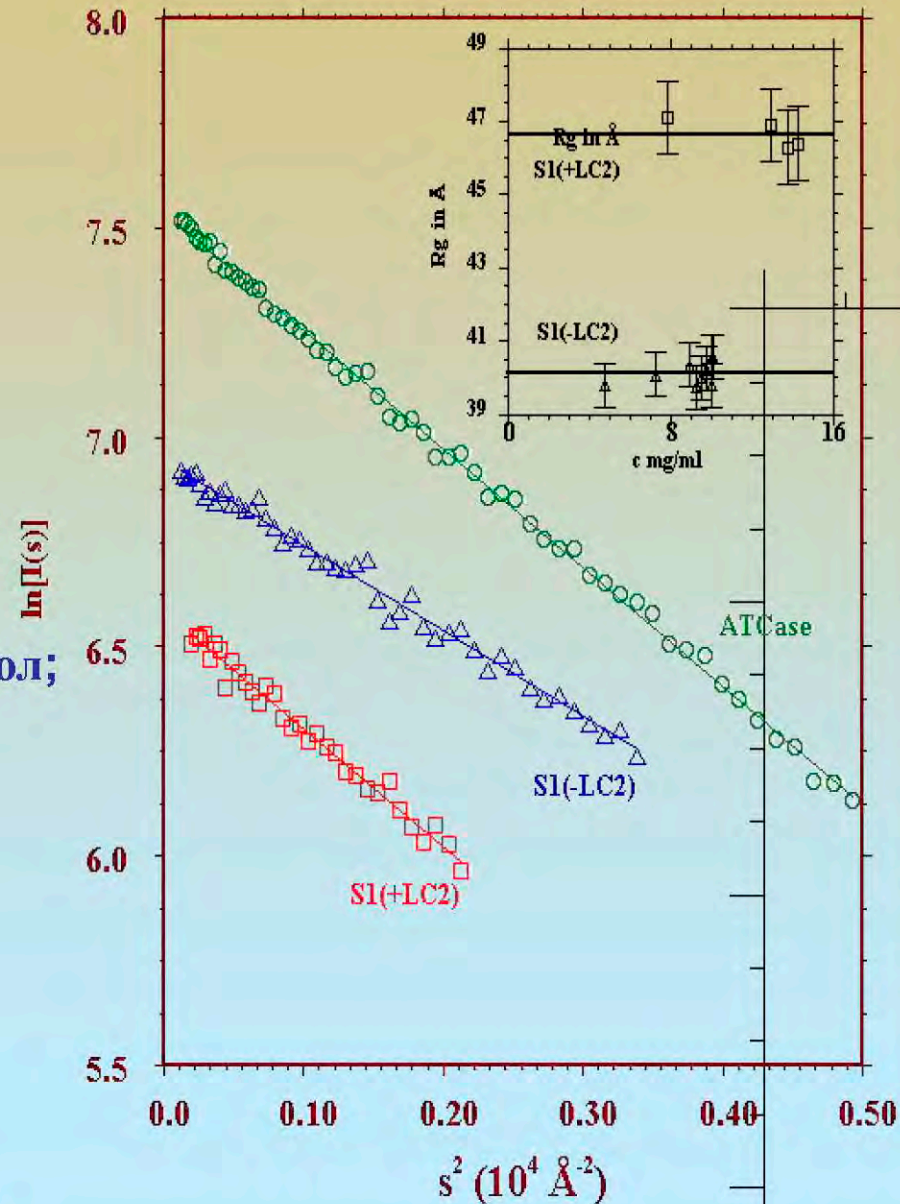
$R_g$  - радиус инерции ( $q = s / 2\pi$ ):

$$R_g^* = \sqrt{\frac{\sum \frac{q_i^2 I_i^2}{\sigma_i^2} \sum \frac{I_i^2 \ln I_i}{\sigma_i^2} - \sum \frac{q_i^2 I_i^2 \ln I_i}{\sigma_i^2} \sum \frac{I_i^2}{\sigma_i^2}}{\sum \frac{q_i^4 I_i^2}{\sigma_i^2} \sum \frac{I_i^2}{\sigma_i^2} - \left( \sum \frac{q_i^2 I_i^2}{\sigma_i^2} \right)^2}}$$

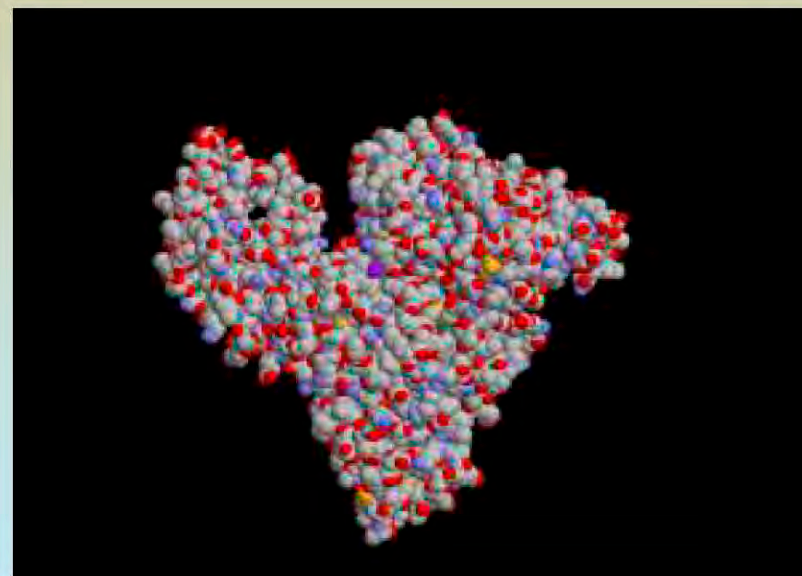
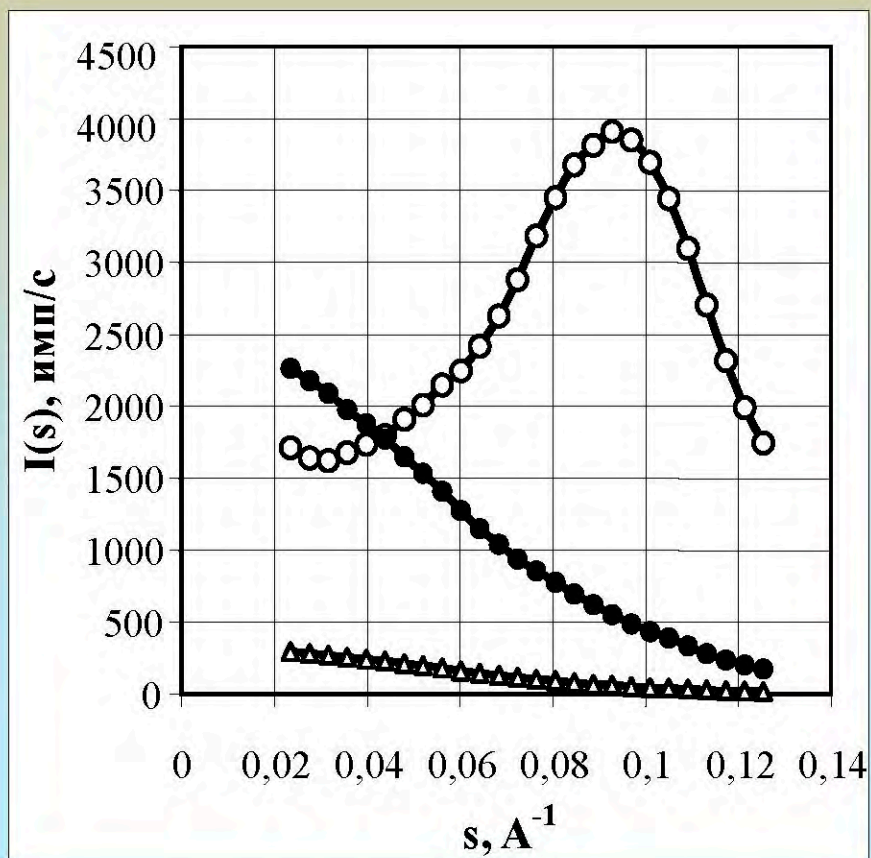
Допустимый диапазон углов :

$$0 < 2\pi R_g s < 1.3$$

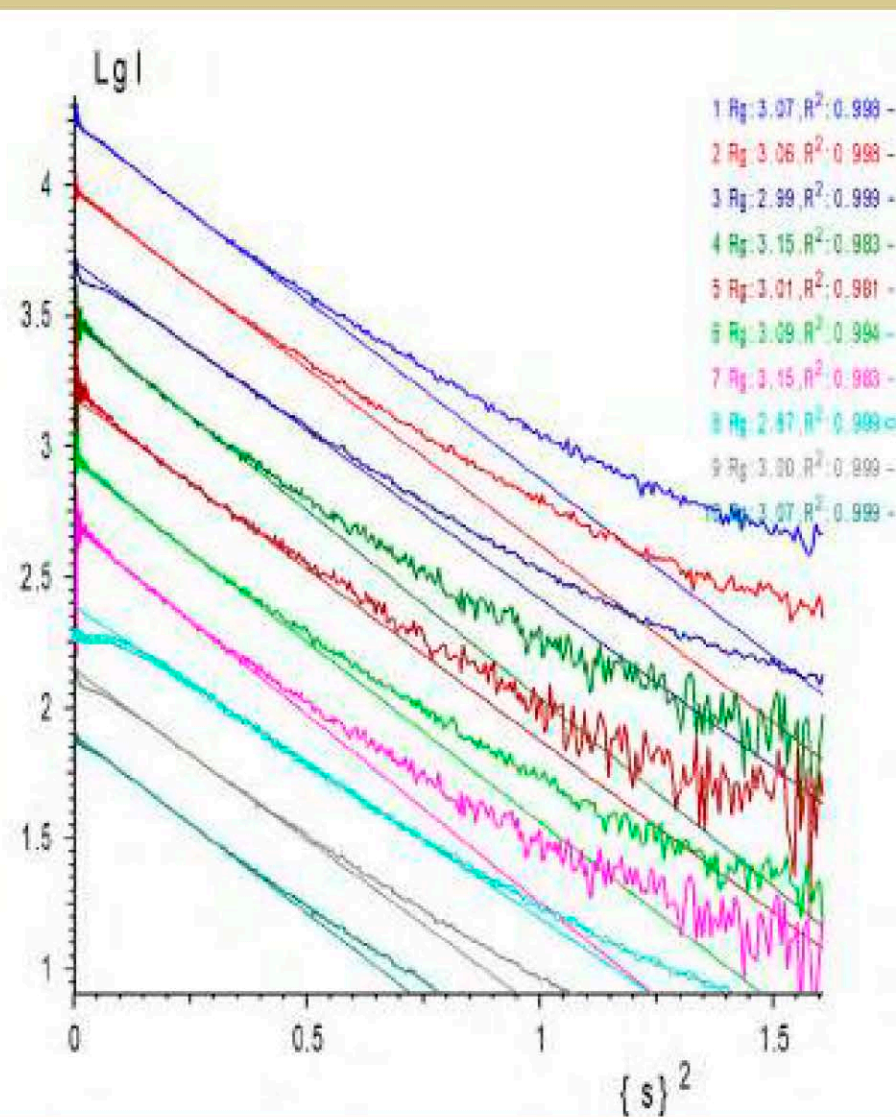
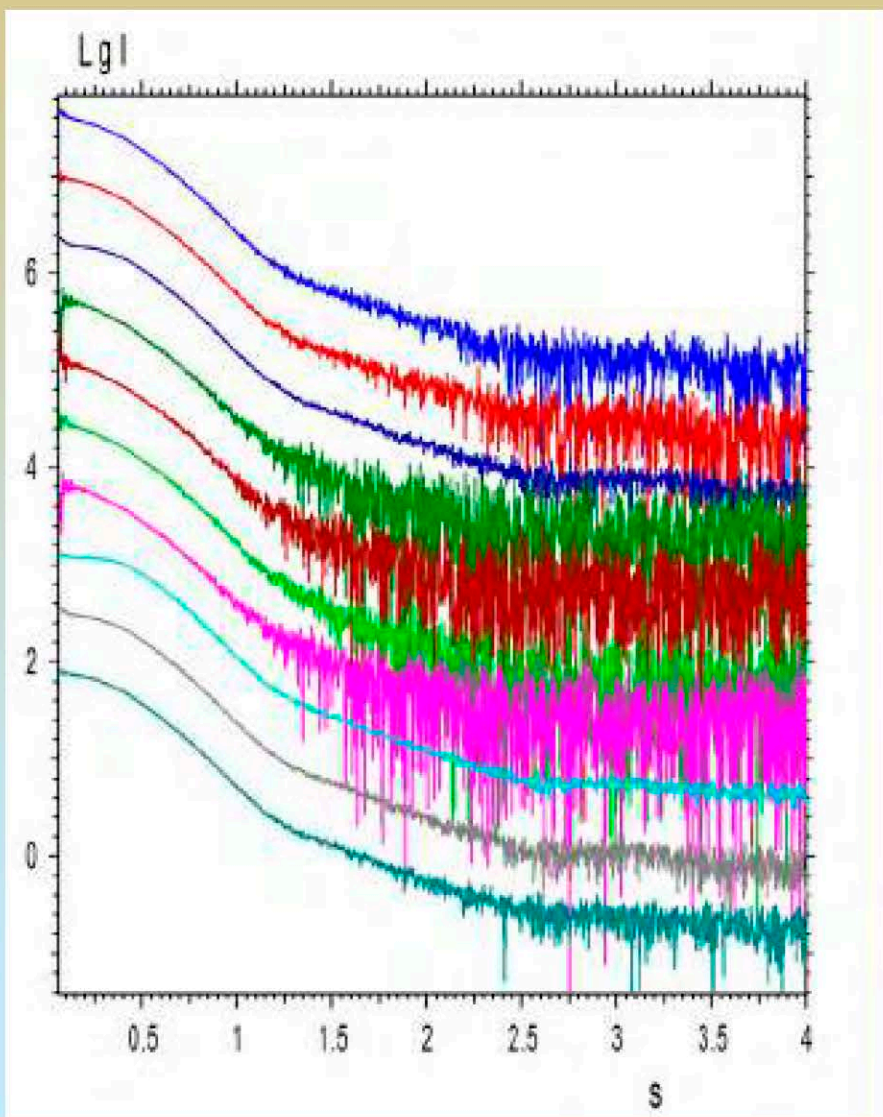
для сферической частицы при точности 1-2%



**Концентрационный эффект:  
интенсивности МУР (а) от растворов бычьего  
альбумина (БСА) при различных концентрациях:  
2,5 ( $\Delta$ ); 25 ( $\bullet$ ); 250 ( $\circ$ ) мг/мл.**

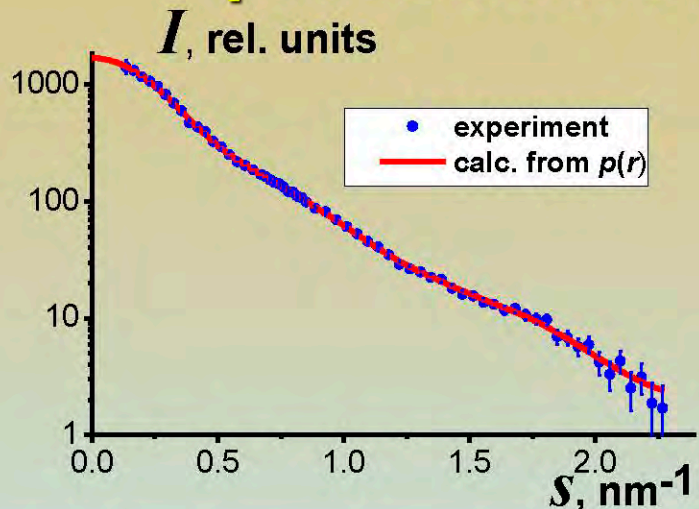


# Пример графиков Гинье: серия измерений раствора бычьего альбумина при разных концентрациях

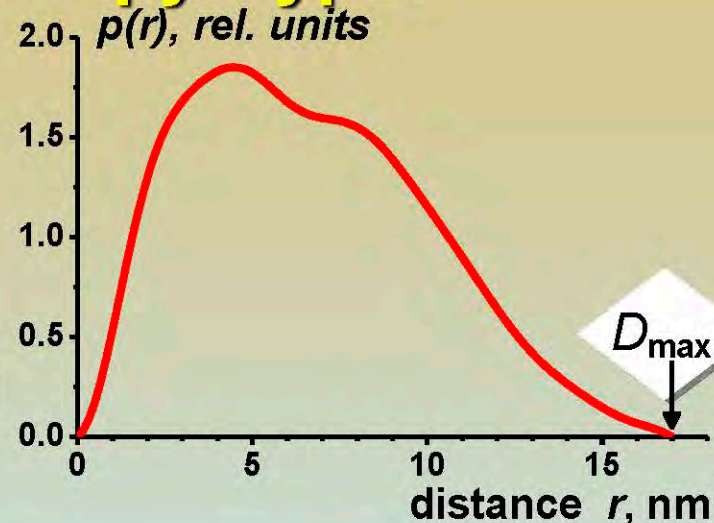




# Однозначная задача: расчет интенсивности рассеяния от известной структуры



Фурье  
↔  
преобразование  
ие  
(однозначно)



$$I(s) = \langle I(s) \rangle_{\Omega} = 4\pi \int_{r=0}^{\infty} r^2 \gamma(r) \cdot \frac{\sin(sr)}{sr} dr$$

$$\gamma(r) = \langle P(\mathbf{r}) \rangle_{\omega} = \frac{1}{2\pi^2} \int_{s=0}^{\infty} s^2 I(s) \cdot \frac{\sin(sr)}{sr} ds$$

$I(s)$  однозначно определена в обратном пространстве Фурье-преобразованием одномерной функции распределения по расстояниям  $\rho(r)$  в прямом пространстве:

$$\rho(r) = r^2 \cdot \gamma(r)$$

$$\gamma(r) \Big|_{r=|r_1-r_2|=const} = \langle \rho(\mathbf{r}_1) \cdot \rho(\mathbf{r}_2) \rangle_{\omega} = \langle P(\mathbf{r}) \rangle_{\omega} = \frac{1}{4\pi} \int_{\omega=0}^{4\pi} P(\mathbf{r}) \cdot d\omega,$$

$$где \quad P(\mathbf{r}) = \int d\mathbf{r}_1 \cdot \rho(\mathbf{r}_1) \cdot \rho(\mathbf{r}_1 - \mathbf{r})$$

$\rho(r)$  (и, соответственно,  $I(s)$ ) однозначно вычисляются для известных структур, что лежит в основе методов определения формы молекул по данным рассеяния

# Косвенное преобразование Фурье: решение системы линейных уравнений с регуляризацией по Тихонову

$$I(s) = 4\pi \int_{r=0}^{\infty} p(r) \cdot \frac{\sin(sr)}{sr} dr$$

Составляем из значений базисных функций матрицу В:

$$B_{ij} = \frac{\sin(q_i r_j)}{q_i r_j}$$

$q_i$  –  $i$ -ое значение угла рассеяния в единицах модуля вектора рассеяния, нм<sup>-1</sup>;

$r_j$  –  $j$ -ое значение расстояния, нм;

$M$  – число угловых отсчетов измерений интенсивности рентгеновского излучения.

$N$  – число точек на графике функции  $p(r)$ ,  $N \leq M$ ,  $r(N) = r_{\max}$  назначают априори исходя из дополнительной информации об объекте или достаточно большим с последующим уточнением после анализа решения.

Нахождение  $p(r)$  сводится к решению системы линейных алгебраических уравнений:

методом наименьших квадратов:  $\min_p \left\{ \left\| I_{\text{exp}} - I_p \right\|^2 \right\}$ :  $p = \left[ B^T \cdot B \right]^{-1} \cdot B^T \cdot I_{\text{exp}}$ ,  $(I_p)_i = 4\pi \sum_{j=1}^N p_j \cdot \frac{\sin(s_i r_j)}{s_i r_j} \cdot \Delta r$

В силу плохой обусловленности задачи решение искажается эффектом обрыва функции  $I_{\text{exp}}(s)$  на больших  $s$  – график  $p(r)$  представляет собой сильно осциллирующую кривую, не отражающую физический смысл. Выход из положения состоит в поиске гладкого решения. С этой целью функцию наименьших квадратов дополняют требованием гладкости, например, минимума интегральной первой производной по решению (исчисляемой в конечных разностях на заданной сетке расстояний  $r_j$ ):

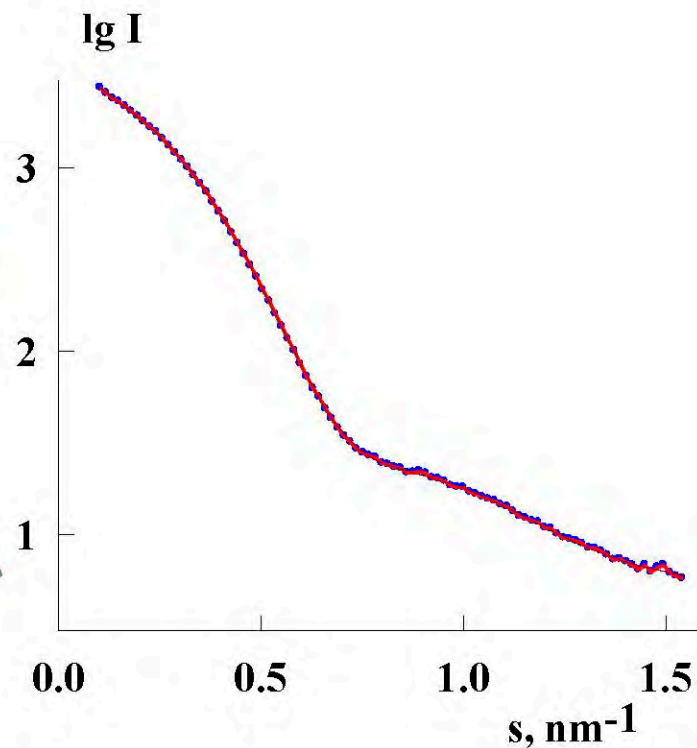
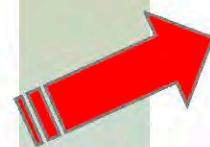
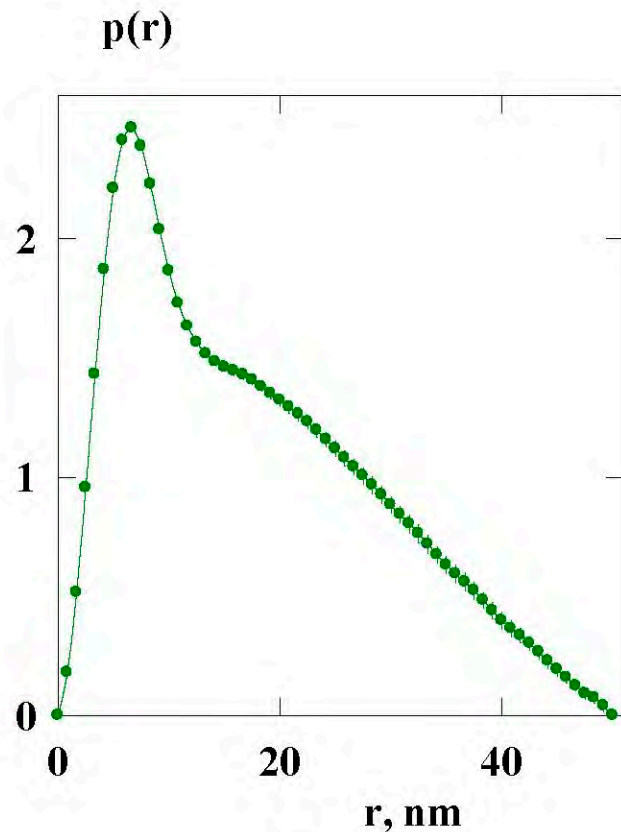
$$\min_p \left\{ \left\| I_{\text{exp}} - I_p \right\|^2 + \alpha \cdot \left\| p' \right\|^2 \right\}: \quad p = \left[ B^T \cdot B + \alpha \cdot R \right]^{-1} \cdot B^T \cdot I_{\text{exp}}$$

где  $R$  – трехдиагональная (в случае регуляризации по первой производной) симметричная матрица регуляризатора,  $\alpha$  – параметр регуляризации.  $\alpha$  необходимо подобрать таким, чтобы разность  $I_{\text{exp}} - I_p$  еще не содержала значимых систематических отклонений, но решение  $p(r)$  было свободно от осцилляций.

На практике ( программа GNOM) поиск оптимального  $\alpha$  осуществляют методом перебора с анализом степени осциллируемости решения, его стабильности относительно искусственного возмущения исходных данных, величины систематического отклонения в  $I_{\text{exp}} - I_p$  и ряда других критериев качества решения [Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria / D.I. Svergun // J. Appl. Cryst. – 1992. – V. 25. – P. 495 – 503].

# Функция парных расстояний: косвенное преобразование Фурье (регуляризация по Тихонову): программа GNOM

$$p(r) = \sum_{n=1}^N c_n \varphi_n(r)$$



$$\min \left\{ \left\| I_{\text{exp}}(s) - I_{\text{calc}}(s) \right\|^2 + \alpha \Omega[p(r)] \right\}$$

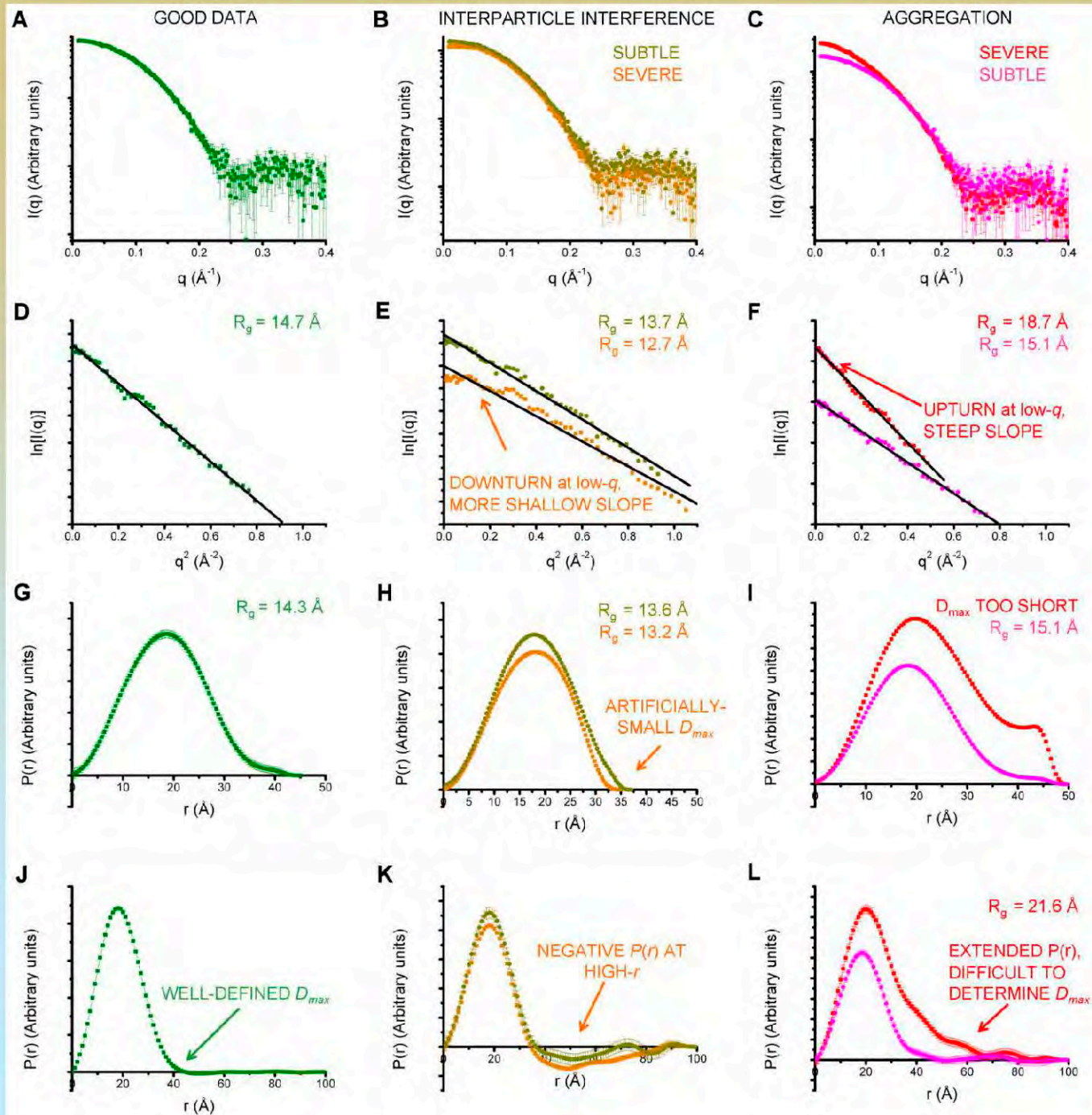
## По функции распределения по расстояниям $p(r)$ можно также устойчиво найти инварианты:

Электронный радиус инерции, максимальный диаметр и интенсивность рассеяния в нулевой угол:

$$R_g^2 = \frac{\int_0^{D_{\max}} r^2 p(r) dr}{2 \int_0^{D_{\max}} p(r) dr} \quad \text{и} \quad I(0) = 4\pi \int_0^{D_{\max}} p(r) dr$$

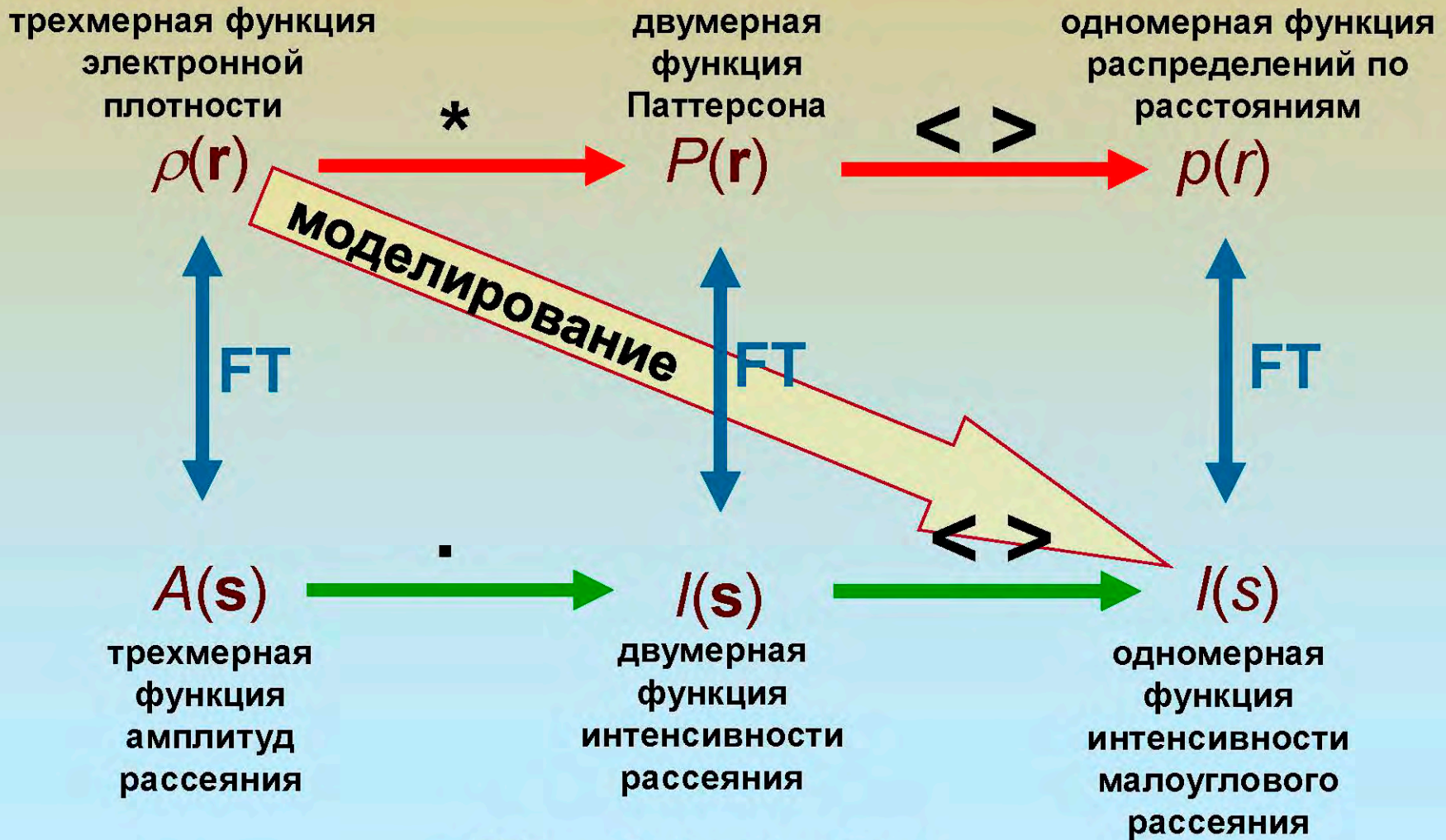
$$R_g = \sqrt{\frac{\sum_{j=1}^{N-1} r_j^2 p_{opt}(r_j) + r_{j+1}^2 p_{opt}(r_{j+1})}{2 \sum_{j=1}^{N-1} p_{opt}(r_j) + p_{opt}(r_{j+1})}}$$

Данный расчет основан на использовании уже всей кривой рассеяния и поэтому более надежен, чем аппроксимация Гинье. Это позволяет надежно распознать случаи нарушения монодисперсности (например, агрегацию) наночастиц.



Предвари-  
тельный  
анализ  
данных

# Функциональные связи между прямым и обратным пространствами



FT: преобразование Фурье

# Формула суммирования Дебая – основа методов моделирования структуры рпо данным рассеяния и дифракции

Если частица представлена дискретным набором из  $N$  элементов (например, атомов), то интенсивность рассеяния от структуры можно представить в виде:

$$i_1(\mathbf{s}) = \sum_{i=1}^N \sum_{j=1}^N f_i(\mathbf{s}) f_j(\mathbf{s}) e^{2i\pi(\mathbf{r}_i - \mathbf{r}_j)\mathbf{s}}$$

где  $f_i(\mathbf{s})$  – амплитуды рассеяния от структурных элементов, а сферически усредненная интенсивность малоуглового рассеяния будет выражена формулой:

$$i_1(s) = \sum_{i=1}^N \sum_{j=1}^N f_i(s) f_j(s) \frac{\sin 2\pi r_{ij}s}{2\pi r_{ij}s} \quad \text{где } r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$$

# Основные шаги структурной интерпретации данных МУР

построение математической модели структуры (**параметризация**),



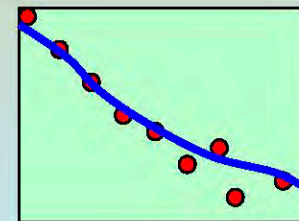
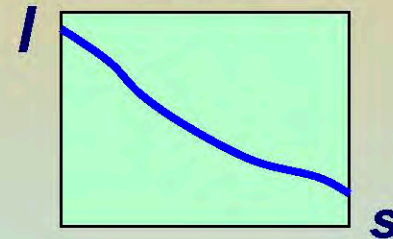
вычисление теоретической интенсивности  $I(s)$  рассеяния от модели (**прямое моделирование**),



сравнение ее с экспериментальными данными (**вычисление критерия невязки с учетом весовой функции, сжимающей диапазон интенсивности до  $\sim 1$  порядка**),

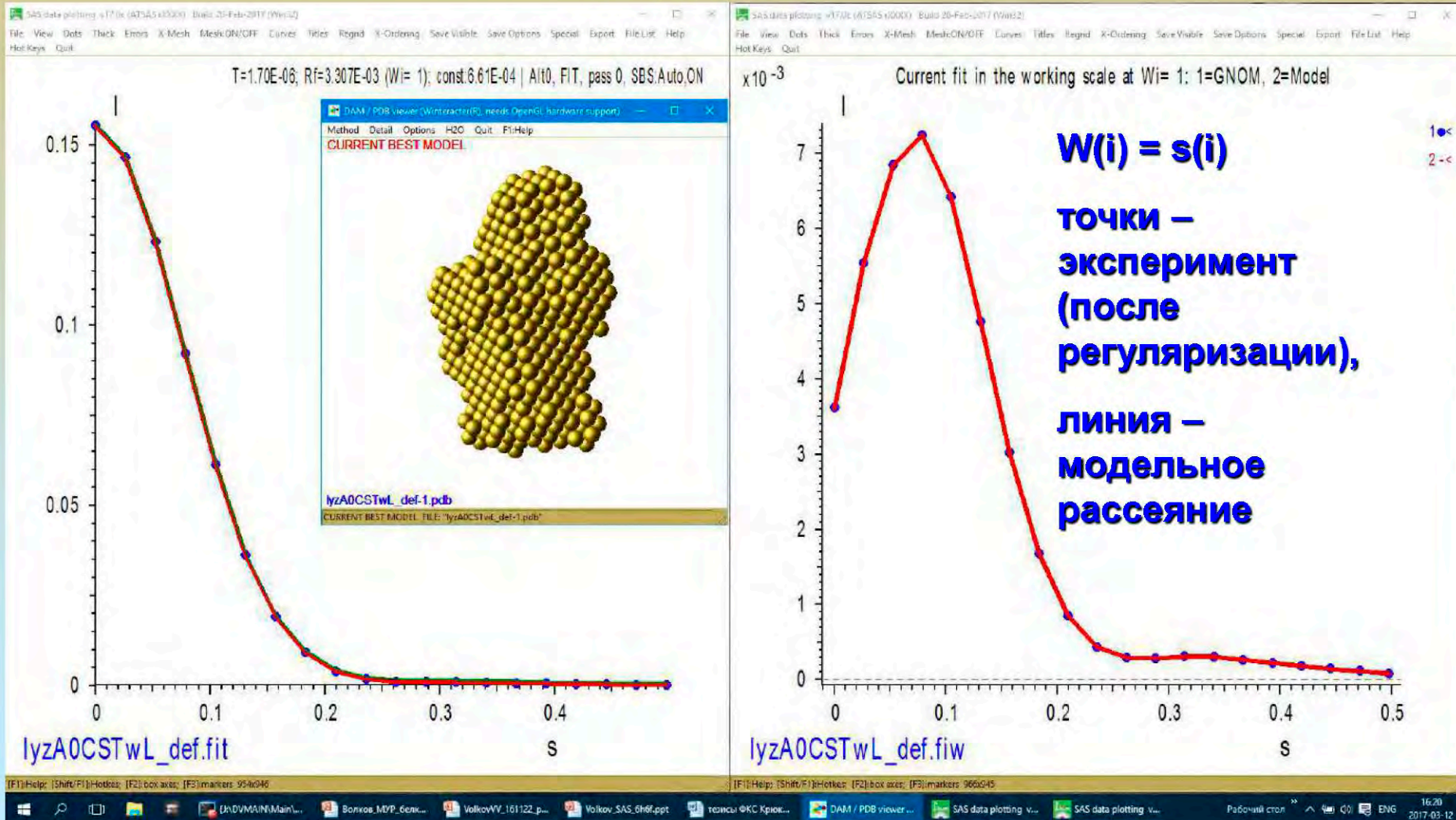


коррекция параметров модели исходя из величины невязки и тенденции ее изменения (**минимизация невязки**) и переход к следующей итерации прямого моделирования. Останов по достижению минимального значения или выполнения других (совокупности) критериев качества

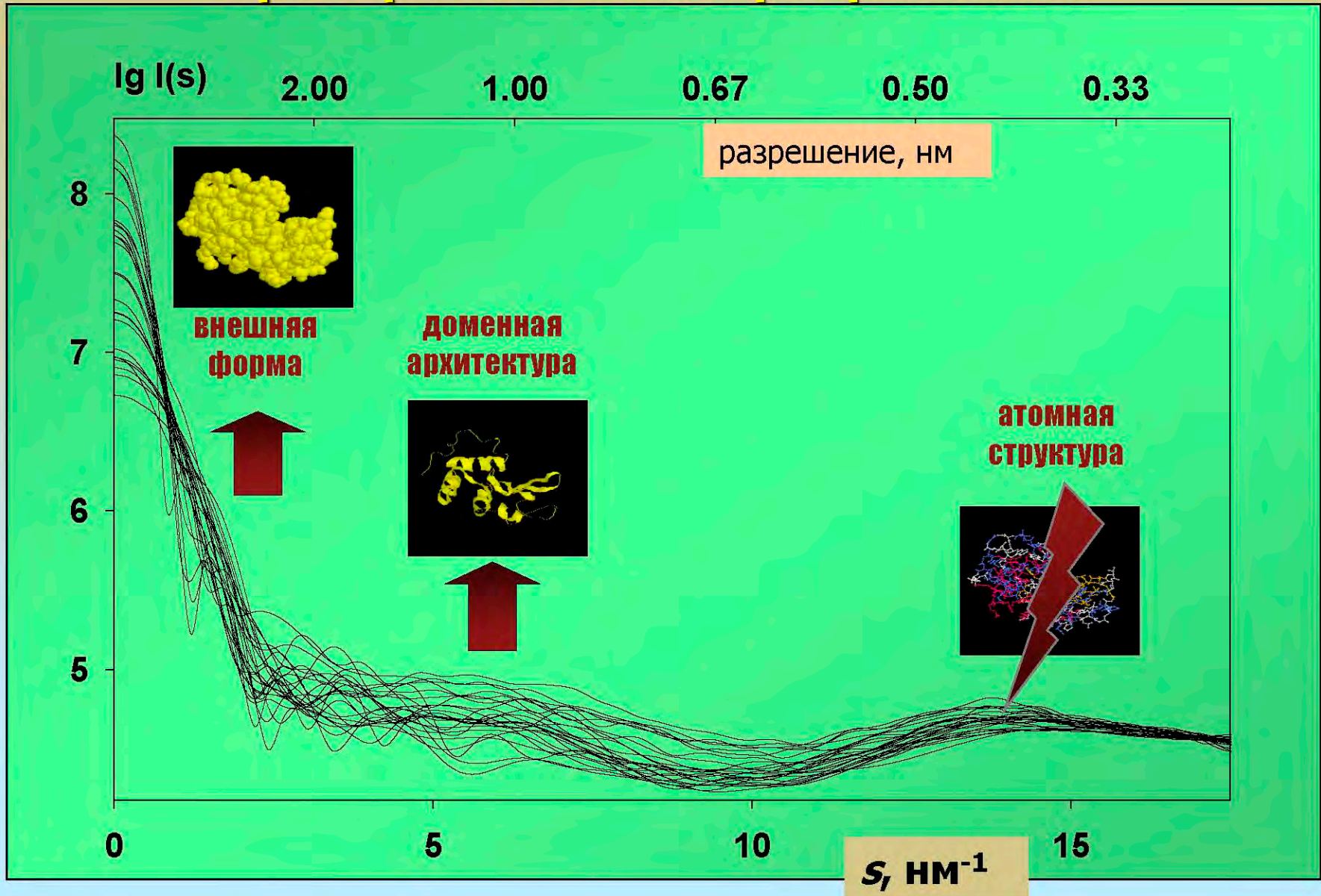




# Пример кривых рассеяния в исходной и рабочей шкале при поиске структурных моделей



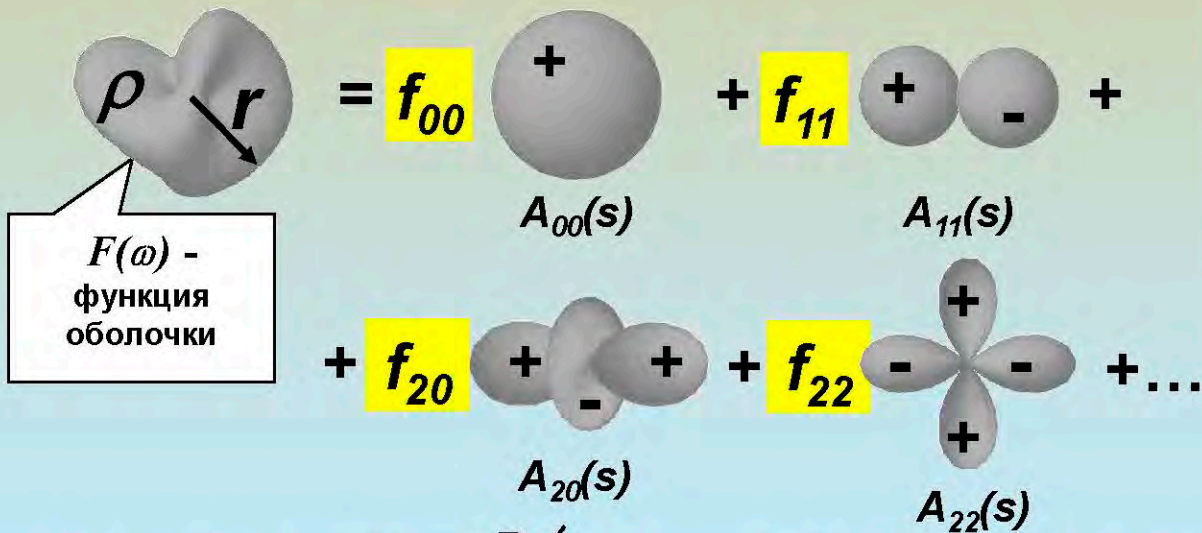
# Данные малоуглового рассеяния и пространственное разрешение



# Представление формы белковой молекулы в виде мультипольного разложения (программа SASHA)

$$F(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega)$$

$$I_L(s) = \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s)|^2, \quad f_{lm}^{(q)} = \int_{\omega} [F(\omega)]^q \cdot Y_{lm}^*(\omega) \cdot d\omega, \quad A_{lm}(s) = (is)^l \cdot \left(\frac{2}{\pi}\right)^{1/2} \times \\ \times \sum_{p=0}^{P_{\max}} \left( (-1)^p \cdot f_{lm}^{l+2p+3} \cdot \left\{ 2^p \cdot p! \cdot (l+2p+3) \cdot [2(l+p)+1]! \right\}^{-1} \cdot s^{2p} \right),$$



[Stuhrmann H. B. Interpretation of small-angle scattering functions of dilute solutions and gases // Acta Cryst. Sect. A. – 1970. – V. 26. – No. 3. – P. 297–306.;

Свергун Д.И., Фейгин Л.А., Щедрин Б.М. Прямой метод интерпретации данных малоуглового рассеяния системами идентичных частиц. // ДАН СССР. –1981. –Т.261. –№4. –С.878–882]

Метод поиска решения:  
алгоритм Левенберга-Марквардта, дополненный схемой переменной метрики (NL2SOL)

разрешение:  $\delta = \frac{\pi R}{(L+1)}$ ,  $R$  – радиус эквивалентной сферы.

Обычно  $L$  выбирают равным 3-4, если не используется ограничение на симметрию формы, и 5-7 в зависимости от симметрии.

Число параметров структурной модели  $f_{lm}$  равно  $(L+1)^2$ .

# Программа SASHA

Программа моделирования формы частиц SASHA находит параметры модели – коэффициенты сферических гармоник  $f_{lm}$  путем минимизации невязки между теоретической и экспериментальной

$$\min_{f_{lm}} \left\{ \frac{\sum_i \left[ \left( I_{\text{exp}}(s_i) - qI(s_i) \right) W(s_i) \right]^2}{\sum_i \left[ I_{\text{exp}}^2(s_i) W^2(s_i) \right]} + \sum_k g_k^2 W_k \right\}$$

$$W(s_i) = s_i^2 / \sigma(s_i)$$

Весовая функция:

«Совмещающий МНК множитель»:

$$q = \left( \mathbf{I}_{\text{exp}} \cdot \mathbf{I} \right) / \|\mathbf{I}\|^2$$

Штрафные члены:

a) за отрицательность расстояний от начала координат до поверхности:

$$g_1 = \int_{\forall F(\omega) < 0} F(\omega) d\omega$$

b) за сдвиг центра тяжести  $[x_c, y_c, z_c]$  из начала координат:

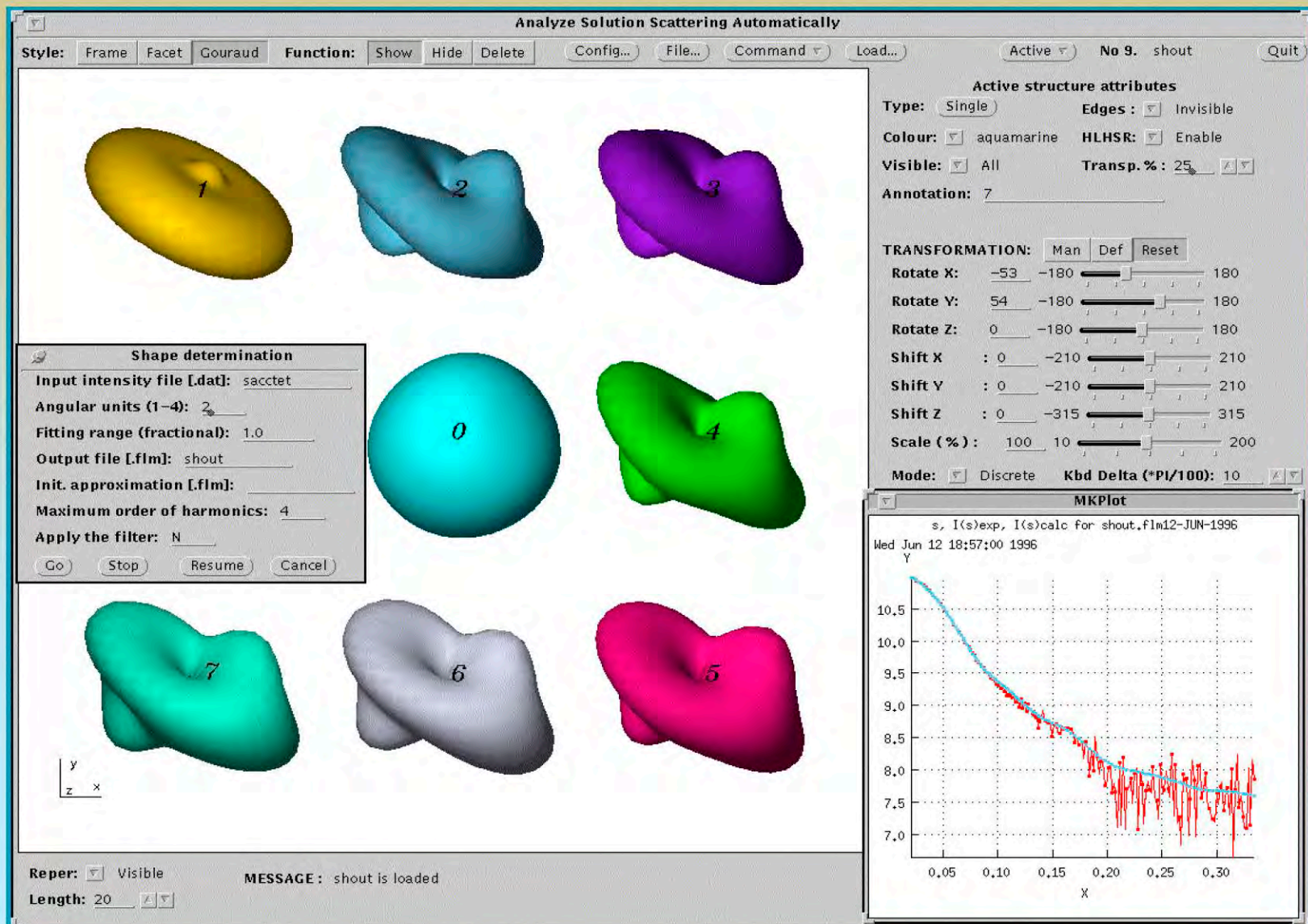
$$g_2 = x_c, \quad g_3 = y_c, \quad g_4 = z_c.$$

c) за «сложность» поверхности ( $W_M$  – молекулярный вес):

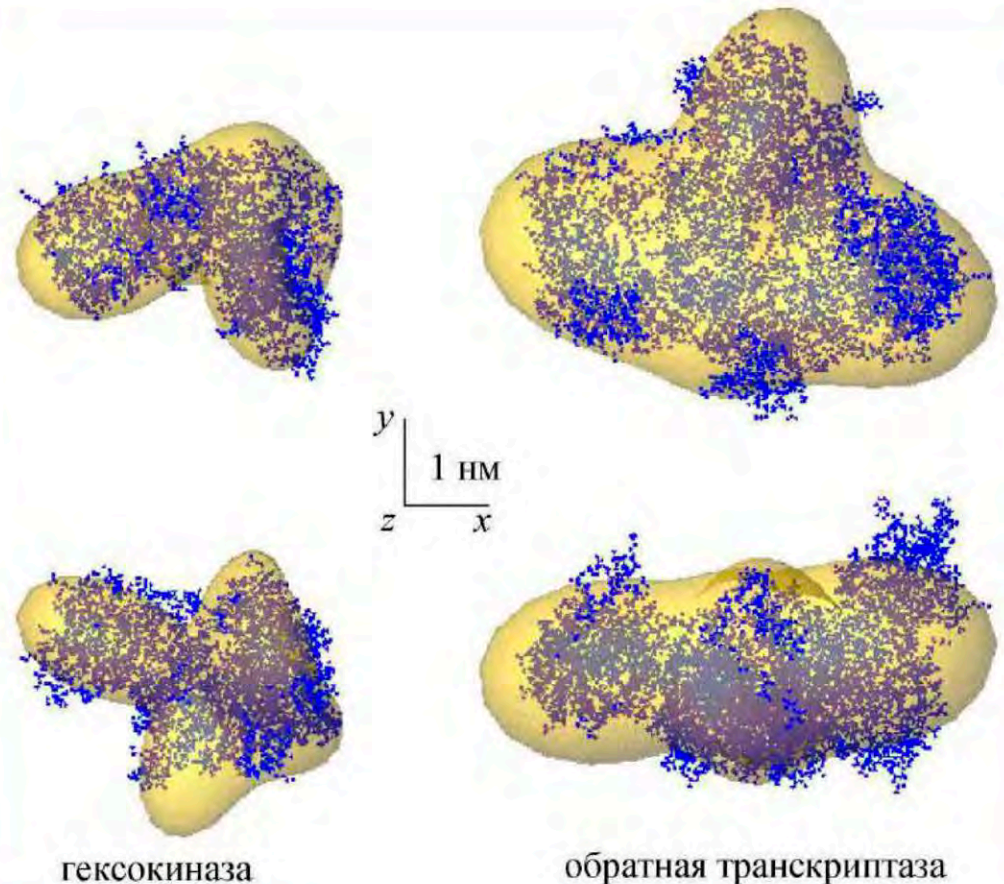
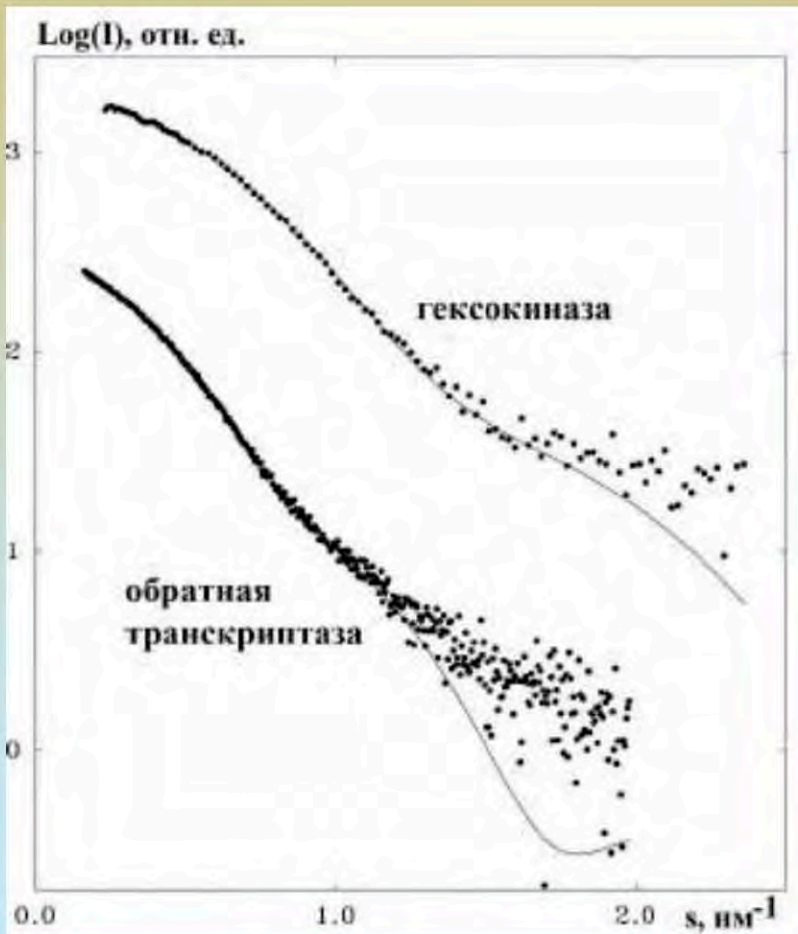
$$g_5 = \frac{(R_g^{\text{exp}})^3}{W_M}$$

(использование дополнительной информации)

# Программа SASNA: определение формы белковых молекул в растворе



# Примеры определения структур белков в растворе



**Внутри найденных оболочек синими точками  
вписаны для сравнения известные  
кристаллические структуры белков**

# Информативность данных малоуглового рассеяния

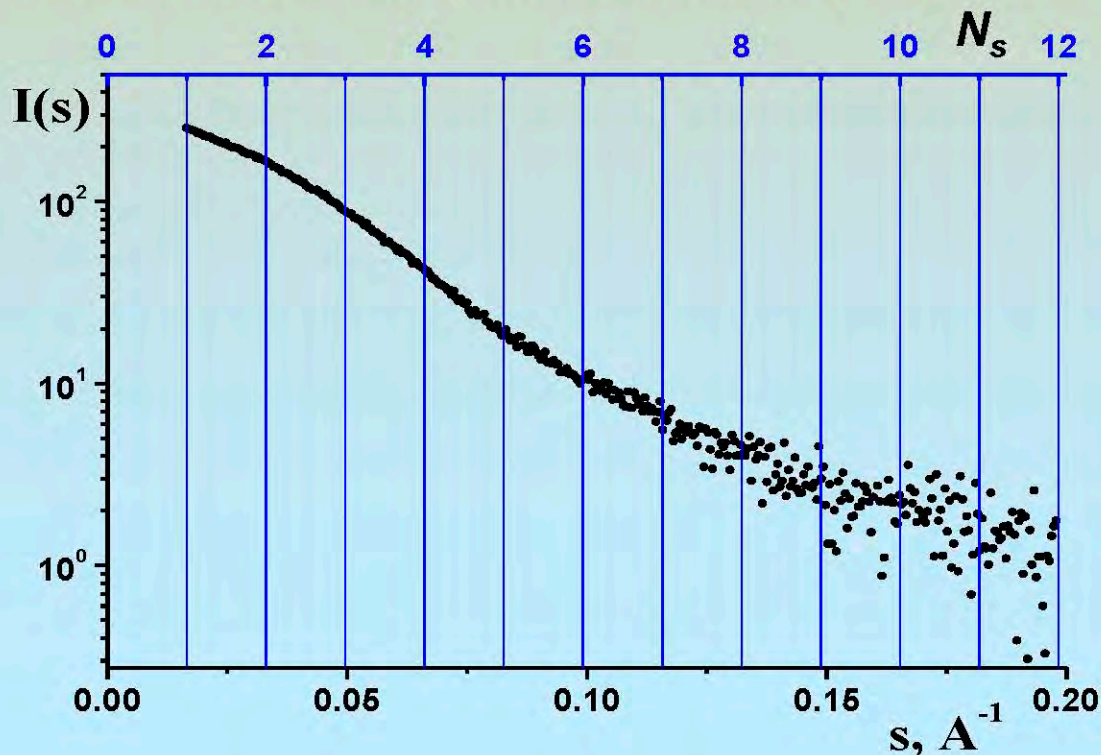
Вывод из теоремы Котельникова-Шеннона:

число независимых параметров, описывающих кривую рассеяния, равен

$$N_s = s_{max} D_{max} / \pi,$$

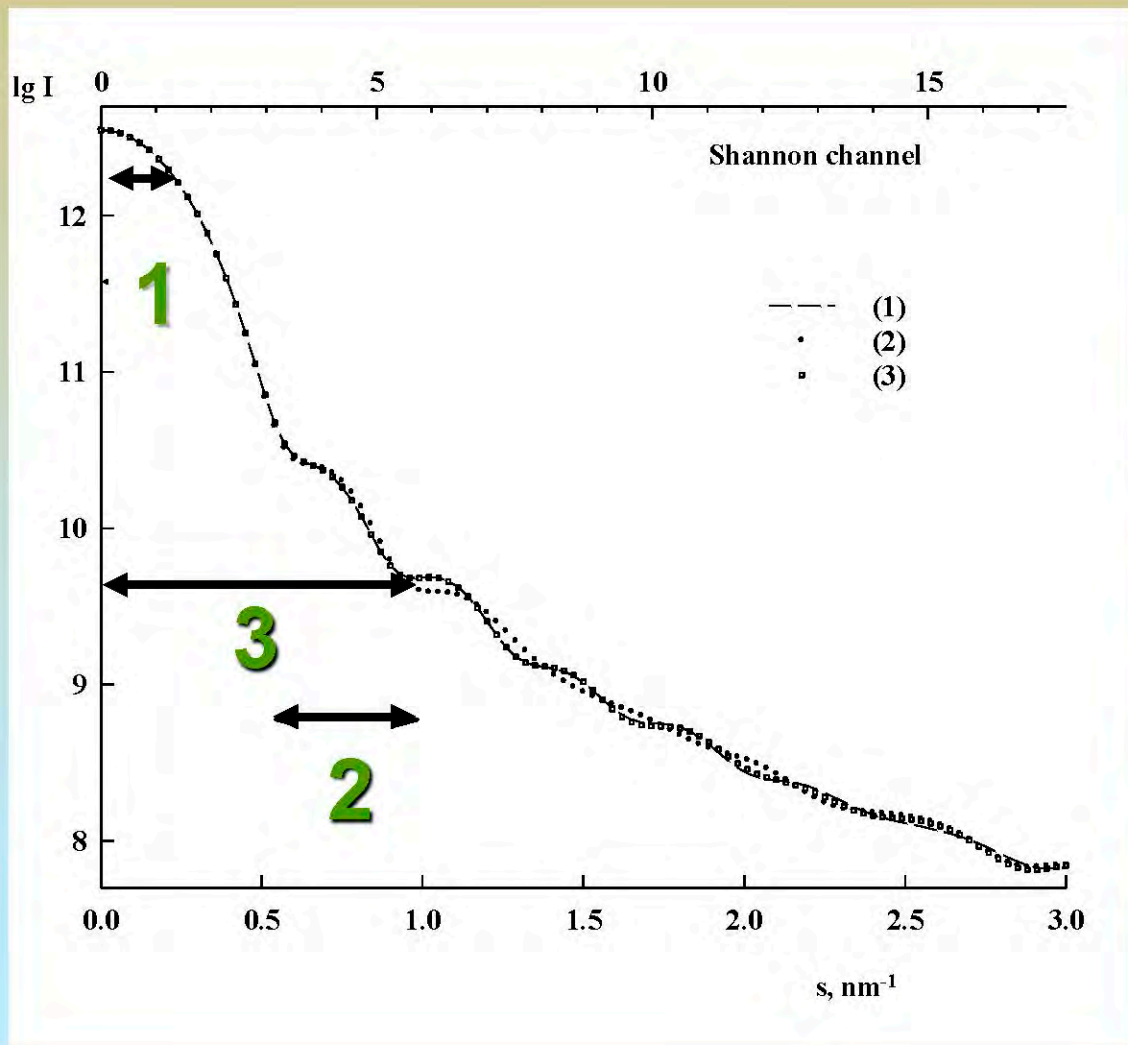
где  $D_{max}$  максимальный размер частицы.

Обычно  $N_s$  невелико:  $10 \div 20$   
 $N_s$  определяет максимально допустимый интервал угловых отсчетов в малоугловом эксперименте:



Следовательно, структура рассеивающей частицы должна быть параметризована небольшим количеством параметров, что ограничивает пространственное разрешение.

# Проверка однозначности и устойчивости решения



Численный эксперимент:  
*ab initio*  
восстановление формы частицы по разным участкам теоретической кривой рассеяния



# Восстановление формы при $L_{\max}=3$ , модельные данные рассеяния

задуманные  
модели

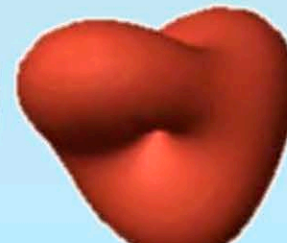
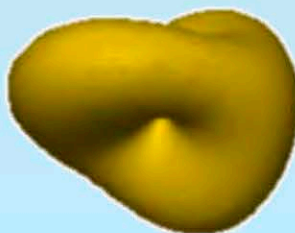
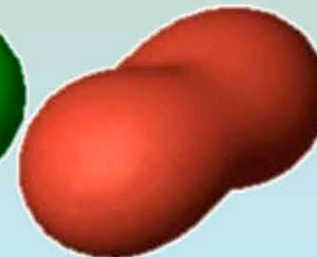
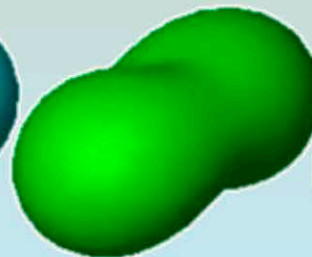
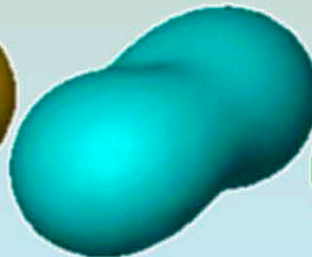
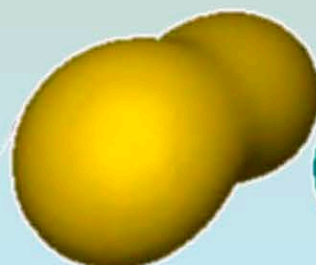
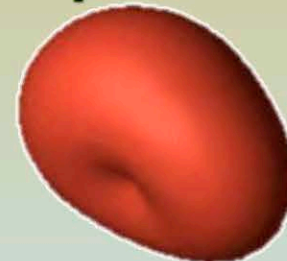
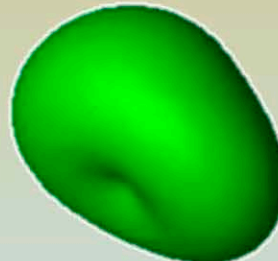
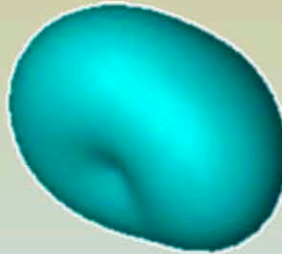
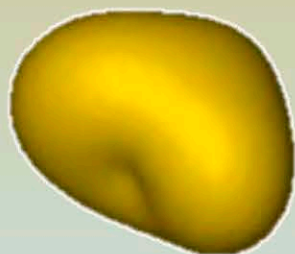
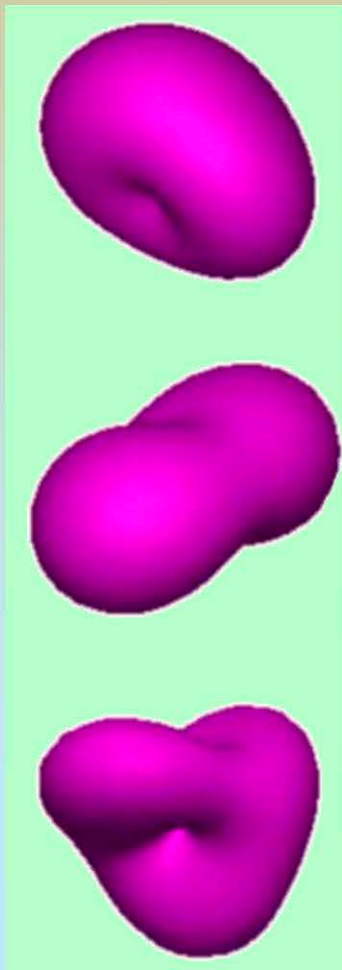
восстановление

1

2

3

вся  
кривая



Восстановление формы при  $L_{\max}=3$ ,  
уровень шумов в данных 1 % отн.

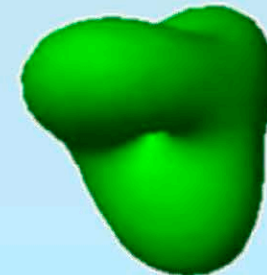
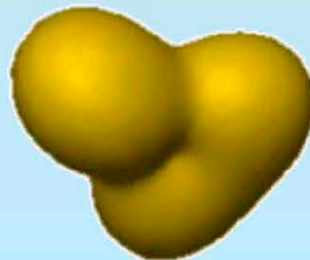
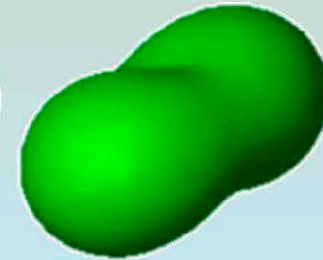
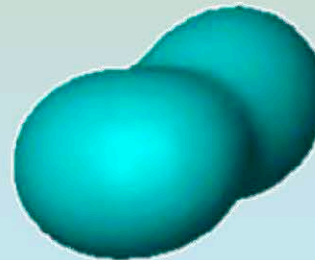
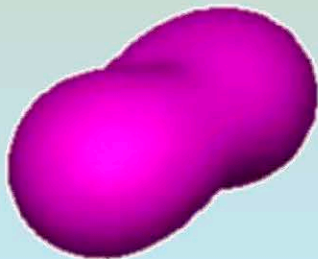
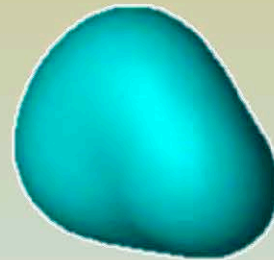
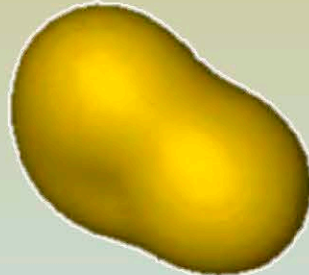
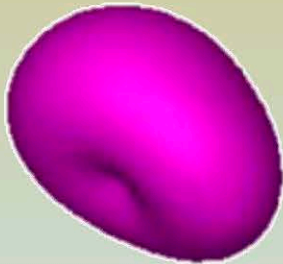
модели

восстановление

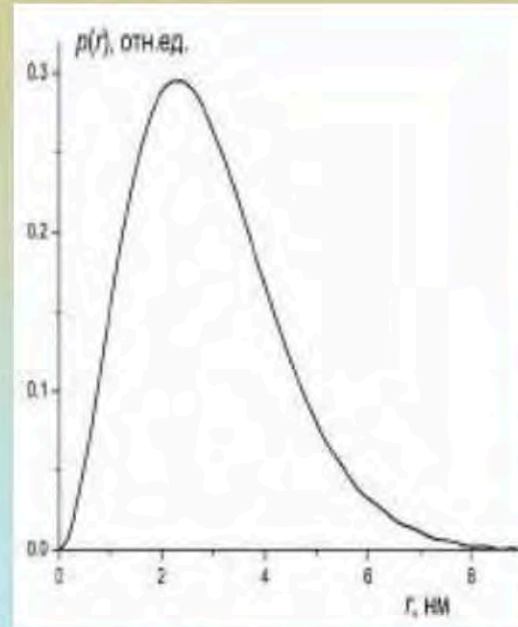
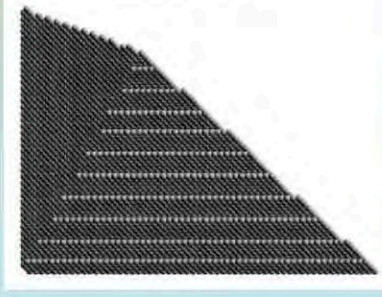
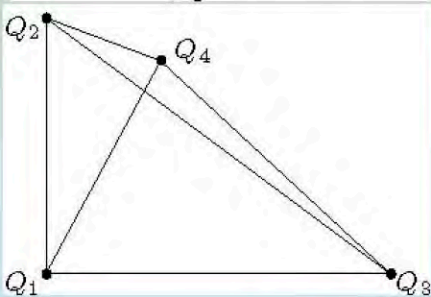
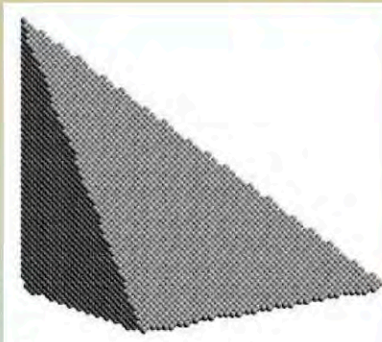
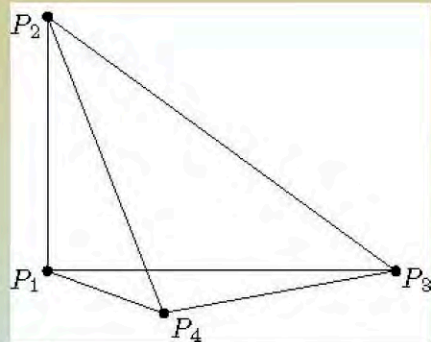
1

2

3



# Проблемы: неоднозначность и неустойчивость решений



**Пример  
неоднозначности:  
одинаковые  
функции парных  
расстояний  $p(r) =$   
одинаковые  $I(s)$**

**Структурные инварианты:**

**Радиус инерции  $R_g$ ,  
объем частицы  $V$ ,**

**максимальный размер  $D_{\text{макс}}$**

**однозначно** определены формой  
кривой рассеяния.

$$p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} s^2 I(s) \frac{\sin(sr)}{sr} ds$$

**Функция парных расстояний  
однозначно** рассчитывается  
по данным рассеяния

# Программа DAMMIN: шариковые модели формы

Программа ищет пространственное расположение шариков, из которых строится модель компактной структуры путем минимизации суммарной квадратичной невязки между экспериментальной и теоретической кривыми рассеяния **методом моделирования отжига**:

$$\Phi(\mathbf{x}) = R^2 [I_{\text{exp}}(s), I(s, \mathbf{x})] + \alpha P(\mathbf{x})$$

где  $R^2$  есть невязка,  $P(\mathbf{x})$  – набор штрафных членов за некомпактность с весами  $\alpha > 0$

**Рассеяние от шариковой модели**

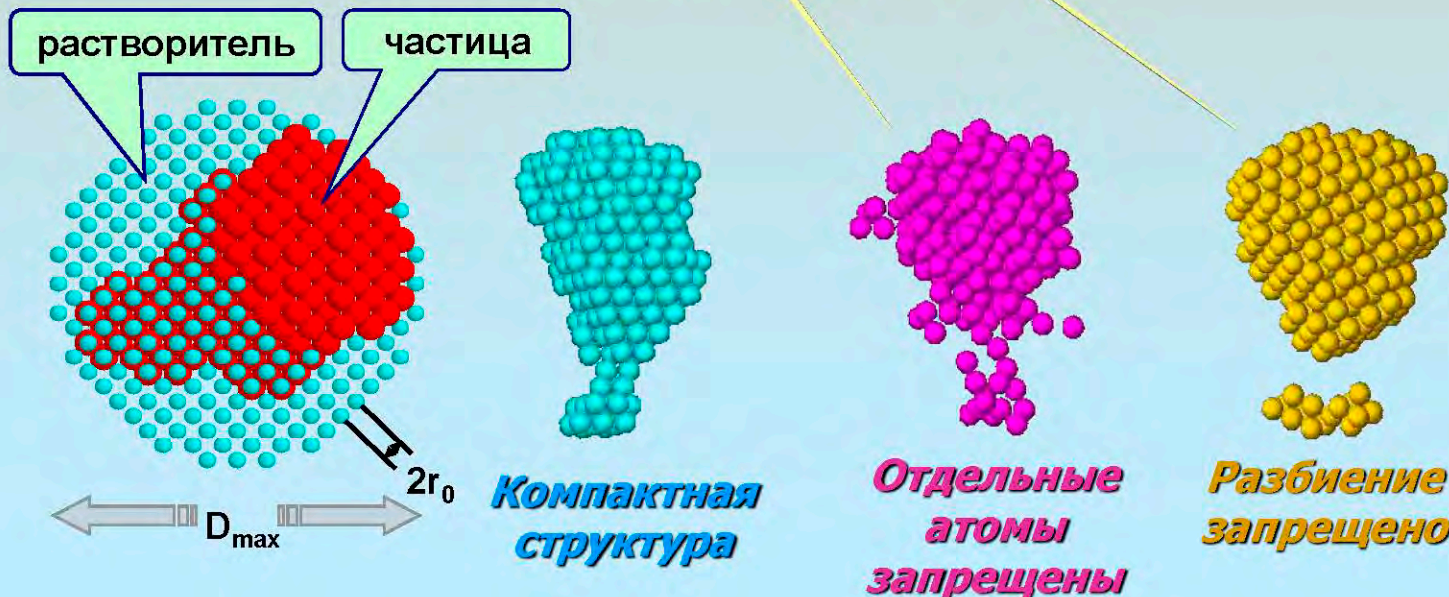
$$I(s, \mathbf{x}) = 2\pi^2 \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s)|^2$$

**Частичные амплитуды**

$$A_{lm}(s) = i^l \sqrt{2/\pi} \sum_j j_l(sr_j) Y_{lm}(\omega_j)$$

**Суммирование амплитуд ведут по всем шарикам, принадлежащим структуре**

На каждом шаге поиска проводится добавление или удаление только одного шарика. Это позволяет быстро пересчитывать рассеяние от всей структуры.



# Строение целевой функции в программе DAMMIN

Минимизируемая целевая функция задачи имеет вид штрафной функции:

$$\Phi(\mathbf{B}) = R^2 + w_D \cdot P_D + w_L \cdot P_L + w_C \cdot P_C + w_G \cdot P_G$$

Разрыв  
тела

Рыхлость  
структуры

Сдвиг из  
начала  
координат

Отклонение  
от радиуса  
инерции

$$R^2 = \left\{ \frac{\sum_{i=1}^N [(I_{\text{exp}}(s_i) - \xi \cdot I_{\text{mod}}(s_i)) \cdot W(s_i)]^2}{\sum_{i=1}^N [I_{\text{exp}}^2(s_i) \cdot W^2(s_i)]} \right\}$$

$$W(s) = \begin{cases} s^n & \text{если } s > s_{\max} [I_{\text{exp}}(s) \cdot s^n] \\ 0.5 \cdot \left\{ s^n + \frac{\max [I_{\text{exp}}(s) \cdot s^n]}{I_{\text{exp}}(s)} \right\} & \text{если } s \leq s_{\max} [I_{\text{exp}}(s) \cdot s^n] \end{cases}$$

(n = 0, 1, 2, 3, 4)

Штраф  $P_D$  вычисляют как отношение общего числа шаров структуры к числу шаров, составляющих домен, образованный непосредственно контактирующими атомами.

$P_L$  вычисляют по формуле  $P_L = 1 - \langle 1 - \exp(-N_e) + \exp(-12) \rangle_N$

где 12 - максимальное число контактов шарика с ближайшими соседями,  $N_e$  - реальное число контактов у данного шарика в структуре,  $\langle \rangle_N$  означает усреднение числа контактов приведенное к одному шарика по всей структуре.

# Программа DAMMIN: определение формы макромолекул и наночастиц по данным МУР

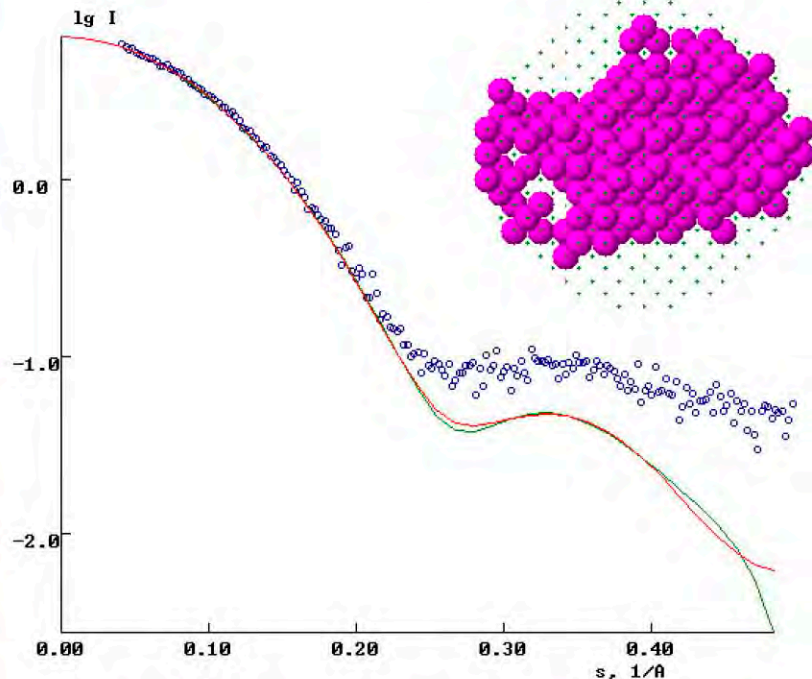
36

dammin43qw

File Edit View State Window Help

Graphics Window

T= 0.316E-04 Rf =0.01440 Los: 0.1126 DisCog: 0.0117 Scale = 0.265E-07



Gnom file : gnolyz.out

12-May-2001 20:04:14

```
a: 276109 CPU: 0.613E+03 SqF: 0.1860
3 Per: 0.4144 Sca: 0.241E-07
a: 303530 CPU: 0.664E+03 SqF: 0.1761
9 Per: 0.4445 Sca: 0.258E-07
a: 331838 CPU: 0.715E+03 SqF: 0.1628
7 Per: 0.4107 Sca: 0.256E-07
a: 361203 CPU: 0.768E+03 SqF: 0.1559
2 Per: 0.4163 Sca: 0.293E-07
a: 391120 CPU: 0.822E+03 SqF: 0.1440
7 Per: 0.4123 Sca: 0.249E-07
a: 421736 CPU: 0.877E+03 SqF: 0.1341
2 Per: 0.3936 Sca: 0.263E-07
a: 452614 CPU: 0.931E+03 SqF: 0.1256
6 Per: 0.3787 Sca: 0.247E-07
a: 484101 CPU: 0.987E+03 SqF: 0.1171
3 Per: 0.4038 Sca: 0.272E-07
a: 516350 CPU: 0.104E+04 SqF: 0.1124
6 Per: 0.3910 Sca: 0.265E-07
a: 549668 CPU: 0.110E+04 SqF: 0.1064
5 Per: 0.3944 Sca: 0.264E-07
a: 583599 CPU: 0.116E+04 SqF: 0.0995
6 Per: 0.3807 Sca: 0.266E-07
a: 617150 CPU: 0.122E+04 SqF: 0.0957
4 Per: 0.3850 Sca: 0.276E-07
a: 651458 CPU: 0.129E+04 SqF: 0.0897
0 Per: 0.3716 Sca: 0.306E-07
a: 686329 CPU: 0.135E+04 SqF: 0.0842
5 Per: 0.3764 Sca: 0.278E-07
a: 721177 CPU: 0.141E+04 SqF: 0.0833
0 Per: 0.3722 Sca: 0.272E-07
a: 756062 CPU: 0.147E+04 SqF: 0.0779
1 Per: 0.3819 Sca: 0.261E-07
a: 791730 CPU: 0.153E+04 SqF: 0.0762
7 Per: 0.3946 Sca: 0.274E-07
a: 827969 CPU: 0.159E+04 SqF: 0.0714
9 Per: 0.3751 Sca: 0.261E-07
a: 864518 CPU: 0.166E+04 SqF: 0.0673
Rf: 0.02357 Los:0.1150 Dis:0.0255 Per: 0.3736 Sca: 0.248E-07
j: 34 T: 0.433E-04 Suc: 10367 Eva: 901587 CPU: 0.172E+04 SqF: 0.0641
Rf: 0.02144 Los:0.1180 Dis:0.0113 Per: 0.3816 Sca: 0.251E-07
j: 35 T: 0.390E-04 Suc: 10367 Eva: 938681 CPU: 0.179E+04 SqF: 0.0607
Rf: 0.01587 Los:0.1120 Dis:0.0113 Per: 0.3965 Sca: 0.251E-07
j: 36 T: 0.351E-04 Suc: 10367 Eva: 976351 CPU: 0.185E+04 SqF: 0.0589
Rf: 0.01959 Los:0.1099 Dis:0.0056 Per: 0.3864 Sca: 0.251E-07
j: 37 T: 0.316E-04 Suc: 10367 Eva: 1014636 CPU: 0.192E+04 SqF: 0.0562
Rf: 0.01440 Los:0.1126 Dis:0.0117 Per: 0.3788 Sca: 0.265E-07
```

Running

# Case study: determination of structure of the myosin head S1

Step 0 Temperature = 0.100E-02 Chi= 36.38

