SFF analysis of a small angle scattering data from phospholipid vesicles systems: online interface and parallel implementation

Bashashin M.V., Zemlyanaya E.V., Kiselev M.A., Lukyanov K.V. Turapbay K.

Joint Institute for Nuclear Research (Dubna, Russia)

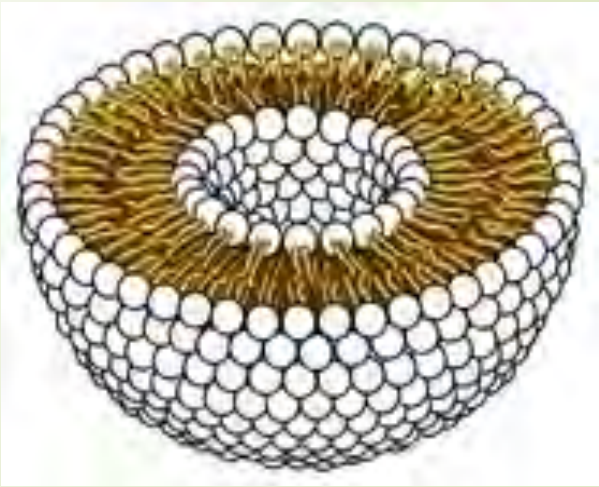
Dubna State University (Dubna, Russia)

National Research Oncology Center (Nur-Sultan, Republic of Kazakhstan)

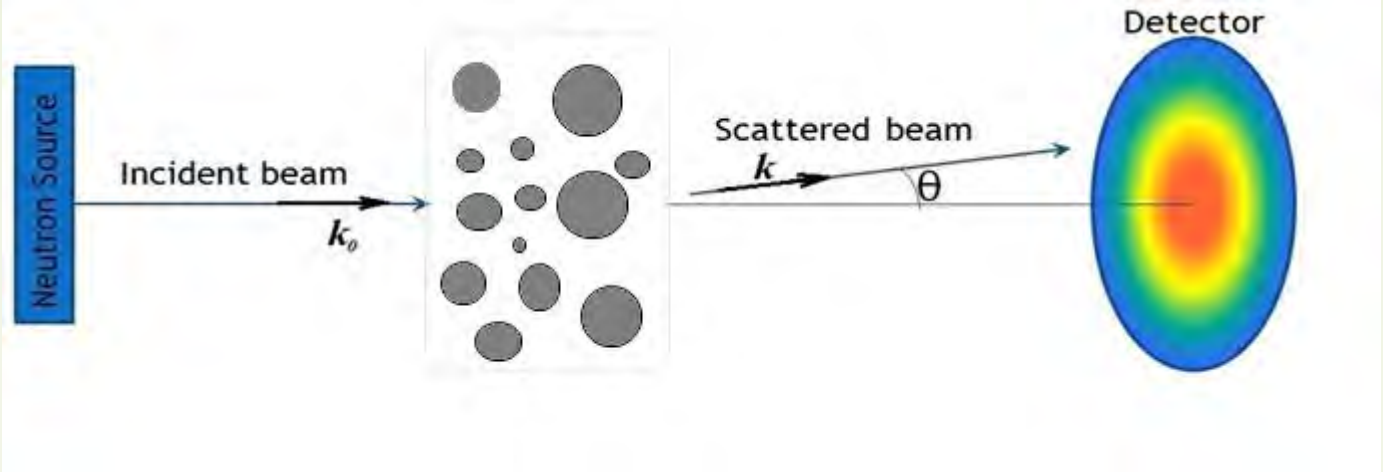
AYSS 2021

Motivation

- Investigation of the structure and properties of phospholipid membranes is an actual problem due to the importance of understanding the basic principles of the organization and functioning of biological and artificial membranes, as well as the prospects for their use in biomedicine and pharmacology.
- In biomedicine, a vesicle is a small structure within a cell, consisting of fluid enclosed by a lipid bilayer. Artificially prepared vesicles are used for transport of drugs.
- One of the effective experimental methods of the phospholipid vesicular systems study is the method of small-angle neutron scattering (SANS). This method allows to determine the size, shape, and also the internal structure of the vesicle bilayer.



Small-angle neutron/X-ray scattering



Small-angle neutron/X-ray scattering (SANS/SAXS) is an experimental technique that uses elastic scattering of neutrons (photons) at small scattering angles to investigate the structure of various substances at a mesoscopic scale of about 1 - 100 nm; the direction of the scattered rays is only slightly (at small angles) deviated from the direction of the incident beam.

The method of separated form factors (SFF) is an efficient method for analyzing data on small-angle neutron scattering for obtaining information on the size and internal structure of vesicular systems in excess water.

Formulation of the problem (SFF). 1

The SFF method is determined by the following formulas:

The macroscopic coherent cross section for scattering of a monodispersed population of vesicles:

$$\frac{d\Sigma}{d\Omega_{mon}}(q) = n \cdot F(q) \cdot S(q)_I \quad S(q,R) = 1 - \frac{8V_v}{v}\left(\frac{\sin(2qR)}{2qR}\right) \tag{1}$$

where q – modulus of the scattering vector, n – number of vesicles per unit volume, $F(q)$ – form factor of neutron scattering by one vesicle, $S(q)$ – structural factor of vesicle population, V_v - volume of vesicles, $v = 1/n$.

For vesicles, with a radius R and a bilayer thickness d_m, under the condition $R >> d_m$, equation (1) can be written with high accuracy in the form:

$$\frac{d\Sigma}{d\Omega_{mon}}(q) = n \cdot F_s(q,R) \cdot F_b(q,d) \cdot \mathrm{S(q)} \qquad F_s(q,R) = \left(4\pi \cdot \frac{R^2}{qR} \cdot Sin(qR)\right)^2 \qquad F_b(q,\ d) = \left(\int_{-d/2}^{d/2} \rho(x) \cdot Cos(qx) \cdot dx\right)^2 \tag{2}$$

where $F_s(q,R)$ – form factor of an infinitely thin sphere of radius R. $F_b(q,d)$ – form factor of a symmetric lipid bilayer of thickness d_m.

Formulation of the problem (SFF). 2

The lipid vesicular system (vesicle population) is not monodispersed, it has a polydispersity. We used the asymmetric Schulz distribution to take into account the polydispersity of the vesicles:

$$G(R) = \frac{R^m}{m!} \cdot \left(\frac{m+1}{\bar{R}} \right)^{m+1} \cdot \exp\left[- \frac{(m+1) \cdot R}{\bar{R}} \right] \tag{3}$$

The macroscopic section $d\Sigma(q)/d\Omega$ has the following form:

$$I_m = \frac{d\Sigma}{d\Omega}(q) = \frac{\int_{R\min}^{R\max} \frac{d\Sigma}{d\Omega}_{mon} (q,R) \cdot G(R,\bar{R}) \cdot dR}{\int_{R\min}^{R\max} G(R,\bar{R}) \cdot dR} \tag{4}$$

The difference between the theoretical I_m and the practical value of the macroscopic cross section $I(q)$ is due to the fact that the resolution function of the spectrometer with respect to q is not a delta function. For the YuMO spectrometer, the resolution function is a symmetric function close to a Gaussian function. Taking into account the resolution function of the spectrometer Δ and the incoherent background IB, we finally have:

$$I(q) = I_m(q) + \frac{1}{2} \cdot \Delta^2 \cdot \frac{d^2 I_m(q)}{dq^2} + IB \tag{5}$$

The program implementation (serial and parallel)

The χ^2 values characterizing an agreement between numerical and experimental data, are calculated as follows:

$$\chi^2 = \frac{1}{N-k}\sum_{i=1}^{N}\left(\frac{I(q_i)-I_{exp}(q_i)}{\delta(q_i)}\right)^2 \quad (6)$$

where $\delta(q_i)$ – experimental data errors; N – number of experimental points; k – number of fits parameters. The program is written in the Fortran programming language.

The analysis of the SANS spectra on the basis of the SFF model was carried out *in the serial mode* using the DFUMIL program from the JINRLIB program library[1] (LIT JINR).

With a large amount of experimental data, computations can take a lot of time, so a parallel MPI-implementation of the SFF approach was developed using PFUMILI[2] – a *parallel version* of the χ^2-minimization.

The parallelization is based on a distribution of the number of experimental points between parallel MPI-processes, with the joining of partial results at each iteration. In our problem, for each experimental point q_i at each iteration in the process of fitting, the theoretical values of $I(q_i)$ are calculated according to the above formulas (2-5), which requires a significant computational load. Therefore, one can expect the acceleration of computations when working in parallel mode, even with a relatively small number of experimental points. It is also clear that the parallel implementation efficiency will increase as the number of experimental points is growing.

1) Dymov S.N., Kurbatov V.S., Silin I.N., Yaschenko S.V. // Nucl Instrum Methods Phys. Res. A. 2000. V. 440. P. 431.
2) A. P. Sapozhnikov, JINRLIB, The second experience of large computational programs parallelization. Parallel version of Fumili program. 2009.

Testing devices

The Govorun supercomputer and the HybriLIT training ground were used for the calculations. The computing node of the Govorun supercomputer is based on two Intel Xeon Platinum 8268 2.9 GHz processors (24 cores). The computing node of the polygon is based on two Intel Xeon E5-2695 v2 2.4 GHz processors (12 cores). Testing was done with Intel MPI compiler.

Methodical calculations 1

Methodical calculations were carried out for the SANS spectra on a polydispersed population of single-layer vesicles water measured at the experimental facilities:

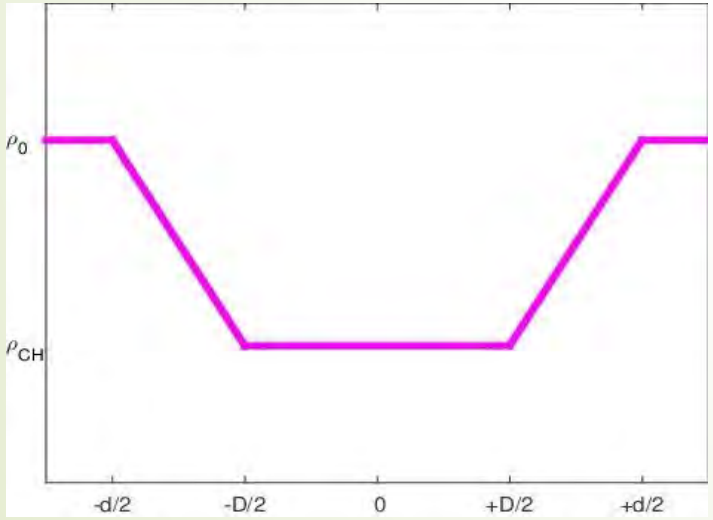
- YuMO small angle spectrometer, JINR, Dubna (60 points), DMPC ULVs in the heavy water
- SANS PSI-1 facility, Paul Scherrer Institute, Villingen, Switzerland

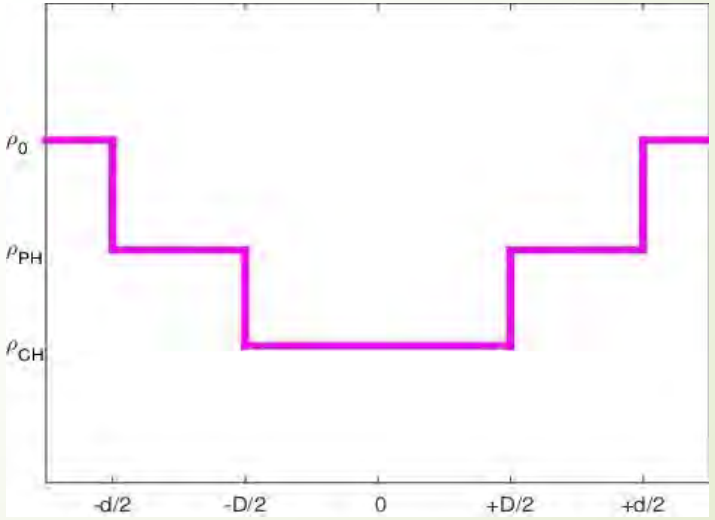
(227 points), DMPC ULVs in the heavy water

- Kurchatov Synchrotron Radiation Source of the NRC “Kurchatov Institute”, Moscow (1115 points), PTNS ULVs in the water solution of the 20% maltose concentration

Methodical calculations 2

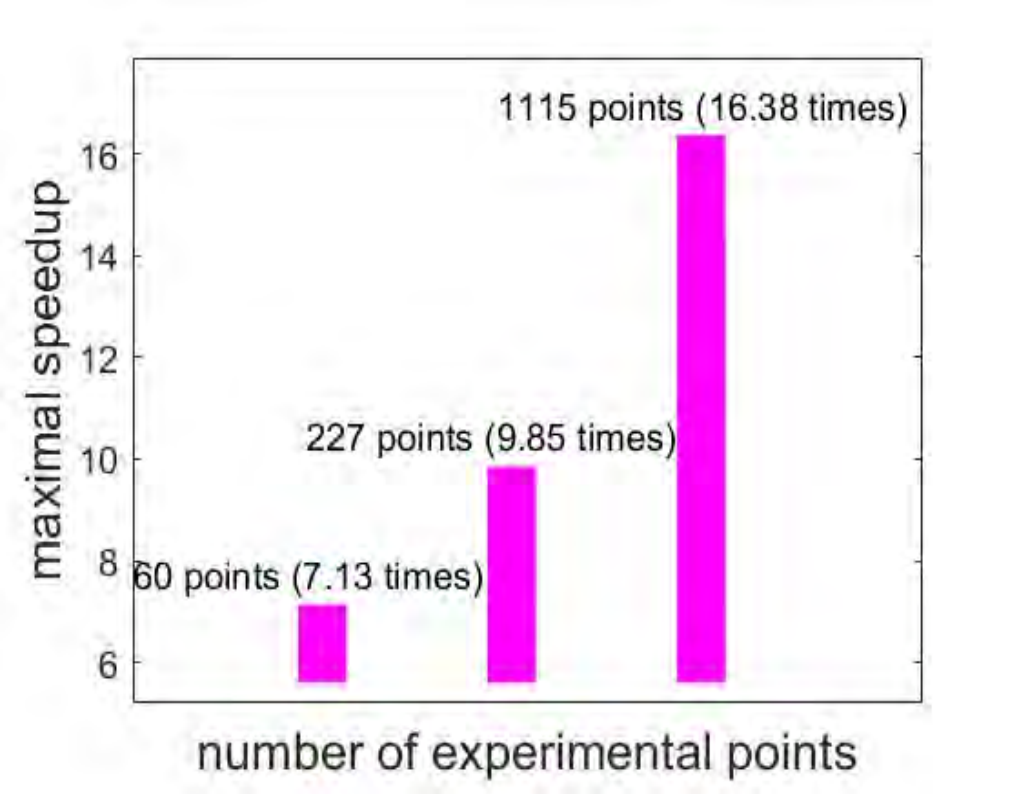
Here, we use two variants of the density function $\rho(x)$. Basic adjustable parameters:
$\bar{R}$ – average radius of vesicles;
m – polydispersity coefficient;
d_b – membrane thickness;
D – thickness of hydrophobic layer;
ρ_{PH} – the polar head group density
(for right figure only).

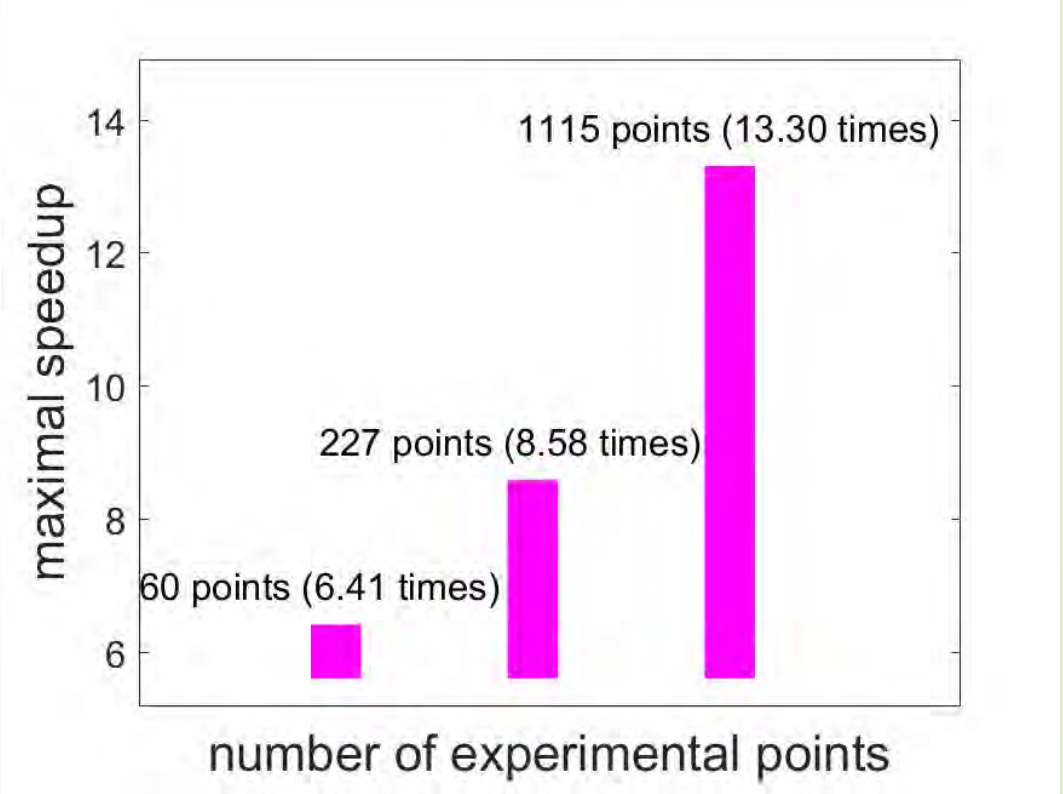




Methodical calculation results: speedup

Maximum speedup with a different number of experimental points:
Left panel Supercomputer "GOVORUN“, right panel HybriLIT education and testing polygon.

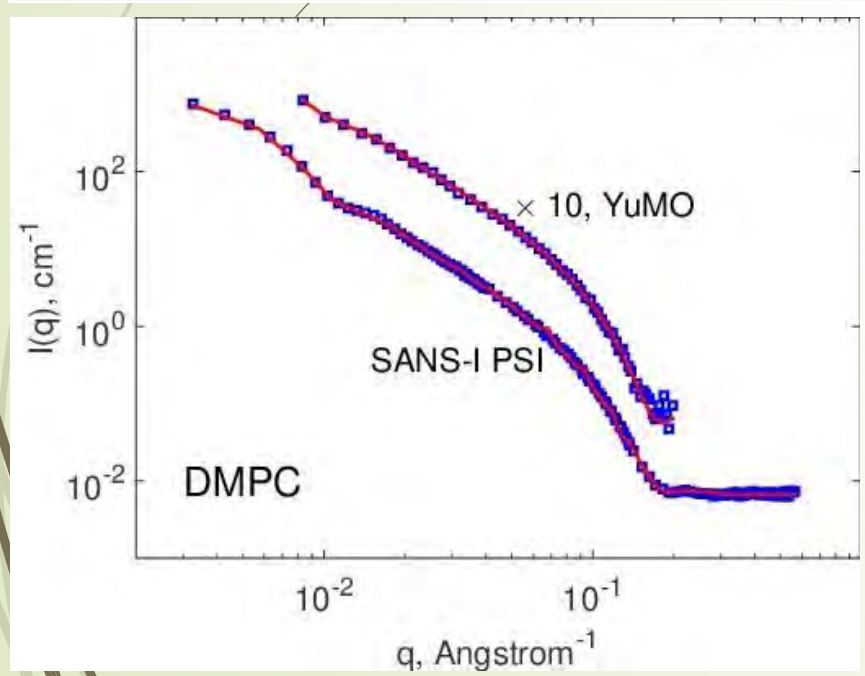


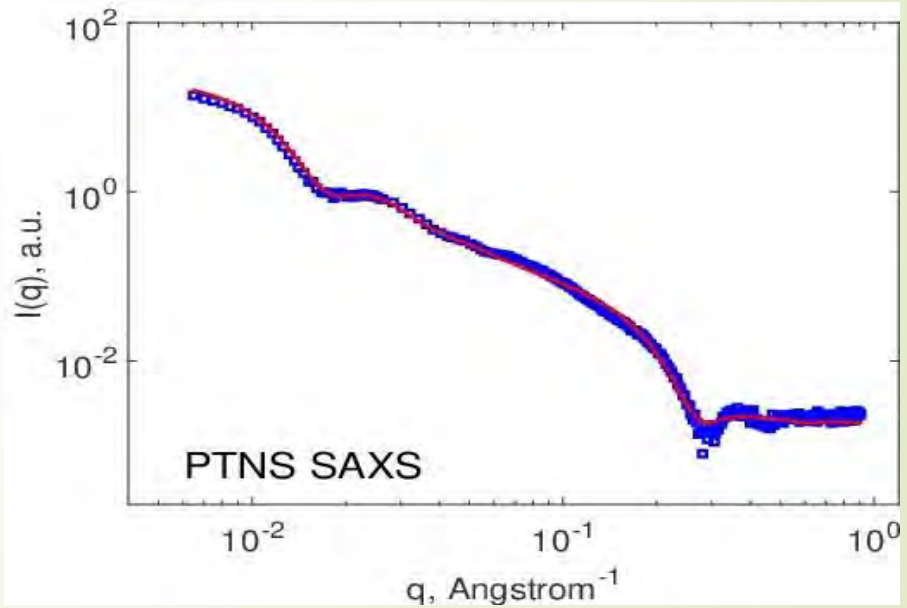


Methodical calculation result: parameters

Parameters of vesicle systems obtained by means of the PFUMILI-based parallel code are consistent with the previously performed serial calculations, which confirms the correctness of the parallel implementation.

Setup	ULV	Maltose	$\bar{R}$, Å	m	d, Å	D, Å	χ^2
YuMO SANS	DMPC	0%	277,0 ± 5,0	9,8 ± 0,7	49,0 ± 2,0	20,0 ± 3,0	1,1
SANS-I PSI	DMPC	0%	275,6 ± 0,5	12,6 ± 0,3	47,8 ± 0,2	20,5 ± 0,4	1,7
KISI SAXS	PTNS	20%	205 ± 3	18 ± 1	44,9 ± 0,4	27,1 ± 0,2	0,7

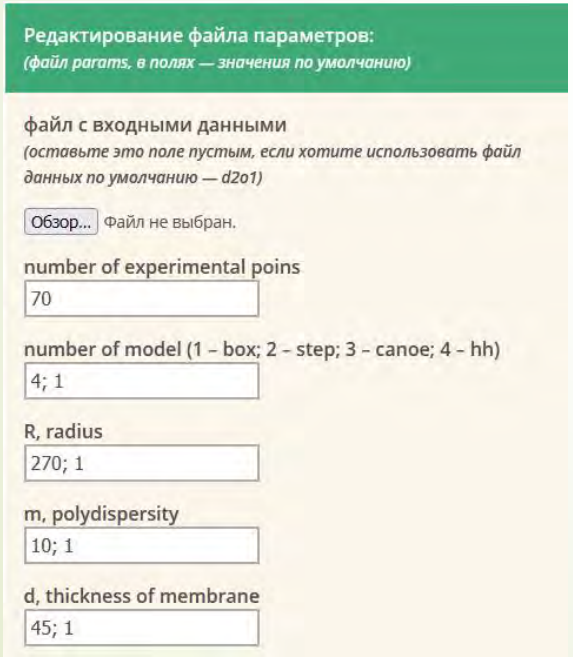




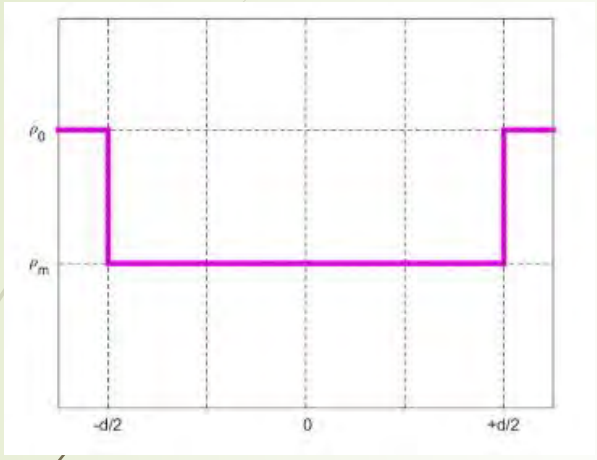
Online interface – data input

To facilitate the RFF-SANS analysis, a Windows-based online interface was created. With the help of this interface, anyone can make calculations using an Internet browser as the only necessary means of working with the program.

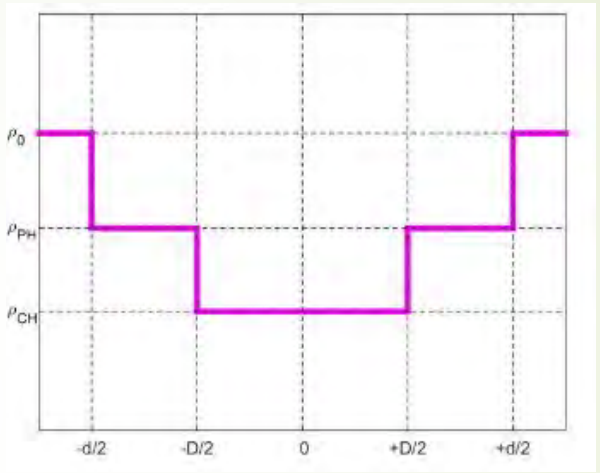
The user needs to fill in the required fields to start the calculation (the figures show the default parameters). We download the experimental points file, choose the bilayer model, introduce the start values of parameters to be fitted.



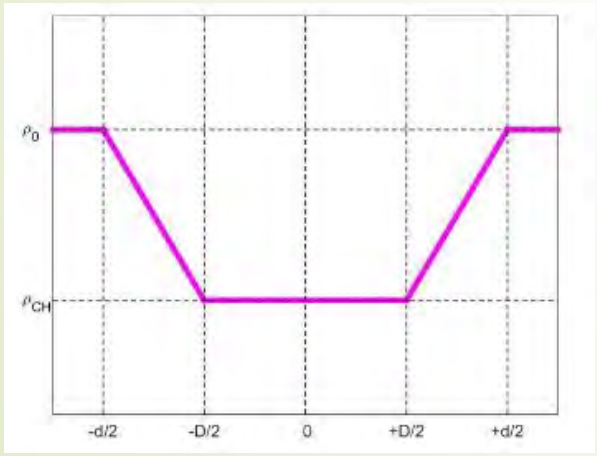
Online interface – The bilayer scattering length density models



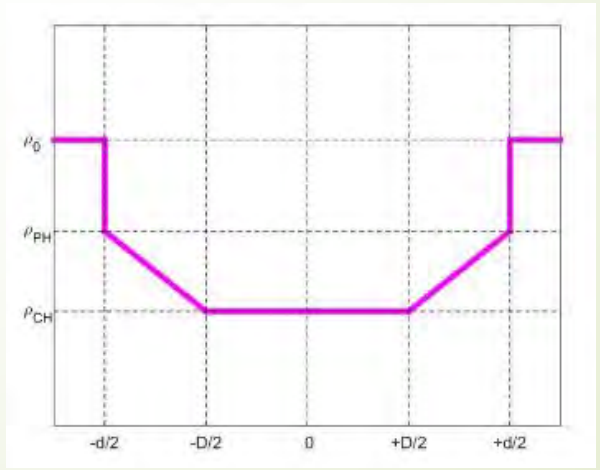
1 - box model



2 - step model



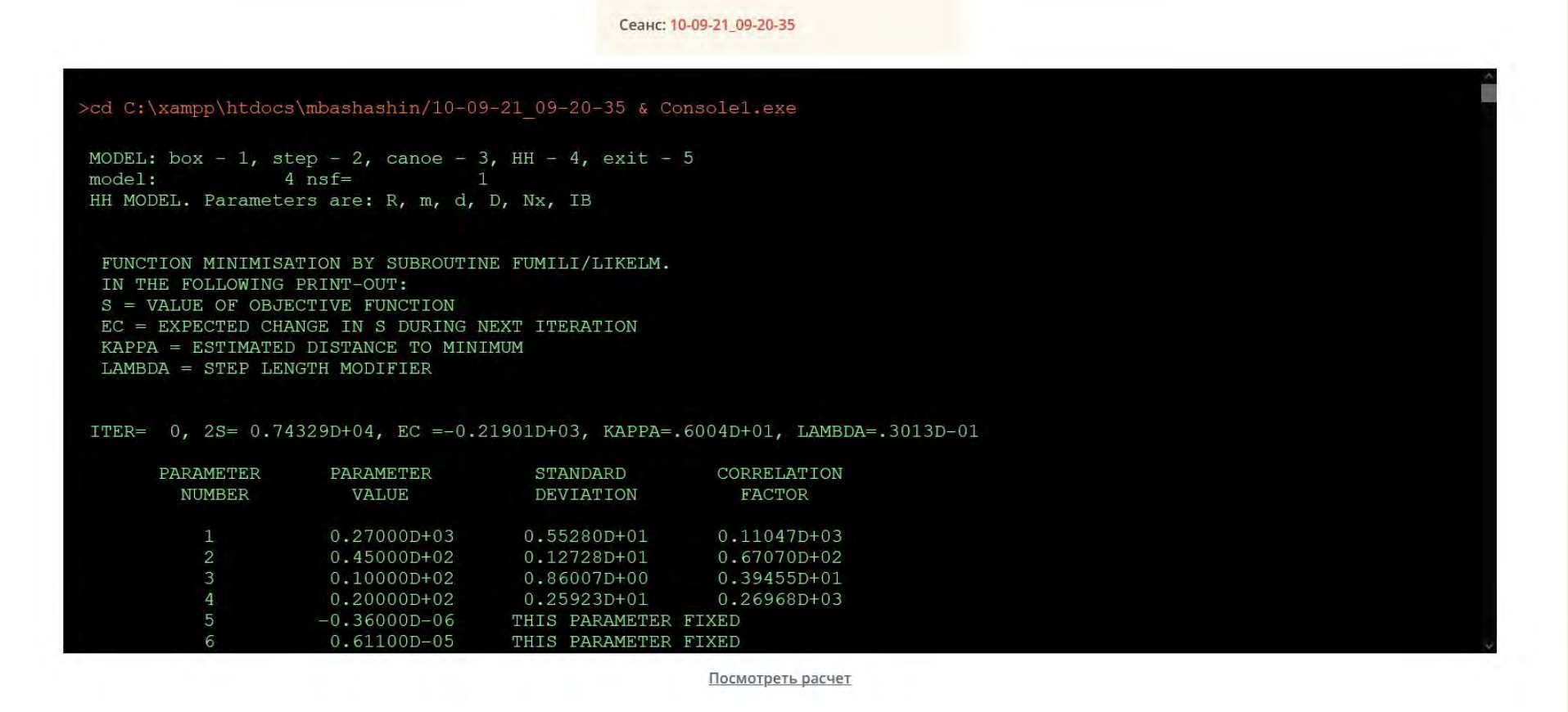
3 - hh- model



4 - canoe model

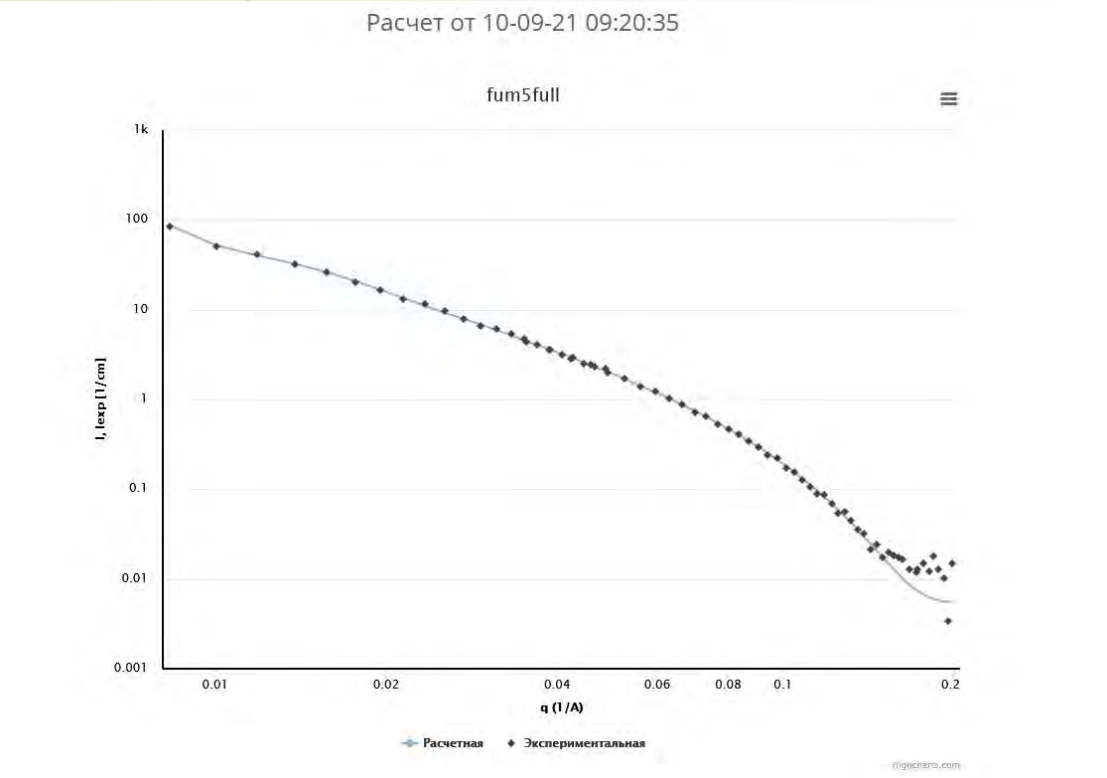
Online interface – output

After performing the calculation, the user can familiarize himself with the listing of the program's work results, which, in addition to technical information, displays the intermediate calculation results.



Online interface - vizualization

By pressing the button "view the calculation" the user can see the resulting graph as a result of the calculation, as well as download the input and output parameters in the form of files and a file with the calculated data.



Input params (скачать)

```
'd2o1'   %       file of experimental poins
70       %       number of experimental poins
4        1       %       number of model (1 - box; 2 - step; 3 - canoe; 4 - hh)
270      1       %       R, radius
10       1       %       m, polydispersity
45       1       %       d, thickness of membrane
1.5e+14  1       %       nx, number of vesicles per unit volume
.00546   0       %       IB, incoherent background
20       1       %       D, internal thickness (model 2,4 )
6.37e+10         0       %       rho 0 (D20)
'data_out'       %       file for data output
'params_out'     %       file for parameters output
```

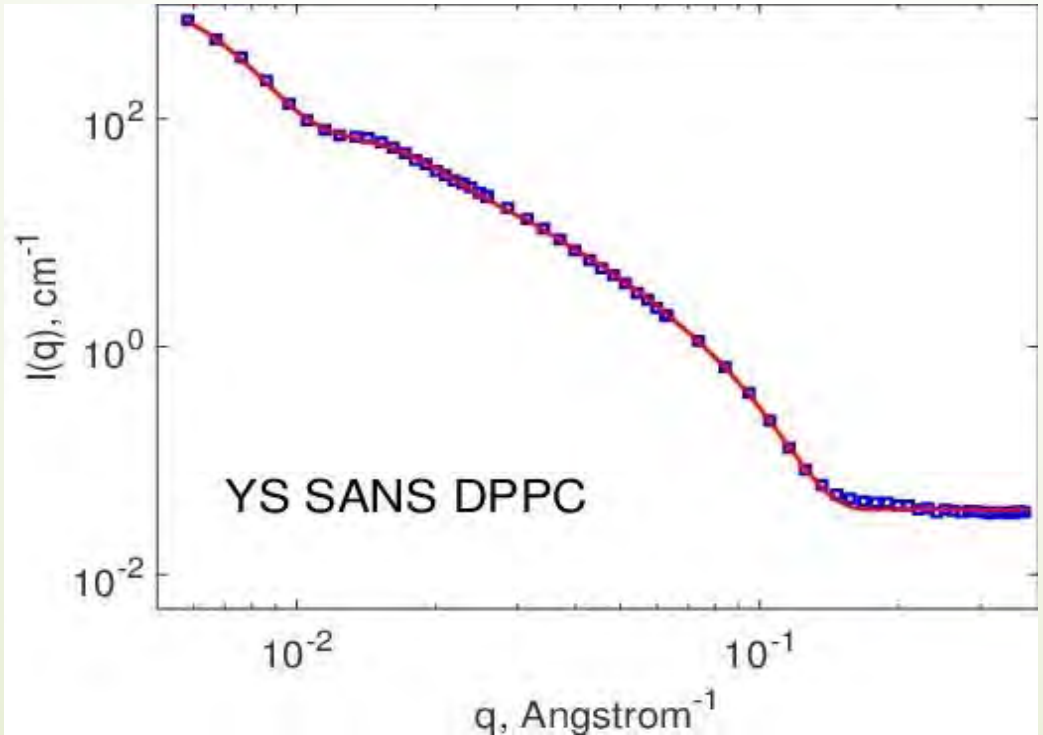
Output params (скачать)

```
 Fitting rezults. HH model. Data file is d2o1

 Number of experimental points:          70

 Paramewters  and errors
 R (A), d (A) ,m, D (A), nx (1/cm^3),IB (1/cm)
   283.674493180777        3.80865697369436
   49.9918682681747       0.780379288557045
   10.3592173296728       0.645023464088385
   12.3141954631474        2.74484599688819
   200517995981905.        10181095639711.3
  5.459999966817008E-003  0.000000000000000E+000

  residual is   1.661050971536448E-004
 Xi^2=   1.71209390410236



 all parameters of fumily
          1   283.674493180777        3.80865697369436
          2   49.9918682681747       0.780379288557045
          3   10.3592173296728       0.645023464088385
          4   12.3141954631474        2.74484599688819
          5 -3.600000013648241E-007  0.000000000000000E+000
          6  6.109999958425760E-006  0.000000000000000E+000
          7  2.005179932050462E-010  1.018109549871918E-011
          8  5.459999966817008E-011  0.000000000000000E+000
          9   63700000000.0000       0.000000000000000E+000
```

An example of using the interface - calculating the parameters of the DPPC vesicular system

The structure of the polydispersed population of DPPC ULVs in heavy water was analyzed on the basis of SANS data obtained from a Yellow Submarine (YS) small-angle spectrometer, Budapest, Hungary. It can be seen from the figure that the theoretical RFF curve reproduces the experimental data well.

Setup	ULV	Maltose	$\bar{R}$, Å	m	d, Å	D, Å	χ^2
YS SANS	DPPC	0%	261,2 ± 1,6	14,1 ± 0,5	55,0 ± 2,4	20,0	3,5



Summary: conclusions and plan

- We have found that the PFUMILI-based parallel implementation of the SFF-parameters fitting procedure can provide the 6-15 times speedup in comparison with the serial calculations.
- The online interface has been developed for the work with the Fortran program that implements the adjustment of the structural parameters of polydispersed vesicular systems to the SANS experimental data on the basis of the SFF approach. This tool can provide a convenient work with the SFF method to a wide range of users, including students of relevant specialties, as well as researchers who do not have programming skills and do not have the appropriate software at their computers.
- Since the computational capabilities using the developed interface are limited, the further plane is a create an information system based on the resources of the HybriLIT platform that implements high-performance processing of small-angle scattering data for analyzing the structure of vesicular systems of various chemical compositions based on the RFF method. The possibility of expanding the functionality of the information system will be implemented by including newly developed modules that implement new models and methods for modeling vesicular systems;

Thank you for the attention