

Assessment of structural changes in proteins and surrounding water molecules in solution according to SAXS and MD data

Alexander V. Smirnov^{1,a}, Andrei M. Semenov^{2,3}, Yuri B. Porozov^{4,5}, Boris A. Fedorov¹

¹ITMO University, Saint Petersburg 197101, Russia

²Alferov University Russian Academy of Sciences, Saint Petersburg 194021, Russia

³Voronezh State University, Voronezh 394018, Russia

⁴Sechenov First Moscow State Medical University, Moscow 119991, Russia

⁵Sirius University of Science and Technology, Sochi 354340, Russia

^asmirnav_2@mail.ru

Corresponding author: A.V. Smirnov, smirnav_2@mail.ru

PACS 61.05.cf, 71.15. Pd, 87.15.B-

ABSTRACT The SASPAR program for calculation of SAXS of proteins in solution uses trajectories of molecular dynamics (MD) and an explicit solvent model. The program allows one to take into account real interactions of solvent molecules both between each other and with the protein molecule. The previously developed SAS-CUBE program (the “cube method”) is also used, it assumes that the protein structures in crystal and in solution coincide, and the water surrounding the proteins is considered as a homogeneous continuum. Using these programs, SAXS curves were calculated for 18 proteins of different molecular weights and then compared with one another and with the corresponding experimental scattering curves. “Vacuum” SAXS curves (i.e., without taking into account the surrounding water) were also calculated for each protein for two approaches: a) based on the coordinates of protein atoms in crystal and b) based on the coordinates of protein atoms for each MD frame with further averaging of the intensities from all the frames. 1) It was shown that for the 14 single-domain proteins considered, the “vacuum” scattering curves calculated by two methods coincide well for almost each protein. Hence, the structure of the studied proteins in a solution is similar to their structure in a crystal and, therefore, the presence of the surrounding water molecules does not alter the protein structure itself significantly. The SASPAR- and SASCUBE-curves coincide well only in two cases (i.e., water is only slightly structured near the protein surface), but in the other cases these curves are markedly different, which indicates the structuredness of the water near the protein surface, although to a different extent. 2) It was shown that for the 4 multi-domain proteins considered, their “vacuum” scattering curves, calculated with the two methods indicated above, differ noticeably, which is an evidence that their crystalline and “water” structures are different. It was also shown that the most of the calculated curves coincide well with the experimental ones.

KEYWORDS small-angle X-ray solution scattering, molecular dynamics, protein structure in solution, water structure.

ACKNOWLEDGEMENTS The authors are grateful to Dr. Alexander Grishaev for providing the experimental SAXS scattering curves of the GB3 protein and to the Master’s student K. S. Mazokha for the assistance with research.

FOR CITATION Smirnov A.V., Semenov A.M., Porozov Yu.B., Fedorov B.A. Assessment of structural changes in proteins and surrounding water molecules in solution according to SAXS and MD data. *Nanosystems: Phys. Chem. Math.*, 2022, **13** (3), 274–284.

1. Introduction

Studies of the proteins structure in a solution is known to be one of the most relevant tasks of molecular biophysics. Among various methods of this research, small-angle X-ray and neutron scattering play an essential role. Numerous works presented over the past fifty years have been dedicated to the development of algorithms that allow one to calculate scattering intensities of proteins in solution based on the information on their structure in crystal. The works of the Luzzati school [1], in which the internal volume of proteins was modelled by closely packed large-sized cubes with a 3Å edge, are considered pioneering in this field. However, in works [2, 3], the edge size was ten times less and the developed “cube method” allowed one to obtain an agreement of calculated curves with quite a few experimental scattering curves of proteins in solutions that was acceptable that time. The poor performance of computers significantly limited the application for this method. Today, the above-mentioned limitations are virtually eliminated. The speed of modern

computers dramatically exceeds the capabilities of the computers from 1970–1980s. Moreover, advances in protein extraction and purification together with the progress in X-ray equipment (primarily, the use of synchrotron radiation) allow one to create a vast database of experimental scattering curves of proteins in solution (SASBDB [4] and BioSis [5] databases). Recently, a new program recreating the “cube method” was presented by [6] The program uses the method which is considerably modified, better justified from the physical point of view and requires essentially less computation time (see [7]).

An important step in the development of this area was made in work [8]. The authors presented the CRY SOL program based on the assumption that there is homogeneous layer of water near the protein surface. The layer electron density differs from the electron density of water in which the protein is immersed. This program uses the method of expansion of exponential functions in spherical harmonics. This is an effective approach for the calculation of spherically symmetric pattern of scattering by free-oriented particles, as it considerably reduces the computation time. It was shown that the CRY SOL program with two adjustment parameters provides a good description of experimental scattering curves for many proteins in solution. This method with a number of different modifications has become widely used in the world ([9–12]).

Until recently, almost all works on improved methods of calculation of scattering intensity of protein in solution were based on the assumption that the structure of the protein in solution is identical to its structure in crystal. Therefore, the chi-square factor was chosen as a quality criterion for the newly developed programmes. It allows one to assess the degree of similarity between the calculated curves based on this assumption and the experimental scattering curves. However, a question arises: to which extent the specified assumption is correct. During the transition of protein into solution, structural changes may occur in the macromolecule itself caused by its interaction with the solvent molecules, as well as changes in the mutual arrangement of molecules of the solvent itself, especially in the areas close to the surface of the protein. These changes can significantly influence on the shape of the scattering curve.

For this reason, works [13–18] can be considered as the next step in the development of calculation methods for the scattering curves of proteins in solutions. Based on the methods of molecular dynamics, these studies consider real interactions between the solvent molecules themselves and with the protein atoms. Using MD allows one to monitor the changes in the protein structure and in the position of the solvent molecules over time, *i.e.*, to obtain a set of coordinates of protein atoms and solvent molecules (frame) at any chosen moment of time. The rigorous formulas presented in works [18] and [13] allow one to calculate the scattering intensity for each frame for specified protein conformation and specified positions of the solvent molecules. These formulas also take into account the free rotation of the whole system in relation with the primary X-ray beam. The last step is averaging the scattering curves over all the chosen frames. Compared with the previously presented methods, the method developed in [18] and later applied in [13] in WAXSIS server is evidently much closer to the realistic description of scattering intensity of proteins in solution.

The present work proposes a new program (SASPAR, see [19]) based on the same principles of joint use of the MD simulations results and strict formulas for the calculation of scattering intensities of proteins in solution. However, the program algorithm, which uses a parallelepiped as the volume containing the protein and the surrounding water molecules, significantly distinguishes the program SASPAR from the WAXSIS algorithm [13]. Another key difference is the method of averaging the scattering intensity in the reciprocal space. Nevertheless, we should note that for a few proteins presented in work [18], the scattering intensities calculated with WAXSIS and SASPAR methods are almost the same.

The further discussion is entirely based on the results of the SASPAR and SASCUBE programs with the use of numerous experimental scattering curves adopted mainly from the SASBDB [4].

Thus, we use the two programs that allow one to calculate the scattering intensity of protein in solution. The first one (SASPAR), based on the MD simulation data, takes into account the mobility of the protein itself and the mobility of the water molecules surrounding the protein. The second program (SASCUBE) considers the protein structure to be fixed, while the water surrounding the protein is regarded as a medium with homogeneous electron density. Further, “vacuum” SAS curves (without taking into account the surrounding water) were calculated for each protein for two approaches: a) based on the coordinates of protein atoms in a crystal from the Protein Data Bank (PDB [20, 21]), and b) based on the coordinates of protein atoms for each MD frame with further averaging of scattering intensities over all the frames.

The main goal of the present work is to answer the following two questions through the comparison of the above-mentioned scattering curves: a) to what extent does the protein structure change during its transition from crystal to solution, and b) to what extent does the structure of water surrounding the protein differ from the homogeneous continuum.

The logic of the presented discussion is as follows. Using the two programmes listed above, the scattering intensities of each protein in solution are calculated, as well as “vacuum” intensity for the crystal structure and “vacuum” intensity averaged over all frames. Then, the comparison of two corresponding “vacuum” curves and two curves obtained with SASPAR and SASCUBE programmes is performed in order to achieve the following conclusions.

1. If the “vacuum” curves are rather similar and the curves obtained with the SASPAR and SASCUBE programs are similar, then a) the protein structures in crystal and in solution are similar and b) the structure of water near the protein surface is close to the homogeneous continuum (protein of type I).

2. If the “vacuum” curves are rather similar while the curves calculated using the SASPAR and SASCUBE programmes are markedly different, then a) the protein structures in crystal and in solution are similar and b) the structure of water near the protein surface is considerably different from the homogeneous continuum (protein of type II).

3. If the “vacuum” curves are markedly different and the curves from the SASPAR and SASCUBE programs also differ substantially, then a) the protein structures in crystal and in solution are essentially different and b) the structure of water near the protein surface can differ from the homogeneous continuum or cannot (protein of type III).

The chi-factor is used for the quantitative assessment of the similarity of both scattering curves and protein structures.

This work considers 18 proteins for which experimental intensities in solution are available in literature. Each protein was analysed as described above. All scattering curves obtained from the SASCUBE and SASPAR programs were also compared with the experimental scattering curves.

2. Calculation methods

2.1. Molecular dynamic simulation

The initial configurations of proteins are from the PDB. The time step of the simulation was 2 fs, the total simulation time was 10 ns, the coordinates were recorded every 10 ps. The simulation was performed at a temperature of 298 K and a pressure of 1 atm. As a solvent, water from the TIP4P model [22] was used. According to [18], this model provides the best balance between speed and accuracy of the calculation. The simulation was performed in an orthorhombic cell of such size that the minimal distance from the protein to the cell boundary is 2.3 nm in all directions. This gap was selected so that the protein stays inside the cell throughout the entire simulation and yet the time of calculation does not grow due to the excessive number of solvent molecules in the cell. The MD simulation for the solvent without protein was carried out in a cubic cell of size 9.0×9.0×9.0 nm. This volume was sufficient to isolate the required volume of pure solvent when calculating the scattering intensities for all proteins. The authors used periodic boundary conditions to correctly describe the aqueous environment of a protein. However, the use of periodic boundary conditions can lead to the interaction of the protein with itself (the “head” of the protein may interact with its “tail”). To eliminate this effect, large enough cells were used.

The simulation was performed using the Desmond program from the software package Schrödinger Biologic Suite 2017-3 [23] and Gromacs 2019-3 software packages [24].

2.1.1. Molecular dynamic simulation in the Desmond program. The simulation was performed in the Desmond program using the OPLS2005 force field [25]. During the simulation, the NPT isobaric-isothermal ensemble was used together with the Nose-Hoover thermostat [26] and the Martyna-Tobias-Klein barostat [27]. Before the simulation, the system relaxation was carried out according to Desmond’s embedded protocol. All settings, except for the simulation time and temperature, were not changed, and the default settings of Desmond were used.

2.1.2. Molecular dynamic simulation in the Gromacs 2019-3 package. In the case of multi-domain proteins with a relatively high molecular weight, the Gromacs 2019-3 package was used, because there are difficulties with large-cell calculations in the Desmond program. In the Gromacs package, molecular dynamics modeling was performed with the AMBER99SB-ILDN force field [28]. The Nose-Hoover thermostat [26] and the Parrinello-Rahman barostat [29] were used to maintain the temperature and pressure. Before the simulation, the relaxation and equilibration of the system was performed in the NVT isochoric-isothermal ensemble as recommended by [30]. Configuration and command files are published at [31] (for proteins) and at [32] (for water solvent).

2.2. Programmes for calculation of scattering intensities of proteins in solutions

2.2.1. SASPAR. The SASPAR program is designed to calculate the scattering intensities of proteins in solutions taking into account the MD simulation data. The protein structure in crystal from the PDB was used as the initial structure. The coordinates of hydrogen atoms were not considered, and their presence was introduced by adding hydrogen electrons to the nearest non-hydrogen atom. The same is true for the water molecule: only the coordinates of oxygen, to which 10 electrons are attributed, were recorded.

2.2.1.1. Geometrical construction of the scattering system. In each frame, the minimum and maximum values of the protein atoms coordinates are selected for each coordinate axis, and the average coordinate, equal to the arithmetic mean of these two values, is determined. A set of three average coordinates is considered the centre of the molecule. A right parallelepiped whose faces are perpendicular to the coordinate axes is circumscribed around the protein molecule. The parallelepiped dimensions are chosen so that the distance from any atom of the protein to the parallelepiped faces is no less than d . This procedure is repeated for all the frames; the sizes of the edges along each coordinate axis are compared, and the maximum edge length is chosen for all the parallelepipeds: A – along x-axis; B – along y-axis; C – along z-axis. The volume of the parallelepiped constructed on the selected edges (ABC parallelepiped) is obviously larger than the volume of the parallelepiped from any frame. Then, the ABC parallelepiped is cut out from each frame so that its centre coincides with the centre of the protein molecule. This parallelepiped, with all the atoms contained in it, is translated until the indicated centre coincides with the coordinate origin.

It is necessary to know the scattering from the protein-free solvent to calculate the intensity of a protein in solution. For this purpose, the MD simulation of the solvent is considered. As before, the ABC parallelepiped is cut out from each frame, and the average electron density (ρ_0) of the solvent is calculated in the volume of each parallelepiped and then averaged over all frames.

The above-mentioned value d is highly important for the determination of the ABC parallelepiped size. It determines the size of the protein border layer, outside of which water no longer influences on the scattering curve of a protein in a solution. In other words, at distance d or farther from the protein surface, the real structure of water can be replaced with a homogeneous continuum with electron density ρ_0 . This value is assessed in [13] as $d \sim 0.7$ nm, which we adhere to in this work.

Thus, two sets of frames are considered as the basis for the calculation of scattering intensity of a protein in solution. Each frame of the first set is represented by the ABC parallelepiped filled with protein and water molecules. Outside the parallelepiped, there is structureless water (homogeneous continuum) with electron density ρ_0 . The centres of the parallelepiped and the protein are located in the coordinate origin. Each frame of the second set is represented by the same ABC parallelepiped that is filled only with water molecules. Its centre is located at the coordinate origin. It is also surrounded by structureless water with the electron density of ρ_0 . (Fig. 1b).

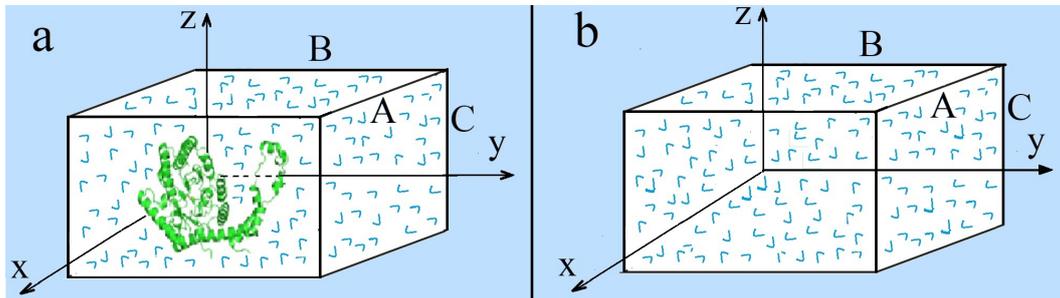


FIG. 1. The ABC parallelepiped cut from the cell with MD simulation: a) for solvent with protein; b) for pure solvent. The solvent outside the parallelepiped is considered a homogeneous medium

2.2.1.2. Calculation of scattering intensity. The scattering intensity $I(\mathbf{q})$ of a protein in a weak solution can be considered as the difference between the scattering intensities of the protein solution and the solvent. In the most general form, based on the model presented above, this intensity $I_{SP}(\mathbf{q})$ can be presented as follows:

$$I_{SP}(q) = \left\langle \left| \sum_k f_k(q) \exp(i\mathbf{r}_k \mathbf{q}) - \rho_0 P_0(\mathbf{q}) \right|^2 \right\rangle_{Fr, \Omega} - \left\langle \left| \sum_l f_l(q) \exp(i\mathbf{r}_l \mathbf{q}) - \rho_0 P_0(\mathbf{q}) \right|^2 \right\rangle_{Fr, \Omega}, \quad (1)$$

where

$$P_0(\mathbf{q}) = 8 \frac{\sin(q_x A/2) \sin(q_y B/2) \sin(q_z C/2)}{q_x q_y q_z}. \quad (2)$$

In this formula, index k numbers all non-hydrogen atoms in the ABC parallelepiped in the current frame of the first set, and l numbers the atoms in the ABC parallelepiped in the current frame of the second set; q_x , q_y , and q_z are the projections of the reciprocal space vector \mathbf{q} on the coordinate axes, $|\mathbf{q}| = q = 4\pi \sin \theta / \lambda$, λ is the X-ray wavelength, 2θ is the scattering angle; $P_0(\mathbf{q})$ is the scattering amplitude from the homogeneous ABC parallelepiped [33], whose centre is at the coordinate origin, with $P_0(0) = 1$. The amplitudes $f_k(q)$ required for the calculation were taken from the periodictable library [34]. If there are only one or several hydrogen atoms and an atomic group is formed (for example, CH_n group), the scattering amplitude of this group is calculated by the following formula

$$f_{\text{CH}_n}(q) = f_C(q) + n f_H(q). \quad (3)$$

Angle brackets $\langle \dots \rangle_{Fr, \Omega}$ in formula (1) indicate two types of averaging of scattering intensity: $\langle \dots \rangle_{Fr}$ is over all frames and $\langle \dots \rangle_{\Omega}$ is over all possible orientations of the parallelepiped with respect to the scattering vector direction. Fedorov et al [6] give a detailed description of the latter algorithm of averaging.

The first term in formula (1) is the scattering intensity of a protein in solution, and the second component is the scattering intensity of a pure solvent. In accordance with Babinet's principle, the term $\rho_0 P_0(\mathbf{q})$ in this formula describes the contribution to the scattering amplitude of the homogeneous solvent with density ρ_0 . This homogeneous solvent fills the entire space except for the volume of the ABC parallelepiped. The ρ_0 parameter is calculated in the following way. For each frame in pure solvent, an ABC parallelepiped is cut out; the electron density is calculated inside it and then averaged over all frames.

The scattering intensity for every single frame has the same weight, but then the number of frames used is $N = 300 - 1000$ both for the protein with solvent and for the solvent itself. The frames are selected with the same step along the timeline. These factors enable to take into account different protein conformations with their real weight.

Formula (1) is another representation of the formula for the scattering intensity proposed in [18] and it is more transparent from the physical point of view.

The computation time essentially depends on the computer capacity. It is possible to speed up the calculations by several orders when using multi-core computers and graphics cards. In our case, SASPAR calculations were performed using one graphics card, 1280 graphics processors and a 4-core computer. The time required to calculate the scattering intensity for one frame is ~ 0.5 s, and this time is virtually independent on the protein size.

After downloading the SASPAR (and SASCUBE) programs to a PC, they can be used “standalone”, i.e. without the Internet. Moreover, SASPAR has an open source code, which can be studied and modified. The “standalone” use of the SASPAR program is its advantage; without access to online services, calculations based on, for example, WAXSIS and AXIS programs are not possible.

One disadvantage of standalone programs (including SASPAR) is that higher computer performance is required. Another probable disadvantage of the SASPAR program is the necessity of using the NVIDIA graphics card that is supported by PyCUDA library.

2.2.2. SASCUBE program. The program is intended for the calculation of scattering intensities of proteins in a structureless solvent, with coordinates of protein atoms from the PDB data. In [6], the algorithm of this program is described in details.

Briefly, the principle of the program operation is as follows. A parallelepiped is circumscribed around the protein molecule with the known coordinates of atoms and is divided into closely packed small cubes (with an edge length of 0.03 nm), and from the coordinates of each cube, it is determined whether it is within the protein molecule or outside of it. Thus, the system of cubes can describe the volume and the form of the protein molecule with high accuracy. The formula for calculation of the scattering intensity $I_{sc}(q)$ by such a system is comparatively simple (it is similar to the first term of formula (1) in the present work) and only requires the knowledge of the effective radii of the atoms of the protein and water molecules. However, it was shown in the work [6] that the variation of these radii within reasonable limits [35, 36] leads to small change of the scattering curve only.

We repeat the remark that the SASCUBE program (as opposed to the SASPAR program) calculates the scattering intensity of a protein that always keeps the crystal structure listed in the PDB. The protein is immersed in a structureless solvent. It means that no specific features of the water structure near the protein surface are taken into account.

2.3. Calculation of “vacuum” scattering intensities and radii of gyration of proteins

For further analysis of the structure of proteins in solvent and their scattering curves, we also need to know their “vacuum” curves $I_{vac}(q)$ and “vacuum” radii of gyration R_g . The corresponding calculations can be performed using the known formulas [37, 38]:

$$I_{vac}(q) = \sum_{i,j=1}^N f_i(q)f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}}, \quad (4)$$

and

$$R_g = \sqrt{\frac{\sum_{i,j=1}^N f_i(0)f_j(0)r_{ij}^2}{2 \sum_{i,j=1}^N f_i(0)f_j(0)}}. \quad (5)$$

Here, r_{ij} is the distance between atoms i and j , while $f_i(q)$ is the scattering amplitude (described above) for the i -th atom of protein, and N is the number of non-hydrogen atoms in the protein.

To calculate “vacuum” scattering intensity $I_{vac}^{crys}(q)$ for the protein crystal structure, formula (4) is used once, but to obtain intensity $I_{vac}^{MD}(q)$ based on the MD data, it is necessary to perform the calculation by this formula for each frame, average the received scattering intensities over all frames, and then calculate the dispersion for each value of q .

The same procedure is used to obtain the values of the radius of gyration: $R_{g,vac}^{crys}$ is calculated by formula (5) once, and $R_{g,vac}^{MD}$ is calculated for each frame and then averaged over all frames.

3. Quantitative assessment of the similarity of scattering curves

For each protein, the calculated curves $I_{SP}(q)$ and $I_{SC}(q)$ were compared using the following formula from [18]

$$\chi_{\log}^{water} = \sqrt{N^{-1} \sum_{i=1}^N [\log I_{SP}(q_i) - \log(f \times I_{SC}(q_i) + c)]^2}, \quad (6)$$

where N is the number of the selected values of q , \log is the decimal logarithm, and f and c are adjustment parameters ensuring the best coincidence of the two curves. In formula (6) and subsequent formulas (7)–(9), parameters f and c are determined by minimizing the value of the corresponding χ . It should be noted that f is not an adjustment parameter in the accepted sense. It only ensures the best coincidence of the curves without distorting them.

A similar formula is used to compare “vacuum” curves $I_{vac}^{crys}(q)$ and $I_{vac}^{MD}(q)$:

$$\chi_{log}^{vac} = \sqrt{N^{-1} \sum_{i=1}^N [\log I_{vac}^{MD}(q_i) - \log(f \times I_{vac}^{crys}(q_i) + c)]^2} \quad (7)$$

In formulas (6) and (7), a logarithmic scale allows one to more effectively take into account the contribution to χ_{log}^{water} and χ_{log}^{vaA} of the medium-angle part of the curve, where the scattering intensity can take value of several orders lower than the intensity in the small-angle part.

The term “medium-angle part of the curve” denotes the region of the curve starting immediately after the relatively abrupt (depending on the protein size) fall of the curve that continues up to $q \sim 10 \text{ nm}^{-1}$; moreover, at larger q , the curve is not very specific to the protein structure [39].

According to the goals of the present work, the medium-angle part of the curve is the most interesting one for the comparison of two scattering curves, as the shape of the scattering curve in the medium-angle part, in accordance with the Bragg formula, is mainly determined by the nature of electron density distribution over relatively short distances. It is these distances that define the internal structure and its change in the protein molecule and in the surrounding water to a large extent.

The scattering intensities obtained from the SASPAR and SASCUBE programmes were compared with the corresponding experimental intensities using the following formulas

$$\chi_{SP} = \sqrt{N^{-1} \sum_{i=1}^N \left(\frac{I_{exp}(q_i) - (f \times I_{SP}(q_i) + c)}{\sigma_{exp}(q_i)} \right)^2} \quad (8)$$

and, respectively,

$$\chi_{SC} = \sqrt{N^{-1} \sum_{i=1}^N \left(\frac{I_{exp}(q_i) - (f \times I_{SC}(q_i) + c)}{\sigma_{exp}(q_i)} \right)^2}, \quad (9)$$

where $\sigma_{exp}(q_i)$ are standard deviations, while f and c , the same as above, are the parameters that provide the best coincidence of the scattering curves.

The similarity of the radii of gyration $R_{g,vac}^{crys}$ and $R_{g,vac}^{MD}$ can be assessed using the following formula

$$\delta R_g = \frac{R_{g,vac}^{MD} - R_{g,vac}^{crys}}{R_{g,vac}^{crys}}, \quad (10)$$

4. Analysis of scattering curves of proteins. Comparison with the experimental curves

The following proteins were considered: the third IGG-binding domain (“GB3”, 1IGD); ubiquitin (1UBQ); cytochrome C (1CRC); neurotrypsin Scavenger receptor cysteine-rich domain 3 (“mmNT-SRCR3”, 6H8M); lysozyme (193L, 6LYZ); ribonuclease pancreatic (1C0B); myoglobin (1WLA); EAL/GGDEF domain (3ICL); truncated alpha-DsbN (6EEZ); carbonic anhydrase (5A25); sensory box/GGDEF domain (3LYX); filamin A Ig-like domains 3–5 P637Q mutant (“FLNa3-5 P637Q”, 6EWL); bovine serum albumin (“BSA”, 4F5S monomer, 4F5S dimer); alcohol dehydrogenase (5ENV, tetramer); glucose isomerase (1OAD, tetramer); apoferritin (1IER, 24-mer). Representative curves for four of these proteins are shown in Fig. 2. Curves for all proteins are presented in supplementary materials.

Five scattering curves are presented for each of the studied proteins in a separate graph: experimental curve $I_{exp}(q)$; SASPAR program curve $I_{SP}(q)$; SASCUBE program curve $I_{SC}(q)$; “vacuum” curve $I_{vac}^{MD}(q)$, calculated with MD and averaged over all frames; and “vacuum” curve $I_{vac}^{crys}(q)$, calculated based on a crystalline structure.

Vertical lines in each graph show the borders within which the corresponding scattering curves were matched with each other. There were certain difficulties with the choice of the left border, so it was selected individually for each protein. As we know, the shape of the experimental curve in the region of the smallest scattering angles can be distorted by interference effects [40] and, probably, by residual impurities of oligomers, while the $I_{SP}(q)$ intensity, apparently, provides the best description of the curve across the entire range of scattering angles for a *single* molecule of protein. Therefore, the q value to the left of which a systematic divergence between the experimental curve and the calculated curve $I_{SP}(q)$ is observed can be considered as the criterion for the choice of the left border.

As for the choice of the right border, generally, it either coincided with the maximum values of scattering angles on the experimental curve or was limited by $q = 15 \text{ nm}^{-1}$, although the experimental curve could be continued further. As mentioned above, for globular proteins, the region of scattering angles with $q > 15 \text{ nm}^{-1}$ (and even with $q > 10 \text{ nm}^{-1}$) is weakly specific to the internal structure of the protein and is defined only by the fact that the polypeptide chains are sufficiently closely packed. In [40], it was shown that the maximum itself at $q \sim 14 \text{ nm}^{-1}$ depends only on the average van der Waals distance between the neighbouring closely packed atoms of protein and disappears during its denaturation.

The values of the above-described parameters for proteins with various molecular weights W are presented in the Table 1. The following information is presented in the table: the protein codes from the PDB (column 1), references to the

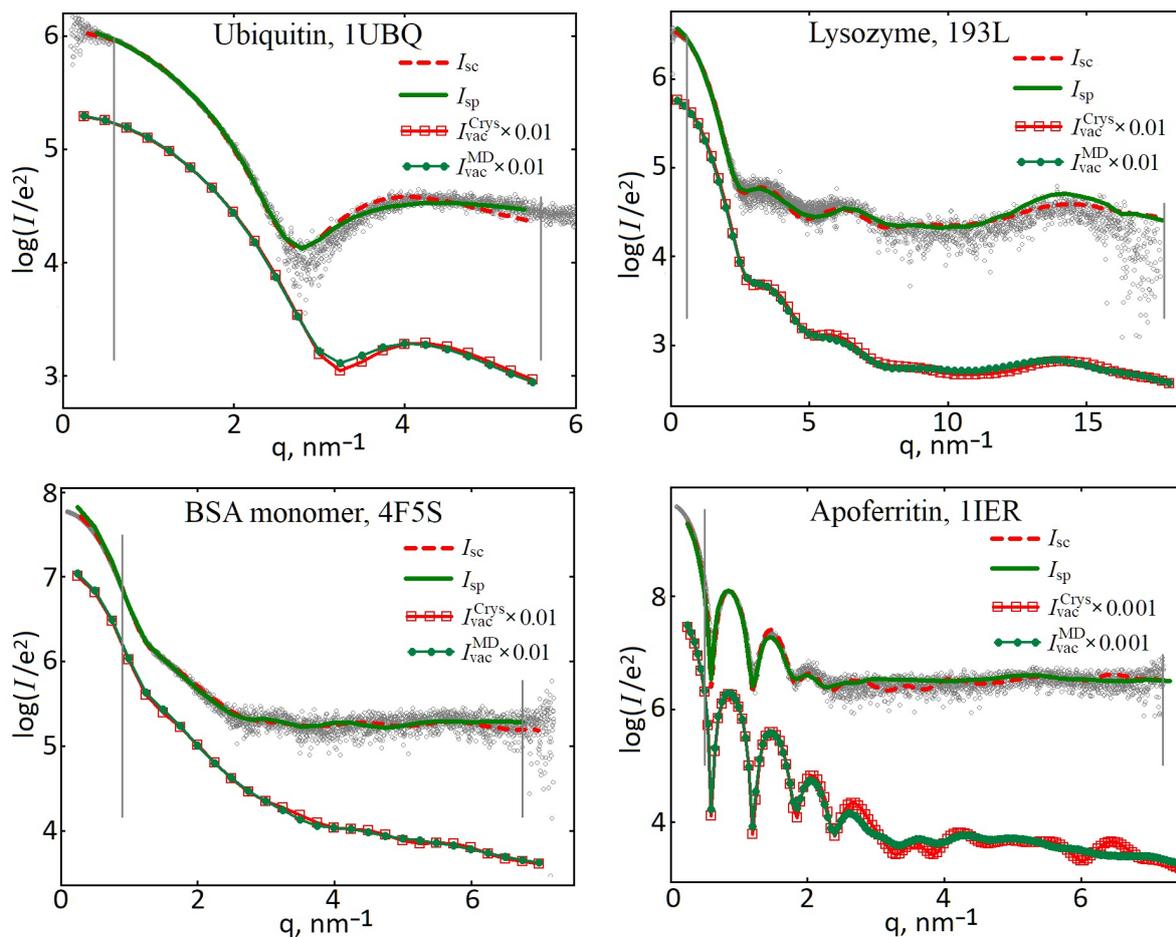


FIG. 2. Experimental and calculated scattering curves for 4 from 18 considered proteins. Experimental curves: $I_{exp}(q)$ (circles); calculated curves: $I_{SP}(q)$ are the SASPAR curves; $I_{SC}(q)$ are the SASCUBE curves; $I_{vac}^{MD}(q)$ are the “vacuum curves” averaged for all MD simulation frames; $I_{vac}^{Crys}(q)$ are the “vacuum curves” for proteins in crystalline form. Hereinafter, the experimental curves were shifted in a logarithmic scale, using a scale factor, to align them with the calculated curves. For each protein, its name and its PDB code are given. The vertical lines mark the boundaries of the region where the calculated and experimental curves are compared

records in SASBDB and references to the SAS experiment (column 2), and the molecular weights of the proteins (column 3). The numerical evaluations of similarity listed in the other columns and their analysis are given in the text of the article.

First, let us consider column 7 listing values χ_{log}^{vaA} (formula 7) that determine the similarity of the “vacuum” scattering curves calculated both from the MD data and for the crystal forms of proteins. If we introduce the criterion of similarity of the curves: $\chi_{log}^{vac} \leq 3\%$, then these curves can be considered “sufficiently similar” for all the 14 single-domain proteins indicated in the first 14 lines of the table. Qualitatively, the similarity of these curves is clearly visible on the respective graphs. Therefore, the first important conclusion is that all the considered single-domain proteins only slightly change their structure during the transition from crystal to solution. At the same time, for the 4 considered multi-domain proteins the following inequality holds: $\chi_{log}^{vac} > 3\%$, which is an evidence that their structure changes to a greater or lesser degree. It should also be noted that all these 18 proteins do not exhibit significant changes in their sizes (column 10), judging by the comparison of the gyration radii of crystal structures (column 9) and the averaged structures according to MD data (column 8). The Lysozyme C protein was considered for two modifications of its structure presented in the PDB (193L and 6LYZ), and each modification has its own line.

Let us now study column 6 presenting the χ_{log}^{water} values that determine the similarity of the scattering curves calculated by the SASPAR and SASCUBE programs (“water” curves). It is evident that the similarity of these curves is a rarer event. If the same criterion of “sufficient similarity” is still to be used ($\chi_{log}^{vac} \leq 3\%$), only two of the studied proteins meets this criterion: 3ICL and 4F5S. Only these proteins can be categorised as type I, i.e., we can assume that if there is the similarity of its structure in crystal and in solution, then the structure of water in the layer adjacent to the protein is only slightly distorted as compared to the homogeneous continuum. As for the other proteins with a “sufficient similarity” of the “vacuum” curves, they show a considerable increase in value χ_{log}^{water} . This indicates that for these proteins,

TABLE 1. The numerical evaluation of the similarity both for “water” and “vacuum” SAXS curves of the studied proteins

protein	ref	W kDa	χ_{SP}	χ_{SC}	χ_{log}^{water} %	χ_{log}^{vac} %	$R_{g,vac}^{MD}$ nm	$R_{g,vac}^{crys}(q)$ nm	δR_g %	Type of protein
1	2	3	4	5	6	7	8	9	10	11
IIGD	(private message) Grishaev <i>et al.</i> (2010)	6.2	2.395	2.875	5.9	2.1	1.076	1.056	1.9	II
1UBQ	SASDAQ2, St. Prot. ¹	9	1.198	1.685	4.4	2.5	1.165	1.173	-0.7	II
1CRC	SASDAB2, St. Prot. ²	12.5	1.092	1.577	9.8	2.2	1.299	1.296	0.2	II
6H8M	SASDES5, [41]	13	1.811	2.228	4.2	2.8	1.304	1.295	0.7	II
193L	SASDAG2, St. Prot. ¹	14	0.942	0.726	4.6	2.6	1.422	1.385	2.7	II
6LYZ	SASDAG2, St. Prot. ¹	14	0.951	0.884	4.6	1.7	1.411	1.398	0.9	II
1C0B	SASDAN2 St. Prot. ¹	16	2.138	1.178	6.1	2.5	1.497	1.455	2.9	II
1WLA	SASDAK2, St. Prot. ¹	17	2.480	1.993	8.2	1.9	1.550	1.532	1.2	II
3ICL	SASDCG6, [42]	19	2.120	2.348	1.2	1.0	1.575	1.526	3.2	I
6EEZ	SASDC38, [43]	21	2.638	2.454	5.6	2.1	1.632	1.616	1.0	II
5A25	SASDFP8, [44]	29	2.097	3.017	6.8	2.7	1.771	1.741	1.7	II
3LYX	SASDCE6, [42]	30	3.100	2.512	4.3	0.9	1.911	1.884	1.4	II
6EWL	SASDEP7, [45]	31	1.440	1.550	5.0	1.5	2.117	2.086	1.5	II
4F5S	SASDFQ8, St. Prot. ³	66	1.099	1.156	3.0	2.3	2.707	2.674	1.2	I
4F5S dimer	SASDFR8, [44]	133	1.625	1.915	3.5	3.5	3.940	3.851	2.3	III
5ENV	SASDFS8, [44]	147	1.707	2.335	8.0	3.5	3.351	3.331	0.6	III
1OAD	SASDCK2, [46]	174	2.531	1.874	6.0	4.3	3.200	3.174	0.8	III
1IER	SASDFN8, [44]	479	5.360	3.726	12.8	9.4	5.300	5.264	0.7	III.

¹Standard proteins Darja Ruskule²Standard proteins Cy M Jeffries³Standard proteins Melissa Graewert, Cy M Jeffries

the differences between the “water” curves are mainly conditioned by the changes (to a greater or lesser degree) in the structure of water in the area adjacent to the protein, as compared to the homogeneous continuum. These 12 proteins are categorised as type II.

Finally, a considerable discrepancy is observed between the “vacuum” curves as well as between the “water” curves for all the studied multi-domain proteins. Thus, the structures of these proteins in solution are different from their structures in crystal and, probably, the water near their surfaces is structured (proteins of type III). The interdomain mobility is a possible reason leading to the difference in their structures in solution and in crystal, while the overall dimensions remain the same.

Column 11 shows the type of the protein according to the classification suggested in the Introduction.

Now let us compare the curves from the SASPAR and SASCUBE programs with the experimental scattering curves. Using formulas (8) and (9) and taking into account the errors specified in the experimental works, chi-factors were calculated for both programs: column 4 (SASPAR) and column 5 (SASCUBE). As one can see, for most of the 18 studied proteins, the obtained chi-factors are indicative of a sufficient similarity between the experimental and the calculated scattering curves. It is also visible in the graphs.

Noteworthy, we did not intend to develop a method that would allow making the calculated curves as close as possible to the experiment. As the table and graphs show, a significant difference between the SASPAR and experimental scattering curves is observed for at least two proteins (1IER and 3LYX). On the one hand, this can be explained by an imperfect theory. Indeed, the use of different packages for the MD simulation (see Fig. 2s in Supplementary) and the use of different water models [18] within the SASPAR programme can result in different scattering curves of proteins if q values are large enough.

On the other hand, the comparison of the calculated scattering curves with the experimental ones has certain difficulties. As it was noted in the review by Hub [47], “the quality of SWAXS data has greatly benefited from better light sources, single-photon counting detectors, and from set-ups coupled to size-exclusion chromatography which have led to reduced statistical noise”. Further, “Common sources for systematic errors in SWAXS are aggregation, inter-particle repulsion, and poor buffer-matching”, as well as vibrations of atoms. Therefore, according to Hub, “the overall uncertainty of the data is by now often dominated by systematic errors”. Nevertheless, as it was mentioned in recent work [48] “noise has the greatest impacts on low signal in the high- q region”.

It can only be assumed that the significant differences between the experimental and calculated curves for some proteins are related either to the factors specified in [47] and [48] or to the errors of subtraction of the buffer scattering from the solution scattering curve at relatively wide angles, at which the solution and buffer scattering intensities are rather similar.

We compared the scattering intensities calculated using the SASPAR program and another program (WAXSIS) with a similar algorithm. We also compared the intensities calculated with the SASPAR program based on the MD simulations performed both with the MD Gromacs 2019-3 software package and the Desmond MD program. Comparison results are presented in supplementary materials. In both cases, there are small discrepancies in the wide-angle range that are within the limits of an experimental error.

5. Conclusion

The main conclusion from the present analysis is that most of the studied one-domain proteins only slightly change their structure during the transition from crystal to solution using the experimental conditions as described for the SAXS experiments. This statement is also confirmed by the similarity of their radii of gyration. However, this does not mean that for some frames, the protein structure in solution cannot significantly differ from its structure in crystal; it is stated only that the average protein structure in a solution is similar to the crystal one. Although this statement follows from the study of the protein structures not in the real space but in the reciprocal one, this conclusion seems quite reasonable, considering the unambiguous relation between these spaces in terms of the Fourier transform. Certainly, this is not the first time the problem of the protein structures’ similarity in crystal and in solution is discussed. There is an extensive literature confirming the indicated similarity of structures using NMR [48–50]. However, our work provides *quantitative evidence* of this statement during the analysis of the considered single-domain proteins based on the combination of two methods: MD and SAXS.

As for the considered multi-domain proteins, their crystalline and “water” structures might differ significantly. A possible reason for this difference is the relative movement of domains. This issue requires additional research, which can be done within the framework of the approach described in this article.

The second important conclusion is that the structure of water in the layer adjacent to the protein surface is usually significantly different from its structure far from the surface. Based on the almost full coincidence of the SASPAR and SASCUBE curves, only one (3icl) of the studied proteins does not have this layer, and all the water surrounding this protein can be considered a homogeneous continuum. However, this protein can be considered an exception. This layer definitely exists for the other proteins, and it is apparently associated with the interaction between the water molecules and the surface atoms of a protein. The electron density within this layer is non-homogeneous, which can significantly influence on the profile of the SASPAR curves. The conclusion about the presence of a hydration layer near a protein

surface is also not new, and many authors introduce this effect into the intensity calculations in some way or another (see, e.g., [8, 52–55]). Our contribution is that we directly confirm the existence of this hydrated layer using the fact that the SASPAR and SASCUBE scattering curves clearly *diverge* from one another.

As for the comparison of the calculated scattering curves with the experimental data, the table and graphs show that for most proteins, especially for their SASPAR curves, there is good agreement with the experiment in the “middle” q -area, which is the most sensitive to intraprotein rearrangements and rearrangements of the water molecules structure near the protein surface. At the same time, for some of the proteins, while there is a sufficiently good qualitative coincidence of their scattering curves, their quantitative coincidence based on the selected criteria is not good enough. The possible reasons for this difference were discussed in the work.

References

- [1] Ninio J., Luzzati V., Yaniv M. Comparative Small-Angle X-Ray Scattering Studies on Unacylated, Acylated and Cross-Linked Escherichia Coli Transfer RNAIVal. *Journal of Molecular Biology*, 1972, **71**(2), P. 217–229.
- [2] Fedorov B.A., Denesyuk A.I. Large-Angle X-Ray Diffuse Scattering, a New Method for Investigating Changes in the Conformation of Globular Proteins in Solutions. *J. Appl. Cryst.*, 1978, **11**(5), P. 473–477.
- [3] Pavlov M.Yu., Fedorov B.A. Improved Technique for Calculating X-Ray Scattering Intensity of Biopolymers in Solution: Evaluation of the Form, Volume, and Surface of a Particle. *Biopolymers*, 1983, **22**(6), P. 1507–1522.
- [4] Valentini E., Kikhney A.G., Previtali G., Jeffries C.M., Svergun D.I. SASBDB, a Repository for Biological Small-Angle Scattering Data. *Nucleic Acids Res.*, 2015, **43**(D1), P. D357–D363.
- [5] Hura G.L., Menon A.L., Hammel M., Rambo R.P., Poole II F.L., Tsutakawa S.E., Jenney Jr F.E., Classen S., Frankel K.A., Hopkins R.C., Yang S., Scott J.W., Dillard B.D., Adams M.W.W., Tainer J.A. Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS). *Nat Methods*, 2009, **6**(8), P. 606–612.
- [6] Fedorov B.A., Smirnov A.V., Yaroshenko V.V., Porozov Yu.B. SASCUBE: An Updated Method of Cubes for Calculation of the Intensity of X-Ray Scattering by Biopolymers in Solution. *BIOPHYSICS*, 2019, **64**(1), P. 38–48.
- [7] <https://sourceforge.net/projects/sascube>
- [8] Svergun D., Barberato C., Koch M.H.J. CRY SOL – a Program to Evaluate X-Ray Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.*, 1995, **28**(6), P. 768–773.
- [9] Bardhan J., Park S., Makowski L. SoftWAXS: A Computational Tool for Modeling Wide-Angle X-Ray Solution Scattering from Biomolecules. *Journal of applied crystallography*, 2009, **42**, P. 932–943.
- [10] Liu H., Hexemer A., Zwart P.H. The Small Angle Scattering ToolBox (SASTBX): An Open-Source Software for Biomolecular Small-Angle Scattering. *J. Appl. Cryst.*, 2012, **45**(4), P. 587–593.
- [11] Svergun D.I., Richard S., Koch M.H.J., Sayers Z., Kuprin S., Zaccai G. Protein Hydration in Solution: Experimental Observation by x-Ray and Neutron Scattering. *PNAS*, 1998, **95**(5), P. 2267–2272.
- [12] Schneidman-Duhovny D., Hammel M., Tainer J.A., Sali A. FoXS, FoXSDock and MultiFoXS: Single-State and Multi-State Structural Modeling of Proteins and Their Complexes Based on SAXS Profiles. *Nucleic Acids Research*, 2016, **44**(W1), P. W424–W429.
- [13] Park S., Bardhan J.P., Roux B., Makowski L. Simulated X-Ray Scattering of Protein Solutions Using Explicit-Solvent Models. *J. Chem. Phys.*, 2009, **130**(13), 134114.
- [14] Oroguchi T., Ikeguchi M. MD–SAXS Method with Nonspherical Boundaries. *Chemical Physics Letters*, 2012, **541**, P. 117–121.
- [15] Köfinger J., Hummer G. Atomic-Resolution Structural Information from Scattering Experiments on Macromolecules in Solution. *Phys. Rev. E*, 2013, **87**(5), 052712.
- [16] Knight C.J., Hub J.S. WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Research*, 2015, **43**(W1), P. W225–W230.
- [17] Grishaev A., Guo L., Irving T., Bax A. Improved Fitting of Solution X-Ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. *J. Am. Chem. Soc.*, 2010, **132**(44), P. 15484–15486.
- [18] Chen P., Hub J.S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-Ray Scattering Data. *Biophysical Journal*, 2014, **107**(2), P. 435–447.
- [19] <https://github.com/andy-biochem/saspar2>
- [20] Berman H., Battistuz T., Bhat T., Bluhm W., Bourne P., Burkhardt K., Feng Z., Gilliland G., Iype L., Jain S., Fagan P., Marvin J., Padilla D., Ravichandran V., Schneider B., Thanki N., Weissig H., Westbrook J., Zardecki, C. The Protein Data Bank. *Acta crystallographica. Section D, Biological crystallography*, 2002, **58**, P. 899–907.
- [21] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research*, 2000, **28**(1), P. 235–242.
- [22] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 1983, **79**(2), P. 926–935.
- [23] Bowers K.J., Chow E., Xu H., Dror R.O., Eastwood M.P., Gregersen B.A., Klepeis J.L., Kolossvary I., Moraes M.A., Sacerdoti F.D., Salmon J.K., Shan Y., Shaw D.E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *In SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, ACM Press, 2006.
- [24] Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., Lindahl E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX*, 2015, **1–2**, P. 19–25.
- [25] Banks J.L., Beard H.S., Cao Y., Cho A.E., Damm W., Farid R., Felts A.K., Halgren T.A., Mainz D.T., Maple J.R., Murphy R., Philipp D.M., Repasky M.P., Zhang L.Y., Berne B.J., Friesner R.A., Gallicchio E., Levy R.M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry*, 2005, **26**(16), P. 1752–1780.
- [26] Hoover W.G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A*, 1985, **31**(3), P. 1695–1697.
- [27] Martyna G.J., Tobias D.J., Klein M.L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.*, 1994, **101**(5), P. 4177–4189.
- [28] Lindorff-Larsen K., Piana S., Palmo K., Maragakis P., Klepeis J.L., Dror R.O., Shaw D.E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics*, 2010, **78**(8), P. 1950–1958.
- [29] Parrinello M., Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics*, 1981, **52**(12), P. 7182–7190.

- [30] Gureev M.A., Kadochnikov V.V. Porozov Yu.B. *Molekulyarnyi doking i ego verifikatsiya v kontekste virtual'nogo skrininga* [Molecular docking and its verification in the context of virtual screening]. St. Petersburg: ITMO University, 2018 (In Russ.)
- [31] https://github.com/andy-biochem/saspar2/tree/master/gmx_prot
- [32] https://github.com/andy-biochem/saspar2/tree/master/gmx_water
- [33] Vainshtein B.K., Vainshtein B.K. *Diffraction of X-Rays by Chain Molecules*. Elsevier, 1966.
- [34] P.A. Kienzle Periodictable V1.5.0. Zenodo. <https://doi.org/10.5281/zenodo.840347>
- [35] Lee B., Richards F.M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.*, 1971, **55**(3), P. 379–400.
- [36] Bondi A.A. *Physical Properties of Molecular Crystals, Liquids, and Glasses*. Wiley, 1968.
- [37] Debye P. Zerstreuung von Röntgenstrahlen. *Annalen der Physik*, 1915, **351**(6), P. 809–823.
- [38] Guinier A. La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.*, 1939, **11**(12), P. 161–237.
- [39] Fedorov B.A., Ptitsyn O.B., Voronin L.A. X-Ray Diffuse Scattering of Globular Protein Solutions: Consideration of the Solvent Influence. *FEBS Lett*, 1972, **28**(2), P. 188–190.
- [40] Guinier A., Fournet G., Yudowitch K.L. *Small-Angle Scattering of X-Rays*, 1955.
- [41] Canciani A., Catucci G., Forneris F. Structural Characterization of the Third Scavenger Receptor Cysteine-Rich Domain of Murine Neurotrophin. *Protein Science*, 2019, **28**(4), P. 746–755.
- [42] Grant T.D., Luft J.R., Wolfley J.R., Tsuruta H., Martel A., Montelione G.T., Snell E.H. Small Angle X-Ray Scattering as a Complementary Tool for High-Throughput Structural Studies. *Biopolymers*, 2011, **95**(8), P. 517–530.
- [43] Walden P.M., Whitten A.E., Premkumar L., Halili M.A., Heras B., King G.J., Martin J.L. The Atypical Thiol–Disulfide Exchange Protein α -DsbA2 from *Wolbachia Pipientis* Is a Homotrimeric Disulfide Isomerase. *Acta Cryst. D*, 2019, **75**(3), P. 283–295.
- [44] Graewert M.A., Da Vela S., Gräwert T.W., Molodenskiy D.S., Blanchet C.E., Svergun D.I., Jeffries C.M. Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data. *Crystals*, 2020, **10**(11), P. 975.
- [45] Haataja T.J.K., Bernardi R.C., Lecointe S., Capoulade R., Merot J., Pentikäinen U. Non-Syndromic Mitral Valve Dysplasia Mutation Changes the Force Resilience and Interaction of Human Filamin A. *Structure*, 2019, **27**(1), P. 102–112.e4
- [46] Trehwella J., Duff A.P., Durand D., Gabel F., Guss J.M., Hendrickson W.A., Hura G.L., Jacques D.A., Kirby N.M., Kwan A.H., Pérez J., Pollack L., Ryan T.M., Sali A., Schneidman-Duhovny, D., Schwede T., Svergun D.I., Sugiyama M., Tainer J.A., Vachette P., Westbrook J., Whitten A.E. Publication Guidelines for Structural Modelling of Small-Angle Scattering Data from Biomolecules in Solution: An Update. *Acta Cryst D*, 2017, **73**(9), P. 710–728.
- [47] Hub J.S. Interpreting Solution X-Ray Scattering Data Using Molecular Simulations. *Current Opinion in Structural Biology*, 2018, **49**, P. 18–26.
- [48] Brosey C.A., Tainer J.A. Evolving SAXS Versatility: Solution X-Ray Scattering for Macromolecular Architecture, Functional Landscapes, and Integrative Structural Biology. *Curr Opin Struct Biol*, 2019, **58**, P. 197–213.
- [49] Sikic K., Tomic S., Carugo O. Systematic Comparison of Crystal and NMR Protein Structures Deposited in the Protein Data Bank. *Open Biochem J*, 2010, **4**, P. 83–95.
- [50] Everett J.K., Tejero R., Murthy S.B.K., Acton T.B., Aramini J.M., Baran M.C., Benach J., Cort J.R., Eletsky A., Forouhar F., Guan R., Kuzin A.P., Lee H.W., Liu G., Mani R., Mao B., Mills J.L., Montelione A.F., Pederson K., Powers R., Ramelot T., Rossi P., Seetharaman J., Snyder D., Swapna G.V.T., Vorobiev S.M., Wu Y., Xiao R., Yang Y., Arrowsmith C.H., Hunt J.F., Kennedy M.A., Prestegard J.H., Szyperski T., Tong L., Montelione G.T. A Community Resource of Experimental Data for NMR / X-Ray Crystal Structure Pairs. *Protein Science*, 2016, **25**(1), P. 30–45.
- [51] Leman J.K., D'Avino A.R., Bhatnagar Y., Gray J.J. Comparison of NMR and Crystal Structures of Membrane Proteins and Computational Refinement to Improve Model Quality. *Proteins*, 2018, **86**(1), P. 57–74.
- [52] Tjioe E., Heller W.T. ORNL_SAS: Software for Calculation of Small-Angle Scattering Intensities of Proteins and Protein Complexes. *J. Appl. Cryst.*, 2007, **40**(4), P. 782–785.
- [53] Poitevin F., Orland H., Doniach S., Koehl P., Delarue M. AquaSAXS: A Web Server for Computation and Fitting of SAXS Profiles with Non-Uniformly Hydrated Atomic Models. *Nucleic Acids Research*, 2011, **39**(suppl.2), P. W184–W189.
- [54] Schneidman-Duhovny D., Hammel M., Tainer J.A., Sali A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophysical Journal*, 2013, **105**(4), P. 962–974.
- [55] Virtanen J.J., Makowski L., Sosnick T.R., Freed K.F. Modeling the Hydration Layer around Proteins: Application to Small- and Wide-Angle X-Ray Scattering. *Biophysical Journal*, 2011, **101**(8), P. 2061–2069.

Submitted 12 April 2022; accepted 14 April 2022

Information about the authors:

Alexander V. Smirnov – ITMO University, Kronverksky Pr. 49, bldg. A, Saint Petersburg 197101, Russian Federation; smirnav_2@mail.ru

Andrei M. Semenov – Alferov University Russian Academy of Sciences, Chlopina ul., 8, Saint Petersburg 194021, Russian Federation; Voronezh State University, Universitetskaya pl. 1, Voronezh 394018, Russian Federation; andy_semenov@mail.ru

Yuri B. Porozov – Sechenov First Moscow State Medical University, Trubetskaya ul. 8, Moscow 119991, Russian Federation; Sirius University of Science and Technology, Olympic ave. 1, Sochi 354340, Russian Federation; yuri.porozov@gmail.com

Boris A. Fedorov – ITMO University, Kronverksky Pr. 49, bldg. A, Saint Petersburg 197101, Russian Federation; borfedorov1@gmail.com

Conflict of interest: the authors declare no conflict of interest.