

BIOMOLECULAR SOLUTION X-RAY SCATTERING AT n²STAR BEAMLINE

Oktay GÖCENLER, Department of Molecular Biology, Koc University, İstanbul, Türkiye, ogocenler18@ku.edu.tr

([ID](https://orcid.org/0000-0003-4310-0928) <https://orcid.org/0000-0003-4310-0928>)

Cansu M. YENİCİ, Department of Molecular Biology, Koc University, İstanbul, Türkiye, cyenici21@ku.edu.tr

([ID](https://orcid.org/0000-0003-0179-6269) <https://orcid.org/0000-0003-0179-6269>)

Kerem KAHRAMAN, Department of Molecular Biology, Koc University, İstanbul, Türkiye, kkahraman19@ku.edu.tr

([ID](https://orcid.org/0000-0001-9714-4892) <https://orcid.org/0000-0001-9714-4892>)

Cengizhan BÜYÜKDAĞ, Department of Molecular Biology, Koc University, İstanbul, Türkiye, cbuyukdag19@ku.edu.tr

([ID](https://orcid.org/0000-0002-1825-0097) <https://orcid.org/0000-0002-1825-0097>)

Çağdaş DAĞ*, Department of Molecular Biology, Koc University, İstanbul, Türkiye

Nanofabrication and Nanocharacterization Center for Scientific and Technological Advanced Research (n²STAR),
Koc University, İstanbul, Türkiye

Koc University Isbank Center for Infectious Diseases (KUISCID), Koc University, İstanbul, Türkiye

cdag@ku.edu.tr, ([ID](https://orcid.org/0000-0003-1595-431X) <https://orcid.org/0000-0003-1595-431X>)

Received: 13.09.2022, Accepted: 02.12.2022

Research Article

*Corresponding author

DOI: 10.22531/muglajsci.1174474

Abstract

Small angle X-ray Scattering (SAXS) is a method for determining basic structural characteristics such as the size, shape, and surface of particles. SAXS data can be used to generate low resolution models of biomolecules faster than any other conventional experimental structural biology tool. SAXS data is mostly collected in synchrotron facilities to obtain the best scattering data possible however home source SAXS devices can also generate valuable data when properly optimized. Here, we examined sample data collection and optimization at home source SAXS beamline in terms of the concentration, purity, and duration of data acquisition. We validated that high concentration, monodisperse and ultra pure protein samples obtained by size exclusion chromatography are necessary for generating viable SAXS data using a home source beamline. At least one hour is required to generate a feasible model from SAXS data, although longer data collection times do not always translate to higher resolutions. We show that with small optimizations during data collection and analysis SAXS can characterize properties such as oligomerization, molecular mass, and overall shape of particles in solution under physiological conditions.

Keywords: SAXS, GST, protein, in solution.

n²STAR MERKEZİNDE BİYOMOLEKÜLER ÇÖZELTİLERİN KÜÇÜK AÇIL X-IŞIMI SAÇILIMI

Özet

Küçük açılı X-ışını Saçılımı (SAXS), parçacıkların boyutu, şekli ve yüzeyi gibi temel yapısal özellikleri belirlemek için kullanılan bir yöntemdir. SAXS verisi ile diğer geleneksel deneysel yapısal biyoloji araçlarından daha hızlı düşük çözünürlüklü biyomolekül modelleri hesaplanabilir. SAXS verileri, mümkün olan en iyi saçılma verilerini elde etmek için çoğunlukla senkrotron tesislerinde toplanır, ancak yerel kaynaklı SAXS cihazları da uygun şekilde optimize edildiğinde değerli veriler üretebilir. Burada, yerel kaynaklı SAXS ışın hattında numune veri toplama ve optimizasyonunu konsantrasyon, saflık ve veri toplama süresi açısından inceledik. Boyut dışlama kromatografisiyle elde edilen yüksek konsantrasyonlu, monodispers ve ultra saf protein numunelerinin, ev kaynaklı laboratuvar tipi ışın hattı kullanılarak uygulanabilir SAXS verilerinin üretilmesi için gerekli olduğunu doğruladık. Daha uzun veri toplama süresi her zaman daha yüksek çözünürlükler üretmez, ancak SAXS verilerinden uygun bir model oluşturmak için en az bir saat gereklidir. Ayrıca, hem veri toplama sırasında hem de daha sonra veri analizi sırasında küçük optimizasyonlarla SAXS, fizyolojik koşullar altında oligomerizasyon, moleküler kütle ve çözeltideki parçacıkların genel şekli gibi özellikleri belirleyebilir.

Anahtar Kelimeler: SAXS, GST, protein, çözelti içinde.

Cite

Göcenler, O., Yenici, C. M., Kahraman, K., Büyükdag, C., Dağ, Ç., (2022). "Biomolecular Solution X-ray Scattering at n²STAR Beamline", *Mugla Journal of Science and Technology*, 8(2), 60-69.

1. Introduction

X-ray scattering is a widely used technique for characterizing chemical and structural properties of a variety of biological macromolecules and complexes. Characterization of chemical composition, structure, and size is mainly achieved by sending X-rays to the sample and measuring its scattering angle and intensity on a detector [1]. If the energy of the x-ray remains the same while the angle changes this scattering is classified as elastic scattering. In contrast, if the energy of X-rays changes after scattering from a particle, then it is classified as inelastic [2]. Elastic scattering utilized for small angle X-ray scattering (SAXS) and wide angle X-ray scattering (WAXS), are common methods for analytical characterization of biological molecules [3]. SAXS is particularly useful for nanoscale resolution characterization of samples, whereas WAXS is used for atomic scale resolution characterization of smaller macromolecules [4].

Small angle X-ray scattering (SAXS) is an established biophysical X-ray imaging method used in nanotechnology, chemical engineering, and structural biology as a tool for analyzing structural components of relatively large molecules by their size, shapes, and morphology [5]. Despite providing very low resolution, SAXS is becoming an important method among structural biologists due to its swiftness and practicality. SAXS is also a reputable X-ray imaging method used for structural characterization of particles ranging from 1 nm to 1000 nm [6,7]. Conceptually SAXS has similar characteristics with other X-ray imaging methods such as X-ray crystallography (XRC) but differs in the condition of the sample. X-ray Crystallography (XRC) requires particles ordered in a crystal lattice, in contrast, SAXS does not require such specific formation but is instead analyzed while target molecules are in solution [8]. This characteristic of SAXS defines the primary difference between SAXS and XRC in terms of data analysis and strategies on the refinement of the data. In XRC, the diffraction data collected from crystallized samples appear in a discreet manner due to highly ordered localization of the molecules in the sample. Scatterings collected from XRC experiments can generate a high resolution 2-D image of a molecule from a certain angle and thus require multiple rotations to have a total 3-D understanding of the sample [9]. Although XRC images come with comparably higher resolutions, the structure obtained from XRC could be misleading since it requires fixing biological molecules with non-biologically relevant interactions under harsh conditions inducing hyperosmolarity of the environment. In contrast, diffraction data obtained by SAXS comes from randomly oriented and biologically free molecules or proteins which, unlike crystals, are not induced into static conditions. Flexibility and accessibility of SAXS as a method provide fast structural analysis of protein structures and certain characteristics of biomolecules such as formation of oligomeric protein structures and protein ligand interactions. Combination of SAXS and NMR spectroscopy for structural studies utilizing each

method's own advantages will result in generation of more accurate models.

Advances in small angle scattering (SAS) data acquisition and a variety of computational methods for analysis of this data significantly increased the volume of SAS published papers. Consequently, a central data bank was required to increase the availability of the published results to interested researchers. In response, The Small Angle Scattering Biological Data Bank (SASBDB) was established to meet the growing storage needs of ever-increasing SAXS/SANS data [10]. A Lacking a universal format of storing SAS data was a problem from the beginning, but it wasn't until 1999 when a decision was made to make it more available to newer researchers in this growing field [11]. In response to this problem, an extension to Crystallographic Information File (CIF) developed which is called small angle scattering Crystallographic Information File (sasCIF) for storing one-dimensional SAS data along with certain parameters of the experiment [12].

After SAXS' initial introduction to structural biology as a tool for basic structure determination, a variety of new methods have been developed for SAXS to analyze structural properties of biological macromolecules in solution and under physiological conditions. Among one of the useful methods for analysis of protein oligomeric structures is the prominent application of SAXS, since in solution properties of proteins differ for individual proteins but could be amplified in presence of multiple proteins expected to form oligomeric structures particularly in solution [13]. As concentrations of each protein in a complex can be adjusted, unlike conventional co-crystallization methods, SAXS allows greater control in important parameters allowing determination of a dynamic self-association equilibrium. As previously mentioned, SAXS data is often combined with higher resolution structural biology tools to obtain sufficient details to generate a 3D model of a protein or a complex. However, a computational protocol called ATTRACT-SAXS was developed which can generate high resolution structures of protein-protein complexes by itself [14]. Intrinsically disordered proteins (IDP) and intrinsically disordered regions of proteins have also been a problem for X-ray crystallography as static models generated by crystallized protein diminish the possibility of structural characterization that are close to actual native conformations of proteins. In that sense, integrative studies with SAXS, NMR and single-molecule Förster Resonance Energy Transfer (smFRET) are capable of uncovering functions of intrinsically disordered regions of protein and more realistic and collective understanding of IDP conformations [15].

Crystallization of large biomolecules like antibodies remains a challenge for structural biologists interested in drug development. Understanding the structural characteristics of antibodies is a critical step for developing more competent therapeutic agents and SAXS could be utilized as a high throughput tool for skimming general structural characteristics of antibodies. In relation to singular characterization of antibodies, real

time analysis of supplementation of additional compounds to determine the effects of these compounds or interacting particles paves the way for development of more efficient antibodies [16]. It is also shown that SAXS is capable of being a high throughput protein structure analysis tool with an automated system, without requiring extensive crystallization of multiple protein samples. Despite providing structures with atomic accuracy, X-ray crystallography has major limitations on determining structures of intrinsically disordered or flexible regions of proteins; however, SAXS provides valuable effort to structural genomic analysis in addition to conventional present tools in this field [17]. SAXS perhaps on its own is not always sufficient for generating highly detailed structures of large RNAs, can be combined with refined NMR structures to validate effectiveness of SAXS for assembling RNAs complexes observed in ribosomes. Structural models obtained from diffraction of ribosome crystals showed similarity between the existing model and thus effectiveness of using SAXS integrated with collection of high resolution NMR structures of RNAs for conformational sampling [18].

During early years of SAXS data collection, a variety of decentralized tools were developed for analysis of scattering data, but ATSAS 3.0 software is the most widely used and the most comprehensive software in terms of available programs for building models and visualization of biological small angle scattering data. Due to growing availability of SAXS data in Synchrotron facilities, more analysis tools have been developed including: Scatter [19], BornAgain [20], SasView [21], and Irena [22]. The available programs and tools in ATSAS are mainly developed for analysis of biological macromolecules and complexes in solution, however some programs are suitable for monodisperse and polydisperse systems in general [23]. In addition to model building, ATSAS package also provides a handy GUI for manual data analysis called PRIMUS primarily utilized for basic data manipulation. Obtained SAS data can be directly loaded into PRIMUS and essential analysis performed by generation of Guinier and Porod plots to derive basic parameters such as radius of gyration [24]. Evaluation of the particle distance distribution function $P(r)$ is critical for ab initio modeling and GNOM is used as an indirect transform program using 1-D SAS data as input [25]. ATSAS offers a variety of programs for ab initio modeling for deriving low-resolution three-dimensional shapes from small angle X-ray scattering of randomly oriented biological macromolecules in solution. DAMMIM, an ab initio modeling program in ATSAS package, uses a dummy atom model for building a multiphase model of a biological macromolecule. Initially a spherical space is filled with spherical dummy atoms in search volume, and the model is reconstructed according to the scattering data via characterizing each dummy atom solvent or scattering belonging to the molecule [26]. Additionally, for rapid ab initio shape determination, DAMMIF is developed by utilization of DAMMIM program for accelerated model generation from small angle scattering data [27]. For molecules with

experimentally determined atomic structures, CRY SOL is employed to fit small angle scattering curves of particles in solution into experimental coordinates of atoms obtained by X-ray crystallography [28] and later structures obtained by NMR [29]. Collective application of these programs provided by ATSAS is sufficient to build a three dimensional model from the original small angle scattering data comparable with higher resolution models obtained from X-ray crystallography or NMR.



Figure 1. n²STAR SAXS/WAXS/GISAXS beamline

Our state of the art research facility (n²STAR) is equipped with Anton Paar SAXSpoint 5.0 (Figure 1) with the Primux 100 high-performance 100 micro Cu microfocus sealed-tube X-ray source and 2D 1M EIGER2 R HPC detector and TrueSWAXS, allowing to choose the optimum q-range for your experiment over a large sample-to-detector distance (SDD) range from ≤ 45 mm to >1600 mm. In this study, usage of X-ray scattering method for biological molecule characterization in solution will be explained with exemplary data generated from a newly constructed small/wide angle X-ray scattering n²STAR beamline at Koç University. Additionally, a protocol for small/wide angle X-ray scattering is developed for future studies utilizing home source SAXS/WAXS beamlines.

2. Material and Methods

2.1. Protein expression and purification

Genes encoding recombinant proteins (UbcH8, UbcH7, E6AP HECT) were cloned into N-terminal (Glutathione S-transferase) GST-tag containing pGEX expression vectors for previous studies and protein purification steps applied as described before [30, 31]. GST tagged proteins are purified with 5 ml GST columns and the elution is treated with Thrombin to cleave the GST from the target protein. Cleaved target proteins are separated from GST by reverse GST column chromatography. Size-exclusion chromatography (Superdex S200 Cytiva) was applied to isolate ultra-pure GST protein. The column was equilibrated using (50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5) and the sample was injected to the column. Fractions containing ultra-pure GST were determined by SDS-PAGE. Same steps applied also for GST-UbcH8 and GST-UbcH7 (data not shown). The fractions were combined and concentrated to 4 mg/ml using 10 kDa

cutoff centrifugal filters for SAXS measurements. Free GST obtained from a variety of samples (UbcH8, UbcH7, E6AP HECT) were mixed altogether in order to increase the final protein concentration for SAXS analysis.

2.2. SAXS beamline preparation

In order to start the device, the main switch which is located on the back of the device is turned on. To gather data, the gap between the sample and the detector must be kept at a low vacuum to avoid air scattering. This air scattering can be apparent when it is compared with the low scattering signal of macromolecules. For this purpose, vacuum pumps are switched on (Figure 2A).

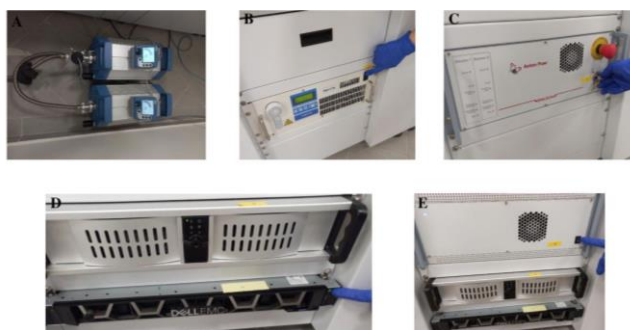


Figure 2. SAXS beamline preparation steps

The low amount of air results in the decrease of the direct interaction of the air with the direct beam coming from the light source. After the vacuum pumps and chillers (Figure 2B) are turned on, which are used to adjust the temperature, the X-ray source/generator (Figure 2C), the detector controller (Figure 2D), computer and the sample unit controller (Figure 2E) are switched on respectively.

2.3. Sample Load

There are two possible setups for biomolecular data collection in solution. First one is the static cell; when a static cell measurement is performed, the loaded sample is exposed to X-rays for some amount of time. This option may be preferred if there is a small amount of sample, the sample is obtained in a difficult way, the sample is expensive, or it is scattering strongly. Second option is the flow cell. During the flow cell measurement, the sample that is exposed to X-ray beams is exchanged with a fresh sample. This measurement type may cost more or necessitate a larger amount of sample, but it provides longer measurements without radiation damage and also it makes it possible to measure the signals that are weakly scattered. In our case, the sample is loaded to the static quartz tube (Figure 3A) using a micropipette. Before loading the sample, the capillary tube is washed with distilled water and dried using dry air for better results as there might be some deposits on the sample cell walls. After the washing, the sample is loaded (Figure 3B). Capillary tube should not contain any air bubbles. Once the sample is loaded, one side of the tube is closed with one finger while closing the other side of the tube. Then the other side of the tube is closed (Figure 3C).

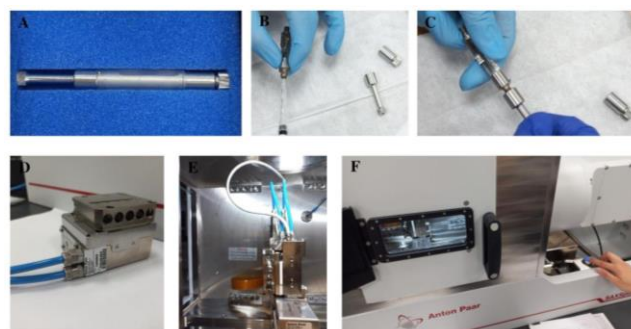


Figure 3. SAXS sample load

The quartz tube is then placed in the sample holder. Heated/Cooled sample holder (Figure 3D) is used to adjust the temperature. Data from sample and buffer (base line) should be collected using the same exposures, in the same cell and in the same position. Additionally, it is recommended to start with the buffer data acquisition and then continue with the biomolecular sample using the same data collection parameters. The system is evacuated after placing the tube into the sample holder. When the required pressure is reached, the system components within the device are aligned for data acquisition.

2.4. Data Acquisition and Processing

Before the acquisition, the temperature is adjusted to 10 °C. Sample/detector distance (SDD) is set to 1600 mm for SAXS experiments. Furthermore, the position of the beam stopper is checked, and it is appropriately aligned with the incoming beam. The acquisition time and the number of frames is indicated.

The transmission measurement is included in the data acquisition. After the data collection, SAXS output file (.dat) is obtained. This file needs to be processed in order to obtain the scattering pattern of the sample. For this purpose, the output file is uploaded to the SAXS analysis program and by clicking on "Source data", the data files that are used are chosen. These files are the output file for the protein sample and the output file for the buffer/background. The first process is created, and all data files are selected to be processed (Figure 4). First tool that is used is "Zero Point by Moments 2D" to select the primary beam coming from the X-ray source. Then "Masking" is performed in order to mask out the beam stop to exclude its influence on the sample data. To include transmission measurement into data acquisition, "Transmission" option is selected in the created "Q Transformation" tool. In "Data Reduction Step", the whole data that is processed is chosen by entering the appropriate range values for x,y and angle parameters.

The data is radially integrated. In the first "Standard Operations 1D" tool, transmittance normalization is done. This way the number of X-rays that are passing through the sample is compared to when there is no sample inside by selecting the "Transmittance" option again. After these steps, another process is created to now process the obtained 1D data.

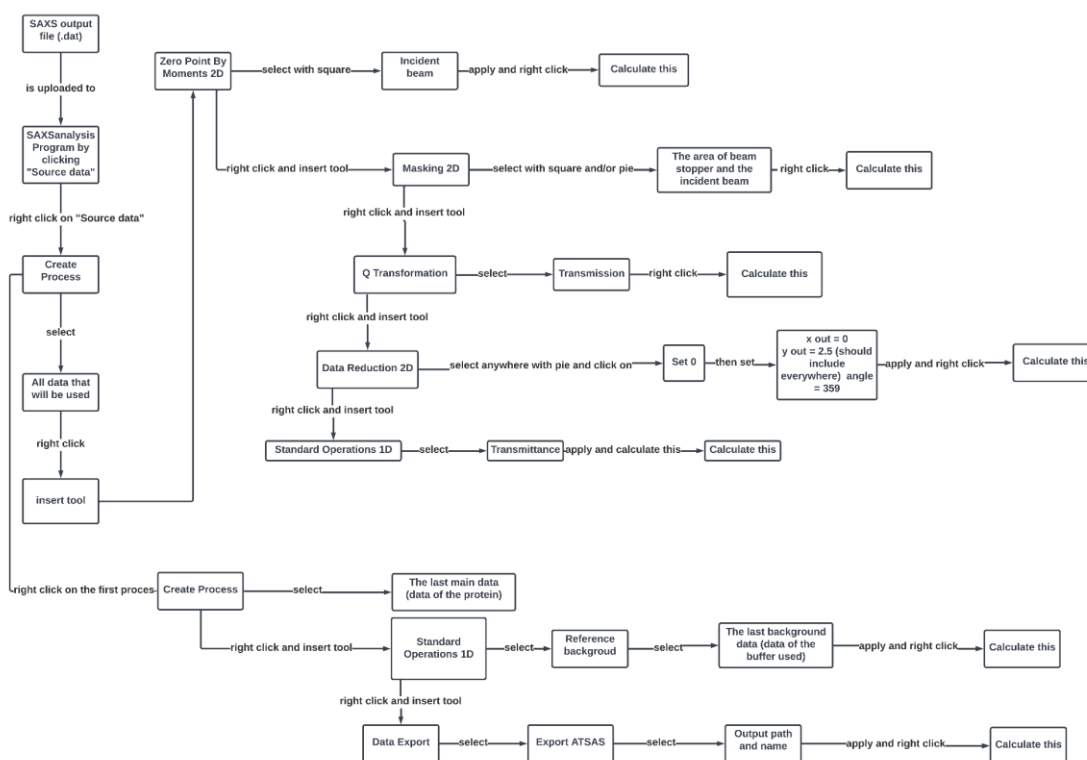


Figure 4. Data processing flowchart

In order to subtract the background scattering from the scattering of the sample in the buffer, "Standard Operations 1D" tool is created once again. The lastly processed background scattering data is selected as the "Reference background". As the last step, the data file is obtained from the SAXS analysis program to be compatible with the ATSAS program. Other file types that are compatible with other processing and analysis programs are available. The appropriate file type should be chosen according to the program that will be used for data processing.

2.5. Data Analysis

The scattering pattern of the protein in the buffer solution is determined by subtracting the scattering pattern of the buffer solution from that of the protein in the buffer solution as indicated in Figure 4. This step is necessary to obtain the distribution belonging only to the protein. The output file containing the scattering information of the protein is saved in a file type that can be analyzed using the ATSAS 3.0 program [23]. Logarithmic graphs are obtained by loading the output file into the PRIMUS package in ATSAS 3.0. This step is performed to understand the possibility of UV damage to the sample and precipitation of the sample in order to increase the final data quality. If there is any radiation damage or precipitation, repeating the experiment is recommended for optimization of experiment parameters. The radius of gyration is computed using Guinier analysis (Guinier, 1939) and Fourier transformation which are included in the PRIMUS package [24]. GNOM is used to determine the particle radius (Dmax) [25]. Packages in PRIMUS can be used to estimate the molecular weight of the molecule of interest. Using the DAMMIF program, the low resolution structure

of the molecule is created by "ab initio" modeling method and DAMAVER is used to obtain an averaged and optimized model of all models created by DAMMIF [27]. Moreover, the CRYSOLO program [32] is used to compare and check the matching of the created models with the structural information that is contained in the protein data bank.

3. Results and Discussion

Preliminary data showed that protein purity is very important for SAXS experiments (data not shown). In order to improve the data quality, gel filtration chromatography applied to samples. Purity and concentration of the samples are validated by SDS PAGE. The protein band in the Figure 5b in the well 6 is weaker than the band in the well 4 because, after 1st sample batch is loaded, the column is washed with the wash buffer to remove all the proteins and this causes dilution of the flowthrough sample. Also, the reason for the thinness of the band in the figure 5b well 7 is washing the column with greater amount of buffer during elution. Ultrapure free GST fractions from gel filtration chromatography were collected and concentrated to 4 mg/ml and 7 mg/ml levels with 10 kDa cutoff centrifugal filter. The optimal concentration for collecting small angle scattering data from free GST was not known and the effect of protein concentration to the quality of SAXS data was aimed to be determined by comparing findings from both concentrations (Figure 5). According to the findings, there was a significant difference in the quality of the unprocessed SAXS data visibly observable on one dimensional data plot but also according to statistical indicators generated by PRIMUS data processing tool (Figure 6). The rest of the analysis was only performed with the data collected from the 7 mg/ml Free GST

sample. Unlike inorganic compounds, biomolecules are highly sensitive to X-ray radiation generated by the light source; the quality of data collected is expected to decrease in a time-dependent manner.

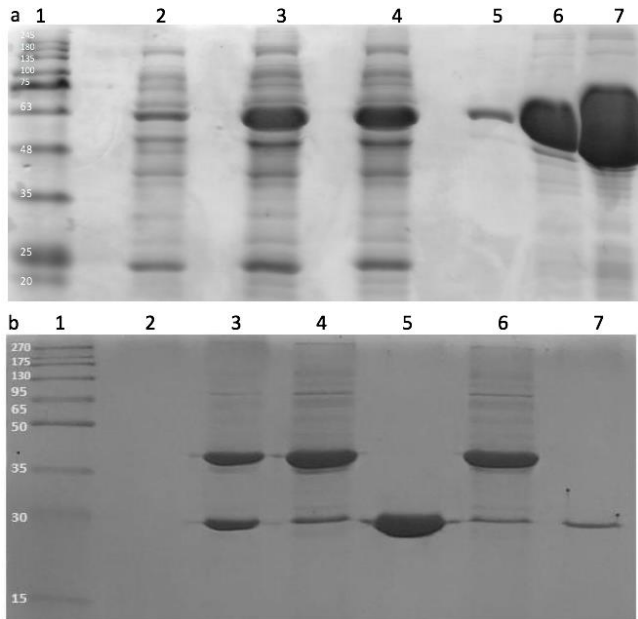


Figure 5. SDS-PAGE analysis of recombinant GST protein a) (1) Marker (2) GST-E6AP HECT pellet (3) load (4) flowthrough (5), (6), (7) elution; b) (1) Marker (2)space before samples (3) E6AP_HECT (42 kDa) after enzymatic cut (4) reverse GST (28 kDa) flowthrough 1 (5) reverse GST elution 1 (6) reverse GST flowthrough 2 (7) reverse GST elution 2

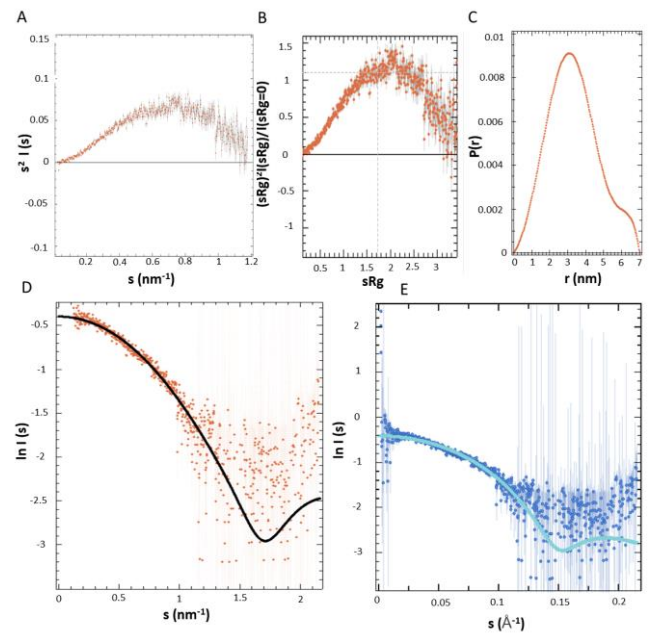


Figure 7. Kratky and CRY SOL analysis of the GST homodimer

(A) Kratky plot of free GST (B) Dimensionless Kratky plot of free GST (C) Distance distribution plot of free GST scattering data calculated by GNOM (D) the fit of the model to the experimental data calculated by GNOM (E) Fit of the calculated rigid body model by CRY SOL to experimental scattering (PDB: 1GNE)

In order to check the effect of radiation damage to data collection, data was collected in 10-minute frames and before it was combined into one data file, possible alterations of the data were checked by superimposing data collected in each interval. During test runs with previous samples, one hour data collection was sufficient for generation of three dimensional models from the obtained data. However, collection time depends on a variety of factors ranging from concentration of the sample to aggregation or oligomerization of the proteins in solution. Nevertheless, one-hour data collection was sufficient for many proteins (data not shown). Radiation damage is not observed with any samples for one hour of data collection compared via 10-minute frames. Due to lack of data availability of protein SAXS data obtained from non-synchrotron-based X-ray devices, there is no established protocol for data collection with weaker beamlines; however, data collection up to 5 hours with 30 minute intervals is possible without significant radiation damage in the sample [33]. SAXS data obtained for the 7 mg/ml GST sample was analyzed using tools provided by PRIMUS in 3.0 ATSAS package. GST is naturally observed as a homodimer according to protein crystal structures thus the dimer form of GST was used for comparison with the model generated by SAXS. As SAXS data alone provides basic structural information according to the scattering profile with Guinier plots and Kratky plots (Figure 7), it was confirmed that GST was in fact in dimer form by comparing the estimated molecular weight from the data to the monomer of GST by PRIMUS. Radius of gyration is a valuable measurement calculated

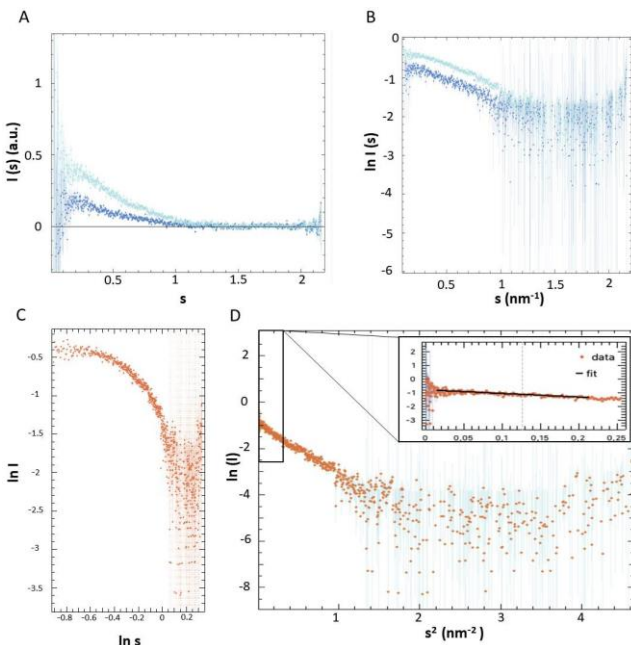


Figure 6. Solution SAXS analysis of the GST protein (A) $I(s)$ vs s plot of the experimental SAXS intensity obtained with 4.1 mg/ml GST (blue) and 7.0 mg/ml GST (turquoise) (B) $\ln I(s)$ vs s plot plot of the experimental SAXS intensity obtained with 4.1 mg/ml GST (blue) and 7.0 mg/ml GST (turquoise) (C) $\ln I$ vs $\ln s$ plot plot of the experimental SAXS intensity obtained with 7.0 mg/ml GST (D) Guinier plot obtained from 7.0 mg/ml GST

by the Guinier analysis as it specifies both weight and size of a particle around an axis without generating a model. In order to verify the dimer conformation of GST, Radius of gyration generated by PRIMUS (R_g : 25.50 Å) was compared to radius of gyration generated by CRY SOL analysis of dimeric form GST (R_g : 22.48 Å) (Table 1). Despite the slight difference between estimated and calculated radius of gyration values, the resemblance between R_g values is sufficient to confirm the small angle scattering is generated by a GST homodimer. Homodimer of free GST with the additional residues remaining from enzymatic cleavage of GST has a theoretical mass of 55.80 kDa. The molecular mass was estimated by ATSAS as 59.3 kDa and by Porod analysis as 58.8 kDa (Table 1). In spite of the difference in the mass estimation compared to the theoretical molecular mass, the error rates are not expected to exceed 10 % [34] in this experiment. The molecular mass estimation error does not exceed 10 % in either estimation, also confirming monodisperse non-aggregated characteristics of our homodimer GST SAXS sample.

Table 1. Molecular Size Parameters Obtained for GST from SAXS Data Analysis

Sample ID	R_g (Å)	$R_{g, \text{CRY SOL}}$	V_p , nm ³	D_{max} , nm	MMI(0) kDa	MM _{aa} kDa	MM _{Porod} kDa
GST	25.50	22.48	80.20	7.03	59.30	55.80	58.80

Considering Radius of gyration and MM estimates were supporting a non-aggregated properly folded homodimer of GST molecules, Kratky plot was generated to investigate flexibility and disclose globular nature of the confirmed GST dimers. Folded and globular proteins expected have a Kratky plot with a well-defined bell-curve that approaches zero after maxima at higher q values [35]. In Figure 7A, the Kratky transformation is applied to GST SAXS data confirming folded and globular conformation of GST dimers as there is a clear bell curve with a peak around ~ 2 sRg. In Figure 7B, dimensionless Kratky plot naturally gives the same result with a distinct bell curve that converges to 0 with increase of s . Pair distance distribution function $P(r)$ is generated as another indicator of globular conformation of the scattered particle (Figure 7C) In Figure 7E, a fitting curve is generated to validate the curve with high resolution models of GST dimers. As high resolution models of GST dimers exist, CRY SOL was run with a GST model from PDB with residual extensions due to cleavage from another protein (PDB:1GNE) (Figure 7E). According to CRY SOL analysis there are considerable similarities between the fitting curve of experimental data and the curve of GST homodimer generated by CRY SOL. As model-free analysis was satisfactory, a three dimensional model from experimental data was generated by using DAMMIF in the ATSAS package. DAMMIF was run 20 times and the average of these models was used to generate the most possible conformation of GST dimer in solution. Three dimensional model of GST homodimer (Figure 8) was aligned with the same model used for CRY SOL analysis. Remarkably, residual loop remaining

after cleavage of GST is also detected by DAMMIF and generated as an extension to the globular model of the GST dimer (Figure 8). The residual loop on GST dimer is obtained from a static crystal lattice and the actual confirmation of this loop could be more flexible in solution, which SAXS is capable of detecting (Figure 8B). SAXS compared to conventional structural biology tools of X-ray crystallography, biological NMR spectroscopy is a tool capable of acquiring structural information about a biological macromolecule's natural conditions in solution, without size limitation, and by matter of hours. Despite providing low resolution, SAXS can be a uniting ground for structural analysis of biological macromolecules as SAXS is capable of compensating the shortcomings of the other structural biology tools capable of generating structures with atomic accuracy.

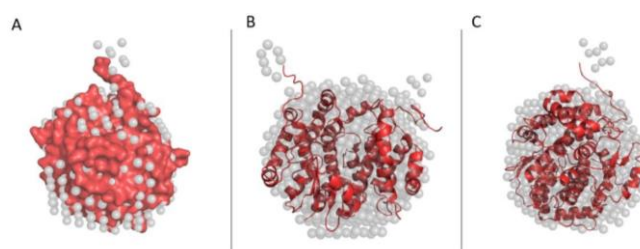


Figure 8. Reconstructed SAXS envelope of GST in dimeric form

(A) 3D reconstruction of dimeric GST superimposed with surface representation of high resolution PDB (PDB:1GNE) structure (B) a counterclockwise rotation of the protein around the vertical axis with 90-degree (C) 3D reconstruction of dimeric GST superimposed with cartoon representation of high resolution PDB structure

X-ray crystallography among structural biology tools is commonly used for generating protein models close to atomic accuracy. In order to obtain proper diffraction from biological molecules, proteins are fixed and ordered in a crystal lattice limiting flexibility and due to changing environmental conditions affecting certain conformational characteristics of protein when they are in physiological conditions [36]. The data collection with SAXS only took an hour compared to necessary time for crystallization of the protein which can take months to achieve. SAXS' role in structural biology would not be replacing X-ray crystallography but change how high throughput structural biology research takes place even in research centers without access to tools necessary for high throughput X-ray crystallography. In a matter of hours, the dimeric form of GST was validated by a simple SAXS analysis, demonstrating effectiveness of SAXS for determining basic structural characteristics of biomolecules. Considering SAXS data was collected with a home-source X-ray, the data generated by this system is still capable of detecting and differentiating subdomains of fusion proteins in addition to oligomerization of proteins (Figure 9 A-N). Atomic resolution structures are certainly useful for a variety of studies in silico, however SAXS models alone could be sufficient to determine most basic characteristics of

proteins if there is no obtaining high resolution structures. Furthermore, integration of the conventional methods in structural analysis with SAXS could generate a considerable amount of data in a significantly short amount of time which could be useful as a high throughput analysis in the growing omics field of molecular biology. In recent years SAXS has been utilized for interdisciplinary structural biology studies and particularly useful as a quick analysis of large biological molecules of the stalk domain of SARS-CoV-2 spike protein [33].

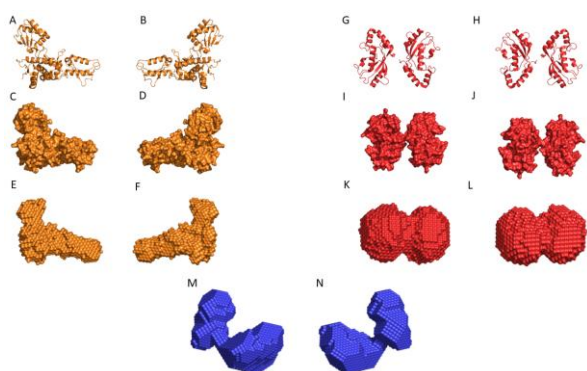


Figure 9. Examples of 3D reconstructions obtained from the data collected at n²STAR beamline (A) and (B) Cartoon representation of target protein 1 PDB structure; (C) and (D) Surface representation of target protein 1 PDB structure; (E) and (F) Surface representation of target protein 1 (multi subdomain) model reconstruction from SAXS data (G) and (H) Cartoon representation of target protein 2 PDB structure; (I) and (J) Surface representation of target protein 2 PDB structure; (K) and (L) Surface representation of target protein 2 (dimeric) model reconstruction from SAXS data; (M) and (N) Surface representation of target protein 3 (fusion) model reconstruction from SAXS data.

The environmental conditions including ligand or cofactor concentrations in solutions, temperature and pH can be manipulated in SAXS samples allowing detection of changes in the conformation of proteins in structural basis. Additional SAXS data including possible environmental changes is a good way of supporting the static X-ray crystal or cryo EM data to generate more reliable results in structural analysis research. The biggest advantage of the home-source SAXS device is, to be able to collect data about your sample without traveling long distances. As data acquisition with home-source X-ray becomes easier with advancements of protocols and data analysis, facilities like n²STAR can become a practical center for local scientists. In terms of quality of the data, heterogeneous samples constitute the biggest problem when compared to the data collected in synchrotrons as it is more difficult to conduct home source SAXS studies with heterogeneous samples. Tuning beam in particular, often causes problems that slow down and hence make data acquisition more expensive. While it takes 50-60 minutes to collect scattering of a sample with a homesource SAXS device, this time is as short as 100-200 ms for synchrotrons;

hence, screening studies with home source is more difficult compared to studies at a synchrotron facility. Furthermore, the quality of the detector is another component of the subject. Synchrotrons, as anticipated, have higher quality detectors and the overall quality of the data is expectedly higher. However, this paper addresses this issue not being a full replacement for data collection rather indicates home-source as a considerable alternative to synchrotron X ray sources.

4. Conclusion

SAXS is still not plug and play, and many parameters need to be optimized before getting successful results with particularly home-source SAXS devices. SAXS gives better results when the sample is ultra-pure, stable and the particles are in the same conformation. The set up for SAXS data collection can be enhanced by including in-line size exclusion chromatography to improve sample purity as a standard approach for home source SAXS data collection [37]. In addition to current data analysis and model construction algorithms, the growing field of deep learning is also contributing to the development of novel methods for more accurate three-dimensional model construction [38]. Despite no deep learning approach being used in this study, it is likely as SAXS becomes more popular and deep learning algorithms developed for scientific data analysis, the approach of SAXS analysis of data can utilize such algorithms. Data entries to SAXS data bases with home source SAXS devices is still low, considering their potential benefits and its availability compared to a synchrotron facility. This number will increase day by day with the development of new generation detector technologies and new generation X-ray sources and better algorithms for analysis of SAXS data. Since structural studies are now focused on more complex systems consistently involving multiple components, SAXS became a more prevalent method for data accusations particularly in environmental diversifications. As one of the biggest advantages of SAXS is elimination of the crystallization stage, with bigger and more fragile complexes, SAXS is more applicable and a faster alternative to conventional crystallography [39].

5. Acknowledgment

We are grateful to Prof. Ziarek at Indiana University-Bloomington for constructive feedback and we appreciate Ziarek Juniors for their understanding. CD acknowledges support from TÜBİTAK (Project No: 120Z594, 221S535, 122Z747). The authors acknowledge the use of the services and facilities of n²STAR-Koç University Nanofabrication and Nanocharacterization Center for Scientific and Technological Advanced Research.

6. References

- [1] Kane, P.P., Kissel, L., Pratt, R.H. and Roy, S.C., "Elastic scattering of γ -rays and X-rays by atoms", *Physics Reports*, Vol. No. 140(2), 75-159, 1986.
- [2] Kane, P.P., "Inelastic scattering of X-rays and gamma rays by inner shell electrons", *Physics Reports*, Vol. No. 218(2) 67-139, 1992.

- [3] Chavas, L.M.G., "Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules", *Acta Crystallographica Section D: Biological Crystallography*, Vol. No. 70(4), 1175-1176, 2014.
- [4] Makowski, L., "Characterization of proteins with wide-angle X-ray solution scattering (WAXS)", *Journal of Structural and Functional Genomics*, Vol. No. 11(1), 9-19, 2010.
- [5] Das, R. and Doniach, S., (Pecora, R. and Borsali, R.), Structural Studies of Proteins and Nucleic Acids in Solution Using Small Angle X-Ray Scattering (SAXS), *Soft Matter Characterization*, 1083-1108, 2008.
- [6] Pauw, B.R., "Everything Saxes: Small-angle scattering pattern collection and correction", *Journal of Physics: Condensed Matter*, Vol. No. 25(38), 383201, 2013.
- [7] Skou, S., Gillilan, R.E. and Ando, N., "Synchrotron-based small-angle X-ray scattering of proteins in solution", *Nature Protocols*, Vol. No. 9(7), 1727-1739, 2014.
- [8] Grant, T.D., Luft, J.R., Wolfley, J.R., Tsuruta, H., Martel, A., Montelione, G.T. and Snell, E.H., "Small angle X-ray scattering as a complementary tool for high-throughput structural studies", *Biopolymers*, Vol. No. 95(8), 517-530, 2011.
- [9] Matthews, B. W., "X-ray crystallographic studies of proteins", *Annual Review of Physical Chemistry*, Vol. No. 27(1), 493-493, 1976.
- [10] Valentini, E., Kikhney, A.G., Previtali, G., Jeffries, C.M. and Svergun, D.I., "SASBDB, a repository for biological small-angle scattering data", *Nucleic Acids Research*, Vol. No. 43(D1), 2014.
- [11] Malfois, M. and Svergun, D.I., "SASCIF: An extension of core crystallographic information file for SAS", *Journal of Applied Crystallography*, Vol. No. 33(3), 812-816, 2000.
- [12] Kachala, M., Westbrook, J. and Svergun, D., "Extension of the sascif format and its applications for data processing and deposition", *Journal of Applied Crystallography*, Vol. No. 49(1), 302-310, 2016.
- [13] Korasick, D.A. and Tanner, J.J., "Determination of protein oligomeric structure from small-angle X-ray scattering", *Protein Science*, Vol. No. 27(4), 814-824, 2018.
- [14] Schindler, C.E., de Vries, S.J., Sasse, A. and Zacharias, M., "SAXS data alone can generate high-quality models of protein-protein complexes", *Structure*, Vol. No. 24(8), 1387-1397, 2016.
- [15] Gomes, G.N.W., Krzeminski, M., Namini, A., Martin, E.W., Mittag, T., Head-Gordon, T., Forman-Kay, J.D. and Gradinaru, C.C., "Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET", *Journal of the American Chemical Society*, Vol. No. 142(37), 15697-15710, 2020
- [16] Belviso, B.D., Mangiatordi, G.F., Alberga, D., Mangini, V., Carrozzini, B. and Caliendo, R., "Structural Characterization of the Full-Length Anti-CD20 Antibody Rituximab", *Frontiers in Molecular Biosciences*, Vol. No. 9, 823174, 2022.
- [17] Hura, G.L., Budworth, H., Dyer, K.N., Rambo, R.P., Hammel, M., McMurray, C.T. and Tainer, J.A., "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)", *Nature Methods*, Vol. No. 6(8), 606-612, 2009.
- [18] Dagenais, P., Desjardins, G. and Legault, P., "An integrative NMR-SAXS approach for structural determination of large rnas defines the substrate-free state of a trans-cleaving Neurospora Varkud Satellite ribozyme", *Nucleic Acids Research*, Vol. No. 49(20), 11959-11973, 2021.
- [19] Förster, S., Apostol, L. and Bras, W., "Scatter: software for the analysis of nano- and mesoscale small-angle scattering", *Journal of Applied Crystallography*, Vol. No. 43(3), 639-646, 2010.
- [20] Pospelov, G., Van Herck, W., Burle, J., Carmona Loaiza, J.M., Durniak, C., Fisher, J.M., Ganeva, M., Yurov, D. and Wuttke, J., "Bornagain: Software for simulating and fitting grazing-incidence small-angle scattering", *Journal of Applied Crystallography*, Vol. No. 53(1), 262-276, 2020.
- [21] Tan, L., Elkins, J.G., Davison, B.H., Kelley, E.G. and Nickels, J., "Implementation of a self-consistent slab model of bilayer structure in the sasview suite", *Journal of Applied Crystallography*, Vol. No. 54(1), 363-370, 2021.
- [22] Ilavsky, J. and Jemian, P.R., "Irena: tool suite for modeling and analysis of small-angle scattering", *Journal of Applied Crystallography*, Vol. No. 42(2), 347-353, 2009.
- [23] Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D., Gruzinov, A., Borges, C. and Jeffries, C.M., "Atsas 3.0: Expanded functionality and new tools for small-angle scattering data analysis", *Journal of Applied Crystallography*, Vol. No. 54(1), 343-355, 2021.
- [24] Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H. and Svergun, D.I., "Primus: A windows PC-based system for small-angle scattering data analysis", *Journal of Applied Crystallography*, Vol. No. 36(5), 1277-1282, 2003.
- [25] Svergun, D. I., "Determination of the regularization parameter in indirect-transform methods using perceptual criteria", *Journal of Applied Crystallography*, Vol. No. 25(4), 495-503, 1992.
- [26] Svergun, D.I., "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing", *Biophysical Journal*, Vol. No. 76(6), 2879-2886, 1999.
- [27] Franke, D. and Svergun, D. I., "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering", *Journal of Applied Crystallography*, Vol. No. 42(2), 342-346, 2009.
- [28] Svergun, D.I.B.C., Barberato, C. and Koch, M.H., "Crysol-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates", *Journal of Applied Crystallography*, Vol. No. 28(6), 768-773, 1995.
- [29] Mertens, H.D. and Svergun, D.I., "Combining NMR and small angle X-ray scattering for the study of biomolecular structure and Dynamics", *Archives of Biochemistry and Biophysics*, Vol. No. 628, 33-41, 2017.
- [30] Tokgöz, Z., Siepmann, T.J., Streich, F., Kumar, B., Klein, J.M. and Haas, A.L., "E1-E2 interactions in Ubiquitin and NEDD8 ligation pathways", *Journal of Biological Chemistry*, Vol. No. 287(1), 311-321, 2012.
- [31] Ronchi, V.P., Klein, J.M. and Haas, A.L., "E6AP/UBE3A ubiquitin ligase harbors two e2~ubiquitin binding sites", *Journal of Biological Chemistry*, Vol. No. 288(15), 10349-10360, 2013.
- [32] Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. and Svergun, D., "Atsas 2.8: A comprehensive data analysis suite for small-angle scattering from Macromolecular Solutions", *Journal of Applied Crystallography*, Vol. No. 50(4), 1212-1225, 2017.
- [33] Živič, Z., Strmšek, Ž., Novinec, M., Lah, J. and Hadži, S., "Structural polymorphism of coiled-coils from the stalk

domain of sars - cov - 2 Spike protein” , *The FASEB Journal*, Vol. No. 36(3), e22199, 2022.

[34] Mylonas, E. and Svergun, D.I., “Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering”, *Journal of Applied Crystallography*, Vol. No. 40(S1), 245-249, 2007.

[35] Burger, V.M., Arenas, D.J. and Stultz, C.M., “A Structure-free Method for Quantifying Conformational Flexibility in proteins”, *Scientific Reports*, Vol. No. 6(1), 1-9, 2016.

[36] Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A., “X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies

in solution”, *Quarterly Reviews of Biophysics*, Vol. No. 40(3), 191–285, 2007.

[37] Graewert, M.A., Da Vela, S., Gräwert, T.W., Molodenskiy, D.S., Blanchet, C.E., Svergun, D.I. and Jeffries, C.M., “Adding SEC and Light Scattering (LS) Devices to Obtain High-Quality SAXS Data”, *Crystals*, 10(11), 975, 2020.

[38] He, H., Liu, C. and Liu, H., “Model Reconstruction from Small-Angle X-Ray Scattering Data Using Deep Learning Methods”, *Iscience*, Vol. No. 23(3), 100906, 2020

[39] Sayers, Z., Avşar, B., Cholak, E., & Karmous, I. “Application of advanced X-ray methods in life sciences”. *Biochimica et Biophysica Acta*, Vol. No. 1861(1), 3671-3685, 2017