BIOLOGICAL SMALL ANGLE X-RAY SCATTERING AT CEITEC-MU

T. Klumpler[1], T. Gerguri[2], O. Jasnovidova[1], L. Krejčí[2], K. Kubíček[1], D. Němeček[1], V. Sedláček[3], R. Štefl[1]

[1]Středoevropský technologický institut, CEITEC-MU, Kamenice 5, 62500 Brno
[2]Biologický ústav, MU, LF, Kamenice 5, 62500 Brno
[3]Ústav biochemie, MU, PřF, Kamenice 5, 62500 Brno
klumpler@sci.muni.cz

Keywords: SAXS, Bio-SAXS, complex, flexible, oligomeric

Abstract

Biological Small Angle X-ray Scattering (Bio-SAXS, SAXS) becomes mature and popular technique for structural studies of the macromolecules and macromolecular complexes in solution. Development of the software tools and advances of synchrotron and „in house" X-ray sources brought Bio-SAXS to routine work-flow of number of structural biologists. SAXS is used for determination of the integral structural parameters, shape reconstruction, determination of the oligomeric and folding state, unraveling the quaternary architecture of the complexes, modeling of molecular flexibility etc. SAXS characterizes macromolecules in solution, *i.e.* close to their native and biologically relevant conditions. It is a low-resolution technique, but in combination with other techniques as X-ray crystallography, nuclear magnetic resonance, electron microscopy, etc., the SAXS becomes powerful and complementary tool of the structural analysis of biological macromolecules.

The Core facility X-ray diffraction and Bio-SAXS of the CEITEC-MU located in Brno facilitate access to the state of art "in house" instrumentation for X-ray structural analysis. Besides the elementary collection of diffraction or scattering data, the facility offers assistance with data processing and interpretation. Year and half after the official opening of the laboratory the most typical SAXS case studies from users community are presented: *ab initio* shape reconstruction experiments, oligomeric state determination, quaternary model building of macromolecular complexes and studies of semi-flexible complexes and intrinsically disordered proteins. Presented SAXS characterization studies should be understood as methodical examples, no biological details are mentioned.

Introduction

Biological Small Angle X-ray Scattering (Bio-SAXS, SAXS) is widely used technique for low resolution structural studies of biomacromolecules and macromolecular complexes in solution. Ongoing development of the software tools, *e.g.* [1,2] dedicated to extract structural information from scattering data of biological systems made SAXS analysis user friendly and relatively fast. Automation of data acquisition and processing, combination with on-line size-exclusion chromatography or rapid mixing stopped-flow allows the high-throughput approach and broader application for structural biologists as rapid screening or time-resolved studies, *e.g.* [3]. In contrast to X-ray crystallography, where diffraction patterns of ordered matter lead to electron density map of up to atomic resolution, SAXS is low resolution technique. In SAXS, randomly oriented macromolecules in ideal solution scatters X-ray isotropically and the scattered intensity depends on the modulus of the momentum transfer $q = 4 \sin$ (where 2 is the angle between the incident and scattered beam):

$$I(q) \quad \langle I(q) \rangle \qquad \left\langle A(q)A^{*}(q) \right\rangle$$

where the scattering amplitude A is the Fourier transform of the excess scattering length density and the scattering intensity is average over all orientations ( ), see [4]. Radially averaged scattered intensity are usually further processed as one dimensional scattering profiles. Each SAXS experiment requires at least two measurements: sample and solvent, when the solvent subtracted, background/instrument corrected intensity is proportional to the scattering of a single particle averaged over all orientations. Several integral structural parameters are directly obtained from the intensity profile, as radius of gyration, volume and maximum diameter of the studied macromolecule or complex.The spherical harmonics representation of the particle or approximation of the Debye formula allows rapid computation of simulated scattering from model structures, for review [4]. *Ab initio* shape reconstruction algorithms *DAMMIN*, *DAMMIN* and *GASBOR* produce models composed of densely packed beads to describe low resolution shape of the macromolecule in solution from its scattering data [5–7]. Beads are representing spherical pieces of electron density of certain radius (dummy atoms) or spherical approximation of electron density of average amino acid (dummy residues). Rigid body methods use atomic structures/substructures known from complementary methods as X-ray crystallography, nuclear magnetic resonance - NMR, molecular dynamics, etc.) [8, 9]. Both, *ab initio* and rigid body methods utilize trial-error model modification algorithms to minimize the discrepancy between the experimental data and the simulated scattering. By using this approach the particle shape is reconstructed, atomic structures are rapidly validated/disproved, quaternary structure of the macromolecular complexes are modeled or ensemble of conformation are selected. SAXS characterize the macromolecules in solution, what allows to observe macromolecules in biologically relevant conditions and effects of altering buffer composition, including *e.g.* screening and optimization of stable complex formation or identification of crystallization conditions.

The Core facility X-ray diffraction and Bio-SAXS of the CEITEC-MU located in Brno facilitates access to the state of art "in house" instrumentation for X-ray structural

analysis: two X-ray diffractometers and Bio-SAXS Kratky camera system. Services of the facility include X-ray diffraction data collection, screening and selection of diffraction quality crystals and SAXS data acquisition. Besides the elementary collection of diffraction or scattering data, the facility offers assistance with data processing, solving the crystal structures, SAXS based sample characterization and modeling. Facility was inaugurated in September 2013. Here the most typical SAXS case studies from users community are presented. Examples of informative *ab initio* shape reconstruction experiments, oligomeric state determination, quaternary structure model building of macromolecular complexes and studies of semi-flexible complexes and intrinsically disordered proteins.

Material and methods

Data collection

The SAXS data were collected on the BioSAXS-1000, Rigaku at CEITEC (Brno, Czech Republic). Data were collected at X-ray beam wavelength = 1.54 Å. Sample to detector (PILATUS 100K, Dectris Ltd.) distance was 0.4 m covering a scattering vector range from 0.008 to 0.65 Å^{-1}. For buffer and sample one two-dimensional image was collected with an exposure time of 60 min per image. Data collection and scattering-derived parameters are summarized in Table 1.

Data reduction and plotting

Data reduction and the buffer subtraction were performed using SAXSLab, Rigaku. Subtracted data was normalized to the protein concentration using *PRIMUS* [10] and truncated individually (mentioned in Results section). Scattering data is plotted using *SASPLOT* [10] with artificial intensity units on logarithmic scale against the modulus of the momentum transfer $q = 4 \sin$ [Å^{-1}] and as Kratky plots, where $q^2 I(q)$ is plotted as a function of q.

Software exploited

Most software tools used are from software package *ATSAS* v.2.5.2 [9]. Evaluation of the solution scattering of the atomic models and the fitting to experimental data was performed by *CRYSOL* [9]. *Ab initio* modeling was performed by *DAMMIN* [5]. Comparison, averaging and filtering of individual *ab initio* models was performed by program suite *DAMAVER* [11]. Superimposition of *ab initio* and atomic models was performed using *SUPCOMB (12)*. The oligomeric state of the protein was evaluated using *OLIGOMER* [10], where the form-factors files were created by *FFMAKER* [10] using atomic structures of most probable multimeric assemblies proposed by *PDBePISA* [13] or generated by crystal symmetry using *UCSF Chimera* [14]. Rigid body modeling was performed using *CORAL* [9] and *EOM* [15]. Homology modeling was performed by *iTASSER* [16].

Biological material

Described SAXS studies of users projects should be understood as methodical examples. No biological details and no full identifiers of biological macromolecules are stated.

Experimental results and discussion

1. *Ab initio* shape reconstruction

Ab initio shape reconstruction is routine step in SAXS data analysis, thus often automated pipelines are used to quickly evaluate the average shape of studied particles. The bead models are usually available for every measured SAXS profile with sufficient data quality. Although *ab initio* models are limited to monodisperse systems and are generally less informative than rigid body methods, examples of informative shape reconstruction are not rare. ProteinO forms octameric complex (8-fold rotational symmetry) of total mass of 301.5 kDa (2720 amino acid moieties). Known crystal structure of the octameric ProteinO lacks 808 amino acid moieties (~ 30%), probably due to flexible nature of the significant portion of the polypeptide chain.

Table 1. Data collection parameters and integral structural parameters derived from scattering data.

	Protein0	Protein0+ ATP+Ca2+	FerA	FerA	ProteinA +ProteinB	CIP	CTD	complex CIP+CTD
Data-collection parameters								
Instrument	BioSAXS-1000							
Wavelength (Å)	1.5418							
q range (Å^{-1})	0.008 – 0.25	0.008 – 0.25	0.008 – 0.3	0.008 – 0.3	0.008 – 0.25	0.008 – 0.25	0.008 – 0.25	0.008 – 0.25
Exposure time (min)	60							
Temperature (°C)	4	4	17	17	20	4	4	4
Concentration (mg ml^{-1})	0.3	0.3	6.0	2.0	2.0	2.5	6.9	2.5
Structural parameters								
I(0) (a.u.) [from Guinier]	0.28	0.53	0.29	0.098	0.321	1.053	0.101	1.696
Rg(Å) [from Guinier]	50.4	42.09	22.,865	21.406	28.7	18.348	25.335	50.986
I(0) (a.u) [from P(r)]	0.28	0.57	0.29	0.10	0.33	1.05	0.10	1.73
Rg(Å) [from P(r)]	49.56	48.38	22.84	21.45	29.74	18.43	26.5	53.,81
Dmax(Å)	144	141	80	75	100	64	83	179
Porod volume estimate	559940	593960	59210	42070	85840	32220	11170	216110
Software employed								
Primary data reduction	*SAXSLab3.0.0r1*							
Data processing	*PRIMUS 2.5*							
Graphic representations	*UCSF Chimera*							

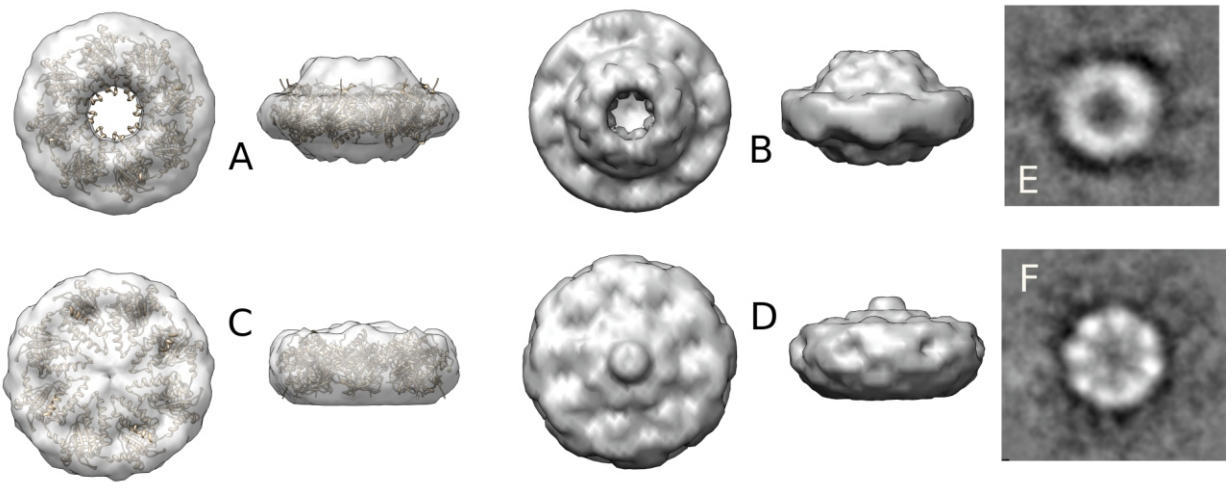

Figure 1. *Ab initio* shape reconstruction of octameric complex of ProteinO. *DAMMIN* dummy atom models contoured at 15Å resolution. Filtered models (A and C) and the crystal structure of the known part of the complex are superimposed. In absence of ATP and Ca^{2+} (first row) the octameric complex resembles the shape of toroidal ring, where the previously not observed portion of the complex forms elevated circumference of the inner inlet of the toroid. In presence of ATP and Ca^{2+} (second row) the octameric complex adopt shape of the flat disc and the flexible portion 'closes' the toroidal ring. Preliminary electron microscopy data from the identical system allows similar deduction (E and F).

Strong conformational change was detected in presence of calcium ions and ATP using circular dichroism spectroscopy. The conformational change of octameric complex of ProteinO was examined by SAXS *ab initio* modeling. Scattering data were collected from apoprotein ProteinO (c = 0.3 mg/ml) and ProteinO in presence of Ca^{2+} and ATP. Integral parameters derived from scattering data differs only slightly in range of experimental error (Table 1).

The *ab initio* reconstruction was performed using *DAMMIN* in 'slow' mode and symmetry was set to 'P8' for both datasets. Ten *DAMMIN* runs were performed for each dataset, individual models were compared and averaged using program suite *DAMAVER*. Filtered and most typical models were examined. *Ab initio* model of ProteinO in absence of ATP and Ca^{2+} is roughly toroidal ring of outer diameter ~140Å (Fig. 1). Previously not observed portion of octameric complex surrounds inner inlet with diameter of ~ 25 Å. The *ab initio* model of ProteinO in presence of ATP and Ca^{2+} in the buffer shows shape of flat disc of similar diameter ~ 140Å (Fig. 1). Hypothetically, the previously not observed portion of the complex undergo conformational change and 'closes' the inner inlet of the toroidal ring of octameric ProteinO in ATP and Ca^{2+} dependent manner. This hypothesis agrees well to preliminary electron microscopy data from the same system, where highly similar phenomenon was observed independently (Fig.1E and 1.F). Structural detail needs to be carefully examined.

2. Oligomeric state determination

Oligomeric state of FerA in solution was examined by SAXS. Data were collected for concentration series of 6; 2 and 1 mg/ml and truncated to q = 0.3 Å^{-1}. Scattering data from 1 mg/ml dilution with unsatisfactory signal to noice ratio and was omitted from further analysis. Porod volume derived from the scattering data (Table 1) and poor *CRYSOL* fit ($_6$ = 4.49 resp. $_2$ = 1.6) of the simulated scattering data from crystal structure of the monomeric FerA (not published) were indications of higher oligomeric state of the enzyme in solution. Simulated scattering from dimeric assembly of FerA indicated by *PDBePISA* gave significantly better *CRYSOL* fit of $_6$ = 1.30 resp. $_2$ = 0.96, Fig. 2. Oligomeric mixtures of FerA in potential monomeric, dimeric and tetrameric forms were examined and disproved using *OLIGOMER* (data not shown). Thus, the proposed dimeric state of FerA in solution was experimentally confirmed by SAXS. Importantly, the crystal structure of the FerA lack N- and C-terminal portion of purified protein. Total sum of 58 amino acid moieties out of 378 (in dimeric state) missing in the crystal structure could be reason for the value deviation from the ideal value = 1.0. Missing portions of crystal structures, usually disordered and flexible regions could be added using other software tools as BUNCH [8] or *CORAL*.

3. Quaternary complex architecture

The quaternary structure of multi-domain protein-protein complex was proposed by SAXS. Formation of the complex is indicated by integral parameters derived from scattering data, Table 1. Complex composed of two-domain ProteinA (22.2 kDa) and two-domain ProteinB (24.1 kDa). Crystal and NMR structures of 3 of 4 individual domains are known, the atomic structure of the fourth was proposed by homology modeling using the *iTASSER* and more, the relative position of the complex core, the subdomain protein-protein interface is known from co-crystal structure. Domain are connected by linkers not present in the atomic structures, representing total unknown structure with significant mass of 8 kDa. Scattering data was collected at total protein concentration of 2 mg/ml and data were truncated to q = 0.25 Å^{-1}. The complex exhibit no significant inter-domain flexibility as indicate well pronounced minimum in the Kratky plot, Fig. 3. The hybrid rigid body modeling as implemented in *CORAL* was used to propose the quaternary architecture of the complex. Relative position of the complex core known from co-crystal structure was fixed, while other parameters were kept default.

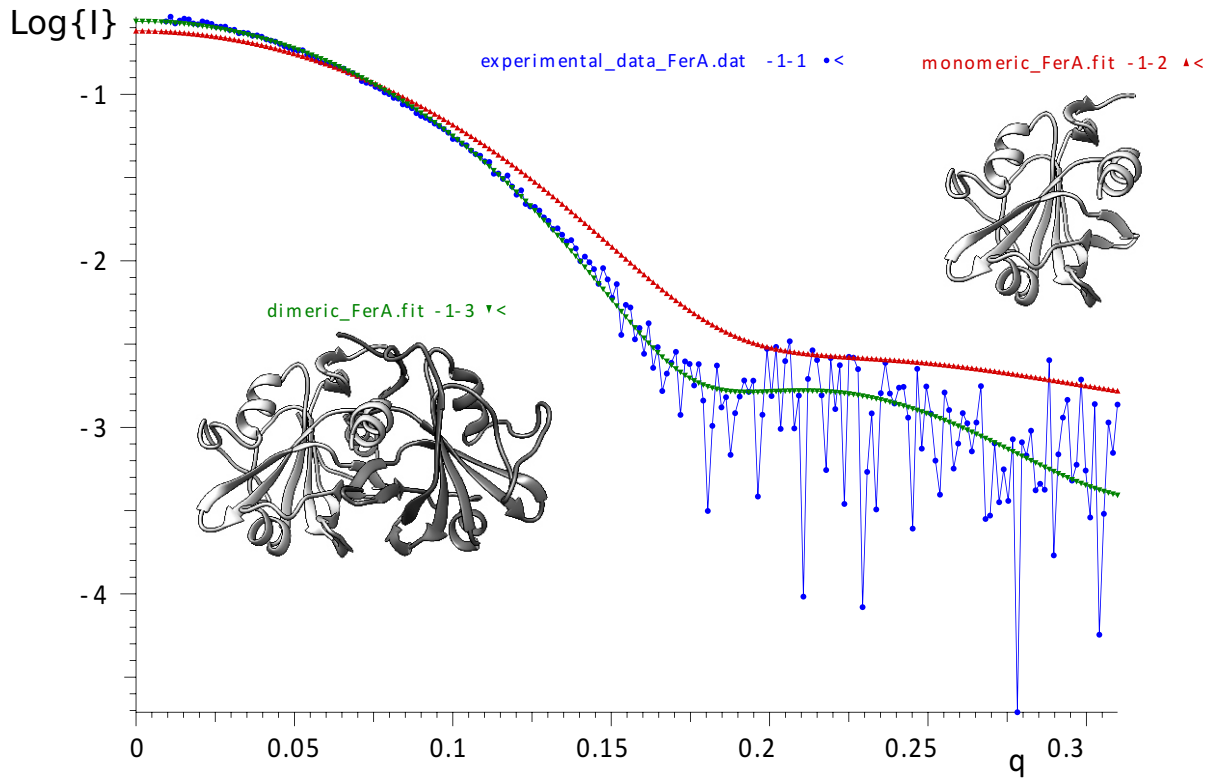

Figure 2. Simulated scattering data of monomeric and dimeric state of enzyme FerA at concentration 6 mg/ml evaluated using *CRYSOL*. Monomeric form ($_6$ = 4.49) is clearly disproved, while the dimeric form ($_6$ = 1.30) of the FerA was experimentally confirmed. Similar results were obtained for FerA at concentration 2 mg/ml ($_2$ = 0.96), data not shown.

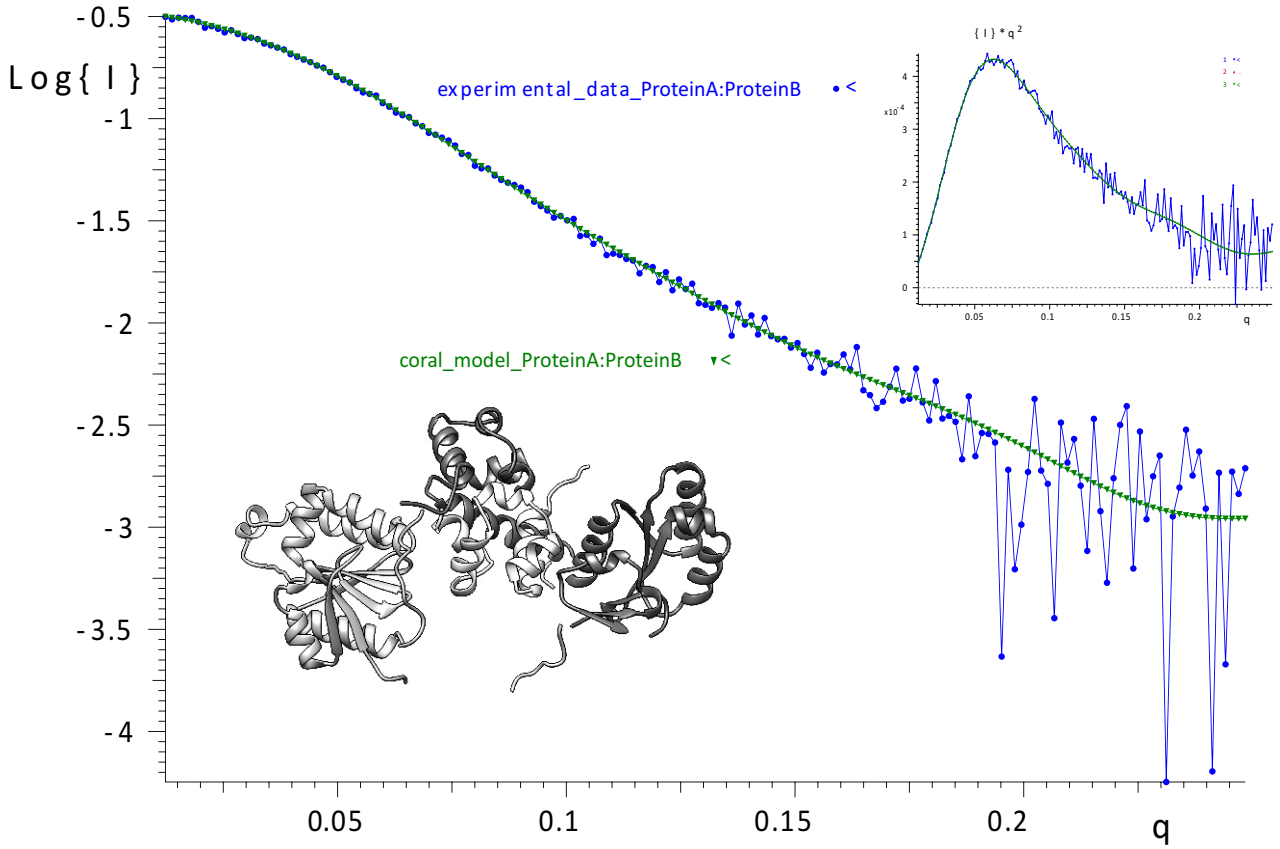

Figure 3. Quaternary structure of the four domain complex of ProteinA:ProteinB proposed by *CORAL*. Crystal structures, NMR solutions, one homology model were kept as rigid body to model complex architecture. Complex include linkers of total mass 8 kDa. Clear minimum in the Kratky plot (upper right corner) indicates no significant inter-domain flexibility. The known core of the complex composed of C-termini of both polypeptide chains was fixed. Final *CORAL* model gave excellent fit of = 0.95 of simulated versus experimental data.

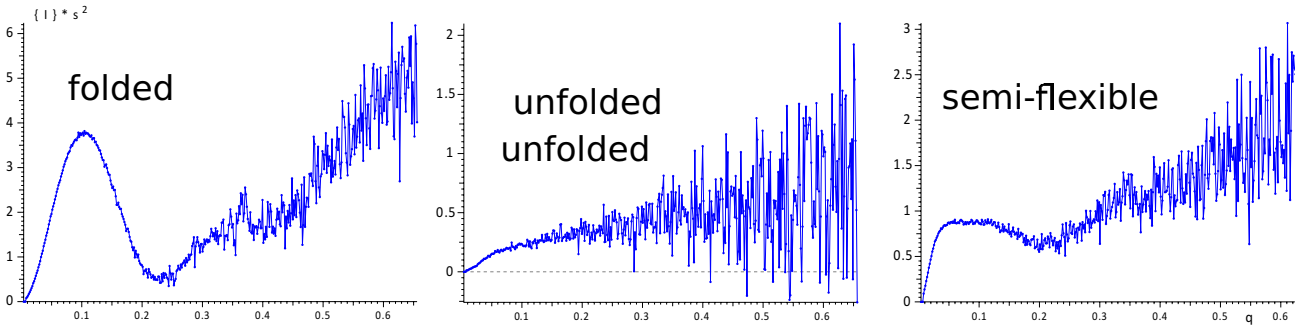

Figure 4. Kratky plots: qualitative evaluation of folding state of individual interacting partners and the complex of CTD of RNA polymerase II and CIP, the CTD interacting domain of ProteinP. (A) CIP - bell shaped curve typical for well folded macromolecules. (B) CTD - plateau and monotonic increase indicate unfolded macromolecules or IDPs. (C). Complex CTD:CIP – intermediate shape of Kratky plot as a indication of semi-flexible system.

The simulated scattering of the final SAXS based model of ProteinA:ProteinB complex fit satisfactory the experimental data (= 0.95), Fig 3. The quaternary structure is being examined by complementary methods, however excellent fit and good agreement with *ab initio* modeling based on the same data are highly promising.

4. Flexible and semi-flexible systems

Intrinsically disordered protein (IDP) behavior of C-terminal domain of RNA-polymerase (CTD) was examined by SAXS and ensemble representation model of semi-flexible complex of CTD with CTD-interacting domain of regulatory ProteinP (CIP) was proposed. Specific serin -phosphorylation pattern of CTD needed for interaction with CIP was mimicked by introduction of glutamic acid moieties. Complex was prepared by mixing of individual proteins and purified by size-exclusion chromatography. Stable complex formation was indicated by Porod volume derived from scattering data (Table 1). Folding states of the individual proteins and the complex were qualitatively evaluated by inspection of Kratky plot, where q^2 *I(q)* is plotted as a function of *q* (Fig. 4). In the Kratky plots, well pronounced minimum of CIP data (16.0 kDa) is typical for well folded globular protein, while CTD (12.7 kDa) exhibit the plateau followed by monotonic increase typical for unfolded behavior. The scattering data of the CTD-CIP complex exhibit "intermediate" behavior in the Kratky plot, typical for semi-flexible systems as proteins with inter-domain flexibility.

Porod volume (Table 1), multiple peaks pronounced in the pair-distance distribution function (Fig. 5) indicate ratio of CTD:CIP as 1:6 or 1:7. Scattering data was truncated to $q = 0.25$ Å^{-1} for further analysis. No unique solution, *i.e.* no single conformation of the 1:6 or 1:7 complex was possible to determined using *CORAL*. This was expected from semi-flexible indications and not stable *ab initio* reconstruction (not shown). Thus, the ensemble search was performed using *EOM*, where known NMR substructure of CIP in complex with CTD fragment (12 amino acid moieties) were kept as the repetitive rigid body and linkers of the length of five amino acid moieties were treated as flexible. Ensemble search finished with four conformation of the complex with excellent fit of = 1.16, Fig. 5. Other CTD:CIP complex ratios (1:5, 1:6, or 1:8) and mixture scenarios were examined by the similar approach with significantly worse fits to experimental data. Ensemble structure of proposed "beads on a string" model is giving the experimental evidence of flexible character and conformational possibilities of the system approximately 180 Å long and of total mass of approx. 108 kDa and it is well consistent with biological role of CTD of RNA polymerase II.

1. Blanchet CE, Svergun DI. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. Annu Rev Phys Chem. 2013 Apr;64(1):37–54.
2. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 2010 Jul 1;38(Web Server):W540–W544.
3. Graceffa R, Nobrega RP, Barrea RA, Kathuria SV, Chakravarthy S, Bilsel O, et al. Sub-millisecond time-resolved SAXS using a continuous-flow mixer and X-ray microbeam. J Synchrotron Radiat. 2013 Nov 1;20(6):820–5.
4. Svergun DI. Small angle x-ray and neutron scattering from solutions of biological macromolecules. First Edition. Oxford: Oxford University Press; 2013. 358 p.
5. Svergun DI. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. Biophys J. 1999 Jun;76(6):2879–86.
6. Franke D, Svergun DI. J Appl Crystallogr. 2009 Apr 1; 42(2):342–6.
7. Svergun DI, Petoukhov MV, Koch MHJ. Biophys J. 2001 Jun; 80(6):2946–53.
8. Petoukhov MV, Svergun DI. Biophys J. 2005 Aug; 89(2):1237–50.
9. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. J Appl Crystallogr. 2012 Apr 1; 45(2):342–50.
10. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. J Appl Crystallogr. 2003 Oct 1; 36(5):1277–82.
11. Volkov VV, Svergun DI. J Appl Crystallogr. 2003 Jun 1; 36(3):860–4.

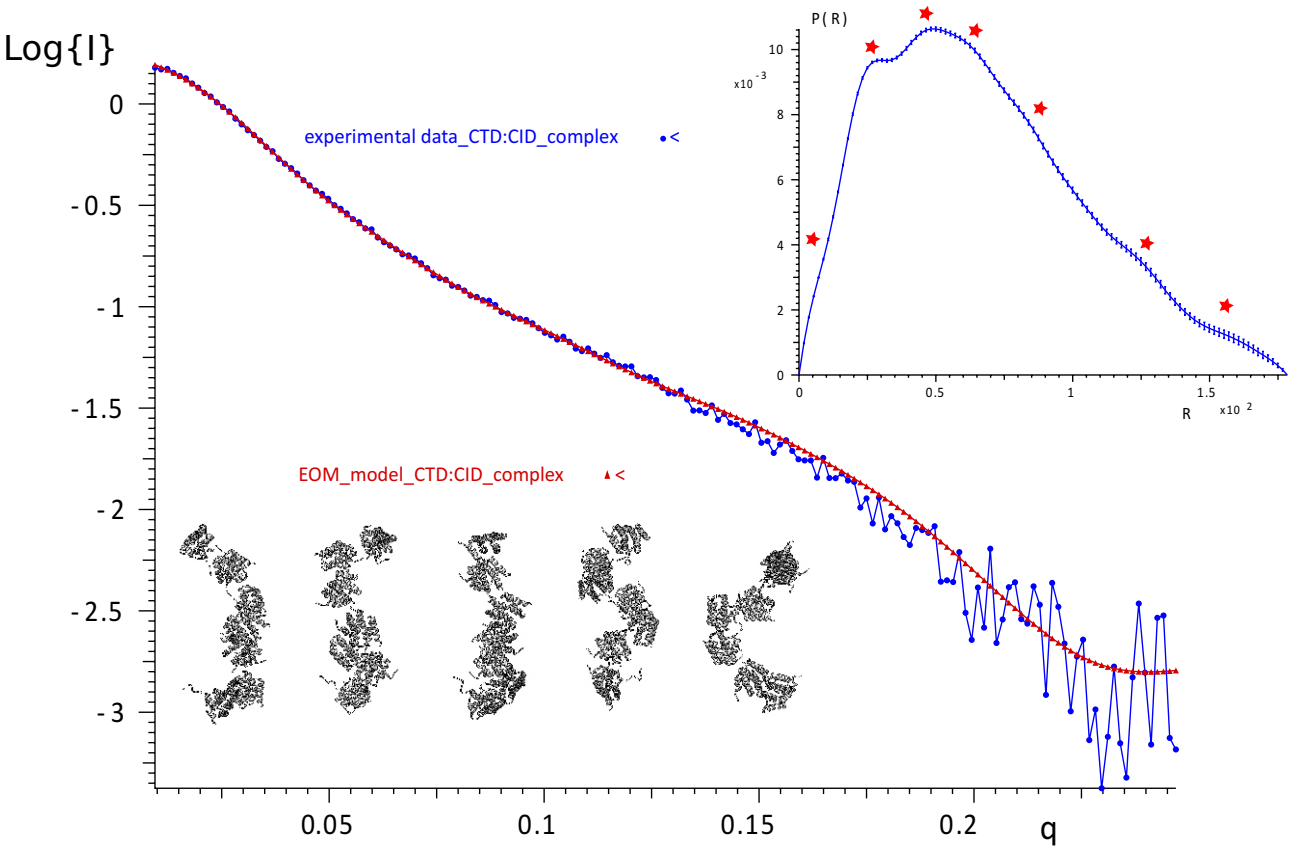

Figure 5. "Beads on a string" flexible model of CTD of RNA polymerase II in complex with CIP, the CTD interacting domain of ProteinP. Model is represented by an ensemble of conformations proposed by *EOM*. Optimized ensemble with final = 1.16 explain well the experimental data. The number of interacting CIP domains was estimated from number of peaks (stars) in the pair-distribution function (upper right corner).

12. Kozin MB, Svergun DI. J Appl Crystallogr. 2001 Feb 1; 34(1):33–41.
13. Krissinel E. Macromolecular complexes in crystals and solutions. 2011 Apr 1; 67(4):376–85.
14. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. J Comput Chem. 2004 Oct; 25(13):1605–12.
15. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. J Am Chem Soc. 2007 May;129(17):5656–64.
16. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. Nat Methods. 2014 Dec 30; 12(1):7–8.

SAXS part of the work was realized in X-ray Diffraction and Bio-SAXS Core Facility of CEITEC- Central European Institute of Technology with research infrastructure supported by the project CZ.1.05/1.1.00/02.0068 financed from European Regional Development Fund and project Employment of Best Young Scientists for International Cooperation Empowerment, reg. number CZ.1.07/2.3.00/30.0037 (to TK) co-financed by the European Social Fund and the state budget of the Czech Republic.