**The structural analysis of the human N2 VWA single-domain collagen VI construct: SEC-SAXS combined with MALLS and DLS.**

**EMBL-P12 SAXS measurements 20 April 2018: Solomon, Jeffries, Wagener**

Small angle X-ray scattering (SAXS) data were measured at the EMBL-P12 bioSAXS beam line1 from a von Willebrand factor (VWA) single-domain fragment, N2, of human collagen VI. The N2 protein underwent size-exclusion chromatography (SEC) SAXS in combination with parallel multi-angle laser light scattering (MALLS), dynamic light scattering (DLS) and refractive index (RI) measurements to obtain solution-based structural parameters and validated molecular weight (MW) estimates2. Subsequent low-resolution structural models have been developed based on the SAXS data to describe the solution state of the protein using a combination of *ab initio*, rigid-body and homology modelling approaches.

This report summarises the SEC-SAXS MALLS/DLS and SAXS modelling results obtained for:

1. Human N2 (Collagen alpha-3(VI) chain: UniProt ID P12111; Amino acid range 1634-1833).

**INSTRUMENT SETUP AND PRIMARY DATA REDUCTION.**

SAXS data collection, *I*(*s*) vs *s*, where *s* = 4sin**/; 2** is the scattering angle and  the X-ray wavelength (1.24 nm; 10 keV) was performed at 20 oC using a GE-Healthcare S75 Increase 10/300 analytical SEC column equilibrated in 20 mM TRIS, pH 7.4, 150mM NaCl 3% v/v glycerol at a flow rate of 0.6 ml.min-1. Glycerol was added to reduce the effects of X-ray radiation damage to the protein sample3. Automated sample injection and data collection were controlled using the *BECQUEREL* beam line control software4. The sample injection volume and load concentration were 25l at 20.mgml-1, respectively. The SAXS intensities were measured as 2400 x 1 s individual X-ray exposures, from the continuously-flowing column eluent, using a Pilatus 6M 2D-area detector. The 2D-to-1D data reduction, i.e., radial averaging of the data to produce 1D *I*(*s*) vs *s* profiles, were performed using the SASFLOW pipeline4, 5 incorporating RADAVER from the ATSAS 2.8 suite of software tools6. The 2400 individual frames obtained for each SEC-SAXS run were processed using CHROMIXS7. Briefly, individual SAXS data frames were selected across the SEC-elution peak of the protein. An appropriate region of the elution profile, corresponding to SAXS data measured from the solute-free buffer, were identified, averaged and then subtracted from the SEC-peak frames to generate individual background-subtracted sample data frames of N2. These individual data frames underwent further CHROMIXS analysis, including the assessment of the radius of gyration (*Rg*) of the protein through the SEC peak. Then the individual frames were scaled (to take into account changes in concentration) and averaged to produce a final 1D-reduced and background-corrected scattering profile. Only those SAXS data frames with a consistent *Rg* through the SEC-elution peak and evaluated as statistically similar through the measured *s*-range (0.024–7.3 nm-1) – post scaling – were used to generate the final SAXS profile of N2 in solution.

The MALLS and DLS data were measured using a Wyatt Technologies Mini-Dawn TREOS with an in-built quasi elastic light scattering (QELS) module coupled to an OptiLab T-Rex refractometer for protein concentration determination. The MALLS system was calibrated relative to the scattering from toluene and, in combination with concentration estimates obtained from refractive index (RI; dn/dc was taken as 0.185 ml.g-1), was used to evaluate the MW distribution of species eluting from the SEC column. The molecular weight estimates from MALLS/RI and the hydrodynamic radius, *RH*, derived from DLS were determined using Wyatt ASTRA7 software.

**RESULTS.**

N2.

The MALLS/RI data and the corresponding protein MW correlation through the SEC elution trace of N2, calculated from static light scattering intensities and protein concentration estimates, are shown in Figure 1. The results show that the sample is extremely pure with only negligible/barely-detctible larger MW species (eluting between 3.75–4 ml). The MWs calculated through the N2 peak fall within a narrow range of 21.4–23.0 kDa, with an average of 22 kDa. The expected MW of N2 calculated from the amino acid sequence is 22.1 kDa. Therefore the protein elutes from the SEC column as a monomer. The average *RH* of N2 determined from the DLS measurements is approximately 1.7 nm, although there is considerable variance on the data due to the small particle size and weak scattering intensities (refer to Supplementary information).

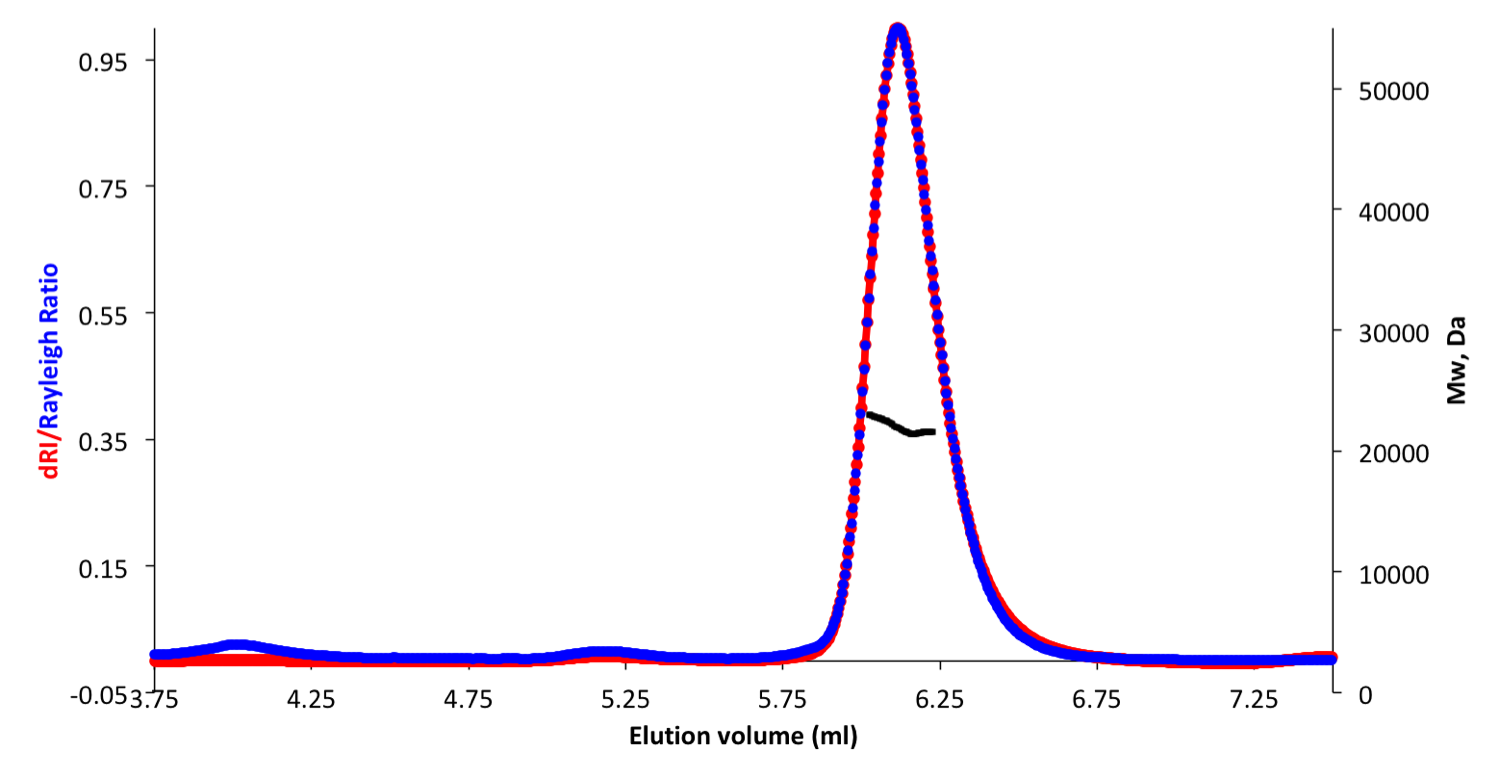


Figure 1. The MALLS, cm-1 (blue) and dRI, ml.g-1 (red) traces obtained for the SEC-MALLS analysis of N2. The MW correlation through the N2 peak is shown as a black line that spans a MW range of 21.4–23 kDa (MW average = 22 kDa).

Figure 2a displays the integrated X-ray scattering intensities vs. SAXS frame number acquired from the same N2 sample measured in parallel to the MALLS/RI/DLS data. After buffer subtraction, the calculated *Rg* of each individual SAXS profile through the N2 peak remains within the range of 1.76–1.8 nm. Figure 2b shows the final SAXS data of N2 obtained from the scaled and averaged individual frames. The maximum useful *s*-value of the SAXS profile i.e., the estimation of how far the information content of the SAXS data extends to, was calculated using SHANUM8. The final working *s*-range of the N2 profile is 0.07–6.14 nm-1 (encompassing 13 Shannon channels).

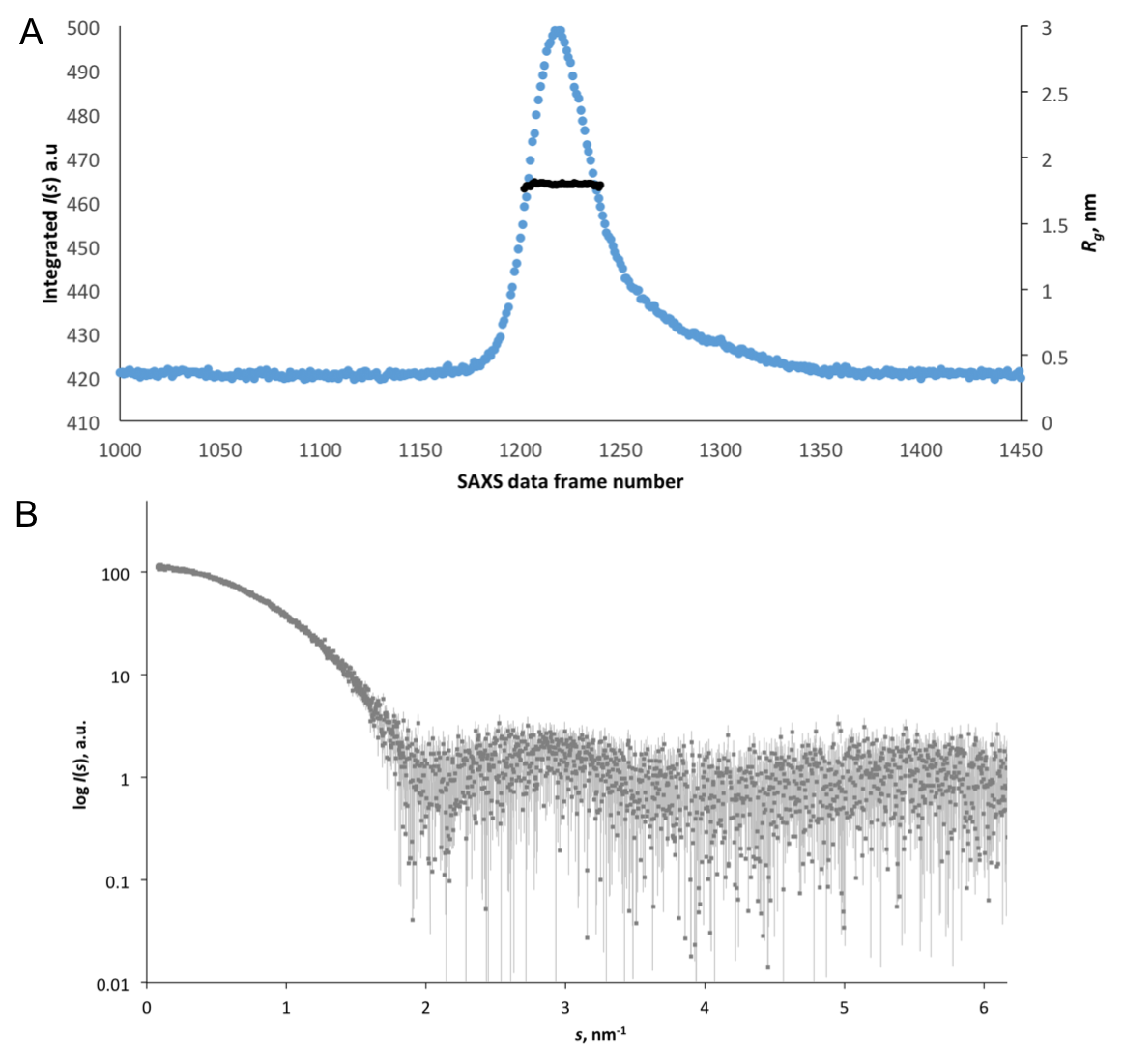


Figure 2. **A**. Integrated SAXS intensities vs. frame number of the N2 sample after SEC and the corresponding *Rg* estimates obtained for each individual data frame after buffer scattering subtraction (black line). **B**. The final SAXS profile of N2 obtained from averaging individual frames through the SEC elution peak plotted to the highest useable value of *s*, i.e., the maximum Shannon limit.

The corresponding Guinier plot9, calculated via a simple transformation of the SAXS data at the lowest of scattering angles (ln*I*(*s*) vs *s*2 to an *sRg* maximum of 1.3), is displayed in Figure 3. The negative slope of the linear plot relates to the *Rg* via:



where *I*(0) is the extrapolated to zero-angle scattering intensity, or forward scattering. The final Guinier *Rg* estimate of the N2 protein using the Guinier estimation is 1.8 nm.

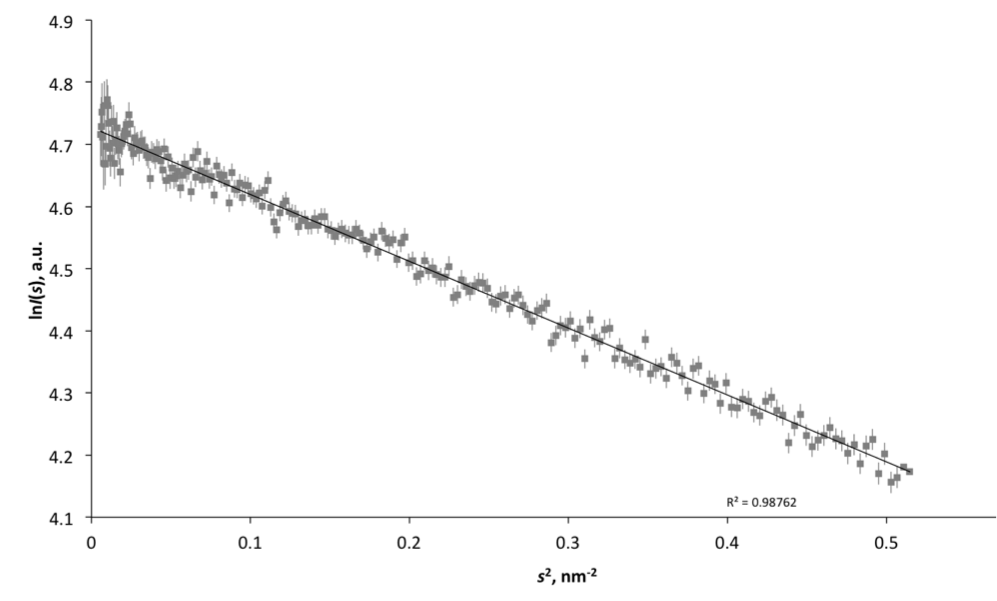


Figure 3. **A**. The Guinier plot of the SAXS intensities measured from N2 at very low angle. The *Rg* (1.8 nm) is calculated from the negative slope.

A Kratky plot representation of the SAXS data, *I*(*s*)*s*2 vs *s*, is shown in Figure 4A. The Kratky plot is consistent with a protein that has a globular/compact structure structure, as shown by the symmetric, bell-shaped main-peak centred at ~1 nm-1. The corresponding probable frequency, *p*, of real-space distances, *r* – or the *p*(*r*) profile – of N2 was calculated from the indirect inverse Fourier transformation of the data using GNOM12 (Figure 4B). The *Rg* determined from the second moment of the *p*(*r*) at 1.8 nm is essentially identical with the Guinier *Rg* estimate. The maximum particle dimension, *Dmax*, is ~6 nm. This *Dmax* is agrees with (automated) predicted *Dmax* estimates from DATCLASS (6.1 nm) and SHANUM (6.7 nm). Overall, the *p*(*r*) profile has a defined symmetric distribution of vector lengths, indicating that N2 has a structure that is globular. The reciprocal space-fit of *p*(*r*) to the SAXS data is shown in the supplementary information. The quality of the fit was assessed using the reduced **2 test (**2 = 1.05) and the correlation map, or CorMap *p*-value (*p* = 0.6)13. (Note: fits within the **2 range of 0.9–1.1 or having a CorMap *p*-values higher than the significance threshold cutoff of =0.01 are considered excellent, i.e., no systematic statistical discrepancies are present between the data and fit.)

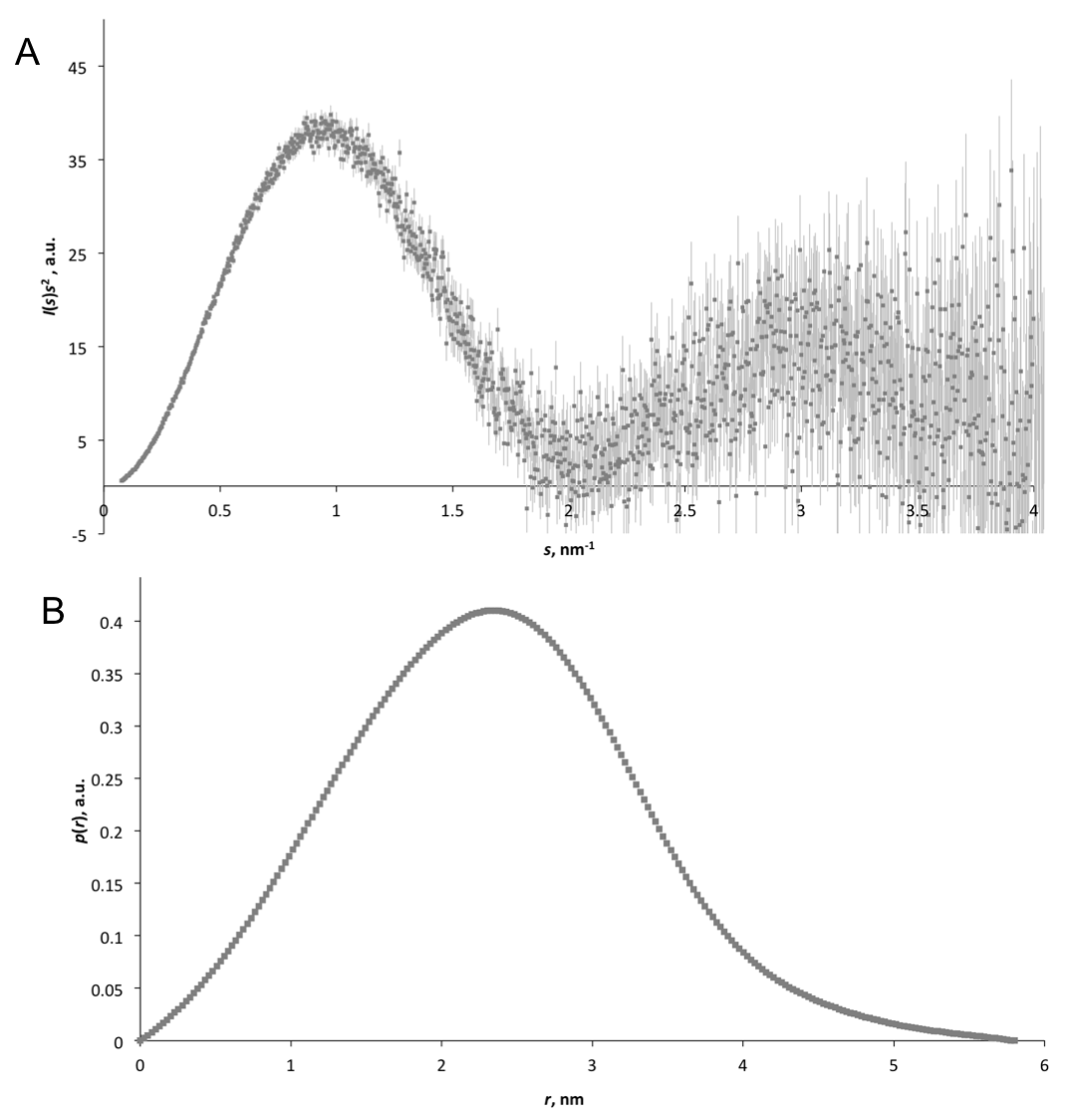


Figure 4. **A**. A Kratky plot representation of the final N2 SAXS data. **B**. The calculated *p*(*r*) vs profile of N2.

Concentration-independent MW estimates, were also evaluated directly from the SAXS data using a number of methods. These methods include: DATMW (Bayesian MW estimator from scattering invariants)14; DATCLASS (shape classification); DARA (PDB nearest neighbours)15; SAXSMoW (normalized Kratky plot)16; and the ‘volume’ of correlation, Vc17. All of these methods yield a particle between 19–21 kDa, that is entirely consistent with the MALLS MW estimate (22 kDa). The combined summary of all structural parameters and MW estimates are reported in Table 1, including the Porod volume (obtained from the SAXS data from area under the Kratky plot).

Of particular note, the shape factor obtained for N2, or the *Rg*/*RH* ratio10, was calculated by combining the *Rg* determined from SAXS and the *RH* measured from DLS. In general, a spherical particle (i.e., a globular-compact protein) has a *Rg*/*RH* ratio of around 0.78, and therefore anything higher than this, e.g., 1.0, suggests some sort of anisotropic/less-compact/more-extended mass distrubution. For N2, the *Rg*/*RH* ratio is ~1, which is higher than expected considering the Kratky and *p*(*r*) analysis that both show the protein is compact with a very symmetric distribution of internal vector lengths. Something interesting is going on with respect to the hydrodynamic behavior of the protein and how this relates to the mass distribution within the proteins volume boundary. As it happens, oblate spheroids have *Rg*/*RH* of 0.88–1, while hollow spheres also have a shape factor that also limits to 1.0, depending on the thickness of the outer shell. Therefore, N2 is likely either equatorially-compressed along one axis (relative to the long axis defined by *Dmax*), or has a region internal to it with a much lower average electron density compared to the outer regions near the periphery, e.g., has an internal cavity or is like a ‘core-shell’ particle. Interestingly, the automated shape-classification of the SAXS data using the program DATCLASS6, 11 classifies N2 as being a ‘compact hollow’ protein, in keeping with the experimentally determined shape factor as well as the *p*(*r*) and Kratky analysis. These combined observations might suggest that the average internal regions of the N2 domain have a somewhat lower average electron density compared to regions constructing the outside-perifery of the protein.

**TABLE 1: SEC-SAXS, MALLS and parameter summary for N2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Instrument Parameters*** |  |  |
|  | **Column Type** | S75 Increase 10/300 |  |
|  | **Flow rate (ml/min)** | 0.6 |  |
|  | **Injection volume (l)** | 25 |  |
|  | **Load Concentration (mg/ml)** | 20 |  |
|  | **Temperature (oC)** | 20 |  |
|  | **Buffer** | 20mM TRIS, pH 7.4, 150mM NaCl 3% v/v glycerol |  |
|  | **sample to detector distance (m)** | 3 |  |
|  | **Exposure time/number of frames** | 1s/2400 |  |
|  | **X-ray wavelength (nm)** | 0.124 |  |
|  | ***Information content*** |  |  |
|  | **#Shannon channels** | 13 |  |
|  | **Highest useable *s*max (nm-1)** | 6.14 |  |
|  | **Predicted** *Dmax* **(nm)** | 6.7 |  |
|  | ***Guinier analysis*** |  |  |
|  | Guinier *I*(0) () | 113.1 (0.2) |  |
|  | *Rg* (Guinier, nm) () | 1.8 (0.01) |  |
|  | *sRg* range/(points used) | 0.14-1.3(1;235) |  |
|  | ***p(r) analysis*** |  |  |
|  | ***I*(0), POR ()** | 113 (0.2) |  |
|  | ***Rg* (POR, nm) ()** | **1.8 (0.01)** |  |
|  | ***Dmax (nm)*** | 5.8 |  |
|  | **Porod volume (nm3)** | 40 |  |
|  | ***MW analysis*** |  |  |
|  | ***Calculated MW, from amino acid sequence*** | **22.1** |  |
|  | **MW from MALLS (kDa)** | **22.0** |  |
|  | **MW (Bayes, kDa)** | **20** |  |
|  | **MW Credibility Interval (kDa)** | **19-21** |  |
|  | MW SAXSMow (kDa) | 21 |  |
|  | MW Vc (kDa) | 21 |  |
|  | MW DatClass (kDa) | 21 |  |
|  | MW DARA (kDa) | 19 |  |
|  | ***Shape classification and hydrodynamics*** |  |  |
|  | Classification/(predicted *Dmax*, nm) | Compact-hollow(6.1) |  |
|  | *Rh* (from DLS) nm | 1.73 |  |
|  | *Rg*/*Rh* ratio | 1.03 |  |
|  | Ambimeter score(*sRgmax*) | 0.477(4.5) |  |
|  | #shape topologies | 3 |  |
|  | Uniqueness | potentially unique |  |
|  |  |  |  |

***Ab initio* modelling of N2**

The assessment of the non-uniqueness of the N2 scattering data was performed using AMBIMETER18. This step was performed to evaluate the overall ambiguity of the SAXS data, as it is possible that multiple shapes with different topologies can fit the same SAXS profile. The AMBIMETER score of 0.48 (calculated to an *sRg* of 4.5) categorises the scattering as ‘potentially unique’, indicating that a somewhat limited set of shapes can be used to describe the scattering profile of N2.

The *ab initio* bead modelling of N2 was performed using the programs DAMMIN19 and GASBOR20 and the results are summarised in Figure 5. Each of the individual *ab initio* models produce excellent fits to the SAXS data (**2 DAMMIN N2 = 1.1; CorMap *p* = 0.43; **2 GASBOR N2 = 1.05; CorMap *p* = 0.33). It appears that the N2 monomer forms an compact, globular, almost spherical, structure, with a slight extension at one end. The N2 domain has overall dimensions of approximately 4 x 4 x 5.7 nm. DAMMIN was run several times and the consistency of the individual DAMMIN models was evaluated using the normalized spatial discrepancy (NSD) metric (where NSD < 0.7 represents spatially similar)21. A final averaged spatial representation of N2 determined from the DAMMIN model cohort (with an excellent NSD of 0.5) was calculated using the DAMAVER set of programs with a resolution estimate of 2.4 nm for the models (obtained using SASRES22.) The averaged low-resolution DAMMIN spatial representation/structure of N2 is shown below in Figure 7 and compared to a constructed homology model of the protein.

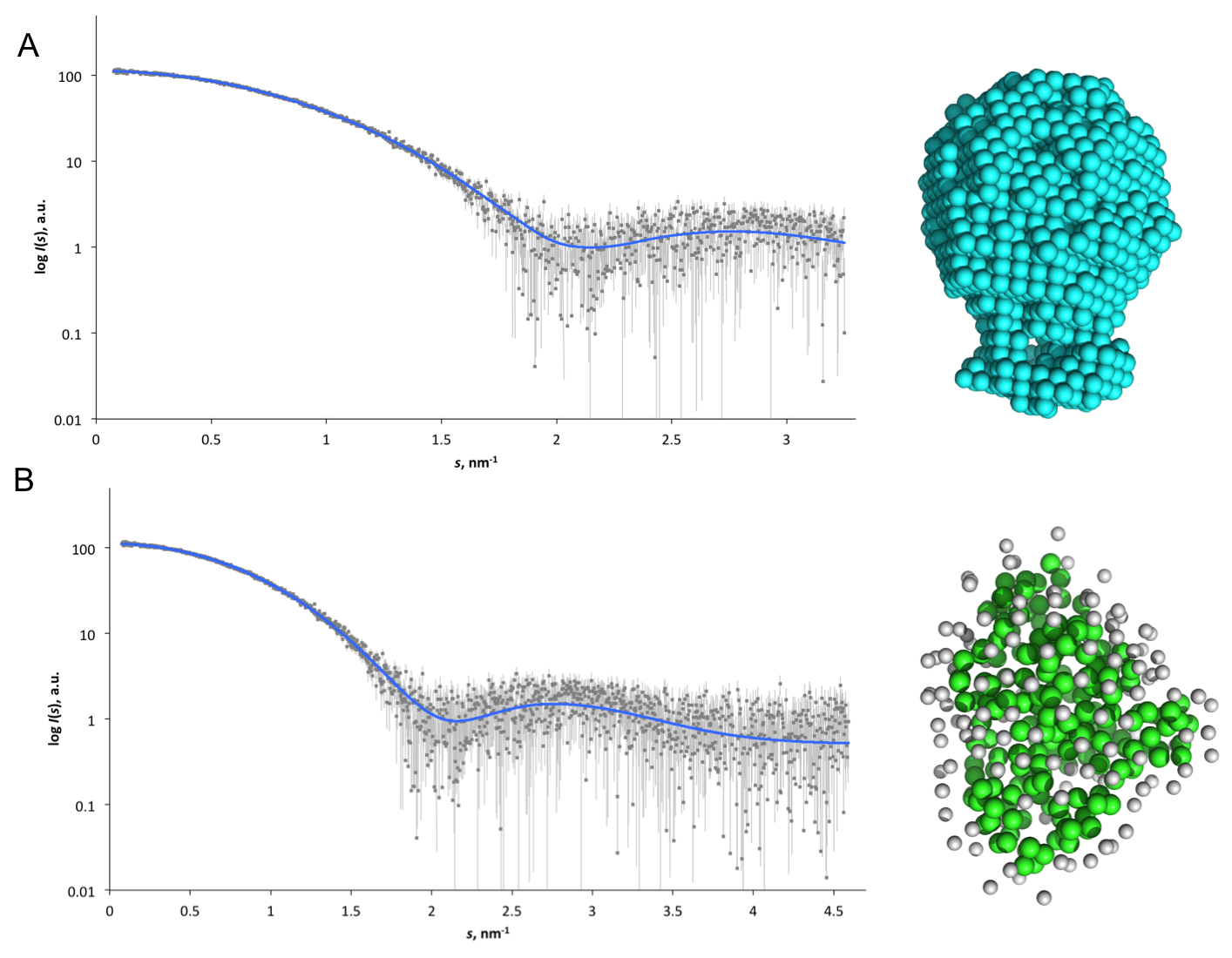


Figure 5. **A**. DAMMIN modelling results showing an individual model fit to the SAXS data (blue line) and the corresponding low-resolution structure of N2 (cyan spheres). **B**. A GASBOR model fit (blue line) to the SAXS data and the GASBOR model of N2 (green spheres). The small white spheres represent a modelled solvent shell.

**Homology modelling of N2.**

The a high-resolution structure of N2 has, at the time of writing, not been published. Therefore, a homology model of N2 was generated using the ITASSER server23 using the amino acid sequence of N2 as input. Five models were generated in total, where the top PDB analogue was identified as 4IGI, the murine N5 VWA domain of collagen IV (~29% sequence identity). The ITASSER model cohort and their corresponding spatial alignment (RMSD C = 0.7) is shown in Figure 6A. When assessing the model fits to the SAXS data (using CRYSOL25; Figure 6B), it is found that model 5 produces the best, and excellent, fit to the experimental profile (**2 = 1.04, CorMap *p* = 0.55).

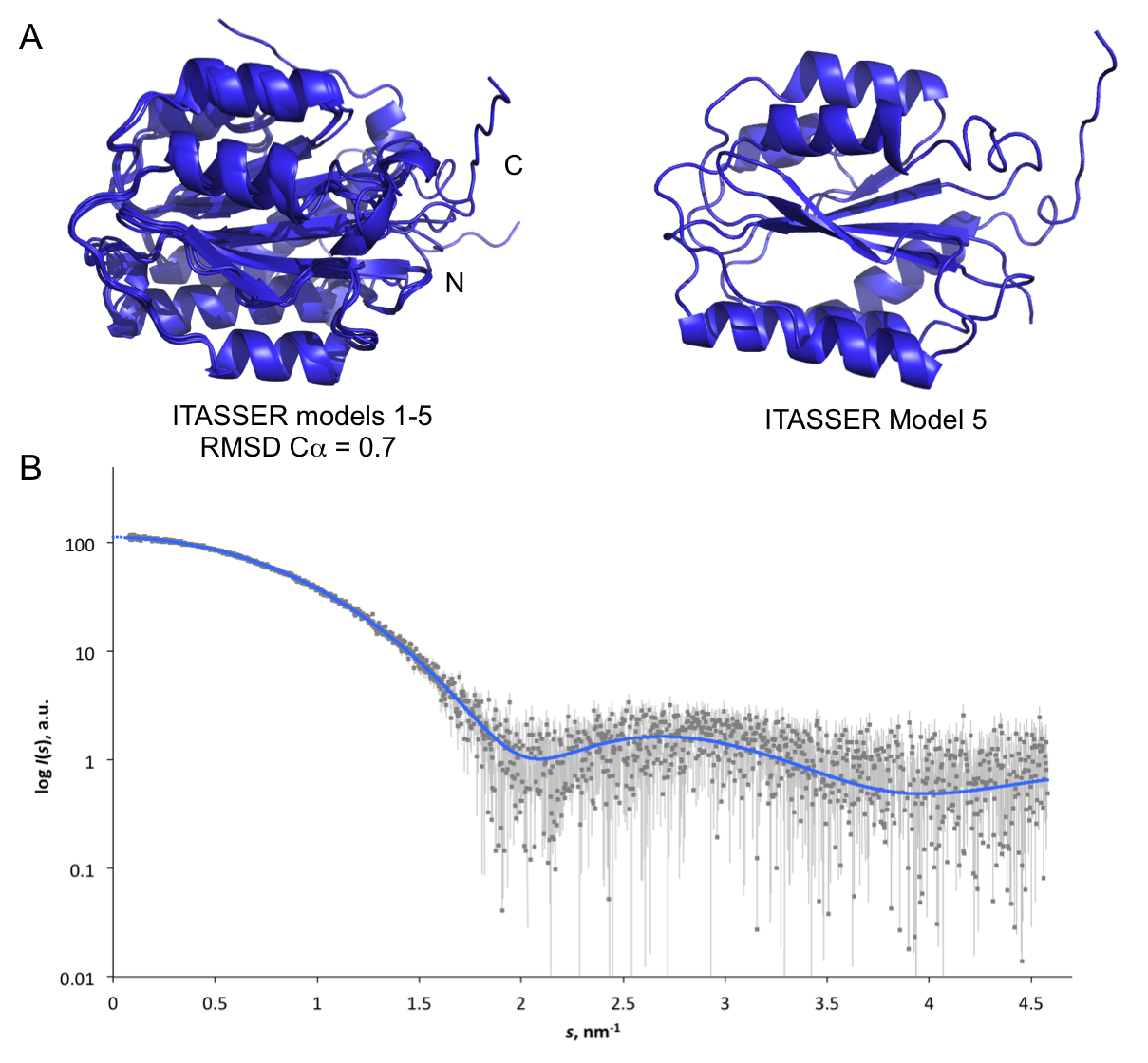


Figure 6. **A**. Five predicted homology models of N2 constructed from the amino acid sequence using ITASSER. The orientation of the N- and C-termini of the models are labeled. It is predicted that the N- and C- termini are spatially positioned close together on the same side of the domain. An individual model, ITASSER model 5, is also displayed. This model fits the experimental SAXS data, **B**, exceptionally well (**2 = 1.04; CorMap *p* = 0.55).

A comparison between the ITASSER homology model of the N2 domain (modle 5) and the best-fit *ab initio* model derived from GASBOR as well as the spatial alignment relative to the averaged low-resolution structure of the protein obtained from DAMMIN is shown in Figure 7.

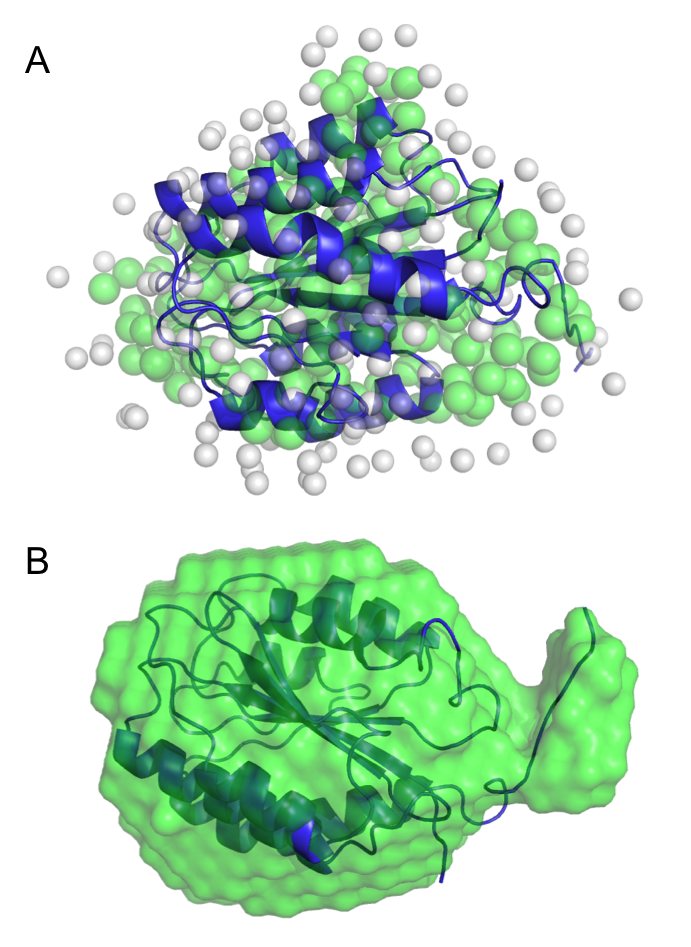


Figure 7. **A**. The spatial alignment between ITASSER homology model (model 5, blue ribbons) with the best-fit dummy amino acid *ab initio* model calculated using GASBOR (green spheres). **B**. The ITASSER model superimposed with the averaged, low-resolution structure of N2 (2.4 nm) obtained from ten individual DAMMIN *ab intio* shape reconstructions.

When analyzing the predicted structure of N2 from ITASSER in terms of the distribution of electron density, there maybe some evidence to suggest why the domain is categorized as a ‘compact-hollow’ structure and why its hydrodynamic behavior deviates from simply being ‘compact’ and more towards a ‘core-shell’ particle. When evaluating the average electron density of the amino acids that go to construct internal secondary structure elements, predominantly a twisted -sheet, it is found that the internal structure does have a predicted lower average electron density (0.421 e-/Å3) compared to the -helices and loops surrounding the -sheet core (0.441 e-/Å3). This, maybe because the ITASSER model seems to have a core-region that is (excessively?) dominated by hydrophobic interactions which are driven by amino acids (e.g., Phe) that are typically less electron dense (Figure 8). In essence the domain appears to have an ‘oily-core’.

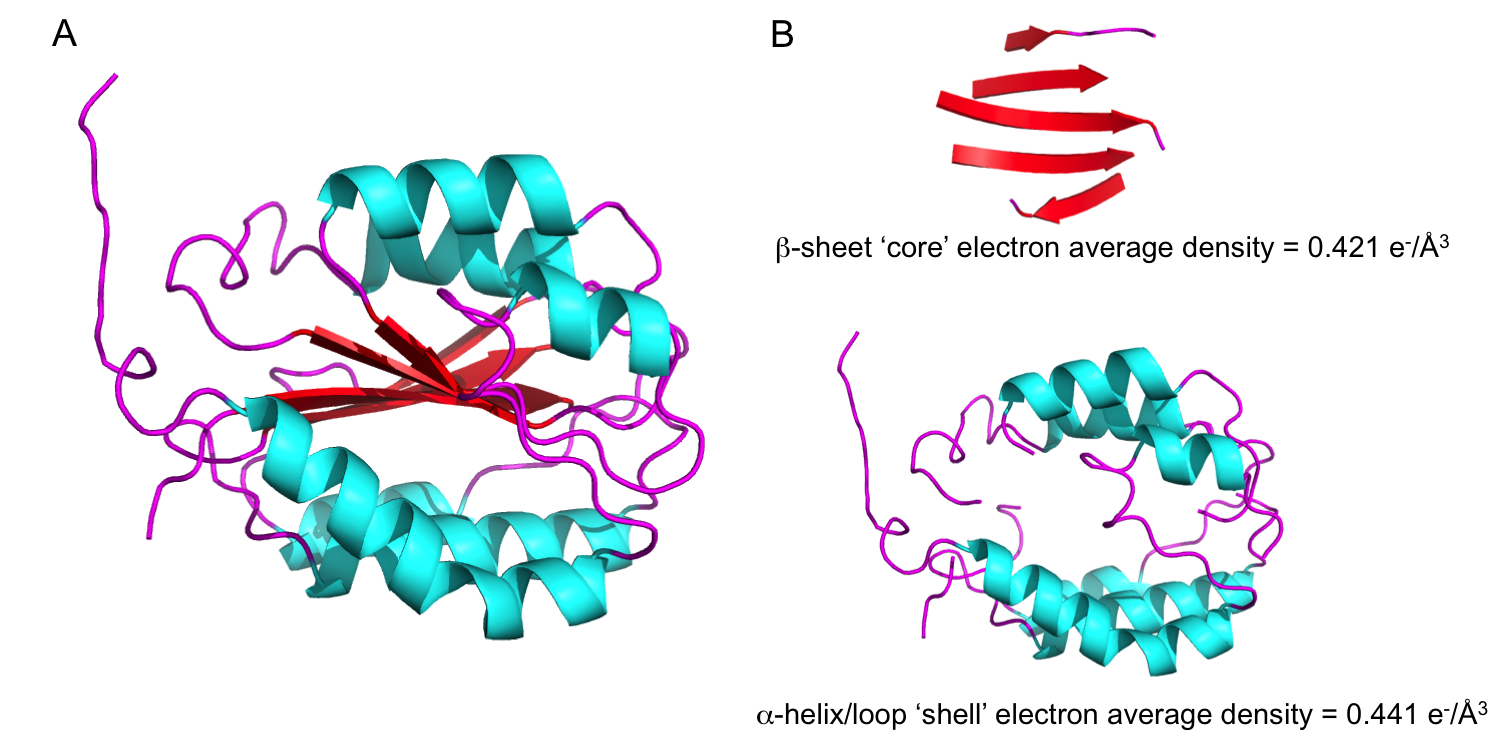


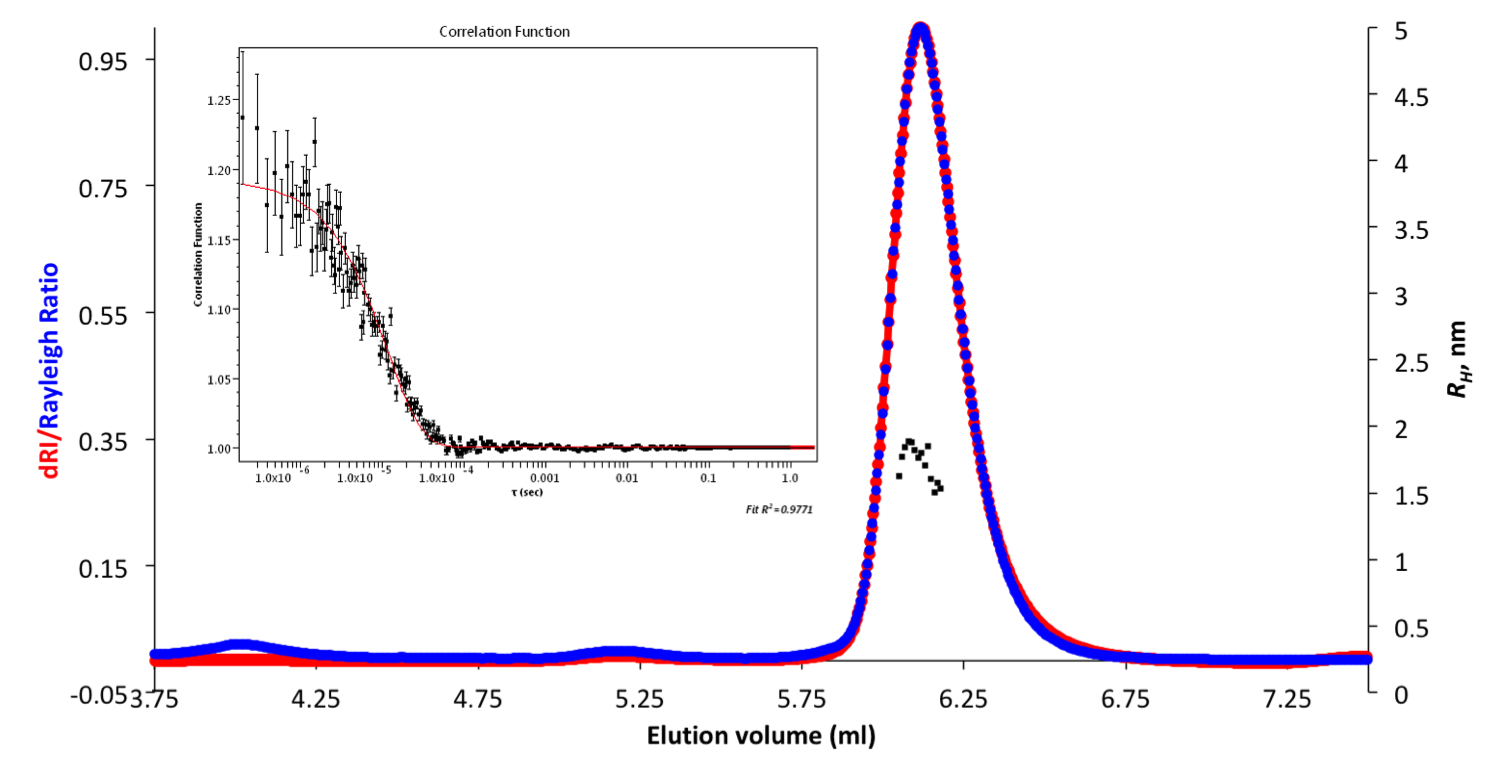
Figure 8. **A**. ITASSER model 5 of the N2 domain coloured by secondary structure (red = -strand; cyan = -helix; magenta = loop). **B**. The average electron density of the b-sheet core is lower than the -helical and loop elements surrounding it. This may be the reason why the N2 domain behaves more like a ‘core-shell’ particle in solution and is classified as ‘compact-hollow’ with respect to its hydrodynamics.

**Summary: N2.**

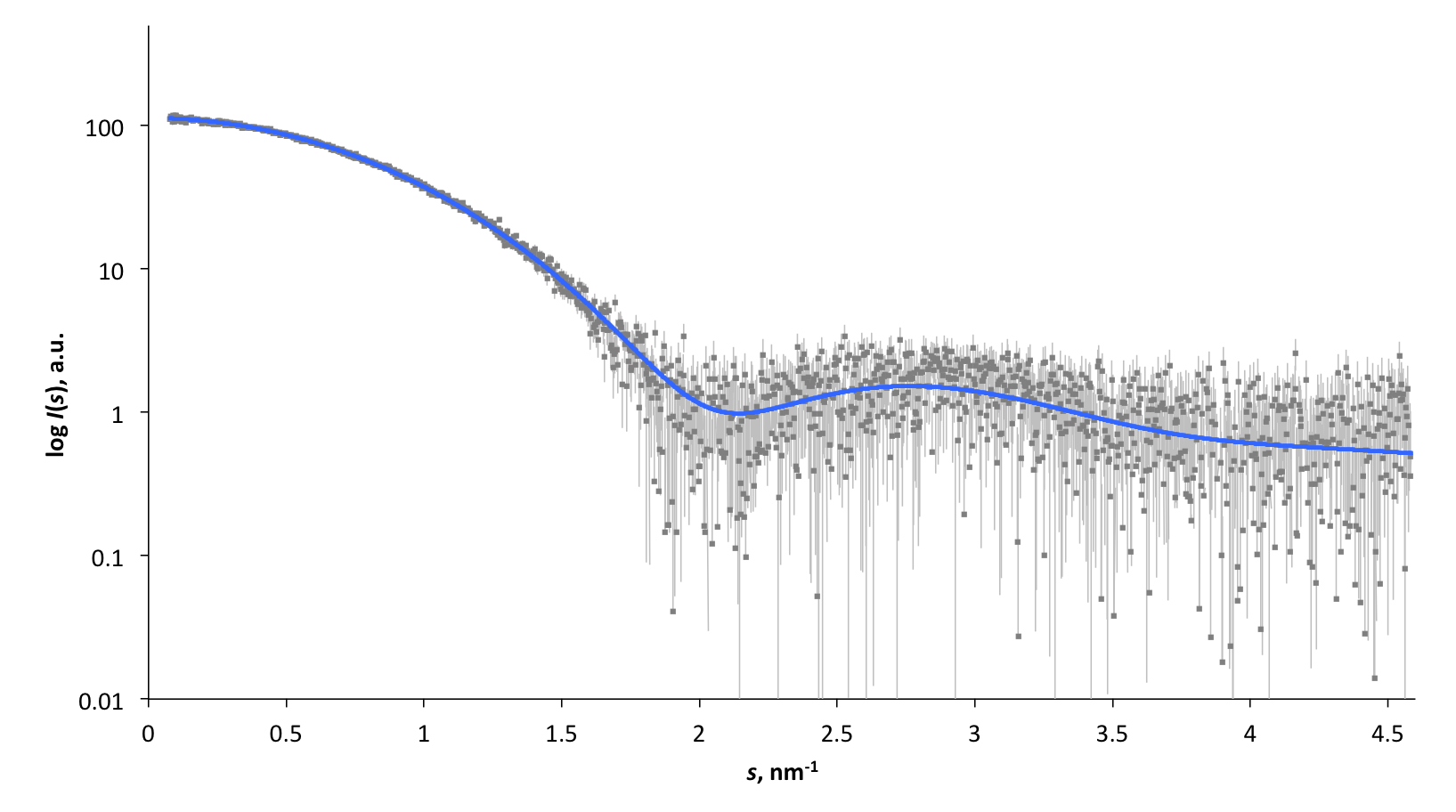
* N2 is monomeric in solution.
* The domain has a compact, almost spherical structure, with a slight extension at one end that is classified as a ‘compact-hollow’ protein.
* It is predicted from ITASSER calculations that the N2 domain adopts a similar structure to the murine N5 VWA domain of collagen IV. Atomistic models generated using this protein as template fit the experimental SAXS data very well.

**Supplementary information.**

Hydrodynamic radius measurements through the N2 SEC elution peak and an example DLS correlation function.



Reciprocal-space fit of the probable real-space distance distribution, *p*(*r*) vs *r*, to the SAXS data (**2 = 1.01; CorMap *p* = 0.8).



**Amino acid sequence information (ITASSER input).**

The amino acid sequence used for ITASSER modelling.

EKKKADIVFLLDGSINFRRDSFQEVLRFVSEIVDTVYEDGDSIQVGLVQYNSDPTDEFFLKDFSTKRQIIDAINKVVYKGGRHANTKVGLEHLRVNHFVPEAGSRLDQRVPQIAFVITGGKSVEDAQDVSLALTQRGVKVFAVGVRNIDSEEVGKIASNSATAFRVGNVQELSELSEQVLETLHDAMHETLCPGVTDAAK

**References**

1. Blanchet CE*, et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr* **48**, 431-443 (2015).

2. Graewert MA*, et al.* Automated pipeline for purification, biophysical and x-ray analysis of biomacromolecular solutions. *Sci Rep* **5**, 10734 (2015).

3. Jeffries CM, Graewert MA, Svergun DI, Blanchet CE. Limiting radiation damage for high-brilliance biological solution scattering: practical experience at the EMBL P12 beamline PETRAIII. *J Synchrotron Radiat* **22**, 273-279 (2015).

4. Hajizadeh NR, Franke D, Svergun DI. Integrated beamline control and data acquisition for small-angle X-ray scattering at the P12 BioSAXS beamline at PETRAIII storage ring DESY. *J Synchrotron Radiat* **25**, 906-914 (2018).

5. Franke D, Kikhney AG, Svergun DI. Automated Acquisition and Analysis of Small Angle X-ray Scattering Data. *Nuc Instr Meth in Phys Research Section A* **689**, 52-59 (2012).

6. Franke D*, et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017).

7. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. *Bioinformatics*, (2017).

8. Konarev PV, Svergun DI. A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ* **2**, 352-360 (2015).

9. Guinier A. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann Phys (Paris)* **12**, 161-237 (1939).

10. Jeffries CM, Graewert MA, Blanchet CE, Langley DB, Whitten AE, Svergun DI. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat Protoc* **11**, 2122-2153 (2016).

11. Franke D, Jeffries CM, Svergun D. Data Mining in Structural Biology: Machine Learning Methods for Data Analysis of Biological Macromolecules in Solution. *Biophysical Journal* **Accepted**, (2018).

12. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* **25**, 495-503 (1992).

13. Franke D, Jeffries CM, Svergun DI. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat Methods* **12**, 419-422 (2015).

14. Hajizadeh NR, Franke D, Jeffries CM, Svergun DI. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci Rep* **8**, 7204 (2018).

15. Kikhney AG, Panjkovich A, Sokolova AV, Svergun DI. DARA: a web server for rapid search of structural neighbours using solution small angle X-ray scattering data. *Bioinformatics* **32**, 616-618 (2016).

16. Fischer H, de Oliveira M, Napolitano HB, Craievich AF, Polikarpov I. The molecular weight of proteins in solution can be determined from a single SAXS measurement on a relative scale. . *J Appl Cryst* **43**, 101-109 (2010).

17. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-481 (2013).

18. Petoukhov MV, Svergun DI. Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr D* **71**, 1051-1058 (2015).

19. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879-2886 (1999).

20. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953 (2001).

21. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small angle scattering. *J Appl Cryst* **36**, 860-864 (2003).

22. Tuukkanen AT, Kleywegt GJ, Svergun DI. Resolution of ab initio shapes determined from small-angle scattering. *IUCrJ* **3**, 440-447 (2016).

23. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).

24. Suhre MH, Gertz M, Steegborn C, Scheibel T. Structural and functional features of a collagen-binding matrix protein from the mussel byssus. *Nat Commun* **5**, 3392 (2014).

25. Svergun DI, Barberato C, Koch MHJ. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* **28**, 768-773 (1995).

26. Petoukhov MV, Svergun DI. Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**, 1237-1250 (2005).

27. Tria G, Mertens HD, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217 (2015).

28. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* **129**, 5656-5664 (2007).