OAK RIDGE National Laboratory

Data analysis for solution scattering of biomacromolecular complexes

Wellington Claiton Leite

Center for Structural Molecular Biology

Neutron Scattering Division

Oak Ridge National Laboratory

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

BER Structural Biology Resources

U.S. Department of Energy

U.S. DEPARTMENT OF ENERGY

Topics to be covered

- Data analysis for solution scattering

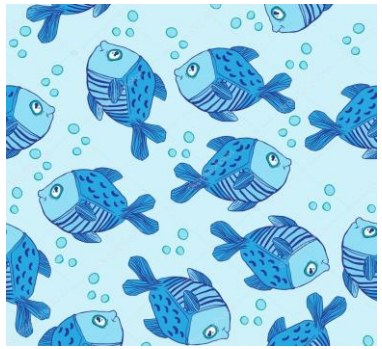
1. Guinier Analysis
2. P(r) calculation
3. Kratky Plot
4. Mw Estimation
5. Fitting experimental data with PDB structures
6. Software for modeling SANS data

Resources

- Data analysis software ATSAS 3.1.1 - https://www.embl-hamburg.de/biosaxs/software.html
  - https://www.embl-hamburg.de/biosaxs/atsas-online/
  - CRYSON
  - GNON
  - SAXSREF/CV
  - OLIGOMER
  - DAMMIN/DAMMIF
  - GASBOR
  - MONSA
  - EOM
  - SUPCOMP
- BioXTAS RAW - https://bioxtas-raw.readthedocs.io/en/latest/index.html
  - GNON
  - DAMMIN/DAMMIF
  - DENSS
- ScÅtter - https://bl1231.als.lbl.gov/scatter/
- Protein data bank - https://www.rcsb.org/
- AlphaFold - https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb
- Pymol – https://www.pymol.org/2/
- Chimera - https://www.cgl.ucsf.edu/chimera/
- FoxSAXS - https://modbase.compbio.ucsf.edu/foxs/

Monodispersed samples for solution scattering

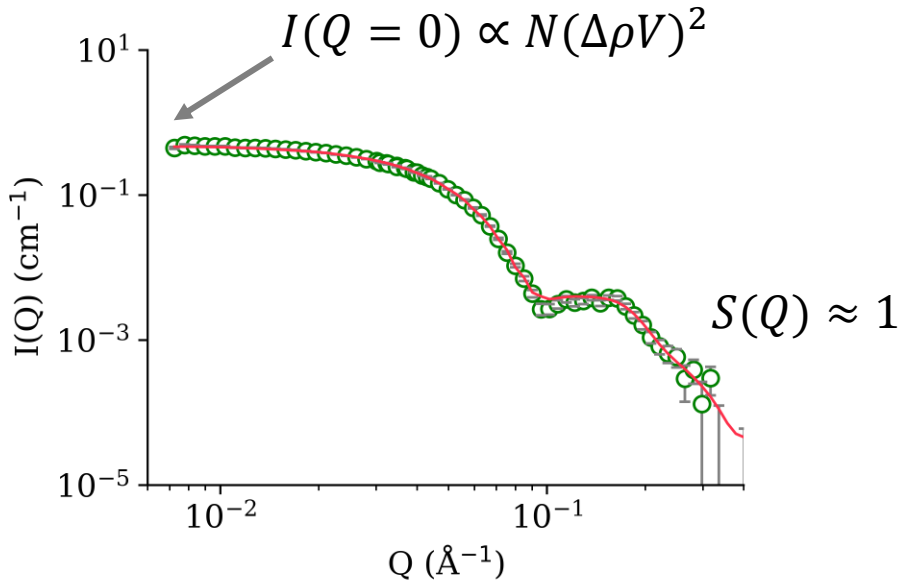
Homogeneous particles



$$I(q) = S(q) \sum_i^n [(\Delta\rho_i V_i)^2 P_i(q)] .$$

1. Pair distance distribution - Dmax
2. Overall size – Rg
3. Conformation – Globular vs extended
4. Oligomerization state
5. Molecular mass – MW
6. Comparison with PDB structures (AlphaFold)
7. Molecular envelope

Scattering 1D profile of monodispersed sample



Experiment details:
- Protiated protein in 100% D_2O Buffer
- Buffer exchange to 100% D_2O applying concentrations/dilution cycles
- Concentration used – 2 mg/mL (41.7 KDa)
- Data collected for 2 h (good S/N)
- Volume used 320 uL
- Oligomer: Hexamer

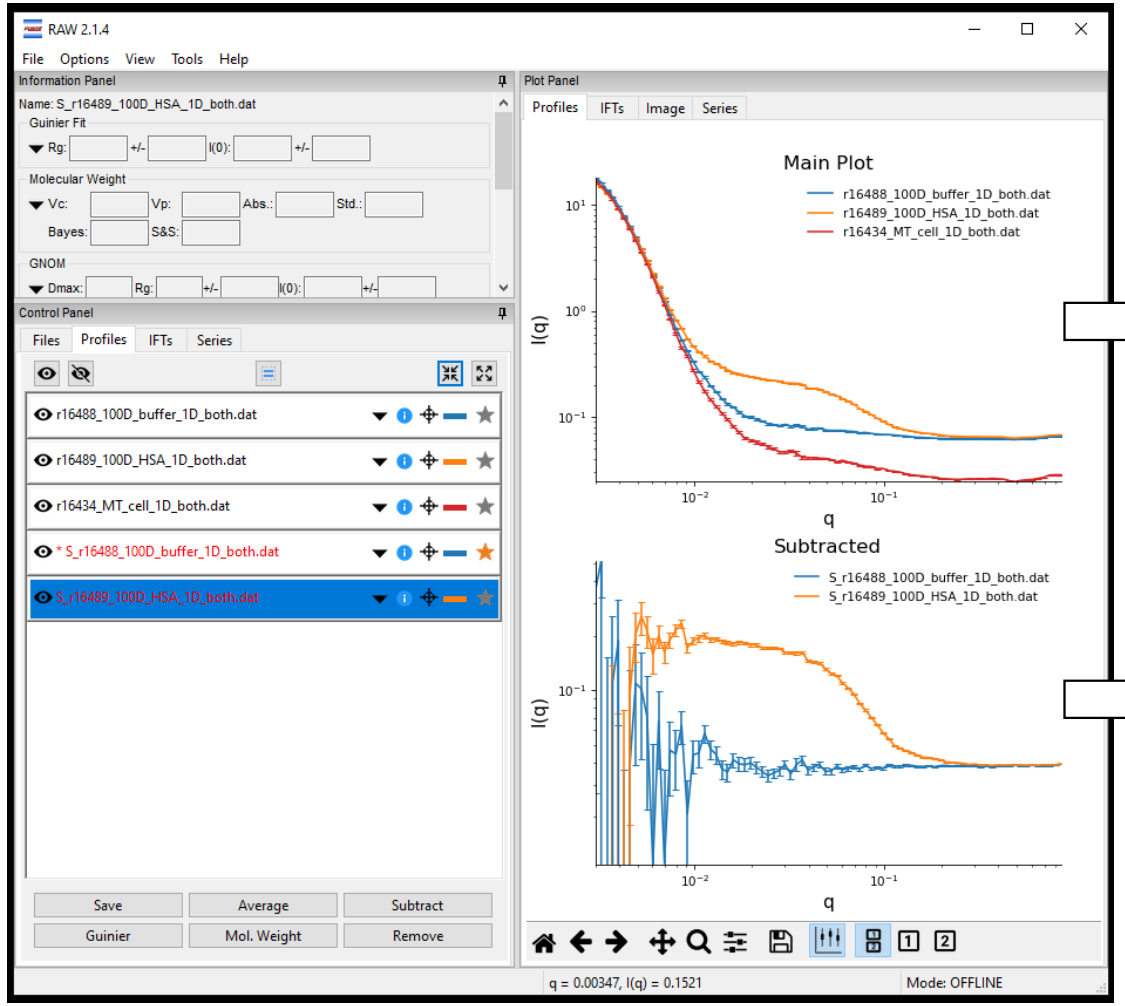
Background subtraction

Data manipulation
- Average
- Subtracted
- Merge
- Rebin

Data collected
- Empty cell
- Buffer
- Sample

Empty cell and buffer are backgrounds

First step - Empty cell subtraction



Before subtraction

After subtraction

Background subtraction

Data manipulation

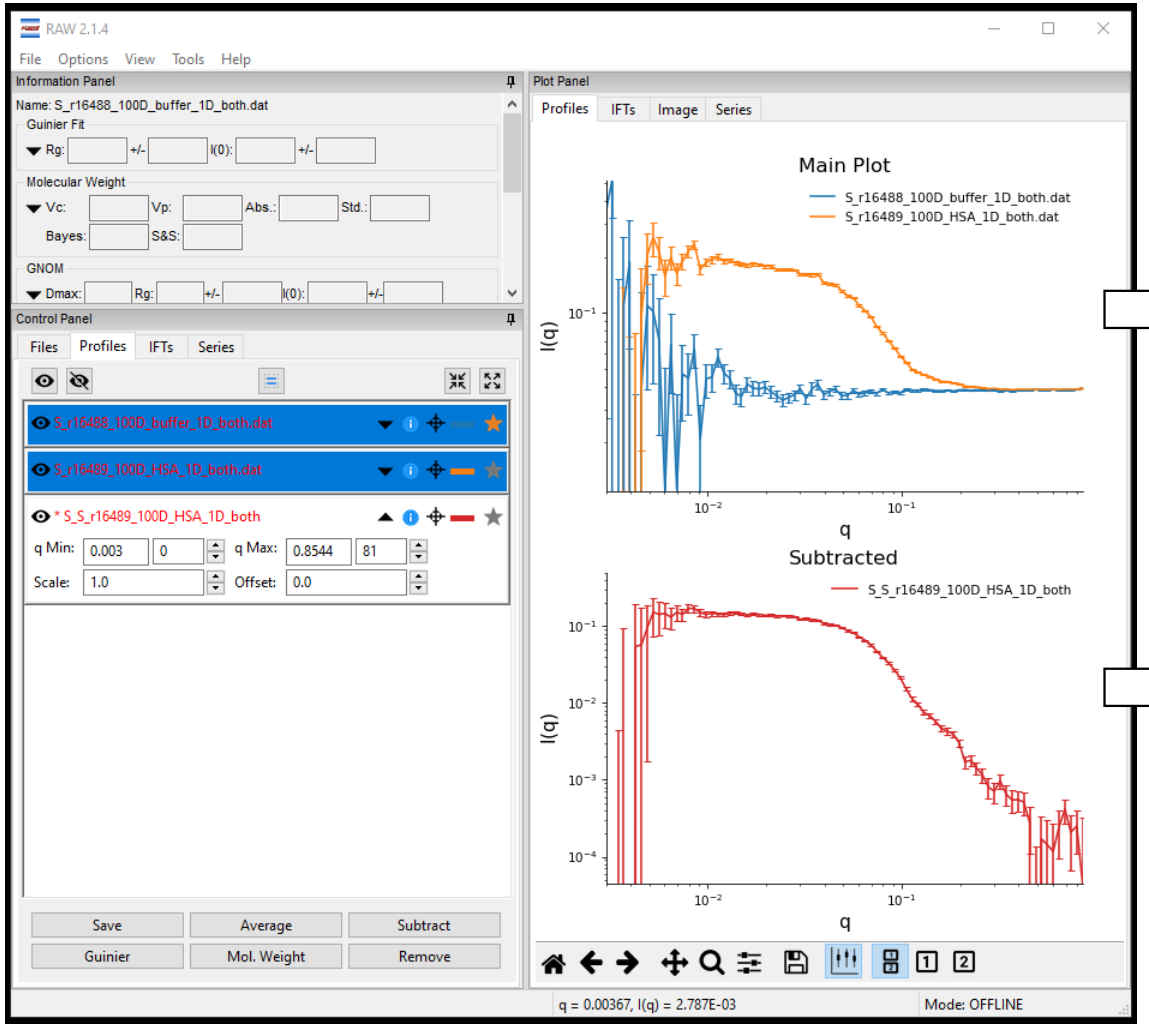
- Average
- Subtracted
- Merge
- Rebin

Data collected

- Empty cell
- Buffer
- Sample

Empty cell and buffer are backgrounds

Second step - Buffer subtraction

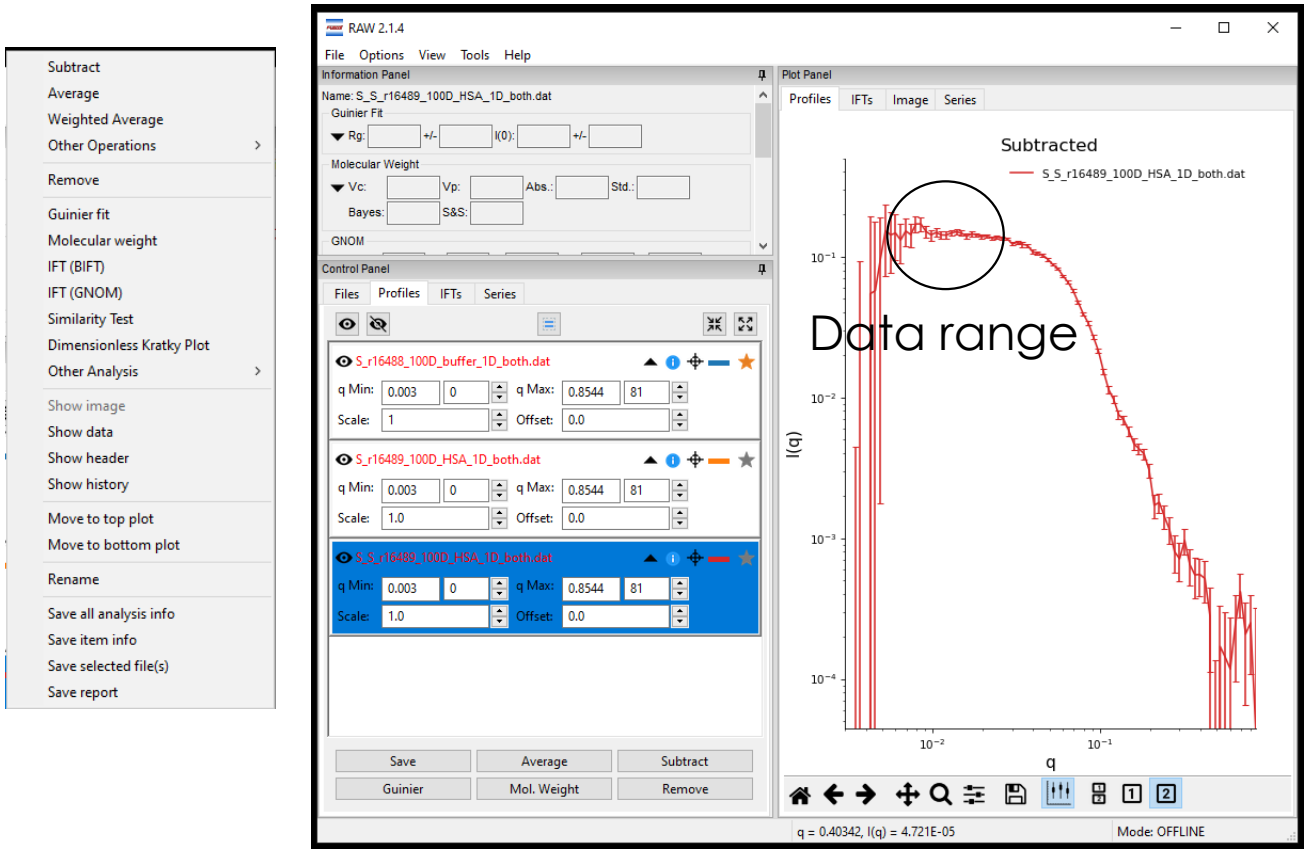


Before subtraction

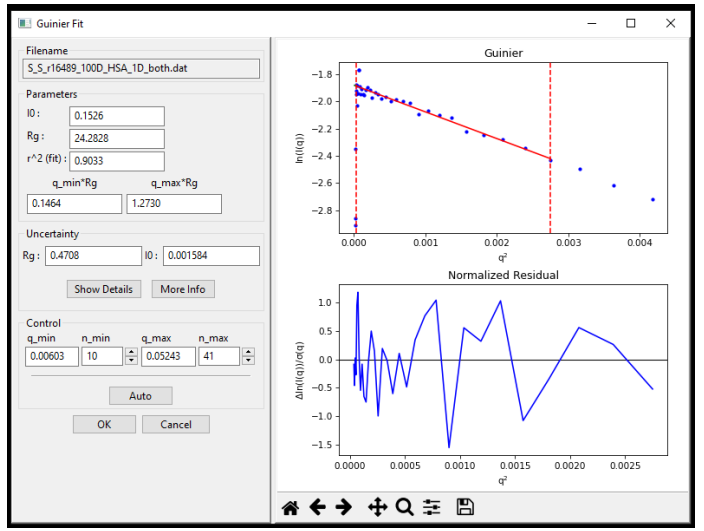
After subtraction

Guinier Analysis

Subtracted Scattering profile



Guinier Analysis



Guinier Analysis ($R_g^{Guinier}$ and $I_0^{Guinier}$)

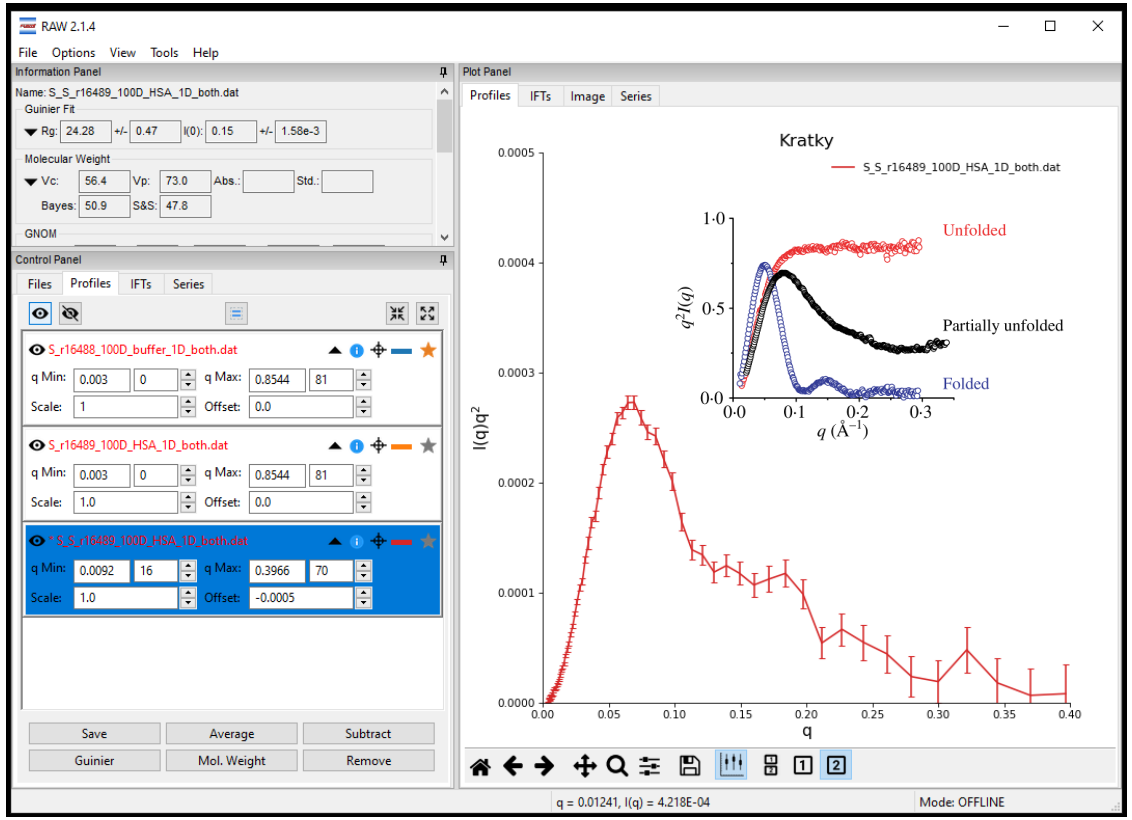
$$I(q) \cong I(0)\exp(\frac{-R_g^2 q^2}{3})$$

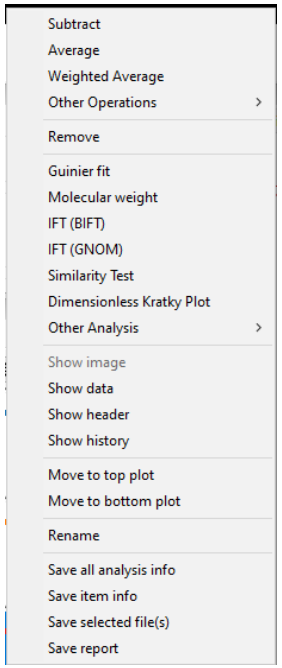
To get reliable Guinier plot / Rg analysis:
$Q_{min}Rg \leq 0.65$
$Q_{max}Rg \leq 1.3$ for globular
$Q_{max}Rg \leq 0.8$ for elongated
$Q_{max}/Q_{min} > 2$

Guinier analysis allows model-free determination of the radius of gyration ($R_g^{Guinier}$) and $I_0^{Guinier}$ (Forward scattering)

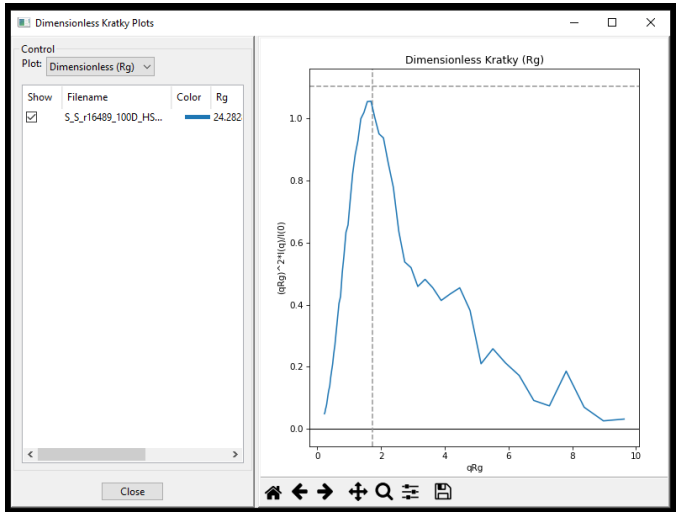
Kratky plot analysis

Kratky plot





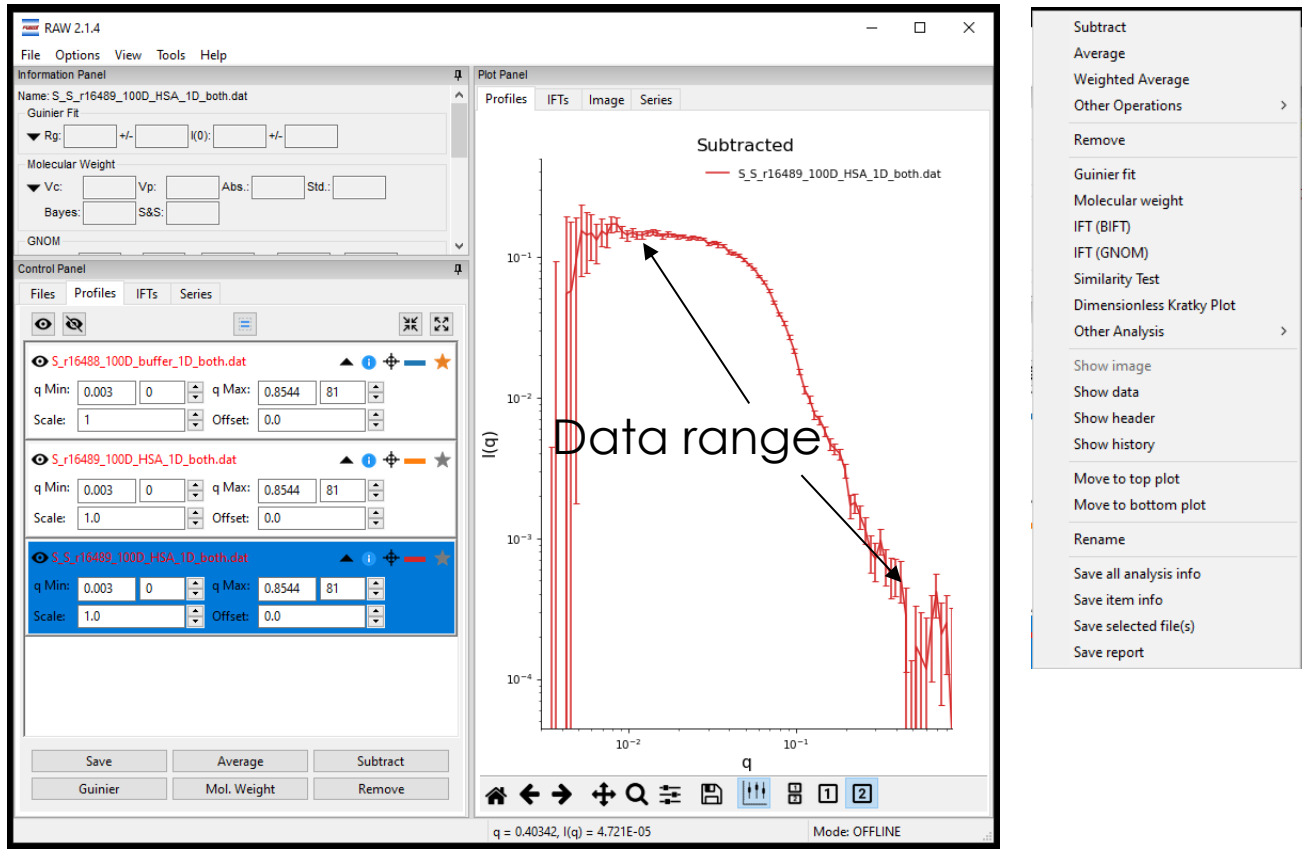
Dimensionless Kratky plot



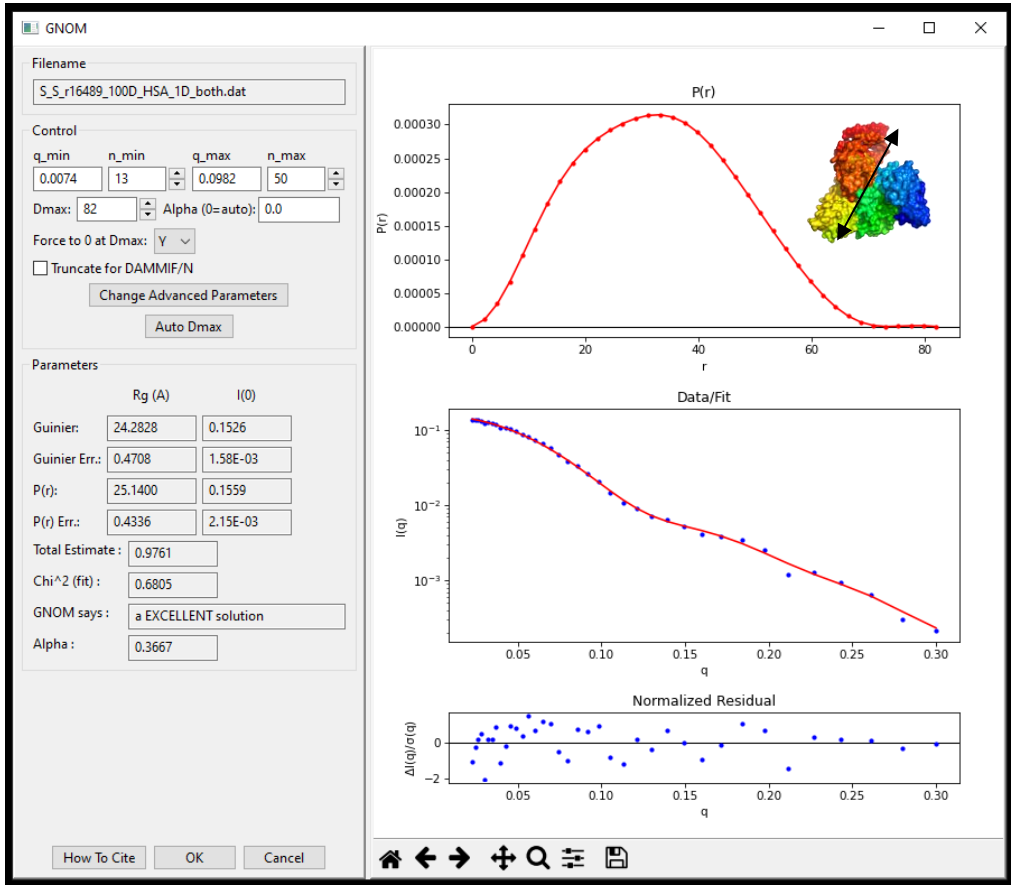
Kratky Plot analysis: A Kratky plot is a plot of I(Q)xQ^{2}vs.Q. Kratky plot can qualitatively assess the flexibility and/or degree of unfolding in samples, as well as distinguish between extended to compact conformations.

Pair distance distribution

Scattering profile



P(r) profile

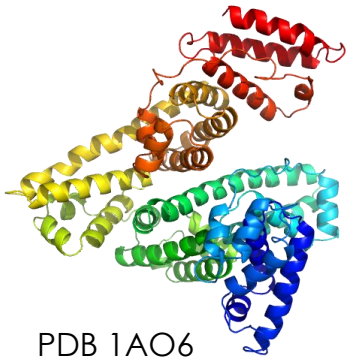


The Inverse Fourier Transform (IFT) transforms experimental (1D I(Q) vs. Q) curves in reciprocal space to produce the pairwise distance distribution profile (P(r) vs. r) of atoms in real space.

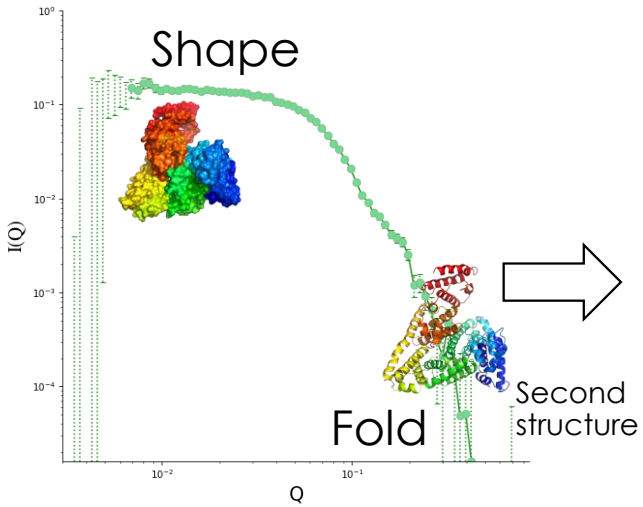
P(r) analysis (D_{max}, R_g^{GNOM} , and I_0^{GNOM})

Comparison with PDB structures

Crystal structure of Human Serum Albumin (HSA)



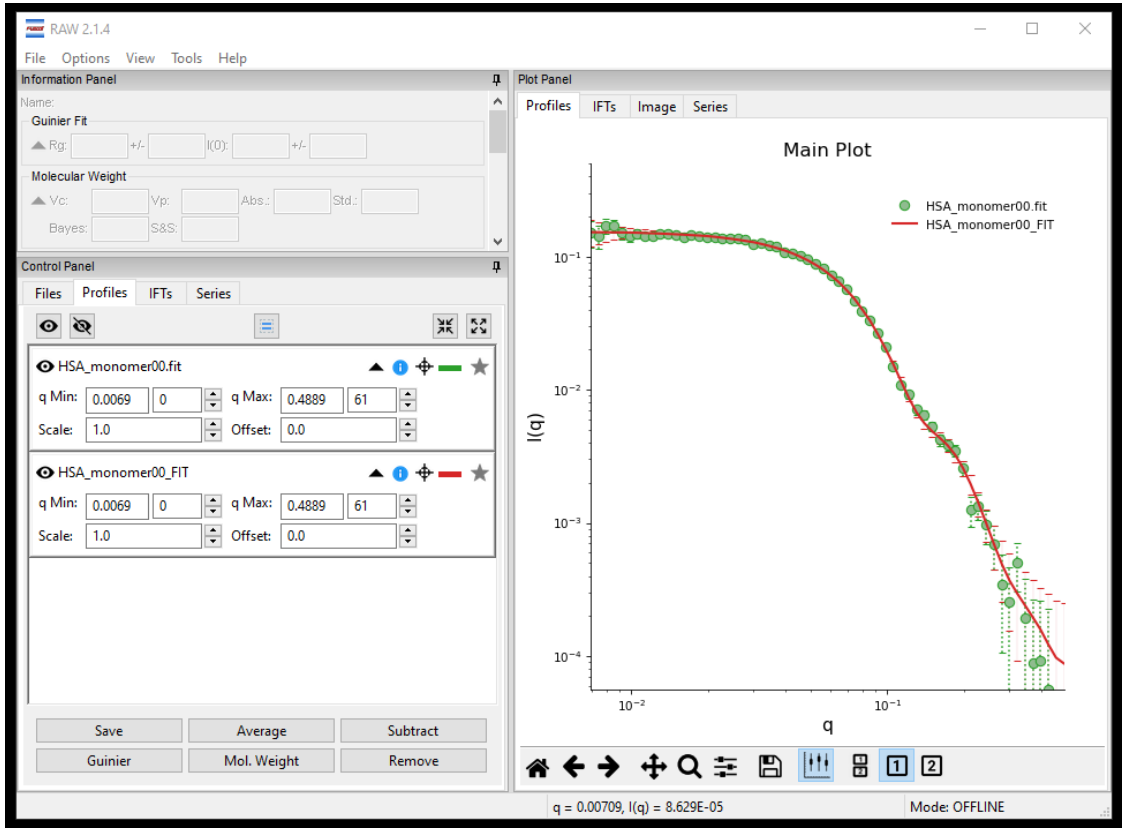
PDB 1AO6



- Crystal structure
- Cryo-EM
- NMR
- AlphaFold
- Homology models

```
Created by CRYSON v. 27 on   07-Sep-2022   21:33:44

Data file name .....................................
Model: HSA_monome  Rg:  25.16  Run: 00  Chi^2:    0.85
```



CRYSON is a program for evaluating the solution scattering from macromolecules with known atomic structure and fitting it to experimental scattering curves from Small-Angle Neutron Scattering (SANS).

Molecular mass calculation

Molecular mass calculation for monodispersed species

Standard equation from SAS

$$MW = \frac{I(0)N_A}{C_w(\Delta\rho v)^2}$$

$I(0)$ = Forward scattering

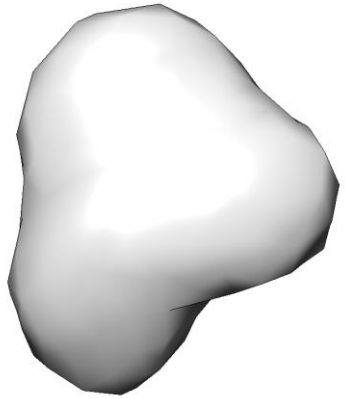
- Guinier Analysis
- Pair distance distribution

Intensity at Q=0	$I(0)$	0.16	cm^{-1}
Avogadro Number	N_A	6.023	$10^{23}\ mol^{-1}$
Molar concentration	C_w	2.6	$10^{-3}\ g/cm^3$
Scattering length Density			
Protein	rho_Prot	2.97	cm^{-2}
Solvent	rho_Solv	6.36	cm^{-2}
Difference	delta_rho	3.39	cm^{-2}
Partial specific volume	v	0.708	cm^3/g
Numerator	$I(0)\ N_A$	0.964	$cm^{-1}\ mol^{-1}$
Denominator	C_w (delta_rho v)^2	14.977	$g^{-1}\ cm^{-1}$
Result	M_w	64.3	kDa

% D2O	Water	hProtein
0	-5.61E+09	1.94E+10
2	-4.21E+09	1.96E+10
4	-2.82E+09	1.99E+10
6	-1.43E+09	2.01E+10
8	-3.54E+07	2.03E+10
10	1.36E+09	2.05E+10
12	2.75E+09	2.07E+10
14	4.14E+09	2.09E+10
16	5.53E+09	2.11E+10
18	6.93E+09	2.13E+10
20	8.32E+09	2.15E+10
22	9.71E+09	2.17E+10
24	1.11E+10	2.19E+10
26	1.25E+10	2.21E+10
28	1.39E+10	2.23E+10
30	1.53E+10	2.25E+10
32	1.67E+10	2.27E+10
34	1.81E+10	2.29E+10
36	1.95E+10	2.31E+10
38	2.09E+10	2.33E+10
40	2.22E+10	2.35E+10
42	2.36E+10	2.37E+10
44	2.50E+10	2.40E+10
46	2.64E+10	2.42E+10
48	2.78E+10	2.44E+10
50	2.92E+10	2.46E+10
52	3.06E+10	2.48E+10
54	3.20E+10	2.50E+10
56	3.34E+10	2.52E+10
58	3.48E+10	2.54E+10
60	3.62E+10	2.56E+10
62	3.76E+10	2.58E+10
64	3.90E+10	2.60E+10
66	4.03E+10	2.62E+10
68	4.17E+10	2.64E+10
70	4.31E+10	2.66E+10
72	4.45E+10	2.68E+10
74	4.59E+10	2.70E+10
76	4.73E+10	2.72E+10
78	4.87E+10	2.74E+10
80	5.01E+10	2.76E+10
82	5.15E+10	2.79E+10
84	5.29E+10	2.81E+10
85	5.36E+10	2.82E+10
86	5.43E+10	2.83E+10
88	5.57E+10	2.85E+10
90	5.71E+10	2.87E+10
92	5.85E+10	2.89E+10
94	5.98E+10	2.91E+10
96	6.12E+10	2.93E+10
98	6.26E+10	2.95E+10
100	6.40E+10	2.97E+10

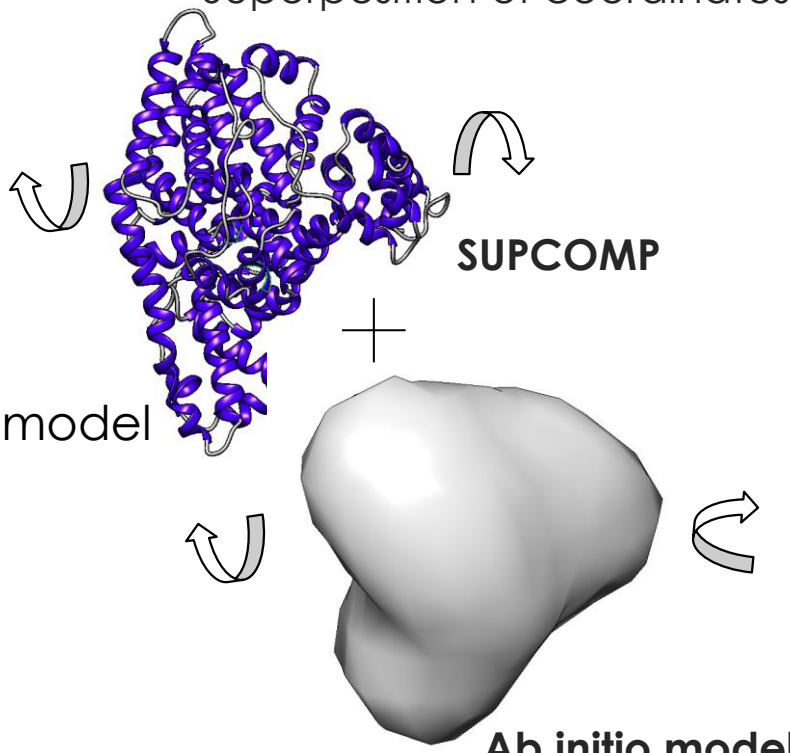
Envelope of a macromolecule using Ab initio modeling

DAMMIN/DAMMIF

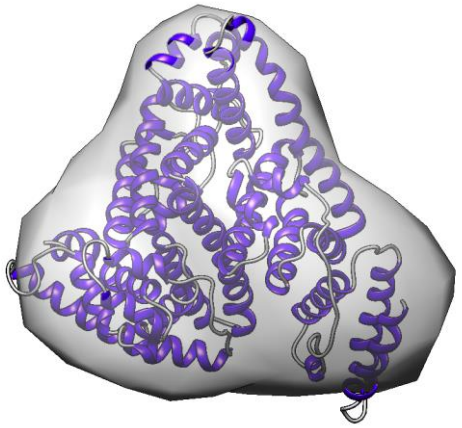


Ab initio model

Superposition of coordinates

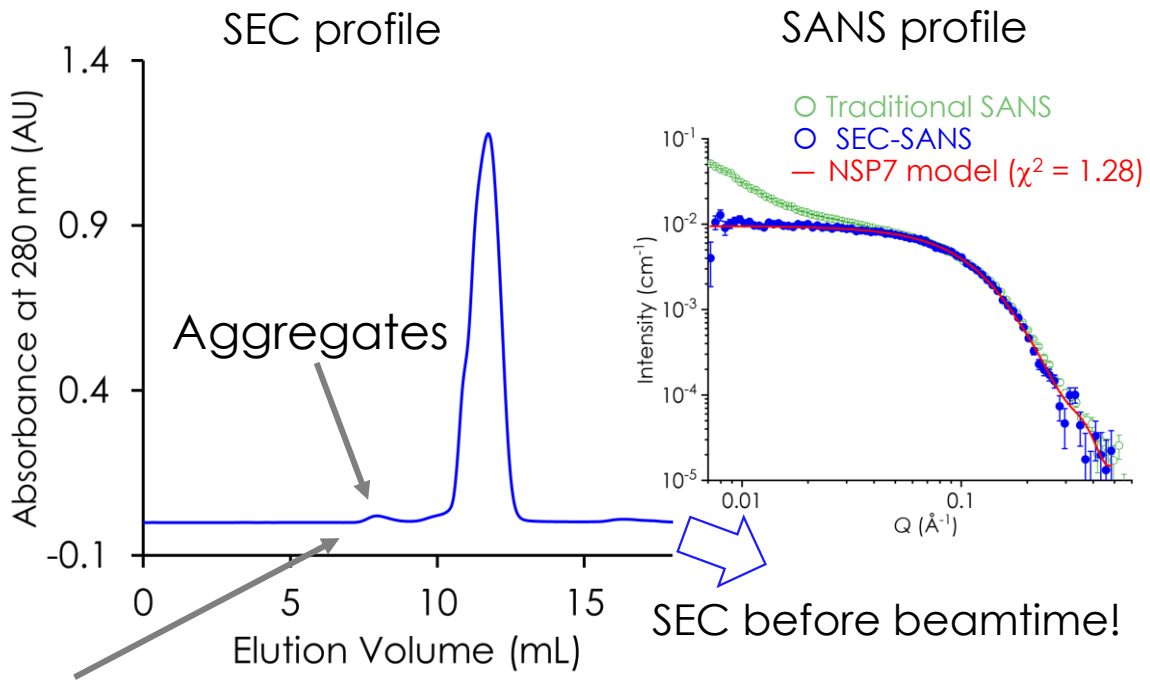


Overall shape or envelope



Interparticle Interferences

Attractive interaction between proteins in solution

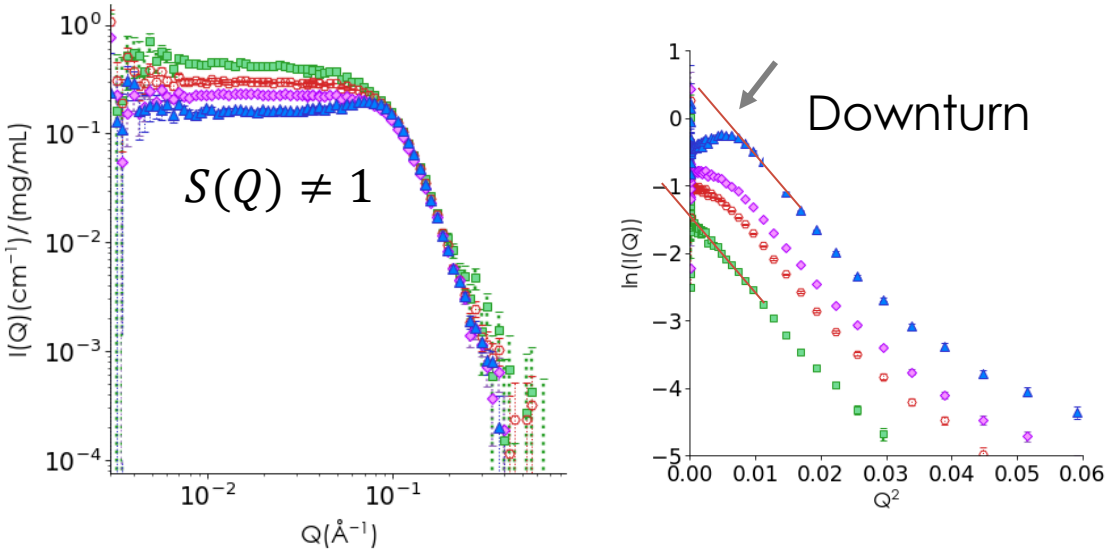


Note: Aggregates have low absorbance at 280 nm, but have high light scattering (MALS)

Experiment details:
- Unlabeled Non-Structural Protein 7 (9.2 KDa)
- Injected ~500µL at ~20mg/mL
- 100% D_2O buffer
- Superdex 75 10/300 GL (24mL CV)
- Data collection triggered manually
- Higher order aggregates removed!
- Nsp7 exists as a dimer (18.4kDa) in solution

Repulsion interaction between proteins in solution

SANS profile and Guinier Analysis

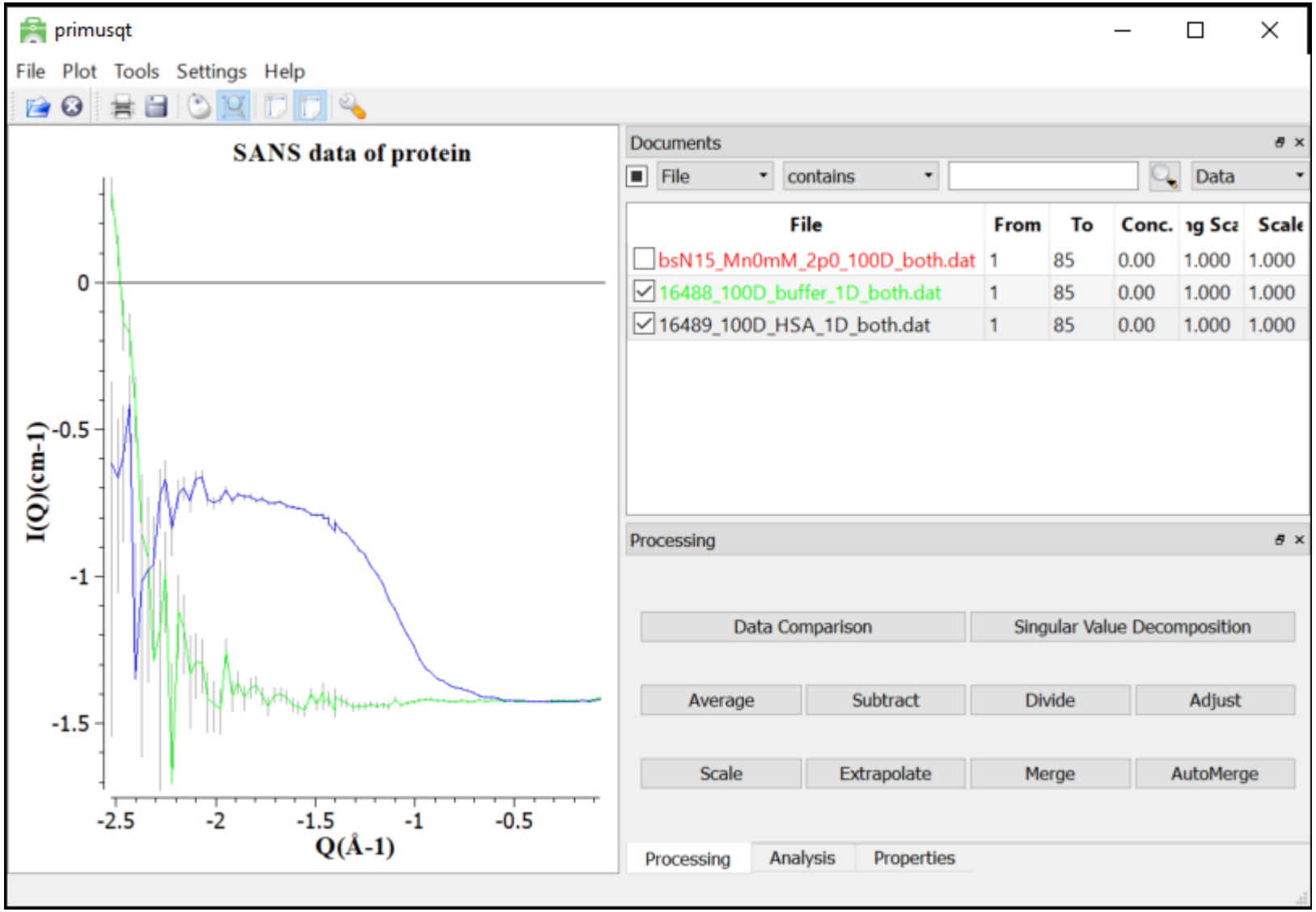


- Membrane protein in micelles
- Sample concentration was reduced 8-fold to reduce attractive interactions

Resources

Data analysis software ATSAS 3.1.1 - https://www.embl-hamburg.de/biosaxs/software.html
Experimental data processing

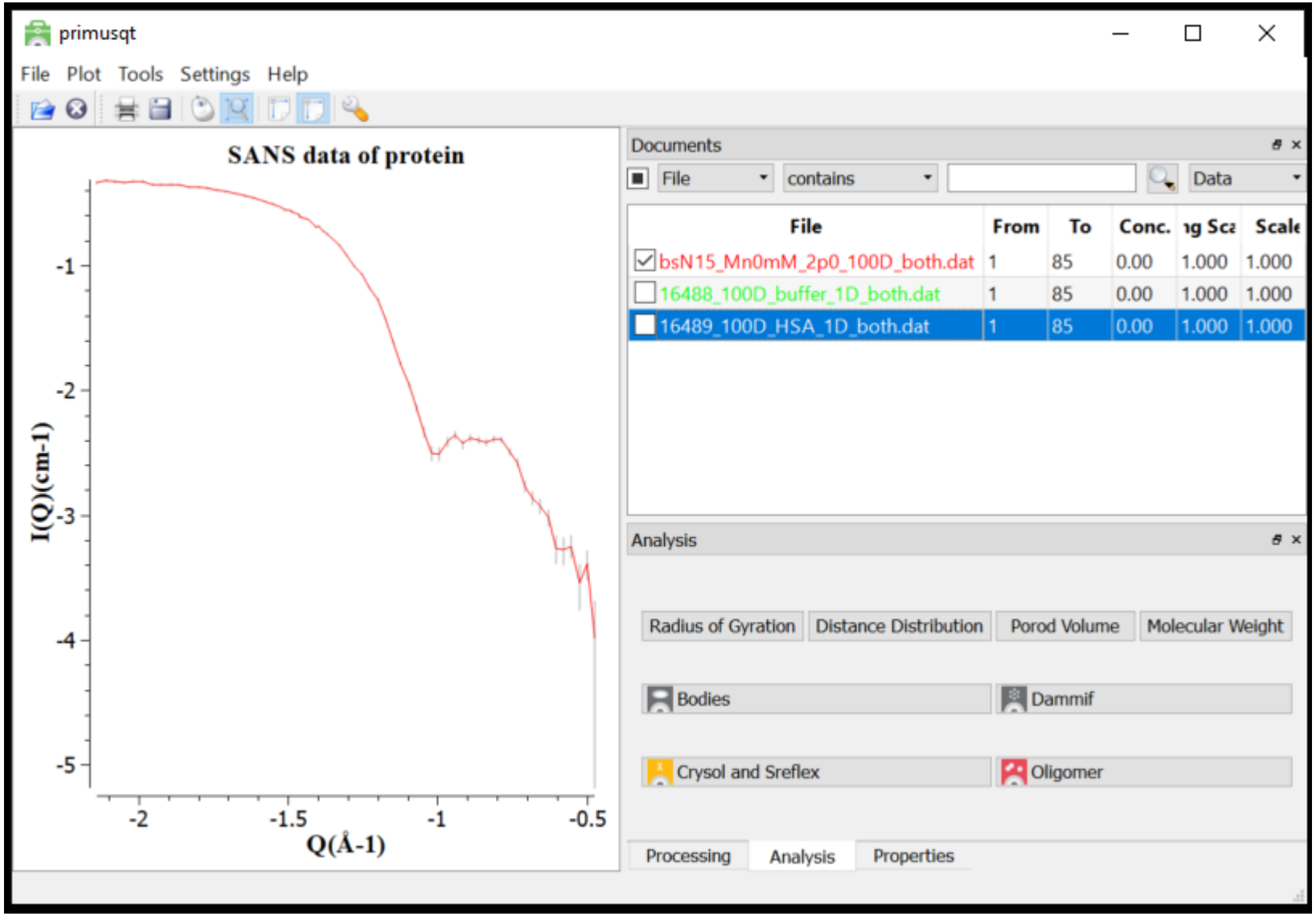
- PRIMUS - GUI for manipulations and primary analysis of experimental 1D SAS data



Resources

Data analysis software ATSAS 3.1.1 - https://www.embl-hamburg.de/biosaxs/software.html
Experimental data processing

- PRIMUS - GUI for manipulations and primary analysis of experimental 1D SAS data



Facility Acknowledgment Statement

- A portion of neutron scattering research presented as examples in this introduction used resources at the High Flux Isotope Reactor or Spallation Neutron Source, DOE Office of Science User Facilities, operated by the Oak Ridge National Laboratory.
- The Bio-SANS of the Center for Structural Molecular Biology at the High Flux Isotope Reactor is supported by the Office of Biological and Environmental Research of the U.S. DOE.