Analysis I: Model-free and model-based fitting

Adrian Sanchez-Fernandez
adrian.sanchez-fernandez@food.lth.se

What is the purpose?

- Scattering in the small-angle arises from **inhomogeneities in the scattering length density profile**, *ρ(r)*.

$$F(q) = \int_V \rho(r) e^{qri} dr$$

$$\frac{d\Sigma}{d\Omega}(q) = \frac{N}{V}\frac{d\sigma}{d\Omega}(q) = \frac{1}{V}\left| \int_V \enclose{circle}{\rho(r)} e^{qri} dr \right|^2$$

- Measured scattered intensity (*I(q)*) relates to the Fourier transform of the scattering length density profile – structure of the scatterer.

- Loss of information – *I(q)* is a function of *q*, not *r*. **Thus it is not always easy to determine *ρ(r)***.

- The combination of **different approaches, multiple contrasts and complementary techniques** lead to a more reliable and robust result.

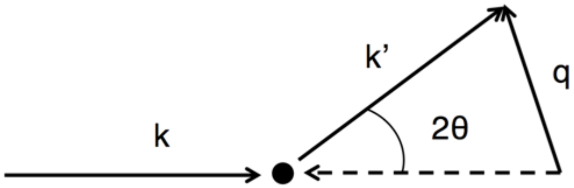
What is the question?

- **What can be measured** with SAS?
  - Probes structures on the 1 to 100’s of nm **length scale**.
  - Features to measure in the right length scale – **q-range**.

- **Contrast, deuteration and composition** – what can be measured with SAXS and SANS?
  - Is there any **contrast** in the sample?
  - Specific **deuteration schemes** and **contrast matching**.
  - Does **isotopic labelling** affect the sample characteristics? (e.g. surfactant CMC or protein hydrogen bonding).
  - **Concentration** range – dilute regime vs concentrated regime.

The scattering vector q

- The scattering vector describes the change of the wave vector: q=k’-k.



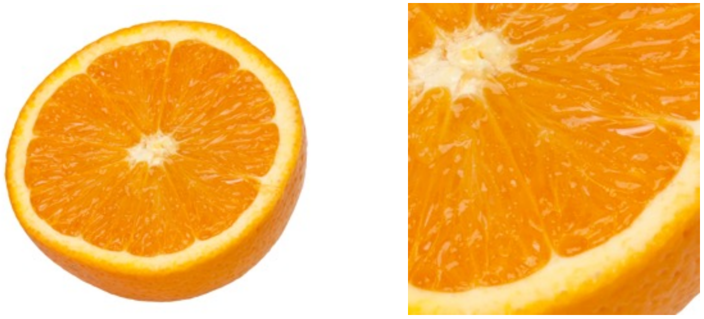
- de Broglie relates the magnitude of the wavevector to the wavelength – elastic scattering lkl=lk’l.

$$|k| = |k'| = \frac{2\pi}{\lambda}$$

$$q = \frac{4\pi\ sin\theta}{\lambda}$$

- **The q-vector standardises the region of interest** – it is a measure of the **reciprocal space**.

$$q \approx \frac{2\pi}{d}$$



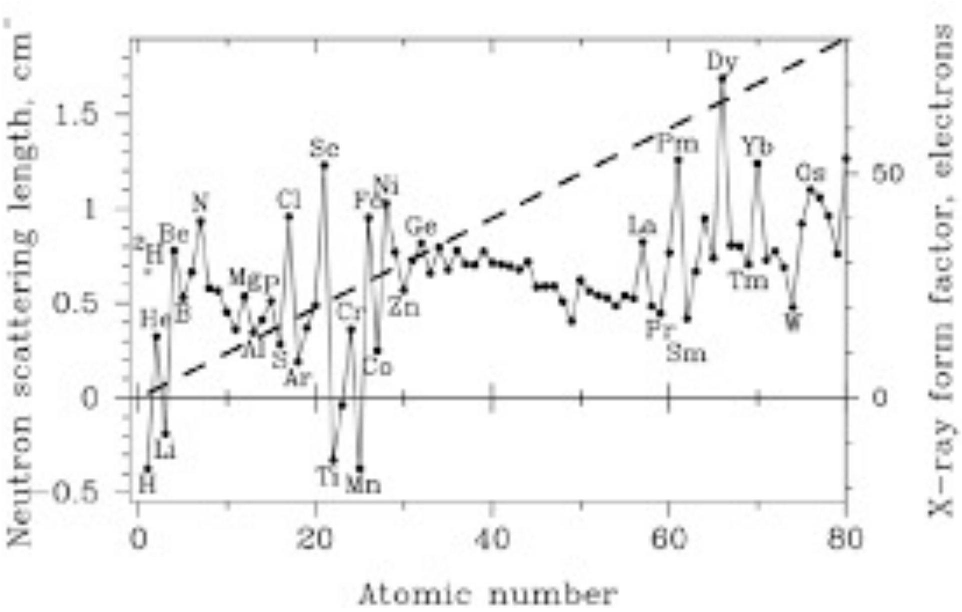
Scattering length density

- The scattering length density (SLD or ρ(r)) quantifies the **scattering power of an ensemble of atoms**.

- Sum of the **atomic scattering lengths** divided by the **volume of the ensemble**.

$$SLD = \frac{\sum_i^n b_i}{V_m}$$

where b_i is the neutron or X-ray **atomic scattering length**.



- Fitting algorithms.
- Resolution functions.
- Model-independent fitting – standard plots.
- Empirical models.
- Indirect Fourier Transformation
- Model-based fitting.
- Fitting strategy

Data analysis hierarchy

- Different approaches – **level of detail and complexity of the analysis**.

Degree of complexity

  - Model-free approaches – quick and rough analysis.
  - Empirical models – identify data trends.
  - Indirect Fourier Transform – real-space data analysis (Analysis II).
  - Model-based fitting – mathematical methods.
  - Advance fitting – simulation-assisted methods.

Algorithms, goodness of fit and resolution

Fitting algorithm and Chi-squared test

- **Approximates a solution** that minimises a function that is the sum of the squares.

$$min_x f(x) = min_x \sum_i {F_i}^2(x)$$

- **Nonlinear squares methods** are used – e.g Levenberg-Marquardt algorithm.

- Includes the statistical weights of each point – **experimental error**.

- Chi^2 is a **statistical parameter** that quantifies the differences between the experimental data and theoretical data (or model).

$$\frac{\chi^2}{n} = \frac{\sum_i^n \frac{\left(y_i - y_{i,theory}\right)^2}{{\sigma_i}^2}}{n}$$

- For a good fit, Chi^2 **tends to zero**.

Marquardt, J Appl Math, 1963.

Resolution function

- The instrument resolution can be **theoretically calculated** from the configuration of the instrument (seen in the previous lecture).

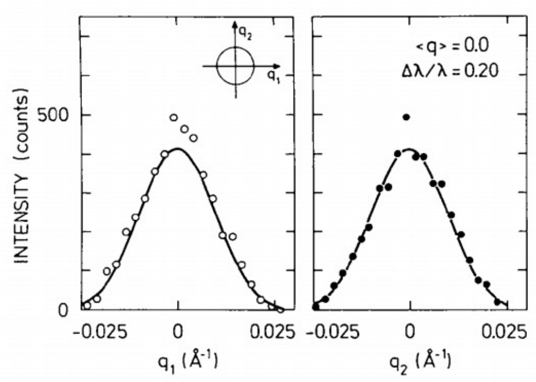
- Data de-smearing is a complicated procedure – **model smearing** is routinely applied.

$$I_{model}(q) = \int R(q,\langle q\rangle)\frac{d\sigma}{d\Omega}(q)dq$$

where R(q,<q>) is the q-dependant resolution function.

- For a pinhole-collimated beam is approximated by a **Gaussian function** with a given dq/q that relates to the FWHM of the function.

Pedersen, J Appl Cryst, 1990.



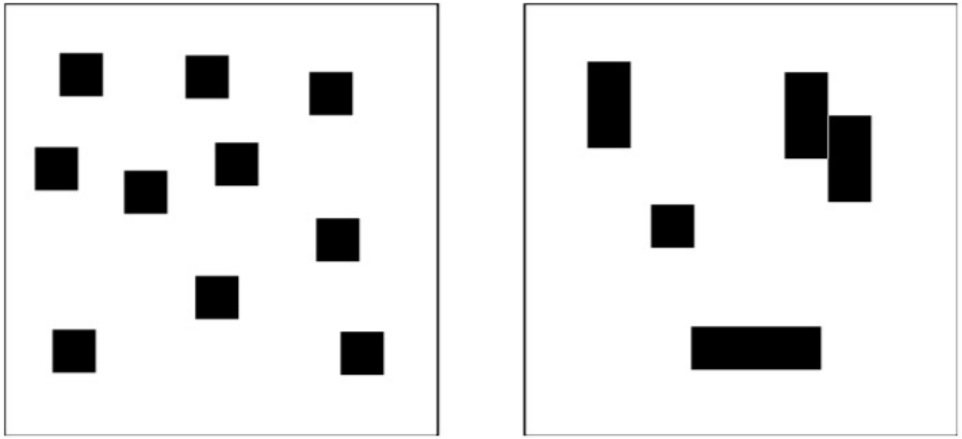
Standard plots and empirical models

Model-free fitting

- Tool for **rapid characterisation** of the scatterer.
- Good **starting point** for data analysis and assessing the quality of the sample/data.
- Assume no interparticle interactions – limited to S(q)=1.

Scattering invariant

- **Integrated scattering cross-section** – Q*.
- Total signal is **independent of the density distribution** of the system.



- This analysis allows for the calculation of the **volume fraction of scatterer** in a two-phase system.

$$Q^{*} = 2\pi^{2}\phi_{p}(1-\phi_{p})\,(SLD_{s}-SLD_{p})^{2}$$

- Requires absolute scaling of the data.

Porod exponent

- At high q (q >> 2π/d) the scattering is dominated by **correlations at interfaces** – for sharp interfaces:

$$I(q) = \frac{A}{q^4} + B$$

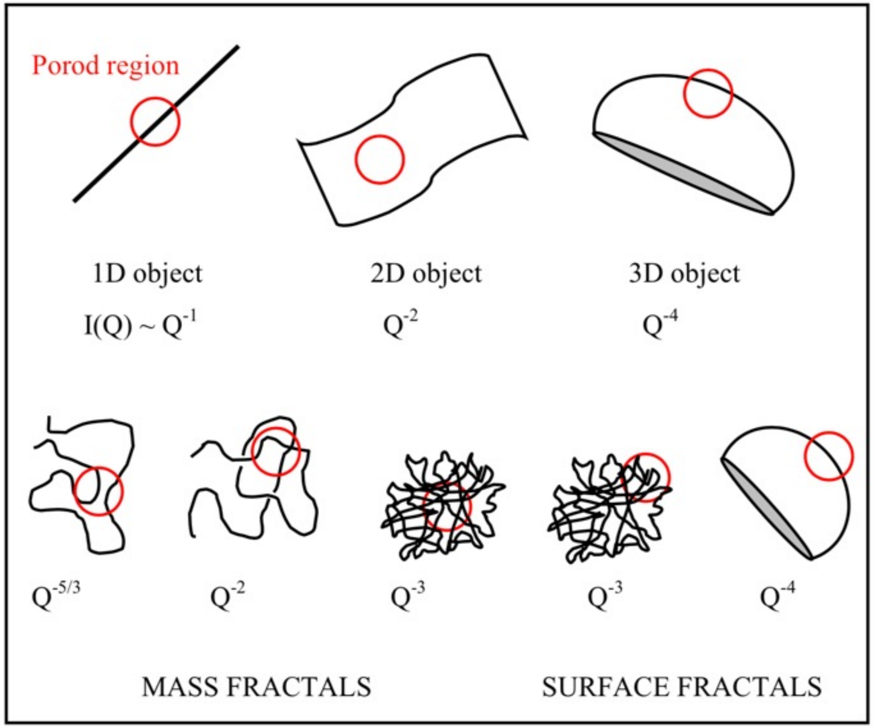
where A is an analytical parameter that relates to the surface-to-volume ratio of the scatterer and B is the experimental background.

The Porod exponent for **different interfaces.**

- ***log(I(q))* vs. *log(q)* plot**.

$$\log(I(q)) = \log(A) - m \log(q)$$

where *m* is the Porod exponent.

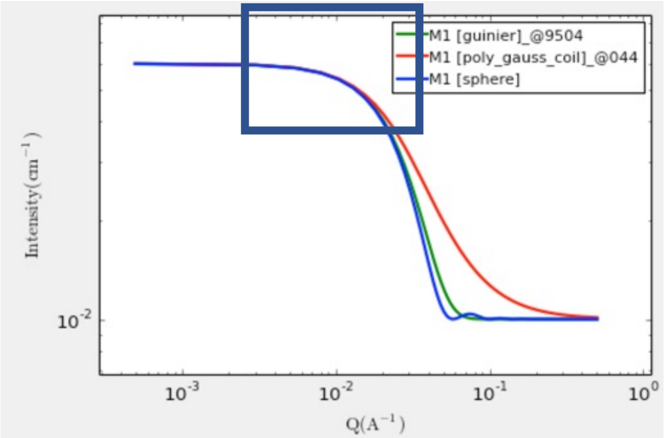


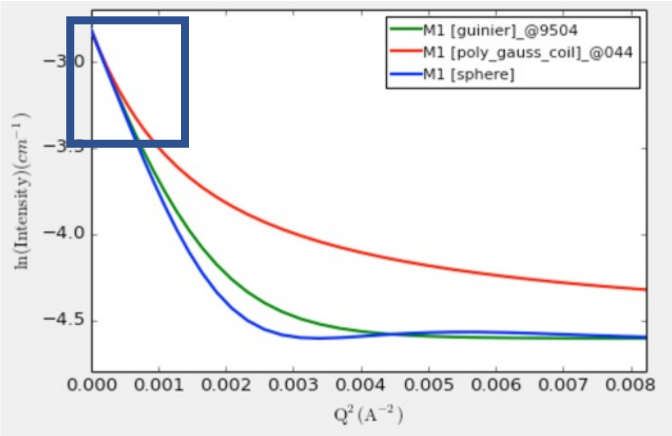
Guinier plot

- The scattering at low q can be describe through the Guinier relationship regardless the morphology of the scatterer.

$$I(q) = I(0)e^{-\frac{q^2 {R_g}^2}{3}}$$

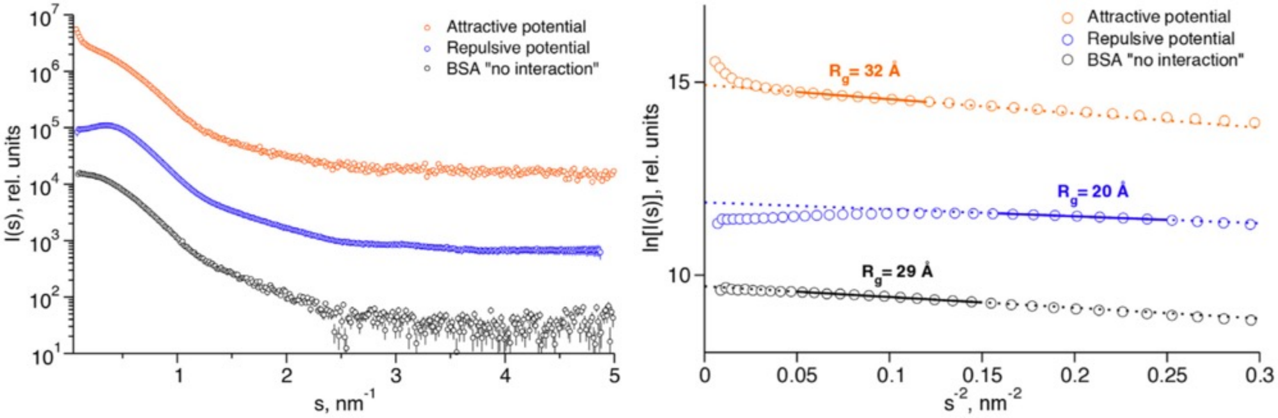
- *$ln(I(q)$* **vs** *q^2.*
- R_g is the radius of gyration of the scatterer and I(0) the extrapolated intensity at zero angle.
- Validity – **$q \times R_g < 1.3$**. Why?





Guinier plot

- The Guinier plot can be used as a tool for a quick **evaluation of the sample characteristics** – look for:
  - Structure factor (decrease in intensity at low q).
  - Aggregation (increase in intensity at low q).
  - Multiple Guinier regions.



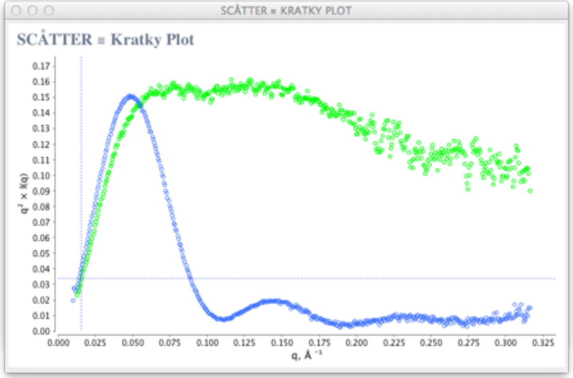
Mertens *et al.*, Arch Biochem Biophys, 2017.

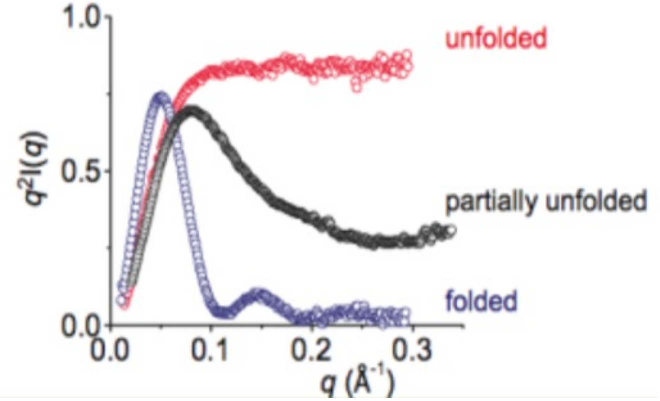
Kratky plot

- **Qualitative assessment of sample morphology** – derived from the theoretical scattering of a Gaussian coil (q^{-2} at high q).
  - Globularity vs anisotropy.

- q^2xI(q) vs q.

- Globularity – decreasing oscillations with q. Random coil – plateau at high q. Intermediate conformations will show somewhere in between those.





- Particularly useful for **protein systems** – highlights conformational changes in the macromolecule.

Periodic structures

- Periodicity/order is characterised by the **presence of peaks.**

- Peak position relates to the **d-spacing** of the crystal.

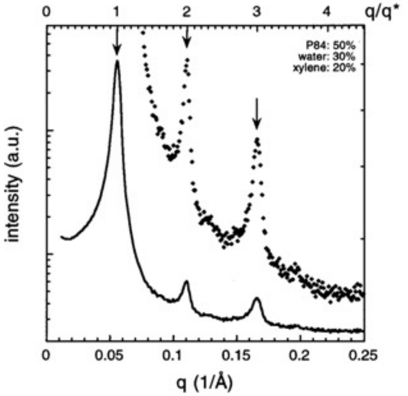
$$d = \frac{2\pi}{q_{max}}$$

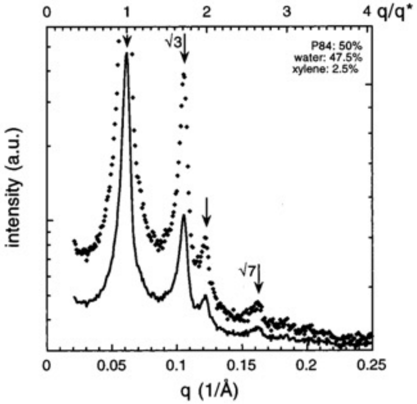
- Relative peak position gives the **lattice structure in reciprocal space**.

$$n = (h^2 + k^2 + l^2)^{1/2}$$

where h, k and l are Miller indices (reflections in reciprocal space).

- Lamellar – 1, 2, 3, 4…; hexagonal – 1, √3, 2, √7…





- SANS is not great for studying periodic structures – resolution!

Analysis III:
Empirical models

Adrian Sanchez-Fernandez

Empirical models

- These models reproduce the **main trends** observed in the SANS data.
- **Intermediate complexity** between standard plots and model-based fitting.
- Correlation length model.
- Gaussian peak model.
- Broad peak model.
- Teubner-Strey model.
- Beaucage model.

Correlation length model

- Approximates the characteristic size and shape of an amorphous system.

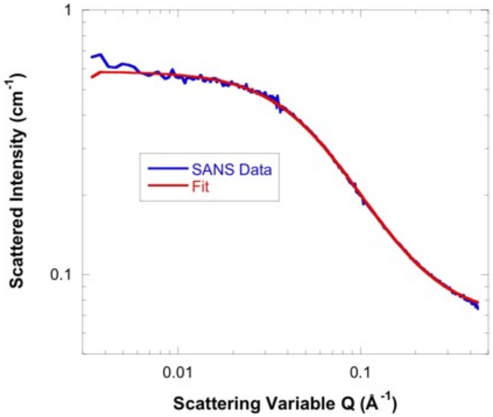
$$I(q) = \frac{C}{1 + (q\xi)^m} + B$$

- where *C* and *B* relate to the limits of I(q) (→0; →∞), *ξ* is a correlation length and *m* is a Porod exponent.

- Applied to polymer systems.

- Correlation length determines the average distance between entanglements. Porod exponent gives information about the structure of the polymer. *C* can be used to determine the MW of the polymer.

Poly(ethylene oxide)/d-water

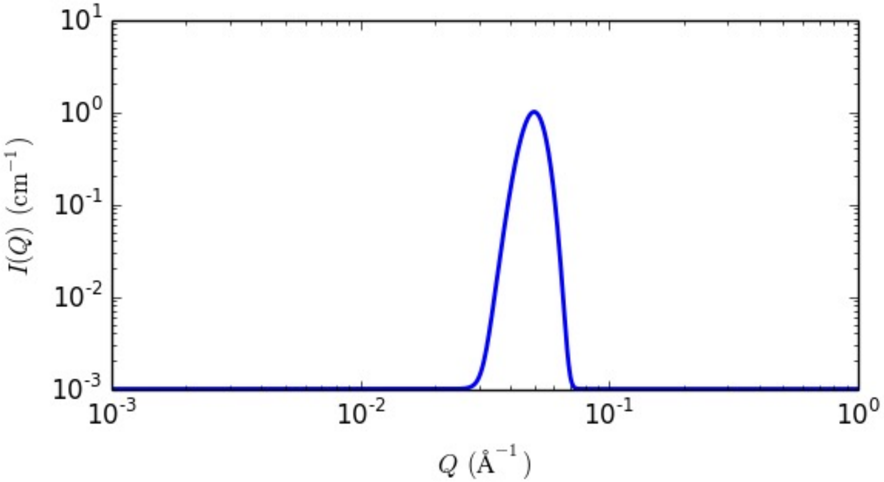


Gaussian peak model

- Describes a Gaussian peak on a flat background.

$$I(q) = S\, e^{-\frac{1}{2}\frac{(q-q_0)^2}{\sigma^2}} + B$$

- where *S* is the scale factor, q_0 corresponds to the center of the peak, σ is the standard deviation (FWHM=2.354σ) and *B* is the background.



- The peak position relates to the d-spacing and the broadening of the peak to the size distribution.

$$d = \frac{2\pi}{q_0}$$

Broad peak model

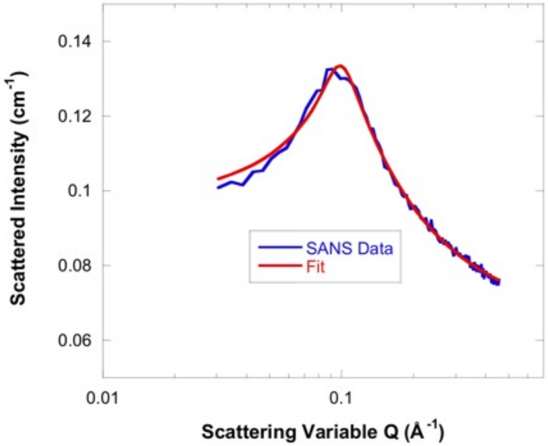
- Determines the characteristic distance between scattering inhomogeneities.

$$I(q) = \frac{C}{1 + (|q - q_0|\xi)^m} + B$$

- where *C* and *B* relate to the limits of I(q) (→0; → ∞), ξ is a correlation length, q_0 corresponds to the center of the peak and *m* is a Porod exponent.

- Applied to polymer, polyelectrolytes and layered systems.

Poly(lysine)/d-water



Teubner-Strey model

- Developed for interacting particle.

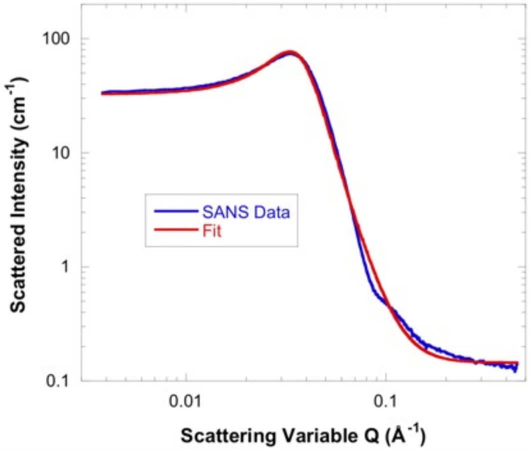
$$I(q) = \frac{1}{a_2 + c_1 q^2 + c_2 q^4} + B$$

$$\xi = \left[\frac{1}{2}\left(\frac{a_2}{c_2}\right)^{1/2} + \frac{1}{4}\frac{c_1}{c_2}\right]^{-1/2}$$

$$d = 2\pi\left[\frac{1}{2}\left(\frac{a_2}{c_2}\right)^{1/2} - \frac{1}{4}\frac{c_1}{c_2}\right]^{-1/2}$$

- Used to determine the correlation length of the micelles and the domain size (periodicity).

P85 pluronic/d-water



Beaucage model

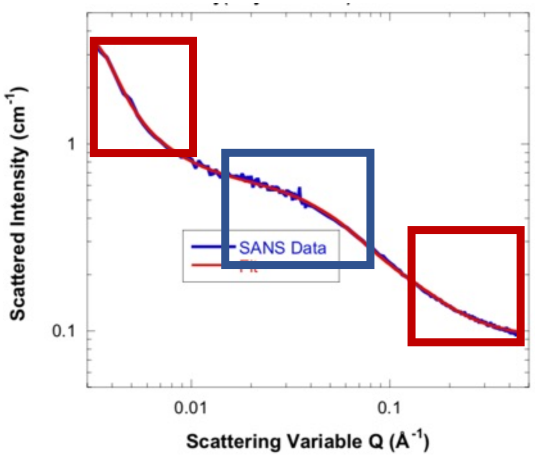
- Used to characterise models with different levels of hierarchical organisation – models several (n) Guinier and Porod regions.

$$I(q) = \sum^{n} \left[\boxed{G_n e^{\frac{-q^2 {R_{g,n}}^2}{3}}} + \boxed{\frac{C_n}{q^{m_n}}} \boxed{\left(erf\left(\frac{qR_{g,n}}{\sqrt{6}} \right) \right)^{3m_n}} \right] + B$$

- *G_n* is related to I(0) and *C_n* is the pre-exponential factor in the Porod Law.

- Fits **fractal and particulate structures** with several layers of structure.

Poly(ethylene oxide) /d-water



Hammouda, J Appl Cryst, 2010.

Geometrical model-based fitting

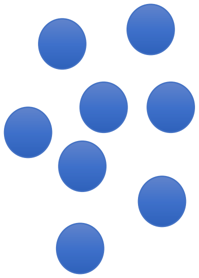
Model-based fitting

- Uses **mathematical models** that simulate the scattering.
- Several variables describe the shape, size, concentration, size distribution and other **characteristics of the scatterers**.
- These are divided in **form factors** and **structure factors**.
- **Preliminary information** about the scatterer and other **constraints** are very valuable.

Form and structure factor

- **Form factor** – describes the intraparticle scattering.
- **Structure factor** – describes the interparticle scattering.

$$I(q) = N_p {V_p}^2 \left(SLD_p - SLD_s\right)^2 P(q) S(q) + B$$

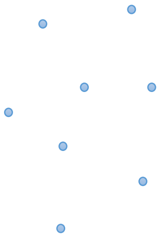


Scattered intensity of a solution

∝

Form factor of a particle

x



Structure factor of the lattice

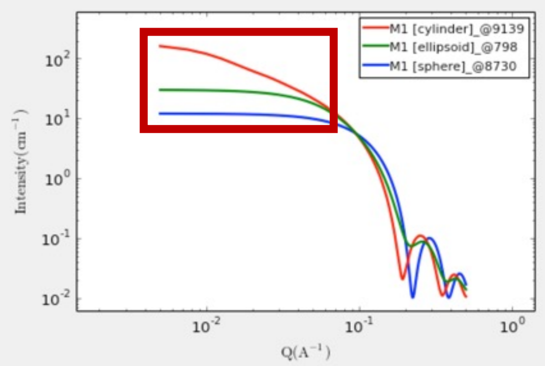
- The form and structure factor are the only **q-dependent functions**.

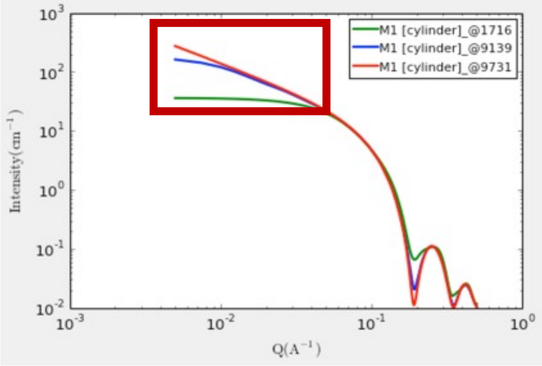
- $N_p V_p^2 (SLD_p - SLD_s)^2$ relates to the **concentration and composition** of scatterers and solvent.

- This is valid for a uniform, centrosymmetric system, but the idea can be extrapolated to more complex systems.

Form factor

- Mathematical models to calculate the scattering from **different shapes**.





- They usually involve **different structural parameters** that describe the morphology of the scatterer – e.g. radius, length, thickness…

- They can build in **different levels of complexity**: uniform shapes, core-shell, complex morphologies.

Form factor

Pedersen, Adv Colloid Interface Sci, 1997.

1. Homogeneous sphere
2. Spherical shell:
3. Spherical concentric shells:
4. Particles consisting of spherical subunits:
5. Ellipsoid of revolution:
6. Tri-axial ellipsoid:
7. Cube and rectangular parallelepipedons:
8. Truncated octahedra:
9. Faceted Sphere:

9x Lens

10. Cube with terraces:
11. Cylinder:
12. Cylinder with elliptical cross section:
13. Cylinder with hemi-spherical end-caps:

13x Cylinder with 'half lens' end caps

14. Toroid:
15. Infinitely thin rod:
16. Infinitely thin circular disk:
17. Fractal aggregates:
18. Flexible polymers with Gaussian statistics:
19. Polydisperse flexible polymers with Gaussian statistics:
20. Flexible ring polymers with Gaussian statistics:
21. Flexible self-avoiding polymers:
22. Polydisperse flexible self-avoiding polymers:
23. Semi-flexible polymers without self-avoidance:
24. Semi-flexible polymers with self-avoidance:

24x Polyelectrolyte Semi-flexible polymers with self-avoidance:

25. Star polymer with Gaussian statistics:
26. Polydisperse star polymer with Gaussian statistics:
27. Regular star-burst polymer (dendrimer) with Gaussian statistics:
28. Polycondensates of A_f monomers:
29. Polycondensates of AB_f monomers:
30. Polycondensates of ABC monomers:
31. Regular comb polymer with Gaussian statistics:
32. Arbitrarily branched polymers with Gaussian statistics:

and many more…

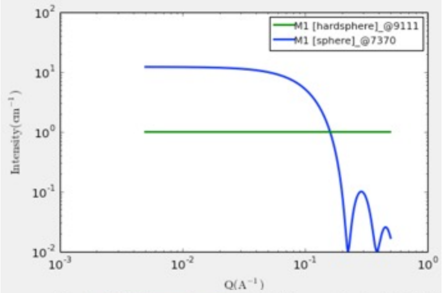
$$F_3(q) = \frac{1}{M_3}\left[\rho_1 V(R_1)F_1(q,R_1) + \sum_{i=2}^{N} (\rho_i - \rho_{i-1})V(R_i)F_1(q,R_i)\right]$$

$$P_{26}(q) = \frac{1}{M_{26}^2}\Big[\rho_e^2 F_5^2(q,R) + N_c\rho_c^2 P_{17}(q,L) + N_c(N_c-1)\rho_c^2\, S_{cc}^{ell}(q) + 2N_c\rho_s\rho_e S_{sc}^{ell}(q)\Big]$$

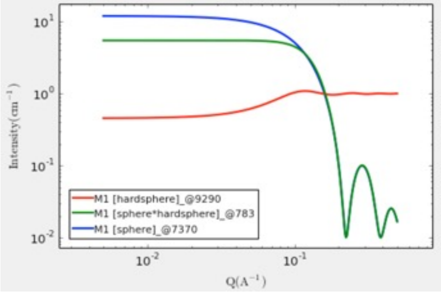
Structure factor

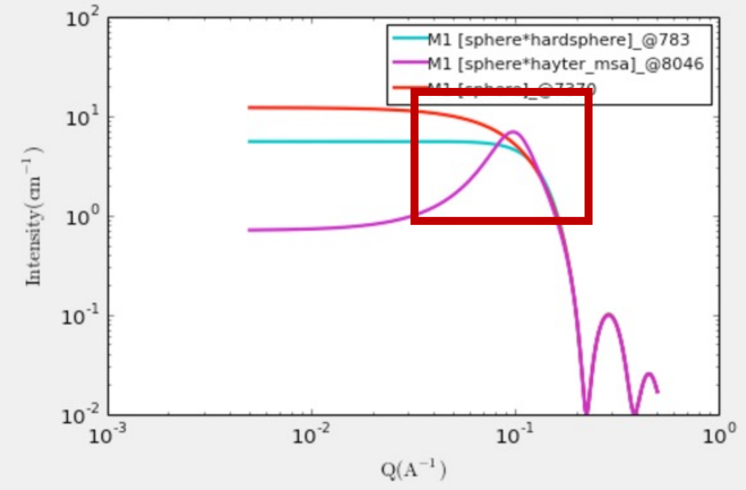
- Mathematical models to calculate the scattering from **the interaction between particles**.
- In the dilute regime (i.e. non-interacting scatterers) S(q)=1.

S(q)=1



S(q)≠1





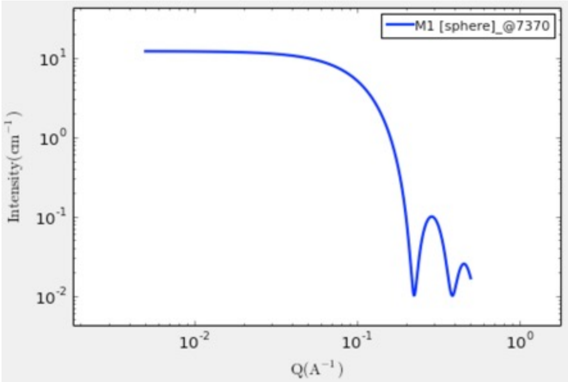
- Hard-sphere – excluded volume; Mean-spherical approximation – particles interacting electrostatically. Both **repulsive potentials**.
- Stickiness and coalescence. **Attractive potentials**.

Structure factor and concentration

- The structure factor is **concentration dependent.**

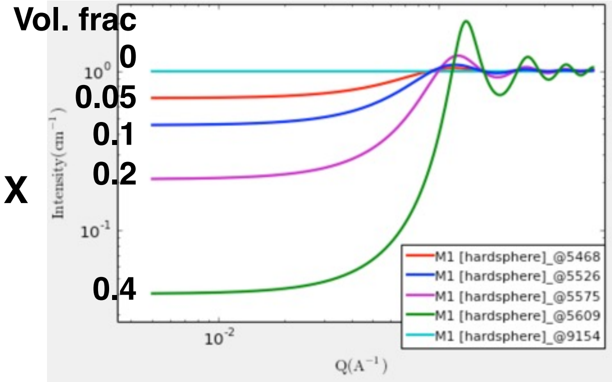
- How does the data is affected in the presence of structure factor?

P(q)



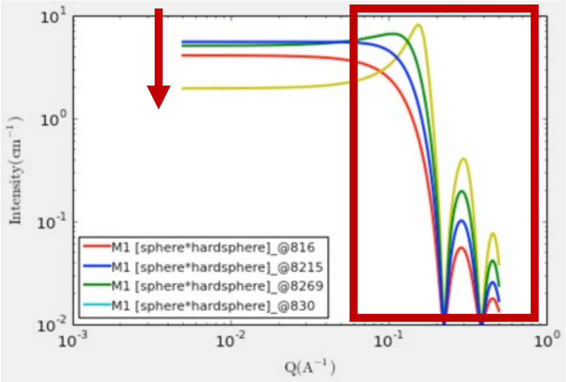
X

S(q)



=

P(q) x S(q)

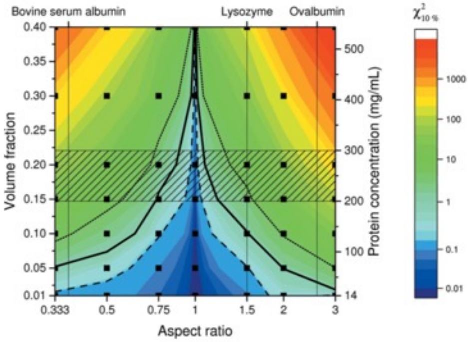


- Structure factors are often derived for **spherical particle**s – anisotropic particles use approximations.
  - Decoupling approximation – polydisperse/anisotropic particles.
  - Random phase approximation – polymers.

$$\frac{d\sigma(q)}{d\Omega} = \Delta\rho^2 V^2 P(q)[1+\beta(q)(S(q)-1)]$$

where

$$\beta(q) = \langle F(q)\rangle_0^2 \,/\, \langle F^2(q)\rangle_0$$



Greene *et al.*, J Appl Cryst, 2016.

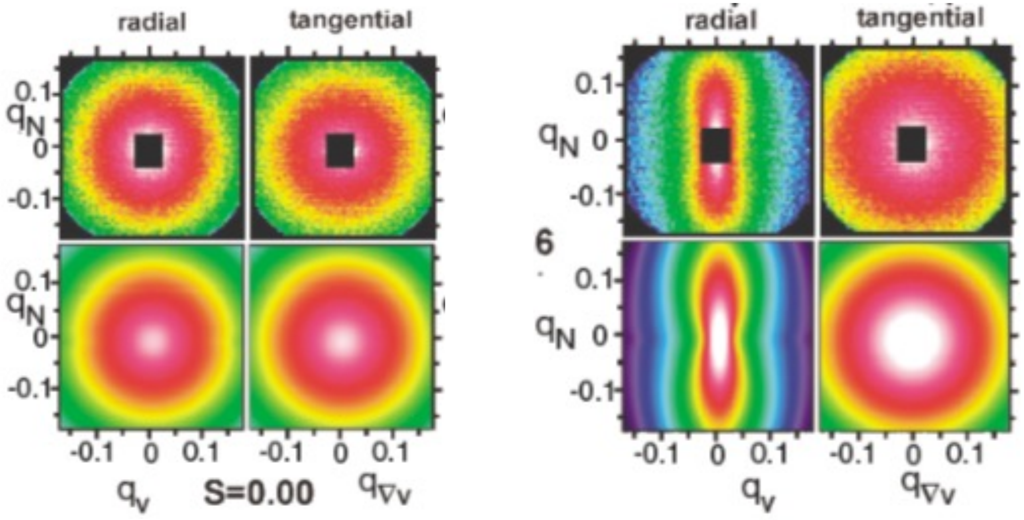
2D fitting

- The previous approaches assumed isotropic scattering – not all systems show that behaviour (e.g. aligned elongated particles).

- Form factor models usually include orientation parameters.

$$P(q,r,AR) = \int_0^{\pi/2} \left[\frac{2B_1(qr\, sin\alpha)}{qr\, sin\alpha} \frac{sin\left(\frac{qL\, cos\alpha}{2}\right)}{\frac{qL\, cos\alpha}{2}} \right]^2 sin\alpha\, d\alpha$$

- When integrated to all possible orientations – isotropic scattering. Oriented bodies - anisotropic scattering.

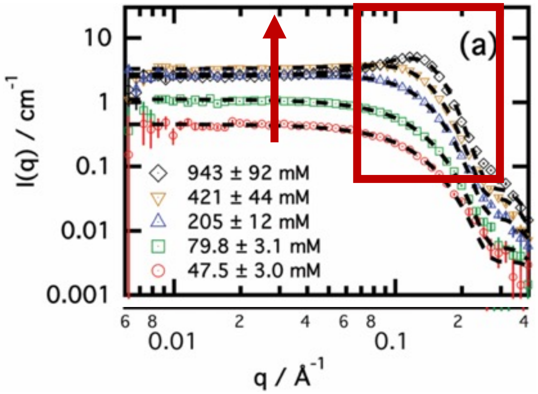


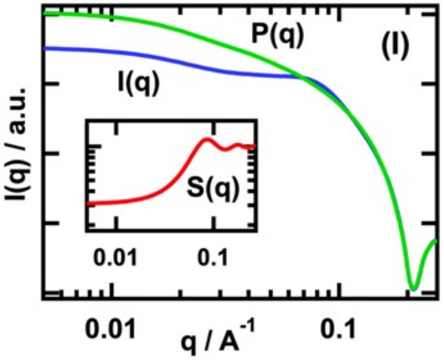
Some tips on data fitting

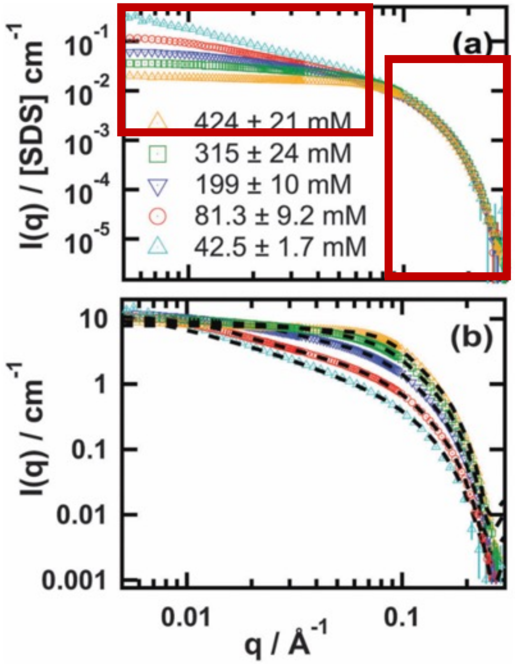
Fitting strategy

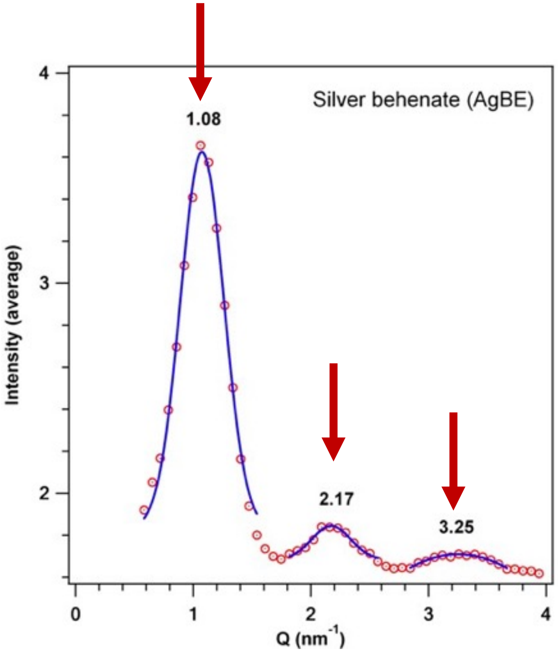
- This is what I WOULD DO, but there are other strategies.

1. **Visual inspection** – look for Bragg peaks, bumps, slope changes, increased/decreased scattering cross section, changes in Guinier region…







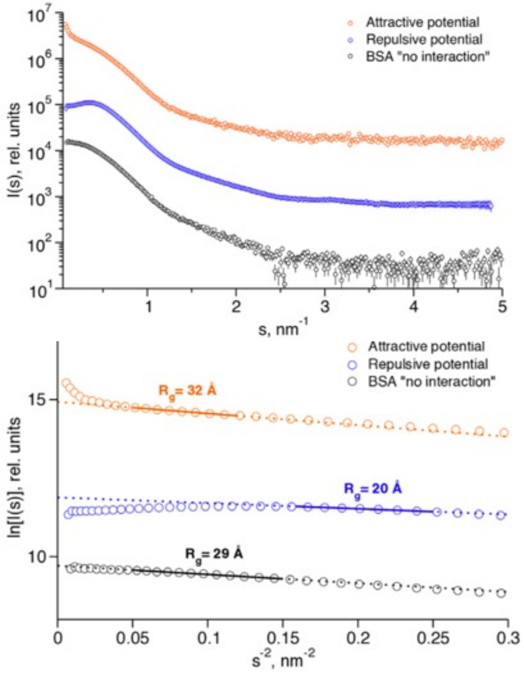


Fitting strategy

- This is what I WOULD DO, but there are other strategies.

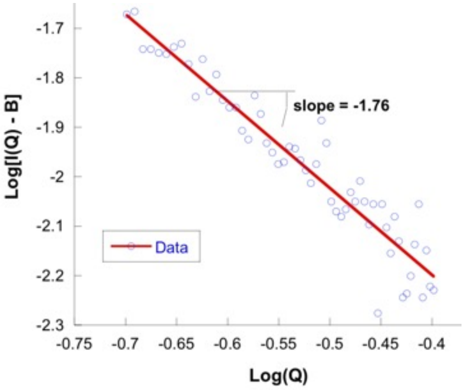
2. **Use standard plots** for a quick evaluation of the data – check for interparticle interaction (structure factor) and particle morphology.

Guinier plot



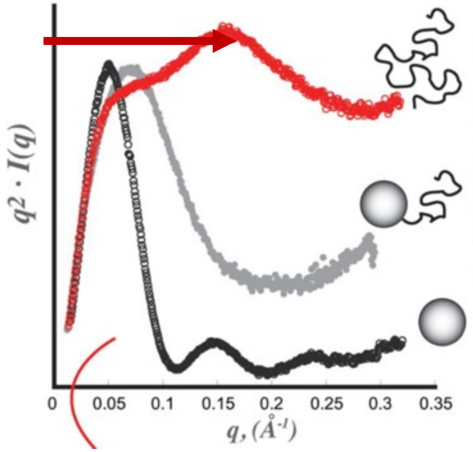
Mertens *et al.*, Arch Biochem Biophys, 2017.

Porod plot



The SANS toolbox

Kratky plot

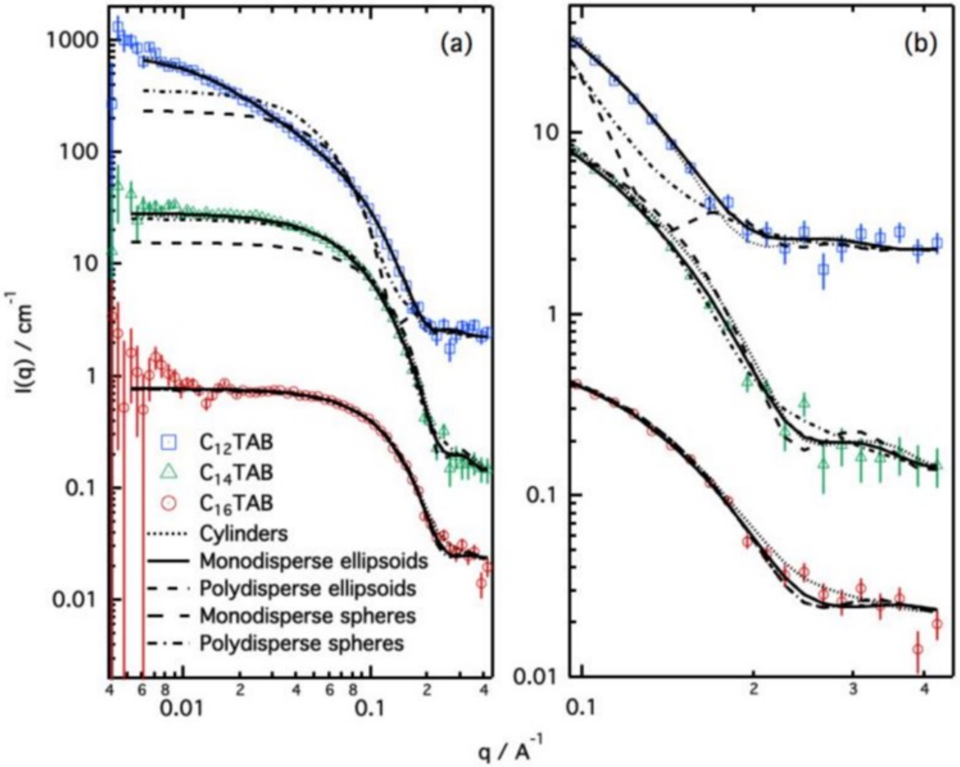


Rambo *et al.*, Biomacromolecules, 2011.

Fitting strategy

- This is what I WOULD DO, but there are other strategies.

3. Test different empirical or mathematical models to **fit a full contrast sample in the dilute regime** – uniform sphere, cylinder, ribbon, lamellar…



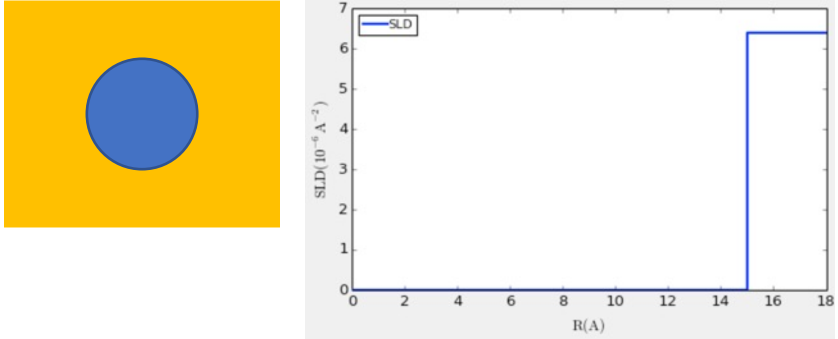
- This will allow to select/discard form factor models.

Fitting strategy

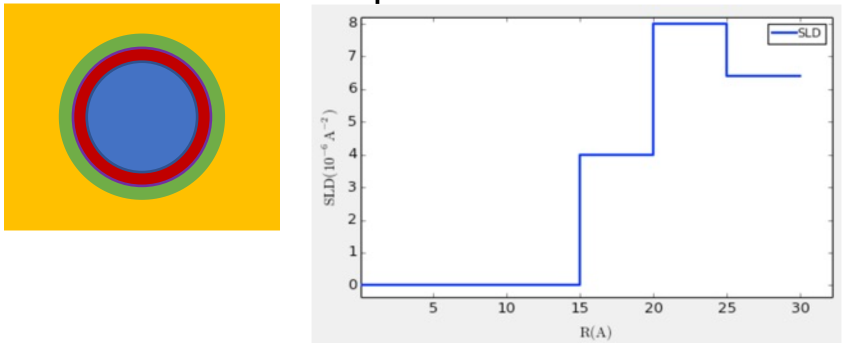
- This is what I WOULD DO, but there are other strategies.

4. Rationalise possible SLD profiles and, considering previous results, elaborate a detailed model – uniform shape, core-shell structure, bicelle, spherical SLD profile…

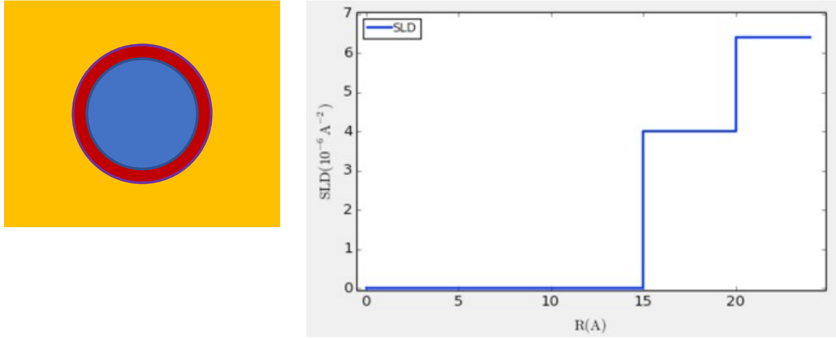
Uniform radial profile



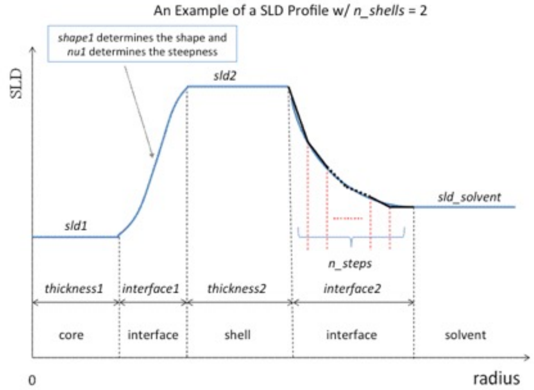
Core-shells radial profile



Core-shell radial profile



Complex interfaces

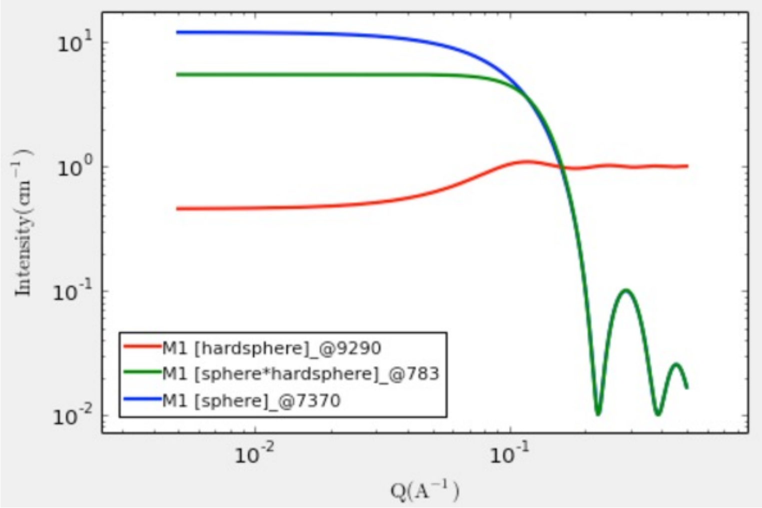


Fitting strategy

- This is what I WOULD DO, but there are other strategies.

5. Is there any **interparticle interaction**? Would you expect it to be electrostatic? Excluded volume? Attractive?

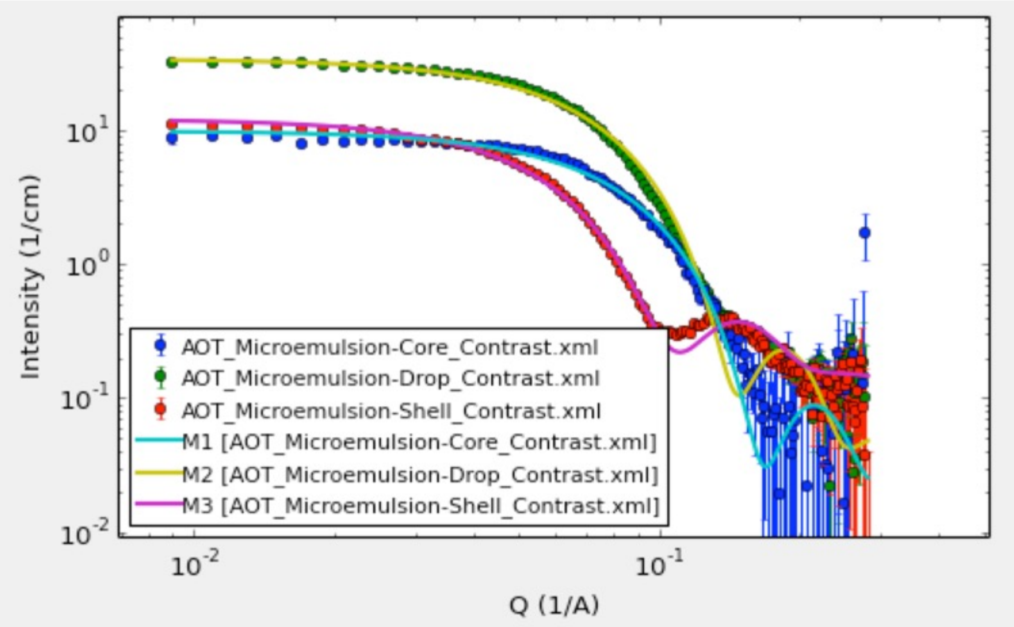
- Incorporate the **structure factor** to the model and fit the data.

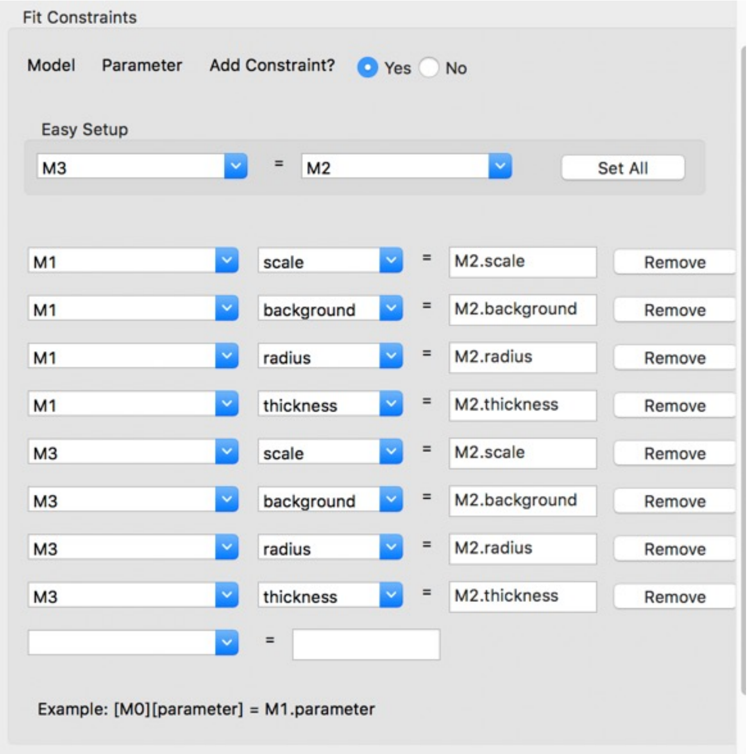


Fitting strategy

- This is what I WOULD DO, but there are other strategies.

6. Once you have the model validated, **simultaneously fit** all the contrasts available.





Add Constraint? ● Yes ○ No

Easy Setup: M3 = M2

M1 scale = M2.scale
M1 background = M2.background
M1 radius = M2.radius
M1 thickness = M2.thickness
M3 scale = M2.scale
M3 background = M2.background
M3 radius = M2.radius
M3 thickness = M2.thickness

Example: [M0][parameter] = M1.parameter

Fitting strategy

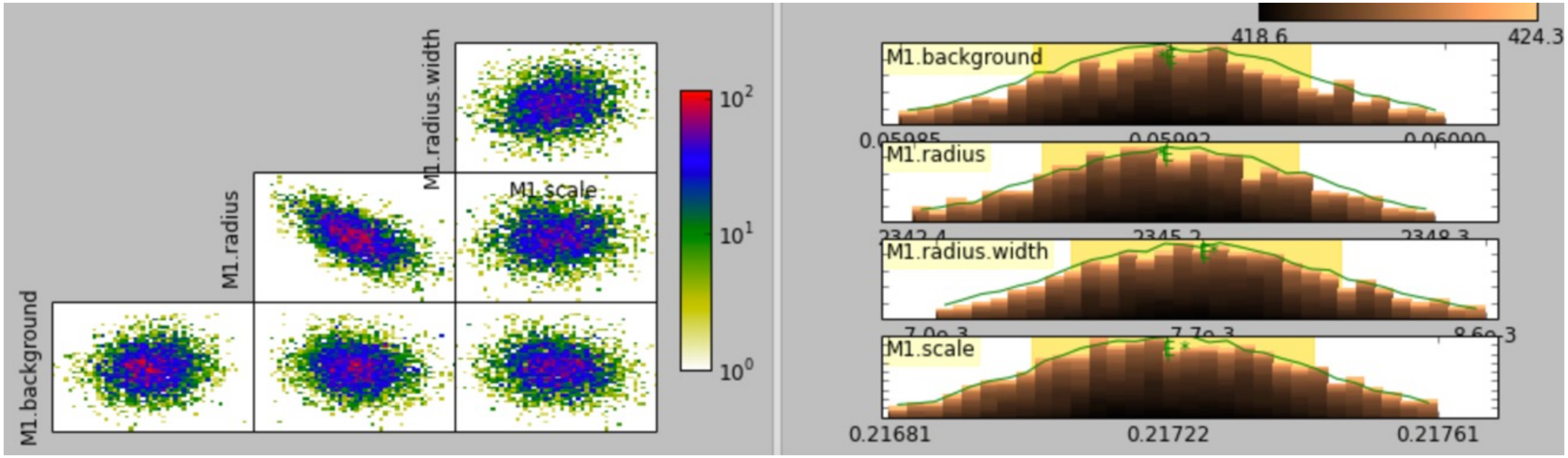
- This is what I WOULD DO, but there are other strategies.

7. **Determine the error** for each fitting parameter.

- The fitting algorithm provide some errors – are these realistic?

- Do they show correlation between different parameters – many local minima for different parameters.

- Use the values from a simple algorithm (e.g. Levenberg-Marquardt) as the input for a more detailed uncertanty analysis (e.g. Markov-Chain Monte Carlo analysis).



Summary

- SANS data analysis may be complex, but the **level of complexity** depends on the amount of info you want to get.

- Stablish a **strategy** to treat the data – from simple approaches to more complex methods.

- Always do your best, but avoid overfitting.

Software and useful links

- Data analysis – model-based fitting: SasView, SASfit
- Data analysis – mainly for biomolecules: ATSAS, BioISIS
- Data analysis – simulation-based fitting: SASSIE

- SLD calculator: https://www.ncnr.nist.gov/resources/activation/
- Biomolecules SLD calculator: http://psldc.isis.rl.ac.uk/cgi_bin/test.py
- Neutron Scattering lengths and cross sections: https://www.ncnr.nist.gov/resources/n-lengths/
- The SANS Toolbox: https://www.ncnr.nist.gov/staff/hammouda/the_sans_toolbox.pdf