An accurate and efficient SAXS/SANS implementation including solvation layer effects suitable for molecular simulations

DIPARTIMENTO DI BIOSCIENZE

UNIVERSITÀ DEGLI STUDI DI MILANO

Federico Ballabio, Cristina Paissoni, Michela Bollati, Matteo de Rosa, Riccardo Capelli and Carlo Camilloni

federico.ballabio@unimi.it

1. Integrating experimental Small Angle Scattering (SAS) data in Molecular Dynamics (MD) simulations.

SAS techniques based on X-rays (SAXS) or neutrons (SANS), are widely used for the characterisation of biomolecules in solution. These methods allow the size, shape, stoichiometry, and dynamics of biomolecules to be assessed under near-physiological conditions, at reasonable concentrations, and without the need of labelling[1]. SAXS/SANS are low resolution techniques, but the size and the disorder level of the system are not a limitation[2]. The integration of SAS data into MD simulations is a powerful approach to increase the resolution of the former and the accuracy of the latter[3].

HOW:

experimental data, *in silico* data

$$E_{FF} + \sum_q K[D_q - f_q(X)]^2$$

The integration requires an accurate and computationally efficient forward model for calculating the experimental observable given a conformation. Our forward model takes advantage of the limited resolution of SAS techniques, by using a single-bead (1B) representation to describe the scattering behaviour of an amino acid, and a three-bead (3B) mapping for a nucleotide, one for the phosphate group, one for the pentose sugar and one for the nitrogenous base, and more importantly, it allows the effective on-the-fly inclusion of solute-solvent scattering corrections at no computational cost.

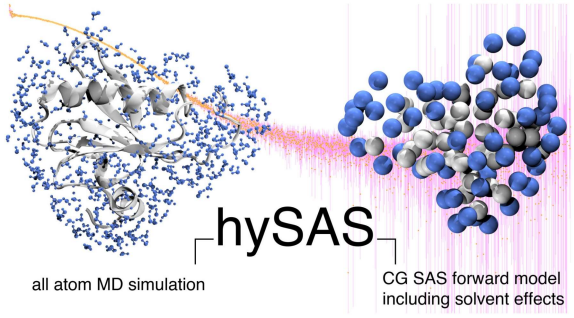
HOW:

$$F_i(q) = \left[\sum_{k\in i}\sum_{l\in i} f_k'^{atomic}(q) f_l'^{atomic}(q) \frac{\sin(qr_{kl})}{qr_{kl}} + \rho^2 \sum_{k\in i}\sum_{l\in i} f_k'^{solvent}(q) f_l'^{solvent}(q) \frac{\sin(qr_{kl})}{qr_{kl}} - \rho \sum_{k\in i}\sum_{l\in i} [f_k'^{atomic}(q) f_l'^{solvent}(q) + f_k'^{solvent}(q) f_l'^{atomic}(q)] \frac{\sin(qr_{kl})}{qr_{kl}}\right]^{1/2}$$

solvent electron density

where $F_i(q)$ is the scattering form factor of the bead i, $f'^{solvent}(q) = v \exp\left(-\frac{q^2 v^{2/3}}{4\pi}\right)$, and $f'^{atomic}(q) = \sum_{n=1}^{4} a_n \exp[-b_n(q/4\pi)^2] + c$.

Although the forward model is computed at a coarse-grained (CG) resolution, the MD simulation maintains all the atomistic details. We call this hybrid approach hySAS and it is implemented in PLUMED[4], allowing it to be used in combination with different MD engines, restraining strategies including metainference[5] and maximum entropy/caliber approaches, or enhanced sampling techniques such as metadynamics and umbrella sampling.

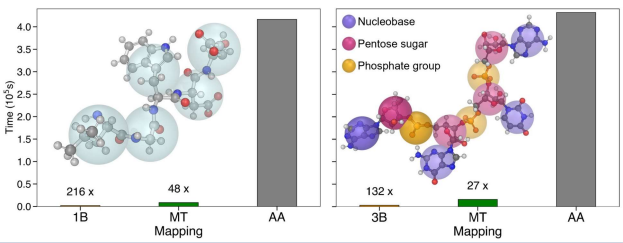


2. 1B mapping for amino acids and 3B mapping for nucleotides are fast ...

The calculation speed of the *in silico* SAS intensity is evaluated at different resolutions: all-atom (AA), Martini (MT), 1B per amino acid / 3B per nucleotide.

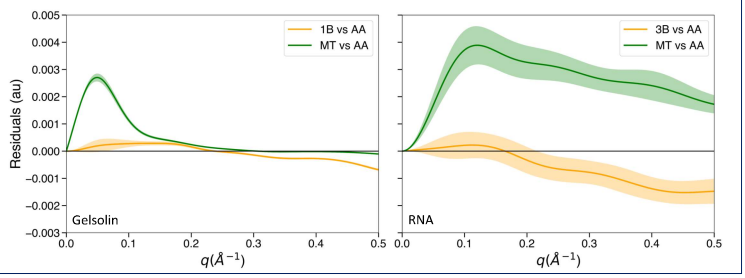
Testing systems:

- 6,500 frames of gelsolin: 11,558 atoms, 1,627 MT beads, 775 1B beads.
- 500 frames of ribosome RNA: 38,287 atoms, 7,796 MT beads, 3,560 3B beads.



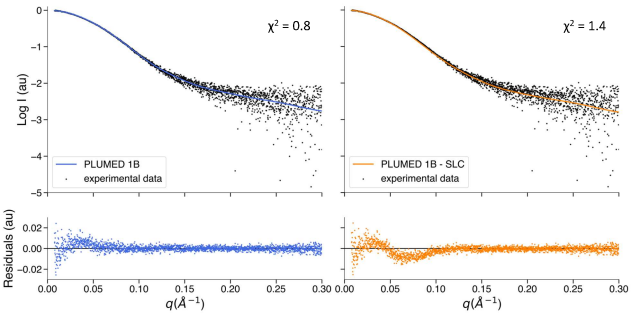
3. ... and accurate for small *q* values.

The accuracy of 1B/3B mappings in calculating scattering intensities is assessed on the same test systems. For each frame, the intensity computed with 1B/3B and with MT representation is compared with the corresponding intensity at atomistic resolution, which is taken as the reference. The gelsolin SAXS intensities calculated with 1B mapping are in better agreement with those obtained with AA resolution than with MT up to 0.3 Å^{-1}. As for the proteins, the calculation of the SAXS intensity on RNA with 3B mapping also proves to be accurate, since the difference (residuals) between 3B and AA is smaller than the difference between MT and AA. These results were obtained without considering the solvation layer contribution (SLC), as the AA reference does not take into account the hydration effects.

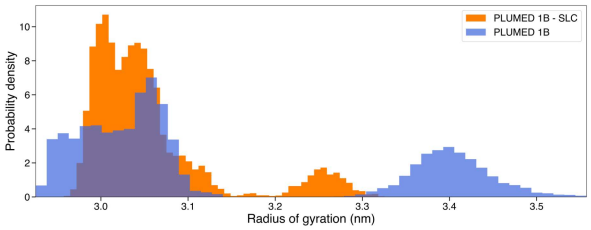


4. Gelsolin conformational ensemble determination: the solvation layer contribution results in a smaller radius of gyration (RG) and an overall decrease in fluctuations.

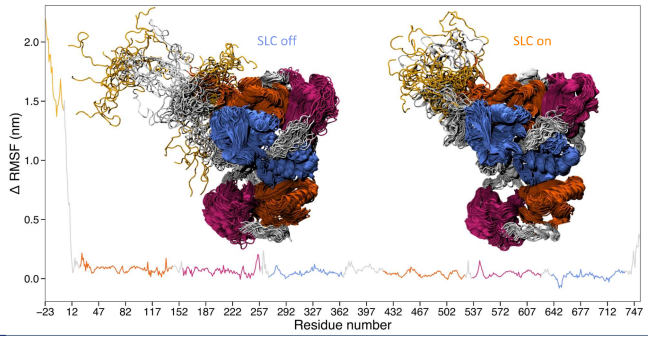
Two conformational ensembles were generated via metainference multi-replica simulations, using SAXS data as restraint and 1B as forward model. One of the two ensembles was obtained by enabling the SLC. An average SAXS profile was determined from each ensemble and compared with the experimental SAXS data.



Both the ensembles show a bimodal distribution of the RG, but the one obtained with the inclusion of the SLC is, as expected, more compact with an average RG of 3.05 nm, compared to the one generated without the SLC, which shows an average RG of 3.14 nm.



A similar behaviour is observed for the root mean square fluctuations (RMSF) of the residues. The ensemble with SLC shows systematically lower fluctuations, with an average RMSF of 0.26 nm, compared to the other ensemble, which has an RMSF of 0.38 nm.



REFS: 1 Tuukkanen et al. 2017, *IUCrJ* | 2 Stelzl et al. 2022, *JACS* | 3 Hub 2018, *COSB* | 4 Tribello et al. 2014, *CPS* | 5 Bonomi et al. 2016, *SA*