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

DIPARTIMENTO DI BIOSCIENZE

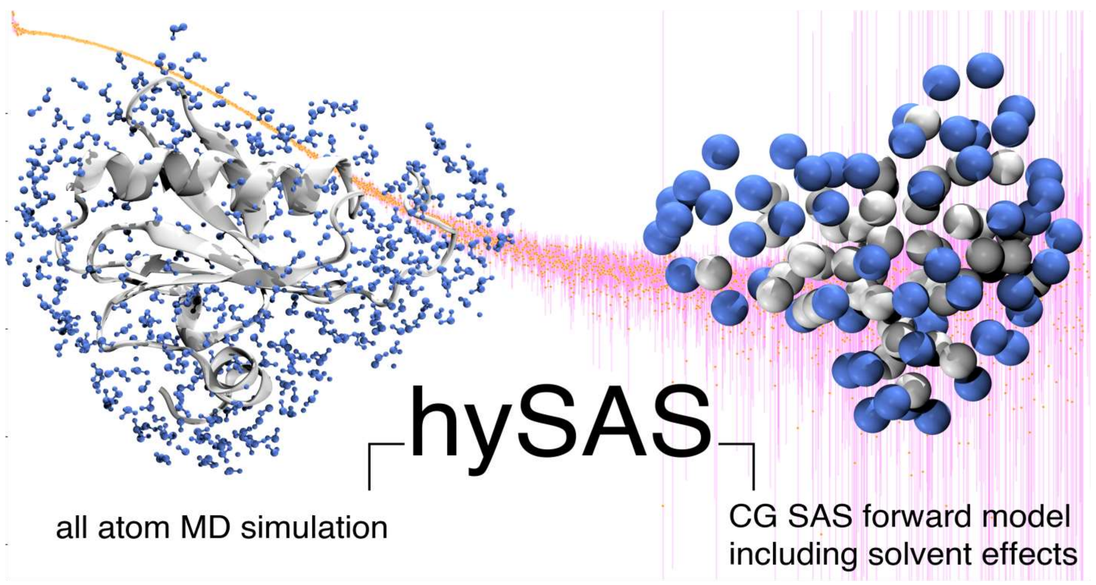
UNIVERSITÀ DEGLI STUDI DI MILANO

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.

Small Angle Scattering is a low-resolution technique based on X-rays (SAXS) or neutrons (SANS) that allows the size, shape, stoichiometry, and dynamics of biomolecules to be assessed under near physiological conditions[1].

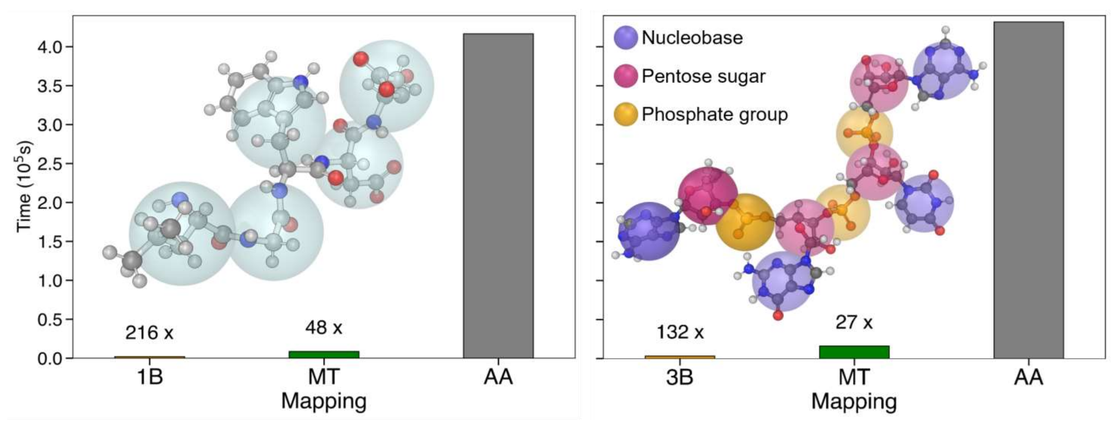


SAS data can be used to restrain all-atom molecular dynamics (MD) simulations and generate conformational ensembles consistent with experimental results. This is achieved by introducing an energy penalty based on the difference between the SAS data and the SAS profiles computed in real time from the system coordinates. SAS profiles are calculated using a coarse-grained (CG) forward model: one bead per amino acid (1B), and three beads per nucleotide (3B). Solvent-exposed beads are corrected on-the-fly to account for solute-solvent scattering effects (SLC) at no additional computational cost. This hybrid approach, hySAS, is implemented in PLUMED[2], making it compatible with different MD engines.

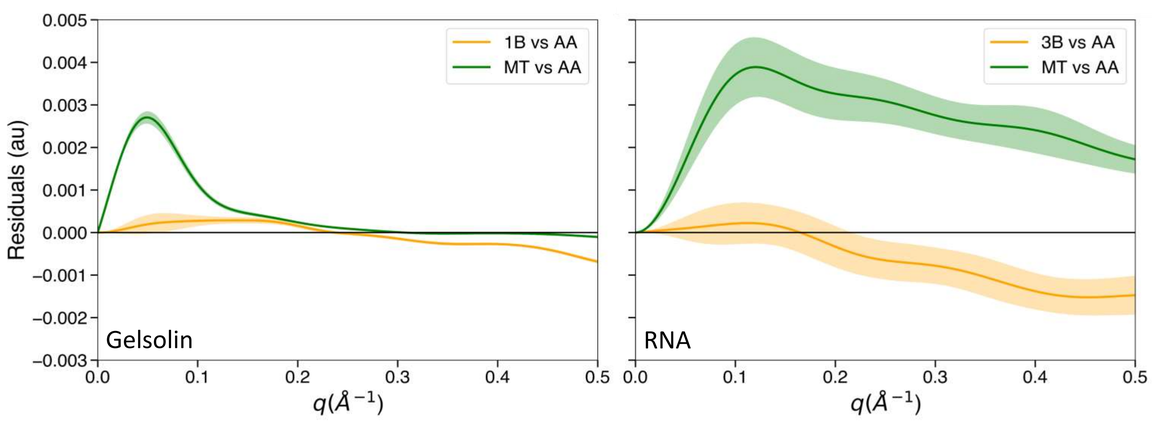
2. 1B mapping for amino acids and 3B mapping for nucleotides are fast and accurate for small *q* values.

The speed of the SAS calculation was evaluated by comparing the time required to determine intensities from MD trajectory frames at different resolutions: all-atom (AA), Martini (MT), 1B per amino acid / 3B per nucleotide.

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



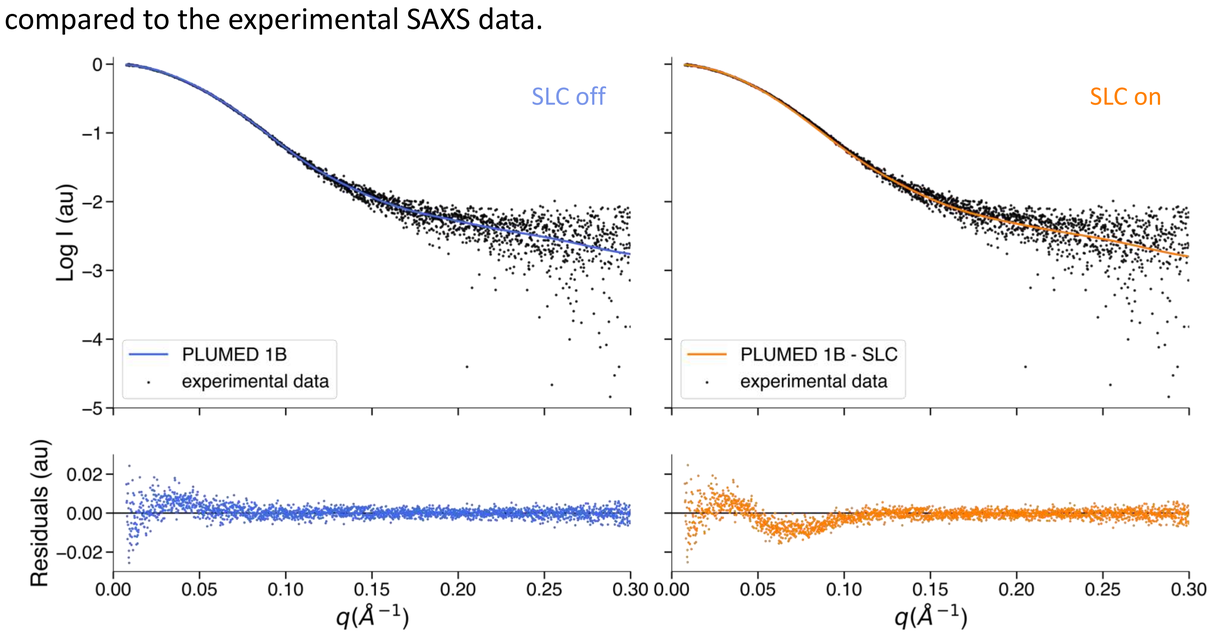
To evaluate the accuracy of the SAS calculation, for each frame of the trajectories, the intensity computed with 1B/3B and with MT representation is compared with the corresponding intensity at AA resolution, which is used as a reference.



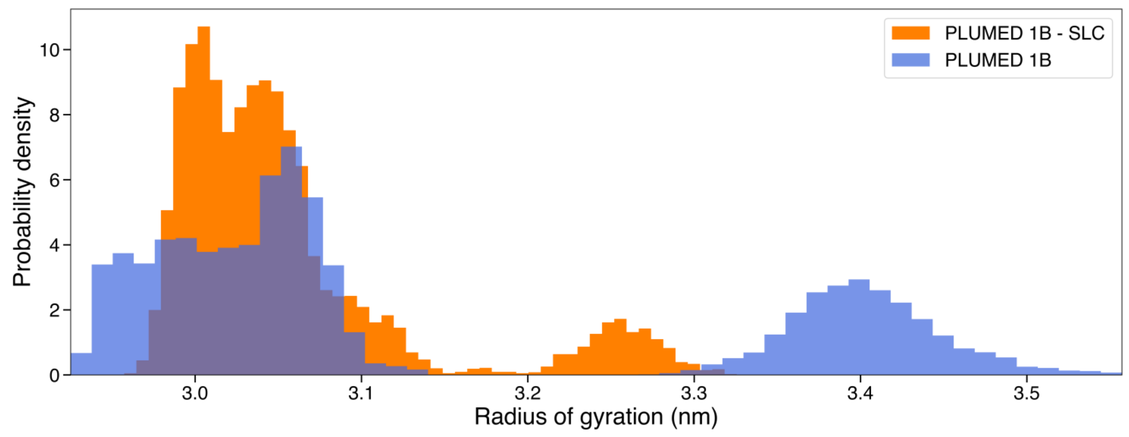
The gelsolin SAXS intensities calculated with 1B mapping show better agreement with those obtained with AA resolution than with MT up to 0.3 Å^{-1}. Similarly, for RNA, the SAXS intensity calculation using 3B mapping proves to be accurate, since the difference (residuals) between 3B and AA is smaller than the difference between MT and AA. These results were obtained without considering the solvation layer contribution (SLC).

3. Case study : determination of gelsolin conformational ensembles.

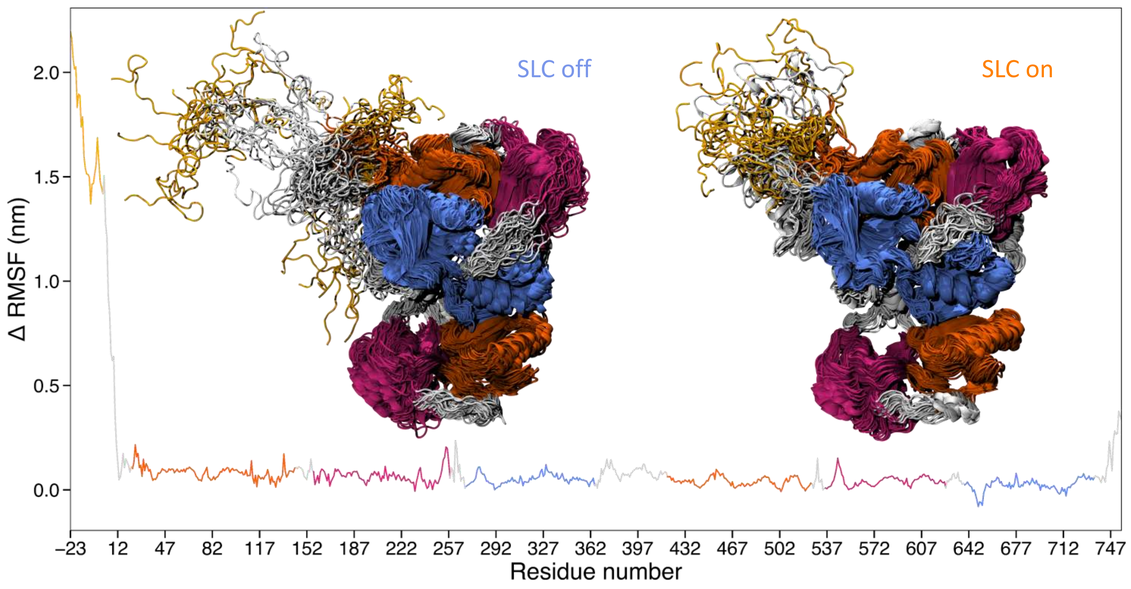
Two conformational ensembles were generated via metainference[3] multi-replica simulations, using SAXS data as restraint and 1B as forward model. One of the two ensembles was obtained by enabling the SLC. An average SAXS profile was determined from each ensemble and compared to the experimental SAXS data.



The probability density distribution of the radius of gyration was calculated over the two ensembles. The area under each histogram integrates to 1.



The flexibility of the ensembles was evaluated by calculating the RMSF difference between the residues of the ensemble without SLC and those of the ensemble with SLC.



Including the SLC results in more compact and stable models.

- SLC on: average RG of 3.05 nm, average RMSF (residues) of 0.26 nm.
- SLC off: average RG of 3.14 nm, average RMSF (residues) of 0.38 nm.

4. Summary.

- Fast and accurate SAXS/SANS forward model.
- Compatible with proteins and nucleic acids.
- Includes implicit solvation effects (and implicit hydrogen-deuterium exchange for SANS).
- Suitable for all-atom MD force fields.

JCTC Publication

Open source

Practical guide

REFS: 1. Tuukkanen A.T. et al. 2017, *International Union of Crystallography Journal* | 2. The PLUMED consortium, 2019, *Nature Methods* | 3. Bonomi M. et al. 2016, *Science Advances*