Accelerating SAXS intensities calculation of in-silico models

DIPARTIMENTO DI BIOSCIENZE

UNIVERSITÀ DEGLI STUDI DI MILANO

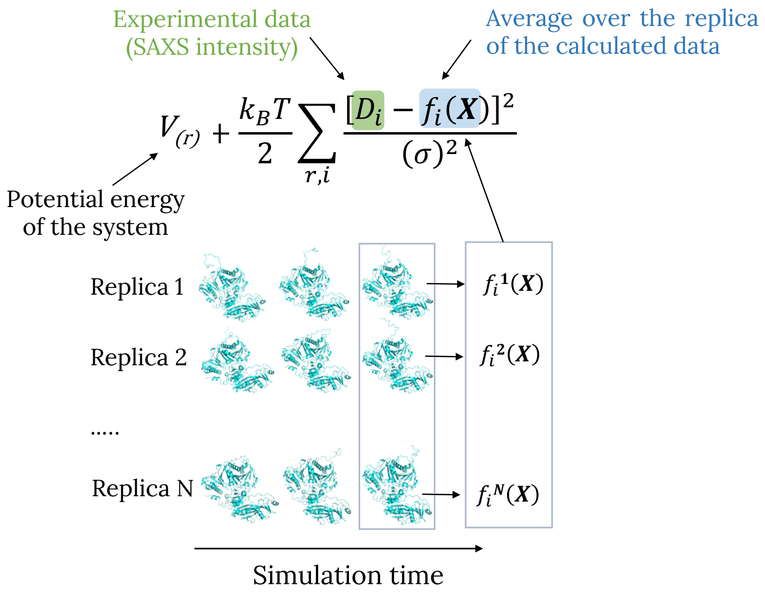
Federico Ballabio, Cristina Paissoni and Carlo Camilloni

federico.ballabio@unimi.it

1.INTEGRATING EXPERIMENTAL SAXS DATA IN MOLECULAR DYNAMICS (MD) SIMULATIONS.

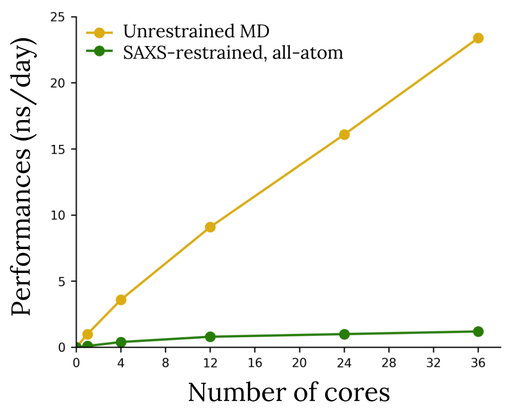
Combining SAXS experiments with MD simulations is an effective strategy in the characterization of biomolecules in solution. MD provides the physical model to interpret at atomistic level the experimental data, which, in turn, helps alleviating the computational inaccuracy.

HOW:



- MD calculation ($V_{(r)}$) is modified by adding a harmonic term which is proportional to the difference between the experimental data and the same data calculated from the simulation: the simulated conformations that are not in agreement with the input experimental data are discouraged;
- to mirror an ensemble averaged condition and to retrieve better statistics, a multiple replica approach is adopted.

DRAWBACKS:



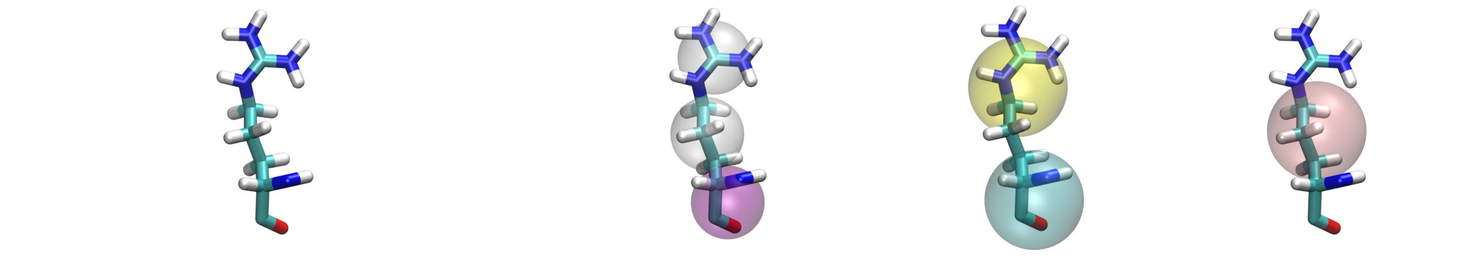
- all-atom SAXS intensity calculation makes a SAXS restrained simulation (green line) unsustainable in terms of performances;
- the computed SAXS intensity is determined without considering the explicit solvent.

2.COARSE-GRAINING THE SAXS INTENSITIES CALCULATION: A WAY TO INCREASE THE PERFORMANCES.

Calculating the SAXS intensities $I(q)$ by considering the system of interest as a collection of beads, each containing a certain number of atoms, can significantly mitigate the performance bottleneck.

WHY:

SOLUTION:



Atomistic representation

Coarse-grain representations

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(q r_{ij})}{q r_{ij}} \propto N_{atoms}^2$$

f(q) = atomic form factor, it reproduces the scattering intensity of an isolated atom. Data available in literature.
r_{ij} = distance between atom $_i$ and atom $_j$.

$$I(q) \sim \sum_{i=1}^{M} \sum_{j=1}^{M} F_i(q) F_j(q) \frac{\sin(q R_{ij})}{q R_{ij}} \propto M_{beads}^2$$

F(q) = bead form factor, it reproduces the scattering intensity of an isolated bead. Must be computed.
R_{ij} = distance between bead $_i$ and bead $_j$.

3.AIM.

Developing an in-silico tool that generates a system-specific bead mapping with a position-dependent *F'(q)* for each residue of the protein. The surface beads *F'(q)* include a correction that takes into account the hydration shell excess of electron density, with no impact on performances.

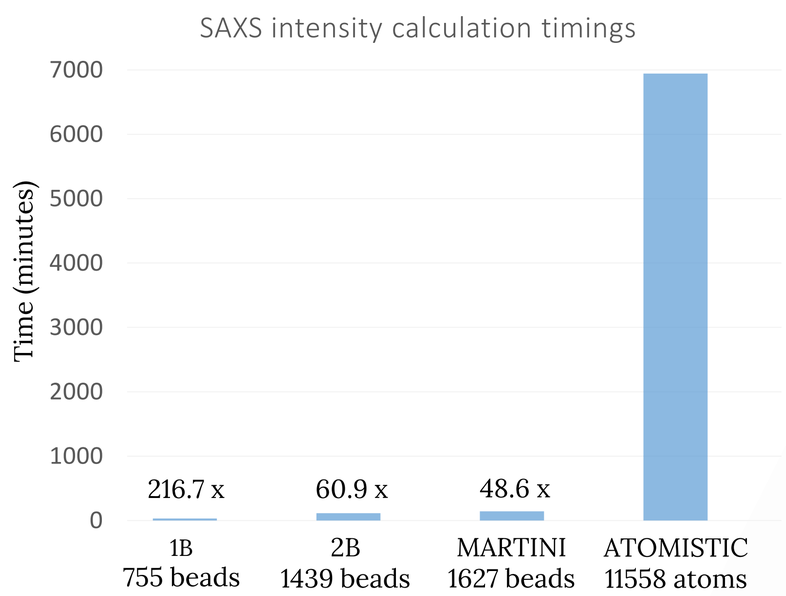
Buried bead

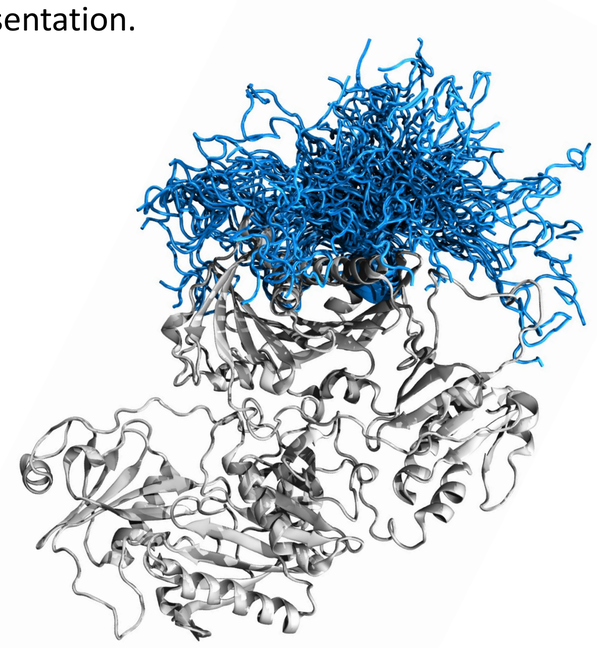
$$F'_i(q) = F_i(q) - solvent\ displacement\ electron\ density + hydration\ shell\ electron\ density\ correction$$

Surface bead

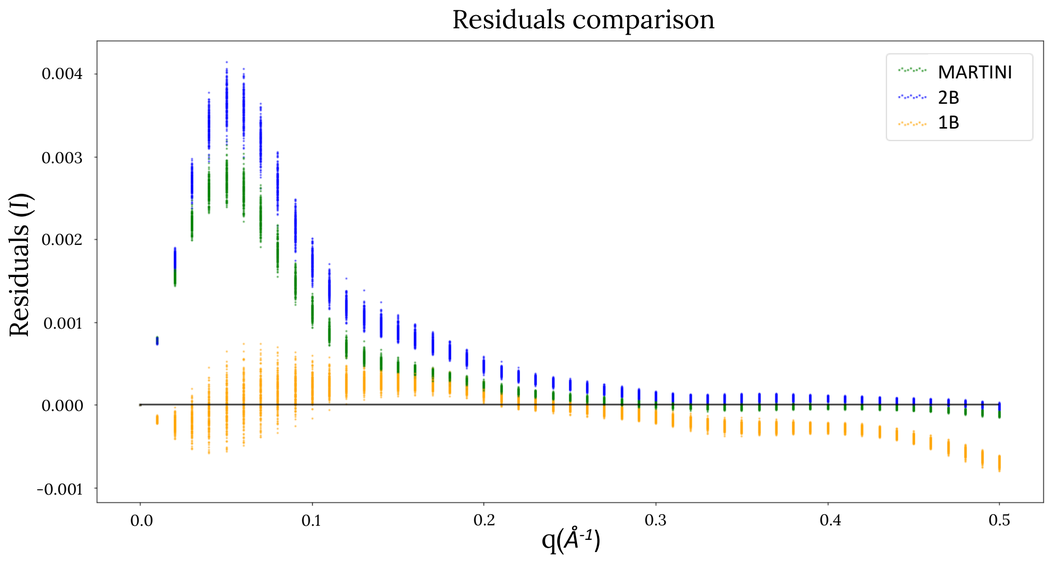
4.TESTING SYSTEM AND PERFORMANCES.

- Gelsolin, 11558 atoms in 755 residues. The protein has a disordered N-term region (22 residues).
- System-specific bead mappings determined by analysing 1μs unrestrained MD simulation.
- SAXS intensity calculation performances have been evaluated on a 6500 frames trajectory, for each bead mapping type and for the atomistic representation.





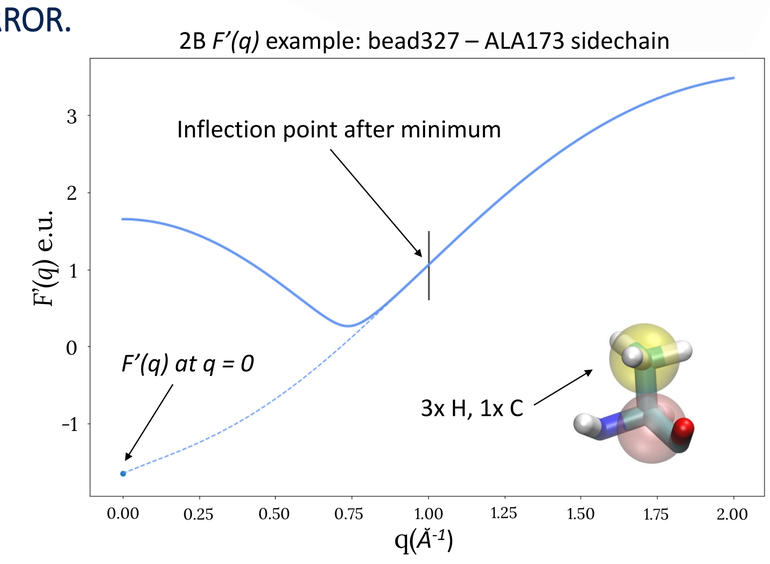
5.INTENSITY CALCULATION ON BEADS: QUALITY EVALUATION.



WHICH BEAD MAPPING TYPE PROVIDES SAXS CURVES CLOSER TO THE ATOMISTIC DETERMINED ONES?

SAXS intensities computed on the 6500 frames trajectory with each bead mapping are compared to the SAXS intensities computed with the atomistic representation (residuals). 1B mapping (yellow), the fastest representation, provides SAXS curves closest to those obtained at all-atom resolution.

6.SOURCE OF ERROR.



- Computing *F'(q)* generates only positive values (solid line), while considering the effect of the solvent electron density displaced by the molecule could lead to negative results, e.g. in case of small beads containing several hydrogen atoms.
- *F'(q)* is corrected by fitting the curve after the inflection point to *F'(q)* at *q = 0* (dash line). The result is an estimation.
- 1B mapping cannot generate beads with negative *F'(q)*.

REFS: Paissoni et al, JAC, 2019 | Bonomi et al, Science Advances, 2016 | Fraser et al, JAC, 1978 | Putnam et al, CSBJ, 2013 | Bonì et al, HumMolGenet, 2018 | Niebling et al, JAC, 2014