Regular Article - Computational Methods

Comparing state-of-the-art approaches to back-calculate SAXS spectra from atomistic molecular dynamics simulations

Mattia Bernetti[a] and Giovanni Bussi[b]

Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, Trieste 34136, Italy

© The Author(s) 2021

Abstract. Small-angle X-ray scattering (SAXS) experiments are arising as an effective instrument in the structural characterization of biomolecules in solution. However, they suffer from limited resolution, and complementing them with molecular dynamics (MD) simulations can be a successful strategy to obtain information at a finer scale. To this end, tools that allow computing SAXS spectra from MD-sampled structures have been designed over the years, mainly differing in how the solvent contribution is accounted for. In this context, RNA molecules represent a particularly challenging case, as they can have a remarkable effect on the surrounding solvent. Herein, we provide a comparison of SAXS spectra computed through different available software packages for a prototypical RNA system. RNA conformational dynamics is intentionally neglected so as to focus on solvent effects. The results highlight that solvent effects are important also at relatively low scattering vector, suggesting that approaches explicitly modeling solvent contribution are advisable when comparing with experimental data, while more efficient implicit-solvent methods can be a better choice as reaction coordinates to improve MD sampling on-the-fly.

1 Introduction

The small-angle X-ray scattering (SAXS) technique is increasingly being applied to gain structural insights into biomolecules directly in solution [1,2]. Indeed, information about size, shape and global dynamics can be obtained from SAXS experiments [1,3]. This technique has historically found main application in the structural characterization of proteins [4,5], and in particular of intrinsically disordered proteins where properly accounting for the dynamics in solution is crucial [6–8]. SAXS experiments can also be used to characterize nucleic acids dynamics [9–11]. Notwithstanding the precious insights that can be achieved, one major drawback of the SAXS technique is its limited resolution. Molecular dynamics (MD) simulations allow exploring the structural dynamics of biomolecules and thus generate conformational ensembles of structures at atomistic resolution [12,13]. Therefore, the use of MD in combination with SAXS experimental data is gradually arising as a valuable instrument for interpreting the experimental information [10,14]. From a practical standpoint, this requires the availability of tools that allow to accurately compute the SAXS spectra from MD-sampled structures. In this respect, the application to nucleic acids can be particularly challenging, as the presence

of the RNA solutes can highly influence the surrounding solvent [15,16].

In this study, we compare different state-of-the-art methods that allow computing SAXS spectra from MD simulations and that model the solvent contribution to a different extent. In particular, we concentrate on their application to RNA molecules using, as a prototype case, the 57-nucleotide long GTPase-associated center (GAC) [17–19]. We have recently studied the conformational dynamics of this RNA molecule through enhanced-sampling methods [20]. Here we analyze unbiased MD simulations where the RNA structure is fixed in order to focus on solvent effects.

2 Methods

2.1 Computational details

Four systems with different ionic conditions were constructed for GAC RNA. These simulations were originally reported in a recent study [20] and are here analyzed in more detail. Specifically, (1) one system comprised the RNA molecule in a buffer solution with 100 mM KCl, (2) in the second system crystallographic Mg^{2+} were included in addition to the first setup, (3) the third system had additional Mg^{2+}, specifically half the amount needed to neutralize the system and (4)

[a] e-mail: mbernett@sissa.it (corresponding author)
[b] e-mail: bussi@sissa.it (corresponding author)

in the fourth system all the amount of Mg^{2+} necessary to fully neutralize the system was added. This resulted in four systems where the amount of Mg^{2+} cations gradually increased. All the four systems were generated from the crystal structure of GAC RNA in its folded state (PDB ID: 1HC8) [21] after removing the bound protein. The AMBER force field for nucleic acids was employed to model the RNA [22–24], while the four-point optimal-point-charge (OPC) model [25] was used for water together with compatible ion parameters [26,27]. A large rhombic dodecahedron simulation box containing approximately 170000 atoms was constructed by specifying edges distant 3 nm from all RNA atoms. The systems were energy minimized and subjected a multi-step equilibration procedure lasting a total of 1.5 ns: thermalization to 300 K in the NVT ensemble was conducted through the stochastic velocity-rescaling thermostat [28], with soft position restraints on the RNA heavy atoms and crystal ions; the restraints were gradually removed in the subsequent NPT simulations using the Parrinello-Rahman barostat [29]. In the production phase, each system was simulated through unbiased MD for 10 ns in the NPT ensemble using soft position restraints on the RNA heavy atoms, so as to exclude solute dynamics effects. Simulations of the pure solvent (*i.e.*, with no solute) having consistent salt concentrations, as required by the WAXSiS and Capriqorn software packages, were performed in the NVT ensemble for 5 ns. All the MD simulations were performed using GROMACS version 2018.4 [30].

2.2 Different approaches to compute the SAXS spectra

We considered four software packages that allow to compute SAXS spectra of biomolecular structures sampled by MD simulations: PLUMED [31], CRYSOL [32,33], WAXSiS [34] and Capriqorn [35]. The former two can be broadly classified as implicit-solvent methods, since the spectra are evaluated using the solute coordinates and introducing corrections to account for solvent effects, while the latter two as explicit-solvent methods, since the contribution of the solvent enters explicitly in the estimation of the total scattering signal. In particular, the subtraction of the solvent contribution is done analyzing a separately-generated pure-solvent trajectory.

In PLUMED, the SAXS spectrum is obtained through the standard Debye equation, where a summation runs over all pairs of N atoms in the solute, resulting in a sum of $\frac{N(N-1)}{2}$ terms. Since the PLUMED implementation is designed to allow SAXS intensities to be restrained on-the-fly [36], this calculation typically needs to be done at every step during an MD simulation. From a computational standpoint, this becomes highly expensive when applied to a structure in an all-atom representation. To mitigate this issue, a coarse-grained representation of the system has been implemented [37], where the summation runs over M beads, with $M \ll N$. For this purpose, bead structure factors

have been determined, which also include the contribution from the excluded volume as formulated in Ref. [38]. Since frames are analyzed independently of each other, when applied on a MD trajectory, the method allows determining the SAXS spectrum of each frame. We here show results obtained with both the atomistic representation and the faster coarse-grained representation.

In CRYSOL, the SAXS spectrum is computed from the solute coordinates through a spherical harmonics expansion procedure. Besides adding a contribution for the excluded volume, the method also introduces an empirical correction to account for the hydration layer around the solute. As in PLUMED, the frame-by-frame spectra can be obtained from an MD trajectory. We here show results obtained using both CRYSOL 2 [32] and CRYSOL 3 [33]. The two versions are different in how the hydration shell and its scattering are evaluated. For both CRYSOL versions, the SAXS spectra for all the MD trajectory frames were computed in the range $q \in [0, 0.3]$ Å^{-1}, with a grid spacing of 0.01 Å^{-1}, a maximum order of harmonics of 20, and default parameters.

In WAXSiS, both solute and solvent coordinates from an MD simulation are used to estimate the SAXS spectrum. As a result, all contributions from the solvent are included explicitly in the computation. The method relies on an envelope that is constructed to include all solute atoms and solvent atoms at sufficient distance from the solute so as to account for solvent effects. Notably, a second simulation of the sole solvent, *i.e.*, where no solute is present, with consistent ionic conditions is also required by the program to accurately conduct solvent subtraction. Indeed, this procedure, including two simulations, ensures a framework which is closest to the experimental setting, where the final spectra is obtained as the difference of two measures, a first one on the sample containing the solute in solution and a second one with only the solvent. By processing the whole MD trajectories, the software returns the average SAXS spectrum. However, it does not provide the frame-by-frame spectra. We here show results obtained for different choices of the envelope width.

The Capriqorn procedure is rather similar to the one implemented in WAXSiS and thus to the experimental conditions, as two separate simulations are also conducted here. In the simulation of the whole system, the solvent within a certain distance from the solute is selected for the computation of the spectrum. In the simplest choice, the value for the radius of a sphere centered in the solute can be specified, tough other options for the geometry are available. Additionally, the width of a shell in the region farther from the solute needs to be specified, which is used to perform a solvent matching step with the simulation containing only the solvent. Contrarily to WAXSiS, besides the final average spectrum, the frame-by-frame SAXS spectra can also be obtained from an MD trajectory using Capriqorn. We here show results obtained for different choices of the sphere radius and of the shell width.

The Guinier fit procedure, that allows computing the R_g^* from the relation $\ln I(q) = \ln I_0 - \frac{(R_g^*)^2 q^2}{3}$, was carried out on the SAXS spectra. In particular, points in the range $q \in [0.02, 0.06]$ Å^{-1} of the spectra in the $\ln I(q)$ vs q^2 form were employed for the linear fit. The statistical functions module of Scipy (scipy.stats) [39] was used to perform the fit. The R_g from the coordinates was computed with MDTraj [40].

2.3 Comparison of SAXS spectra

The available experimental SAXS spectra were in the range $q \in [0.021, 0.269]$ Å^{-1}. To allow for a comparison with the spectra computed from the MD trajectories on the same q grid, the points at $q < 0.021$ Å^{-1} were extrapolated applying the Guinier fitting procedure as described above. The experimental SAXS curves were further smoothed through Gaussian Kernel regression to reduce the effect of the experimental noise when comparing with the spectra predicted using the different software packages. The Kernel regression was performed through a python script making use of Numpy [41]. The comparison was finally conducted in the range $q \in [0, 0.27]$ Å^{-1} with a grid pace of 0.01 Å^{-1}. In particular, distance matrices between pairs of i and j spectra were determined by estimating the distance d as:

$$d(i,j) = \sqrt{\frac{\sum_q (\ln \tilde{I}_i(q) - \ln \tilde{I}_j(q))^2}{N_q}} \qquad (1)$$

where $N_q = 28$ and $\tilde{I}$ were the spectra aligned at $q = 0$ Å^{-1}, obtained via

$$\ln \tilde{I}_i(q) = \ln I_i(q) - \ln I_i(q_0) \qquad (2)$$

Here $I_i(q)$ represent the spectrum obtained by analyzing an entire trajectory with one of the used methods.

For methods capable of computing the spectrum of an individual frame, the comparison was also performed frame-by-frame. In this case, the distance was computed as

$$d(i,j) = \sqrt{\frac{\sum_{q,t} (\ln \tilde{I}_i(q,t) - \ln \tilde{I}_j(q,t))^2}{N_q N_t}} \qquad (3)$$

where N_t was the number of frames in the trajectory, and spectra were aligned using the average intensity:

$$\ln \tilde{I}_i(q,t) = \ln I_i(q,t) - \ln I_i(q_0) \qquad (4)$$

Principal component analysis (PCA), as implemented in the Scikit-learn library [42], was also conducted on the pool of the different SAXS spectra. The first two components, accounting for 99.27% of the variance, were retained and used in the analysis.
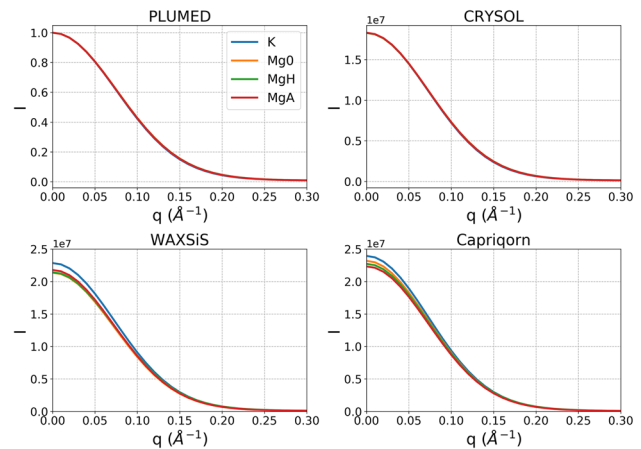

Fig. 1 Comparison of SAXS spectra computed from the MD simulations using four different software packages. Specifically, the four panels show spectra obtained with PLUMED using the Martini beads representation, CRYSOL version 2, WAXSiS with envelope width of 10 Å, and Capriqorn with radius = 40 Å and shell width = 7 Å, as indicated in panel titles. For each software package, results from the four ionic conditions having an increasing number of Mg^{2+} ions (K, Mg0, MgH and MgA, respectively) are displayed. Results are shown in the respective scales as obtained from the different software packages

3 Results

In a previous work, we performed biased and unbiased MD simulations on GAC RNA under varying ionic conditions [20]. In particular, four different setups were considered for unbiased simulations, where the concentration of Mg^{2+} cations was gradually increased. The simulations were started from the GAC folded state (see Sect. 2). Notably, relevant solute dynamics was excluded by applying soft position restraints on the heavy atoms of the RNA molecule, while allowing for adequate sampling of the systems' solvent. Herein, further analyses were conducted on the available MD trajectories to achieve a more comprehensive comparison of different existing methods to compute SAXS spectra from MD simulations. Specifically, we examined results obtained from the software PLUMED [31], CRYSOL [32,33], WAXSiS [34] and Capriqorn [35].

The SAXS spectra computed from the MD trajectories of the four different considered systems, using each one of the different software packages under examination, are displayed in Fig. 1 in the classical I vs q form, where the scattering intensity I is in arbitrary units and the scattering vector q is in Å^{-1}. For the implicit-solvent methods, namely PLUMED and CRYSOL (Fig. 1, top panels), coincident SAXS spectra were obtained for all the systems. This is expected, as only the solute coordinates enter in the computation of the spectra, and possible differences due to solute dynamics were limited by the restraints applied to the RNA. Conversely, both methods that take explicitly into account the contributions from the solvent, i.e., WAXSiS and

Capriqorn, displayed some discrepancies in response to the varying ionic conditions in the region at $q < 0.1$ Å^{-1}. Most notably, WAXSiS and Capriqorn demonstrated a consistent behavior, with SAXS spectra computed in sole presence of K$^+$ and no Mg^{2+} having a slightly higher intensity than those obtained in presence of Mg^{2+}. In particular, Capriqorn resulted more sensitive to the presence of Mg^{2+} in different amounts, while in the case of WAXSiS, the spectra were nearly indistinguishable. Nevertheless, in all cases, the different ionic conditions explored in the present study did not produce remarkable discrepancies in the overall computed SAXS spectra.

The Kratky form of the spectra, where Iq^2 is displayed against q, is indicative of the degree of compactness of the solute structure [11]. In particular, the shape of the curve is typically monitored to assess folded and unfolded states of the solute [11,44]. Herein, solute dynamics was excluded to focus on the effect of the solvent, thus the observed discrepancies can be associated with diverse solvent conditions. Figure 2a shows the Kratky plots obtained for the system with only K$^+$ using the different methods. As can be observed from the figure, differences were present and became more marked in the region at $q > 0.1$ Å^{-1}, where the contribution from the solvent is larger. In particular, all implicit-solvent methods, i.e., PLUMED and CRYSOL in both versions 2 and 3, were similar to each other, displaying higher values for Iq^2. Conversely, explicit-solvent methods resulted in profiles with lower Iq^2. The Guinier fit procedure (see Sect. 2), which allows determining the solute's radius of gyration (R_g^*, where * denotes that R_g was computed from the SAXS spectra), was also performed on the same spectra. The fits on the explicit-solvent spectra returned higher values of R_g^* of 16.8 Å and 16.7 Å, for Capriqorn and WAXSiS respectively. Interestingly, CRYSOL version 2 also resulted in a R_g^* of 16.7 Å, while the one computed through PLUMED was significantly lower (16.1 Å). CRYSOL version 3 sat in between of PLUMED and explicit-solvent approaches, with a R_g^* of 16.4 Å. As a reference, the R_g computed from the solute coordinates was 16.2 Å.

The difference between all the software packages is summarized in a pair-wise distance matrix (Fig. 3a). As a reference, we also included experimental SAXS spectra measured in presence of K$^+$ and Mg^{2+} from a previous work [43]. As already noted, differences existed between the different methods for computing the spectra. In particular, all software produced spectra that were remarkably different from the experimental one obtained with K$^+$. This was true also for the experimental spectra measured in presence of Mg^{2+}, although in this case the differences were less marked. Notably, for each method, no remarkable dependence on the varying ionic conditions was apparent, as already observed. We also performed a principal component analysis (PCA) on the compared methods. The projection on the first two components (Fig. 3b), which accounted for more than 99% of the total variance, located the experimen-
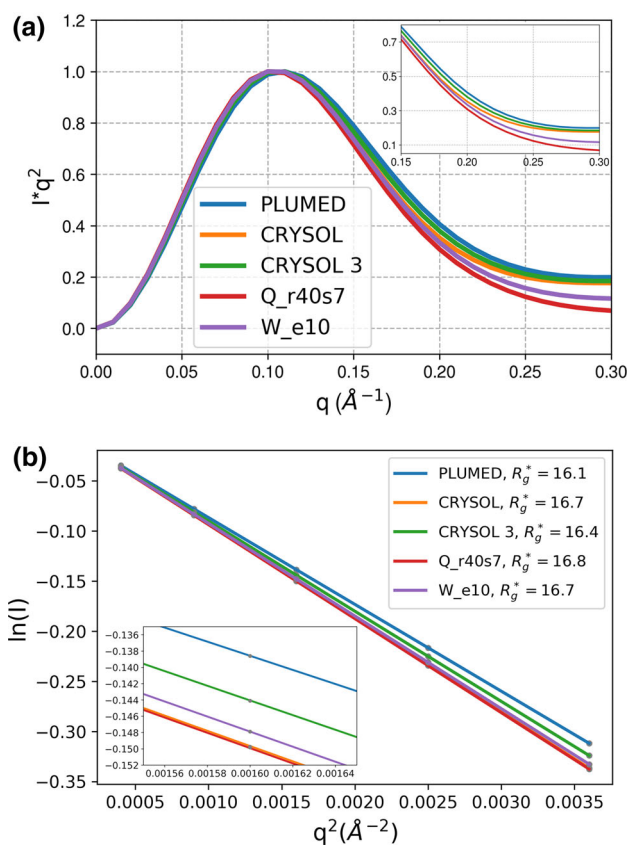

Fig. 2 Comparison of **a** SAXS spectra in the Kratky form and **b** the corresponding Guinier fits using different methods for computing the spectra (color codes are consistent between **a** and **b**). The insets show a focus where overlaying lines in the corresponding main plots can be better distinguished. The Guinier fit (colored lines) was conducted on points (gray dots) in the very low q regime of the spectra in the $\ln I$ vs q^2 form, where the Guinier approximation ($\ln(I(q)/I(0)) \approx -q^2 R_g^{*2}/3$) holds. The q range has been chosen as the one in Fig. S3 of Ref. [43]. WAXSiS spectra were computed using an envelope width of 10 Å (W_e10), while radius=40 Å and shell width = 7 Å were used for Capriqorn (Q_r40s7). Lines in panel **a** were produced with the same data shown in Fig. 3b of Ref. [20], except for CRYSOL 3 data which were computed for the present work

tal K$^+$ spectrum the farthest from all the computed spectra, at PC1≈1.5 and PC2≈1.75. The experimental Mg^{2+} spectrum, located at PC1≈0.5 and PC2≈0.75, was still evidently separate but closer to the SAXS spectra computed through the different software packages, and being in no particular vicinity with respect to any of them. Notably, all the spectra computed with the different methods for the system with K$^+$ only were grouped in the region centered at about PC1 = 0.8 and PC2 = − 0.5, with PLUMED and CRYSOL being almost coincident. Differently, in the case of the system with the highest amount of Mg^{2+}, Capriqorn, WAXSiS and CRYSOL were close in the region center at about PC1≈0.0 and PC2≈0.9, while PLUMED was in sepa-
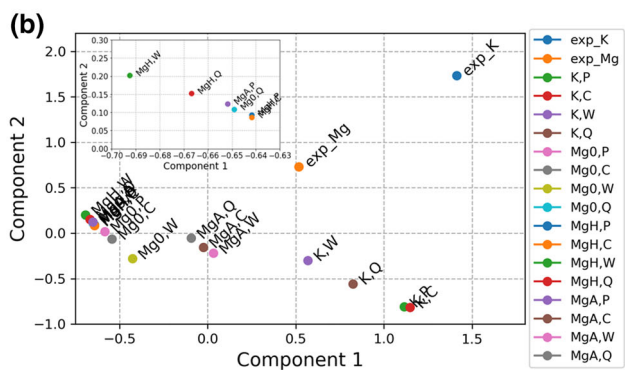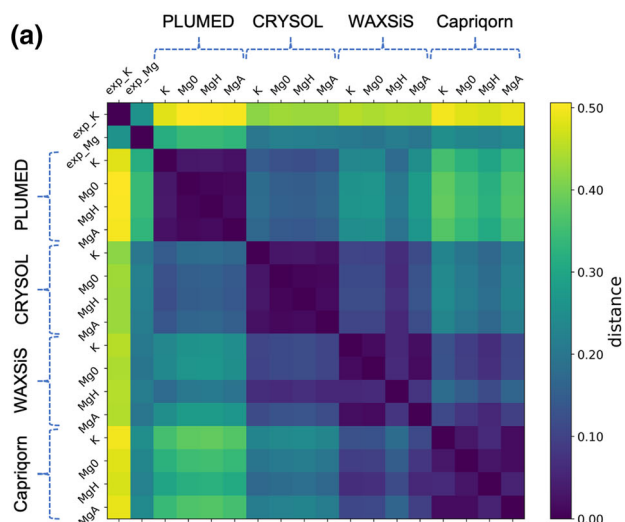
(a)

(b)

Fig. 3 Comparison of different methods and systems. **a** The distance matrix reports the root square difference between pairs of SAXS spectra, computed using different software packages (labels on top of the dashed curly brackets) and considering systems at varying ionic conditions (labels on the matrix axes: K, Mg0, MgH and MgA have increasing Mg^{2+}, see systems 1–4 in Sect. [2]) for each method. In particular, the spectrum for each system was computed with PLUMED using the Martini beads representation, CRYSOL version 2, WAXSiS with envelope width of 10 Å, and Capriqorn with radius=40 Å and shell width=7 Å. Note that the experimental spectra obtained in presence of K^+ only or K^+ and Mg^{2+} cations (first two elements of the matrix) were also included in the analysis. **b** Projection along the first two components obtained from principal component analysis (PCA) on the SAXS spectra compared in the distance matrix. System labels are consistent with the above matrix, while P, C, W and Q stand for PLUMED, CRYSOL, WAXSiS and Capriqorn, respectively

rate region. Most of the other spectra were concentrated in the region at PC1≈-0.5 and PC2≈0.25.

Finally, for the same system containing only K^+ we compared in a distance matrix all of the different methods to compute SAXS spectra from MD trajectories, while in particular exploring different choices of the parameters that can be controlled by the user (Fig. [4]). The matrix showed that, in general, implicit-solvent
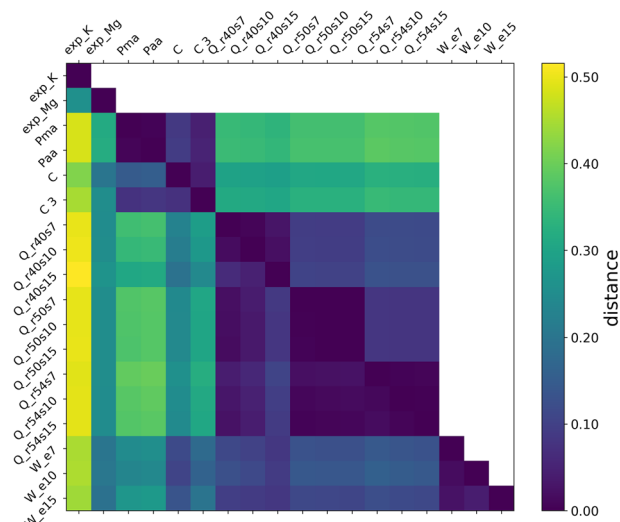

Fig. 4 Comparison of different methods and parameters. The distance matrix reports the root square difference between pairs of SAXS spectra for the same system (K^+ only, no Mg^{2+}) computed using different software and combinations of software parameters (labels on the matrix axes: Pma and Paa are PLUMED with Martini beads and all-atom; C and C3 are CRYSOL version 2 and 3, respectively; Q_r40s7 is Capriqorn with radius=40 Å and shell width=7 Å, with variants having corresponding meanings; W_e10 is WAXSiS with envelope width of 10 Å, with variants having corresponding meanings). The lower half matrix shows the comparison on the average spectra, while in the upper half the analysis was conducted on the frame-by-frame spectra of the MD trajectory. White regions correspond to cases where the frame-by-frame comparison was not possible, specifically the experimental spectra and methods that do not provide the frame-by-frame SAXS spectra, i.e., WAXSiS

methods displayed higher similarity between them than to the methods that included the solvent in the computation of the spectra (lower half of the matrix in Fig. [4]). Additionally, the picture did not change when conducting the comparison on the frame-by-frame spectra for the same MD trajectory, where possible (upper half of the matrix in Fig. [4]). Furthermore, for each software package, the overall picture was independent of the specific choice of the parameters.

4 Discussion

In this study, we compared the results obtained from different methods that allow computing SAXS spectra from MD simulation trajectories. Specifically, we focused on the implementations of the PLUMED [31], CRYSOL [32,33], WAXSiS [34] and Capriqorn [35] software packages and computed the spectra for four different MD simulations of the same system under diverse ionic conditions. Most importantly, the structure of the solute, namely the GTPase-associated center (GAC) RNA, was kept fixed through soft position restraints,

so as to neglect solute dynamics and focus on solvent effects. Raw spectra and a python notebook that can be used to reproduce the figures reported in the Results section can be found at https://github.com/bussilab/comparison-saxs.

For a given MD simulation, the comparison of the SAXS spectra in the low-q regime ($q < 0.3$ Å^{-1}) highlighted an appreciable difference between implicit-solvent and explicit-solvent methods in computing the SAXS spectra. Indeed, implicit-solvent approaches, i.e., PLUMED and CRYSOL, produced similar results and thus grouped together (Fig. 4). Explicit-solvent methods, i.e., WAXSiS and Capriqorn, also generated compatible spectra that were clearly distinguishable from the implicit-solvent ones. A clear advantage of implicit-solvent schemes, where the SAXS spectra are estimated using the solute coordinates and accounting for solvent effects by introducing corrections, is the speed of computation. This can be of critical relevance within an MD framework in those cases where the spectra need to be computed on-the-fly during the simulation. In such a scenario, a further boost can be achieved through a coarse-grained representation of the system, as implemented in PLUMED. Notably, while the all-atom representation is used for the MD simulations, the system is projected into a Martini bead representation for the sole purpose of computing the SAXS spectrum. This procedure was observed to produce coincident results with spectra computed from an all-atom representation in the low-q regime ($q < 0.3$ Å^{-1}) for a variety of systems [37], and was herein confirmed with an independent result (Fig. 4). Contrarily, explicit-solvent approaches, that include the contribution from solvent molecules explicitly in the calculation, are computationally more demanding. However, the procedure is designed so as to be as close as possible to the experimental one, since an additional simulation with no solute is typically required in order to conduct the solvent subtraction. It is worth noticing that, in such a scenario, the explicit-solvent framework gives the possibility to explicitly model different solvent conditions. Nevertheless, in the case of the varying ionic conditions with fixed RNA structure that we inspected herein, the effect observed was rather minimal (Fig. 3). Therefore, the observed discrepancies between implicit and explicit-solvent approaches can be mainly ascribed to the inclusion of the water contribution in the computation of the SAXS spectra. In particular, we can hypothesize this to be related to a structured solvation shell resulting from the highly charged RNA biomolecule [20].

As already stressed, to focus on the role of the solvent, we intentionally neglected the effect of structural dynamics in our simulations through the application of position restraints to the RNA molecule. However, when comparing the computed SAXS spectra with experimental ones, such effect can become particularly critical. Herein, we also compared the results from the different software packages with experimental SAXS data obtained for GAC RNA in presence of K$^+$ or Mg^{2+} [43]. In particular, the experiments suggested that K$^+$ favored more extended conformations of GAC, while Mg^{2+} favored the folded state. Notably, our systems were constructed with the RNA molecule in its folded conformation, as found in crystal structures [21,45]. Nevertheless, it is interesting to note that none of the solvent conditions that we examined found agreement with neither of the experimental SAXS spectra. Indeed, in a separate work, we have observed that a mixture of compact and extended conformations was necessary to correctly reproduce the experimental SAXS spectrum obtained in presence of Mg^{2+} [20]. When different conformations are accessible in solution, experimental SAXS measurements return the average signal resulting from such conformational ensembles. As such, reasoning in terms of structural ensembles, and thus allowing the solute to sample different conformations during the MD simulations, can be of pivotal importance to conduct a meaningful comparison with reference experimental data. However, achieving an exhaustive sampling of solutes' conformational space through unbiased MD simulations is most times intractable. Thus, resorting to enhanced sampling approaches can be a successful strategy [46]. In this respect, using implicit-solvent spectra, which can be computed on-the-fly without severely affecting the MD simulation performance, as collective variables can be of striking support. Notwithstanding their limited accuracy, the exploration of a broad variety of structures corresponding to heterogeneous SAXS spectra [20] or the direct enforcement of experimental data during the MD [36] can produce valuable starting ensembles on which conducting more accurate reweighting procedures [20,47].

Concerning the simulations inspected in this work, where only compact conformations were considered for the solute and conformational dynamics was excluded, we have observed that the radius of gyration computed form the SAXS spectra through Guinier fit (R_g^*) was higher for explicit-solvent methods with respect to the pure-solute approaches implemented in PLUMED and CRYSOL 3 (Fig. 2b). Interestingly, the CRYSOL 2 was able to produce an R_g^* value consistent with explicit-solvent methods, despite the SAXS spectrum demonstrated higher similarity with the one computed with the other implicit-solvent methods (Fig. 4). Nevertheless, this effect might be non-systematic and could be non-trivially predictable for extended structures [20].

Finally, for each software package, we explored diverse choices for the parameters that can be specified by the user. Indeed, while implicit-solvent approaches rely solely on the solute structure factors and possible corrections to account for solvation effects, explicit-solvent methods have a wider variety of parameters that can be tuned (see Sect. 2). As a result, we observed that the results weakly depended on the set of parameters that we considered.

In conclusion, our comparison of state-of-the-art software packages that are available to back-calculate SAXS spectra from MD simulations highlighted how the effects of the solvent can be appreciable even in the low-q regime. Therefore, we suggest that the choice should be adapted depending on the specific scientific scope. In particular, when aiming at a comparison

with experimental SAXS data, explicit-solvent methods should be employed. Conversely, when a wider exploration of the solute conformational space is sought, less accurate but computationally more effective implicit-solvent approaches might be a more suitable option.

Author contributions

M.B. and G.B. conceived the presented idea. M.B. performed the computations. M.B. and G.B. discussed the results and contributed to the final manuscript.

Data availability statement This manuscript has associated data in a data repository. [Authors' comment: That can be found at https://github.com/bussilab/comparison-saxs.]

References

1. M.H. Koch, P. Vachette, D.I. Svergun, Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q. Rev. Biophys. **36**(2), 147 (2003)
2. S. Skou, R.E. Gillilan, N. Ando, Synchrotron-based small-angle X-ray scattering of proteins in solution. Nat. Protoc. **9**(7), 1727–1739 (2014)
3. A.T. Tuukkanen, A. Spilotros, D.I. Svergun, Progress in small-angle scattering from biological solutions at high-brilliance synchrotrons. IUCrJ **4**(5), 518–528 (2017)
4. P. Bernadó, E. Mylonas, M.V. Petoukhov, M. Blackledge, D.I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. **129**(17), 5656–5664 (2007)
5. H.D. Mertens, D.I. Svergun, Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. **172**(1), 128–141 (2010)
6. P. Bernado, D.I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol. Biosyst. **8**(1), 151–167 (2012)
7. A.G. Kikhney, D.I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsi-

cally disordered proteins. FEBS Lett. **589**(19), 2570–2577 (2015)
8. M. Kachala, E. Valentini, D.I. Svergun, in *Application of SAXS for the structural characterization of IDPs, in Intrinsically Disordered Proteins Studied by NMR Spectroscopy.* ed. by I. Felli, R. Pierattelli (Springer, Cham, 2015)
9. J. Lipfert, V.B. Chu, Y. Bai, D. Herschlag, S. Doniach, Low-resolution models for nucleic acids from small-angle X-ray scattering with applications to electrostatic modeling. J. Appl. Crystallogr. **40**(s1), s229–s234 (2007)
10. Y. Chen, L. Pollack, SAXS studies of RNA: structures, dynamics, and interactions with partners. Wiley Interdiscip. Rev. RNA **7**(4), 512–526 (2016)
11. W.A. Cantara, E.D. Olson, K. Musier-Forsyth, Analysis of RNA structure using small-angle X-ray scattering. Methods **113**, 46–55 (2017)
12. R.O. Dror, R.M. Dirks, J. Grossman, H. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology. Annu. Rev. Biophys. **41**, 429–452 (2012)
13. J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R.A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N.G. Walter, M. Otyepka, RNA structural dynamics as captured by molecular simulations: a comprehensive overview. Chem. Rev. **118**(8), 4177–4338 (2018)
14. J.S. Hub, Interpreting solution X-ray scattering data using molecular simulations. Curr. Opin. Struct. Biol. **49**, 18–26 (2018)
15. A.A. Chen, D.E. Draper, R.V. Pappu, Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. J. Mol. Biol. **390**(4), 805–819 (2009)
16. M.T. Ivanović, L.K. Bruetzel, R. Shevchuk, J. Lipfert, J.S. Hub, Quantifying the influence of the ion cloud on SAXS profiles of charged proteins. Phys. Chem. Chem. Phys. **20**(41), 26351–26361 (2018)
17. D. Moazed, J.M. Robertson, H.F. Noller, Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature **334**(6180), 362–364 (1988)
18. Y.-G. Gao, M. Selmer, C.M. Dunham, A. Weixlbaumer, A.C. Kelley, V. Ramakrishnan, The structure of the ribosome with elongation factor G trapped in the posttranslocational state. Science **326**(5953), 694–699 (2009)
19. A. Weixlbaumer, H. Jin, C. Neubauer, R.M. Voorhees, S. Petry, A.C. Kelley, V. Ramakrishnan, Insights into translational termination from the structure of RF2 bound to the ribosome. Science **322**(5903), 953–956 (2008)
20. M. Bernetti, K.B. Hall, G. Bussi, Reweighting of molecular simulations with explicit-solvent SAXS restraints elucidates ion-dependent RNA ensembles. Nucleic Acids Res. **2021**, gkab459 (2021)
21. G.L. Conn, A.G. Gittis, E.E. Lattman, V.K. Misra, D.E. Draper, A compact RNA tertiary structure contains a buried backbone-K+ complex. J. Mol. Biol. **318**(4), 963–973 (2002)
22. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and

organic molecules. J. Am. Chem. Soc. **117**(19), 5179–5197 (1995)

23. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T.E. Cheatham III, C.A. Laughton, M. Orozco, Refinement of the amber force field for nucleic acids: improving the description of α/γ conformers. Biophys. J. **92**(11), 3817–3829 (2007)

24. M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T.E. Cheatham III, P. Jurecka, Refinement of the cornell et al nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theory Comput. **7**(9), 2886–2902 (2011)

25. S. Izadi, R. Anandakrishnan, A.V. Onufriev, Building water models: a different approach. J. Phys. Chem. Lett. **5**(21), 3863–3871 (2014)

26. I.S. Joung, T.E. Cheatham III, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J. Phys. Chem. B **112**(30), 9020–9041 (2008)

27. O. Allnér, L. Nilsson, A. Villa, Magnesium ion-water coordination and exchange in biomolecular simulations. J. Chem. Theory Comput. **8**(4), 1493–1502 (2012)

28. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. J. Chem. Phys. **126**(1), 01401 (2007)

29. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. **52**(12), 7182–7190 (1981)

30. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX **1**, 19–25 (2015)

31. G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: new feathers for an old bird. Comput. Phys. Commun. **185**(2), 604–613 (2014)

32. D. Svergun, C. Barberato, M.H. Koch, Crysol-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. **28**(6), 768–773 (1995)

33. D. Franke, M. Petoukhov, P. Konarev, A. Panjkovich, A. Tuukkanen, H. Mertens, A. Kikhney, N. Hajizadeh, J. Franklin, C. Jeffries et al., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. **50**(4), 1212–1225 (2017)

34. C.J. Knight, J.S. Hub, WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. Nucleic Acids Res. **43**(W1), W225–W230 (2015)

35. J. Köfinger, G. Hummer, Atomic-resolution structural information from scattering experiments on macromolecules in solution. Phys. Rev. E **87**(5), 052712 (2013)

36. M. Bonomi, C. Camilloni, Integrative structural and dynamical biology with PLUMED-ISDB. Bioinformatics **33**(24), 3999–4000 (2017)

37. C. Paissoni, A. Jussupow, C. Camilloni, Martini bead form factors for nucleic acids and their application in the refinement of protein-nucleic acid complexes against SAXS data. J. Appl. Crystallogr. **52**(2), 394–402 (2019)

38. R. Fraser, T. MacRae, E. Suzuki, An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules. J. Appl. Crystallogr. **11**(6), 693–694 (1978)

39. P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright et al., Scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods **17**(3), 261–272 (2020)

40. R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C.X. Hernández, C.R. Schwantes, L.-P. Wang, T.J. Lane, V.S. Pande, MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys. J. **109**(8), 1528–1532 (2015)

41. C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith et al., Array programming with numpy. Nature **585**(7825), 357–362 (2020)

42. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

43. R. Welty, S.A. Pabit, A.M. Katz, G.D. Calvey, L. Pollack, K.B. Hall, Divalent ions tune the kinetics of a bacterial GTPase center rRNA folding transition from secondary to tertiary structure. RNA **24**(12), 1828–1838 (2018)

44. S. Doniach, Changes in biomolecular conformation seen by small angle X-ray scattering. Chem. Rev. **101**(6), 1763–1778 (2001)

45. R. Welty, M. Rau, S. Pabit, M.S. Dunstan, G.L. Conn, L. Pollack, K.B. Hall, Ribosomal protein L11 selectively stabilizes a tertiary structure of the GTPase center rRNA domain. J. Mol. Biol. **432**(4), 991–1007 (2020)

46. R.C. Bernardi, M.C. Melo, K. Schulten, Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim. Biophys. Acta Gen. Subj. **1850**(5), 872–877 (2015)

47. A. Cesari, S. Reißer, G. Bussi, Using the maximum entropy principle to combine simulations and solution experiments. Computation **6**(1), 15 (2018)