



Integrating an Enhanced Sampling Method and Small-Angle X-Ray Scattering to Study Intrinsically Disordered Proteins

Chengtao Ding¹, Sheng Wang² and Zhiyong Zhang^{1*}

¹MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, National Science Center for Physical Sciences at Microscale, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, ²Tencent AI Lab, Shenzhen, China

Intrinsically disordered proteins (IDPs) have been paid more and more attention over the past decades because they are involved in a multitude of crucial biological functions. Despite their functional importance, IDPs are generally difficult to investigate because they are very flexible and lack stable structures. Computer simulation may serve as a useful tool in studying IDPs. With the development of computer software and hardware, computational methods, such as molecular dynamics (MD) simulations, are popularly used. However, there is a sampling problem in MD simulations. In this work, this issue is investigated using an IDP called unique long region 11 (UL11), which is the conserved outer tegument component from herpes simplex virus 1. After choosing a proper force field and water model that is suitable for simulating IDPs, integrative modeling by combining an enhanced sampling method and experimental data like small-angle X-ray scattering (SAXS) is utilized to efficiently sample the conformations of UL11. The simulation results are in good agreement with experimental data. This work may provide a general protocol to study structural ensembles of IDPs.

Keywords: IDPs, biological function, MD simulation, sampling, integrative modeling

INTRODUCTION

It has been recognized that a large segment of the human proteome comprises intrinsically disordered proteins (IDPs) that lack stable secondary and tertiary structures under physiological conditions (Colak et al., 2013; Kulkarni and Uversky, 2019). IDPs play important roles in a multitude of crucial biological functions despite their lack of a stable structure, such as cell cycle regulation, molecular recognition, and signal transduction (Dunker et al., 2005; Uversky et al., 2005). According to previous work, IDPs are involved in the majority of human cancer (Iakoucheva et al., 2002) and many chronic diseases like cardiovascular disease (Cheng et al., 2006), neurodegenerative diseases (Uversky, 2009; Uversky, 2014), and type 2 diabetes (Du and Uversky, 2017).

Although researchers continue to discover the functional importance of IDPs, it remains difficult to explore the structure-function relationship because getting the high-resolution structures of IDPs remains elusive. Since an IDP is generally not stable in one conformational state, these classical technologies of structural biology, including X-ray crystallography and cryo-EM, cannot determine its atomic-resolution structure. Alternatively, structural information on the ensemble average of the

OPEN ACCESS

Edited by:

Yong Wang,
University of Copenhagen, Denmark

Reviewed by:

Haiguang Liu,
Beijing Computational Science
Research Center, China
Andreas Larsen,
University of Oxford, United Kingdom

*Correspondence:

Zhiyong Zhang
zzyzhang@ustc.edu.cn

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 25 October 2020

Accepted: 08 February 2021

Published: 15 April 2021

Citation:

Ding C, Wang S and Zhang Z (2021)
Integrating an Enhanced Sampling
Method and Small-Angle X-Ray
Scattering to Study Intrinsically
Disordered Proteins.
Front. Mol. Biosci. 8:621128.
doi: 10.3389/fmolb.2021.621128

IDP is available by techniques like nuclear magnetic resonance (NMR) (Dunker and Oldfield, 2015), small-angle X-ray scattering (SAXS) (Bernado and Svergun, 2012), and Förster resonance energy transfer (FRET) (LeBlanc et al., 2018).

In order to obtain structural details of IDPs, atomistic molecular dynamics simulation is a useful and complementary method for illuminating the molecular nature of IDPs' conformational ensembles because it can provide spatial and temporal resolution unavailable from experiments (Potoyan and Papoian, 2011; Burger et al., 2014; Granata et al., 2015; Bhowmick et al., 2016). Despite the significant progress made, a sampling problem remains in MD simulations of IDPs. The conformational space of an IDP is generally very large, so conventional MD simulations at a timescale of microseconds (μ s) cannot capture all the states adequately. To tackle this problem, many enhanced sampling methods have been developed, which achieve good sampling by modifying potential energy function (Hamelberg et al., 2004) or increasing the temperature of barrier regions (Zhang et al., 2003; Hu et al., 2012). In recent years, a new kind of sampling techniques has been proposed, which are built on iterative multiple independent MD (MIMD) simulations (Harada and Kitao, 2013, Harada and Kitao, 2015; Shkurti et al., 2019; Yuan et al., 2020; Zhang and Gong, 2020). Such a method generally contains many cycles, and each cycle consists of a number of short MIMD simulations starting from selected seed conformations. The sampling efficiency would depend on the strategy of selecting seeds, and different criteria have been tried (Harada and Shiget, 2018).

Many studies have shown the possibility of combining experimental data and computational simulations to interpret structural dynamics of large biomolecules in a solution that is called integrative modeling (Braitbard et al., 2019). There are various integrative modeling techniques for the interpretation of different structural data (Bonomi et al., 2017; Saltzberg et al., 2019; Orioli et al., 2020), which can be divided into two categories: refining-while-sampling and the screening-after-sampling (Zhang et al., 2015). A refining-while-sampling method directly adds an extra pseudo energy term based on the experimental data and then a conformation or an ensemble is simulated by optimizing the energy (Zheng and Tekpinar, 2011; Bjorling et al., 2015). In a screening-after-sampling method, a structure pool of the biomolecule is firstly sampled without experimental restraints, and then a reweighting method acts on these conformations to optimize their weights in order to fit the experimental data well (Bottaro et al., 2020). An ensemble containing a small number of conformations selected from the pool could be determined (Bernado et al., 2007; Curtis et al., 2012).

In this work, we propose a general strategy to study the conformations of IDPs. After choosing a suitable force field and water model for simulating IDPs, an integrative modeling procedure combining an enhanced sampling method based on iterative MIMD and SAXS data is used to sample conformations of IDPs efficiently. We present a case study on an IDP called unique long region 11 (UL11), an RNA-binding protein that is one of the conserved outer tegument components from herpes simplex virus 1 (HSV-1) (Bowzard et al., 2000; Metrick et al., 2020).

HSV-1 contains a unique tegument layer sandwiched between the capsid and lipid envelope, including 24 tegument proteins

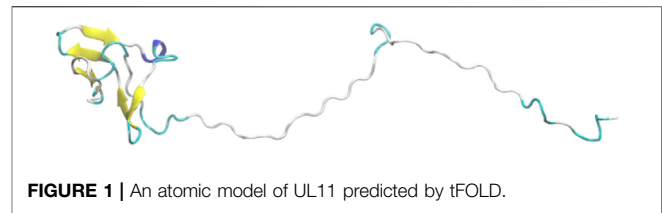


FIGURE 1 | An atomic model of UL11 predicted by tFOLD.

(McLauchlan and Rixon, 1992). UL11 is the smallest tegument protein with only 96 amino-acid residues (MacLean et al., 1989; Bowzard et al., 2000). UL11 and its homologs have been found to play crucial roles in efficient viral replication (MacLean et al., 1992; Baird et al., 2010) and tegument assembly (Owen et al., 2015). However, the mechanistic understanding of its role in these processes is limited due to the lack of knowledge of its biochemical and structural properties. A recent article (Metrick et al., 2020) has suggested that UL11 is an IDP in solution, which can undergo liquid-liquid phase separation (LLPS) *in vitro*. Analysis of experimental SAXS data showed that the protein is highly dynamic. Here, we aim to construct an atomic structural ensemble of UL11 that is in agreement with the available experimental data.

MATERIALS AND METHODS

An Initial Atomic Model of UL11

The UL11 construct used in this work is called UL11-StII (Metrick et al., 2020), which is the UL11 sequence (96 residues) plus a small C-terminal Strep-tag II (StII) including eight residues (WSHPQFEK). We used this 104-residue construct, on which the SAXS experiment was conducted. In the following, we call this construct UL11 for simplicity.

According to a prediction from the FoldUnfold server (<http://bioinfo.protres.ru/ogu>), many residues of UL11 are predicted to be disordered, except for some N-terminal residues that are natively folded (Metrick et al., 2020). We predicted an atomic model of UL11 using the tFOLD server (<https://drug.ai.tencent.com/console/cn/tfold>) (Figure 1). There are some β -strands at the N-terminus (residues 11–14, 17–20, 24–27, 39–41, and 44–47), while the other regions are disordered till the C-terminal end. The tFOLD model is consistent to the prediction of the disorder, so we used it as a starting structure for simulations.

Simulation Details

In this work, all-atom conventional MD (cMD) simulations and accelerated MD (aMD) simulations were conducted using the Amber20 package.

Conventional MD (cMD) Simulation

It has been recognized that, in MD simulations using those traditional force fields and water models, IDPs may become over-compact. Therefore, combinations of new force fields and water models have been proposed to address this issue (Kuzmanic et al., 2019). In this work, we used the A99SB force field in combination with a 4-point OPC water model (Izadi et al., 2014).

It has been reported that this A99SB/OPC combination is suitable for simulating conformations of IDPs (Shabane et al., 2019).

The system was built via the LEaP module (Case et al., 2005). The OPC waters (Izadi et al., 2014) were added to a truncated octahedral box with a minimal distance of 10.0 Å between the solute and the box boundary. 102 Na⁺ and 98 Cl⁻ ions were added by replacing water molecules to balance the charge on the system and bring the salt concentration to about 100 mM NaCl. The box size is $1.66 \times 10^6 \text{ \AA}^3$, with 205,909 atoms in total. To remove bad contacts, the waters and ions were initially minimized for 2,000 steps using the steepest descent method for the first 1,500 steps and then the conjugate gradient for the last 500 steps, with the position of protein fixed (force constant was $500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). In the second energy minimization, the restraints on the protein were removed. This stage was conducted for 2,500 steps, using the steepest descent method in the first 1,000 steps and then the conjugate gradient algorithm for the last 1,500 steps. After that, a heat-up MD was run at a constant volume. The system was heated from 0 to 300 K for 100 ps with a weak restraint of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ on the solute. A free MD simulation of 150 ns was carried out under the NPT condition utilizing the GPU-accelerated pmemd.cuda code. The temperature was regulated using the Langevin dynamics with a collision frequency of 1.0 ps^{-1} (Pastor et al., 1988). Pressure was controlled with isotropic position scaling at 1 bar with a relaxation time of 2.0 ps. All the bonds involving hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert et al., 1977). A 2 fs integration step was used. Van der Waals interactions outside the cutoff distance were approximated via a continuum model (vdwmeth = 1) (Izadi et al., 2014; Izadi and Onufriev, 2016). The long-range electrostatic interaction was calculated using the PME method (Muller et al., 1996) with a 10 Å cutoff for the range-limited nonbonded interaction.

Accelerated MD (aMD) Simulation

The aMD (Muller et al., 1996) introduces a boost potential, $\Delta V(r)$, to the original potential energy $V(r)$ when the latter is below a threshold energy E :

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E, \\ \left[\frac{(E - V(r))}{\alpha + (E - V(r))} \right]^2, & V(r) < E. \end{cases} \quad (1)$$

where α is a factor that tunes the depth of the modified energy basins. Boosting potentials were applied to both the total potential and the individual dihedral energy term. The aforementioned 150 ns cMD simulation was used to estimate the aMD parameters. In the cMD trajectory, the average total potential energy was $-641,138 \text{ kcal mol}^{-1}$ and the average dihedral energy was $1,068 \text{ kcal mol}^{-1}$. UL11 has 104 residues and the simulated system consists of 205,909 atoms. The following parameters were set based on the above information:

$$E(\text{tot}) = -641,138 \text{ kcal mol}^{-1} + (0.2 \text{ kcal mol}^{-1} \text{ atom}^{-1} \times 205,909 \text{ atoms}) \approx -599,956 \text{ kcal mol}^{-1}$$

$$\alpha(\text{tot}) = 205,909 \text{ atoms} \times 0.2 \text{ kcal mol}^{-1} \text{ atom}^{-1} \approx 41,182 \text{ kcal mol}^{-1}$$

$$E(\text{dih}) = 1,068 \text{ kcal mol}^{-1} + (3.5 \text{ kcal mol}^{-1} \text{ residue}^{-1} \times 104 \text{ residues}) \approx 1,432 \text{ kcal mol}^{-1}$$

$$\alpha(\text{dih}) = 0.2 \times (3.5 \text{ kcal mol}^{-1} \text{ residue}^{-1} \times 104 \text{ residues}) \approx 73 \text{ kcal mol}^{-1}$$

With these parameters, a 150 ns aMD simulation was conducted. All the other parameters were the same to the aforementioned cMD simulation.

The Strategy of Integrative Modeling

We have previously developed a method called SAXS-oriented ensemble refinement (SAXS-ER) (Cheng et al., 2017), and the flowchart is as follows (Figure 2). The code is available at <https://github.com/pcheng27/SAXS-ER/tree/v1.1>.

- 1) Set up the system starting from an initial structure of the biomolecule, and perform a preliminary simulation. Any simulation method can be utilized, such as atomistic MD simulations, enhanced sampling techniques, or coarse-grained modeling. In this work, we are studying an IDP, and the sampling is challenging. Therefore, aMD simulations are carried out using the most updated code of pmemd.cuda in the Amber20 package.
- 2) Calculate the scoring function and obtain an ensemble of conformers with the best score. The number of conformers in the ensemble is N_{es} . In this work, the scoring function is χ^2 between the calculated SAXS profile of the ensemble and the experimental SAXS profile. More details will be introduced in the “Ensemble Optimization Method” section.
- 3) Starting from the N_{es} conformers selected by scoring function, $N_{\text{sim}} (=N_{\text{es}})$ -independent simulations are carried out. Multiple independent short-time simulations may achieve a better sampling than a single long-time simulation. All the trajectories are combined.
- 4) Repeat steps 2 and 3 for N cycles. Analyze all those cycles with the saturated scoring function.

SAXS Data

The SAXS data of UL11 were taken from SASBDB (www.sasbdb.org) with the ID SASDEX4. All the experimental details and analyzed results can be found in the database and the published article (Metrick et al., 2020). In this work, we took the data points with q from 0.009 to 0.206 \AA^{-1} ($q = 4\pi \sin \theta / \lambda$, where 2θ is the scattering angle and λ is the wavelength of 1.246 \AA), and the signal-noise ratios in this range are essentially larger than 2.0 (Figure 3A). The radius of gyration (R_g) of the protein was estimated to be $24.1 \pm 1.7 \text{ \AA}$ by Guinier analysis using the autoRg program in the ATSAS package (Franke et al., 2017). The pair distance distribution function (PDDF) was calculated by GNOM (Semenyuk and Svergun, 1991) using the maximum dimension (D_{max}) of 89.0 \AA as input. The normalized PDDF is asymmetrical and tailed off to a large distance (Figure 3B), which resembles the shape of an elongated ellipsoid (Mertens and Svergun, 2010). Therefore, the protein should be able to take extended conformations in the solution that can be disordered. The Kratky plot (Figure 3C) also supports that the protein is an IDP with partially folded regions.

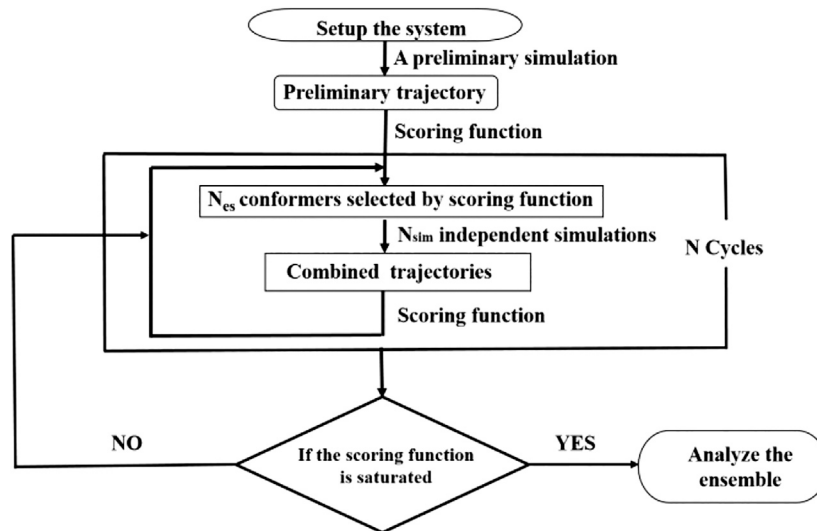


FIGURE 2 | Flowchart of integrative modeling that is a modification from **Figure 1** in (Cheng et al., 2017).

CRY SOL (Svergun et al., 1995) was used to compute the theoretical SAXS profile of a known atomic structure in PDB format, and then autoRg was run on the SAXS profile to estimate the R_g of the structure. The CaPP software, available at github.com/Niels-Bohr-Institute-XNS-StructBiophys/CaPP, was used to calculate PDDF from these PDB files.

Ensemble Optimization Method

A structural ensemble was obtained by the ensemble optimization method (EOM) (Bernado et al., 2007). EOM was used to select a small number of representative conformations from a pool containing lots of conformations of UL11 in order to fit the experimental SAXS data. The scoring function of EOM is as follows:

$$\chi^2 = \frac{1}{K-1} \sum_{i=1}^K \left[\frac{\mu I(q_i) - I_{\text{exp}}(q_i)}{\sigma(q_i)} \right]^2, \quad (2)$$

where K is the number of data points in the SAXS profile and $\sigma(q)$ are experimental errors. For every conformation in the ensemble, its theoretical scattering profile is computed. $I(q)$ is the average of them, and μ is a scaling factor.

A new version of EOM called EOM2 (Tria et al., 2015) was used to compute the scoring function (Eq. 2) and pick the ensembles. In the original SAXS-ER using EOM2 (Cheng et al., 2017), the program automatically determined the ensemble size in each cycle that was generally small. An IDP should be represented by an ensemble containing more conformers than folded proteins. Therefore, in this work, we used an option of fixing the ensemble size to a relatively large number like 24 when running EOM2 in each cycle.

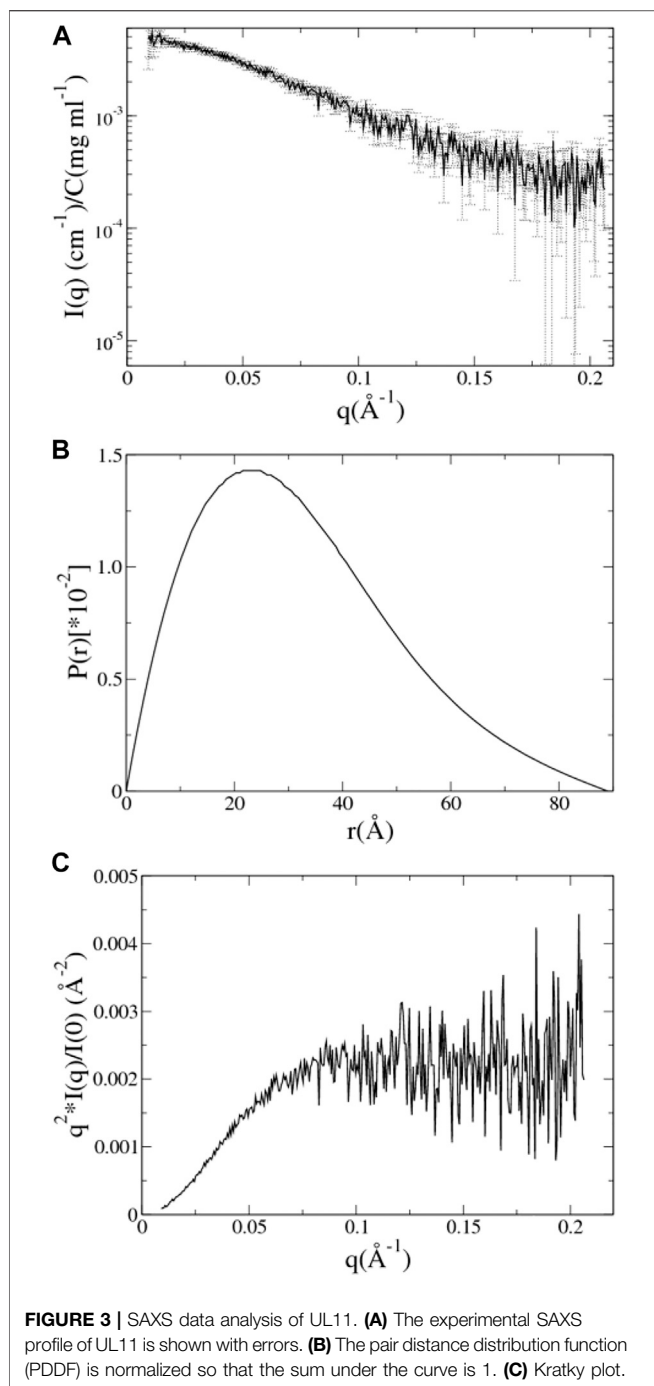
RESULTS AND DISCUSSION

aMD of UL11 without Integrating the SAXS Data. Three independent aMD simulations, each of 150 ns, were conducted.

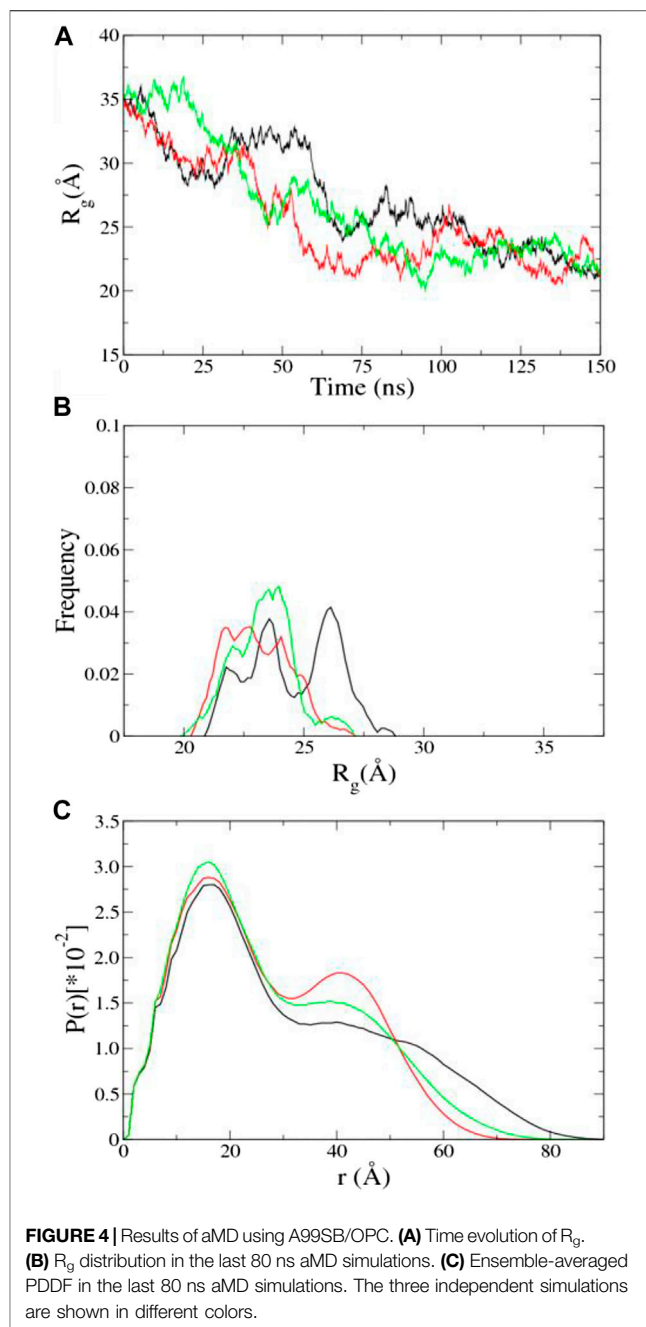
We converted a trajectory into sequentially individual PDB files; then CRY SOL and autoRg were run to obtain R_g of each atomic structure as described in the “SAXS Data” section. The initial structure of UL11 (Figure 1) is extended with R_g of 35.2 Å. In the first 70 ns of the aMD simulations, the protein is equilibrating with a clear tendency of R_g decrease (Figure 4A), and then the R_g values essentially fluctuate between 21.0 and 27.5 Å in the remaining simulations. According to the R_g distribution of the conformations in the last 80 ns (Figure 4B), they seem to show agreement with the experimental R_g of 24.1 ± 1.7 Å. We calculated the PDDF of each conformation in the last 80 ns of one trajectory and then plotted the ensemble-averaged PDDF (Figure 4C). The shape of the three ensemble-averaged PDDF curves is obviously not similar to that of the experimental PDDF (Figure 3B). That is to say, the aMD simulations at the time scale of 150 ns cannot adequately sample solution conformations of the IDP, which is the cause for the discrepancy between the simulated and the experimental PDDF. A straightforward way is to simply run longer simulations so that the protein could expand again and sample diverse conformations. However, it is not sure how long would be long enough to give a representative picture of the IDP. Therefore, we performed integrative modeling of UL11.

Integrative Modeling of UL11. Starting from the same structural model (Figure 1), we conducted integrative modeling of UL11 using the protocol introduced in Figure 2. A cycle consisted of $N_{\text{sim}} = 24$ independent 200 ps aMD simulations using A99SB/OPC. In each aMD simulation, a conformation was recorded every 1 ps, so a structural pool containing 4,800 conformations was generated in one cycle. By fitting the experimental SAXS data of UL11, EOM2 selected an ensemble with the size of $N_{\text{es}} = 24$ from the pool. Starting from these conformations, the next cycle of multiple independent simulations was run. We carried out 30 cycles, so the total simulation time was 144 ns (200 ps \times 24 aMD \times 30 cycles).

The χ^2 and the average R_g ($\langle R_g \rangle$) of the ensemble are plotted against the cycle number (Figure 5A). The initial model of UL11

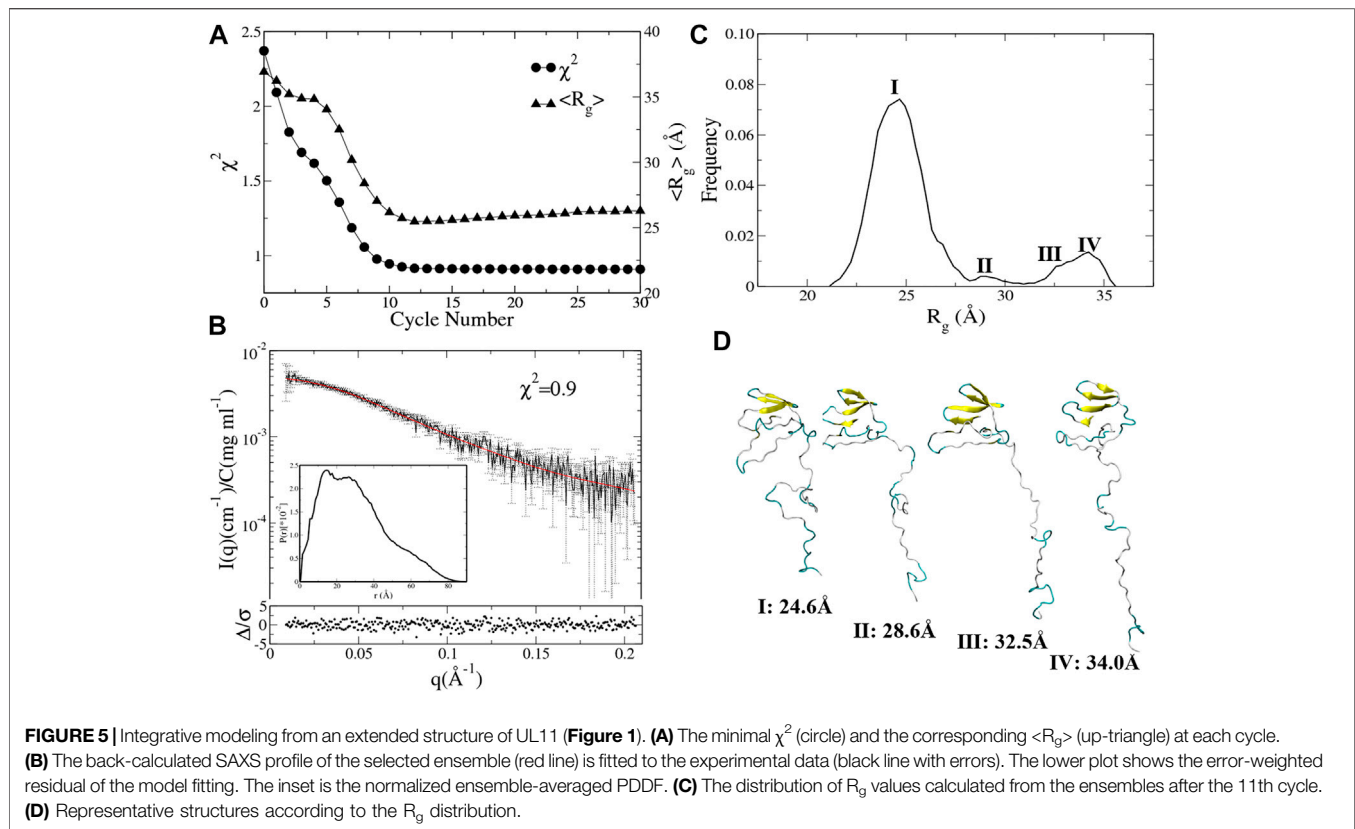


is very extended (**Figure 1**); the EOM ensemble generated at cycle 0 cannot fit the experimental SAXS data well, with a χ^2 of 2.3. It is found that χ^2 decreases relatively fast in the first eight cycles (from 2.3 to 1.0), and then it slowly converges to about 0.9 after the 10th cycle (**Figure 5A**, circle). When looking at the time evolution of the $\langle R_g \rangle$ (**Figure 5A**, up-triangle), it converges to 25.5 Å after 12 cycles, that is in good agreement with the estimated R_g (24.1 ± 1.7 Å) from the experimental SAXS data (**Figure 3A**). Therefore, we plotted the calculated SAXS profile of the ensemble at the 12th cycle and its error-weighted residual



(**Figure 5B**). The residuals are defined as $(I_{exp}(q) - I_{calc}(q)) / \sigma_{exp}(q)$, corresponding to the difference between the experimental and the computed intensities weighted by the experimental uncertainty (Carter et al., 2015; Trewhella et al., 2017). The residual difference plot is flat, which indicates that the results are in good agreement with the data. The inset is the normalized average PDDF of the ensemble, which has a similar shape to the experimental PDDF (**Figure 3B**).

To characterize conformations consistent with the SAXS data, we analyzed the R_g distribution of all the ensembles after the 11th cycle (**Figure 5C**). There is a major peak with the R_g value around 24.6 Å, a minor peak located between 27.5 and 30.0 Å, and two



more peaks with the R_g values larger than 30.0 Å that do not appear in the 150 ns aMD simulations (Figure 4B). A representative structure of each peak is shown in Figure 5D. One can clearly see several states of UL11, which correspond to relatively compact, intermediate, and extended conformations, respectively.

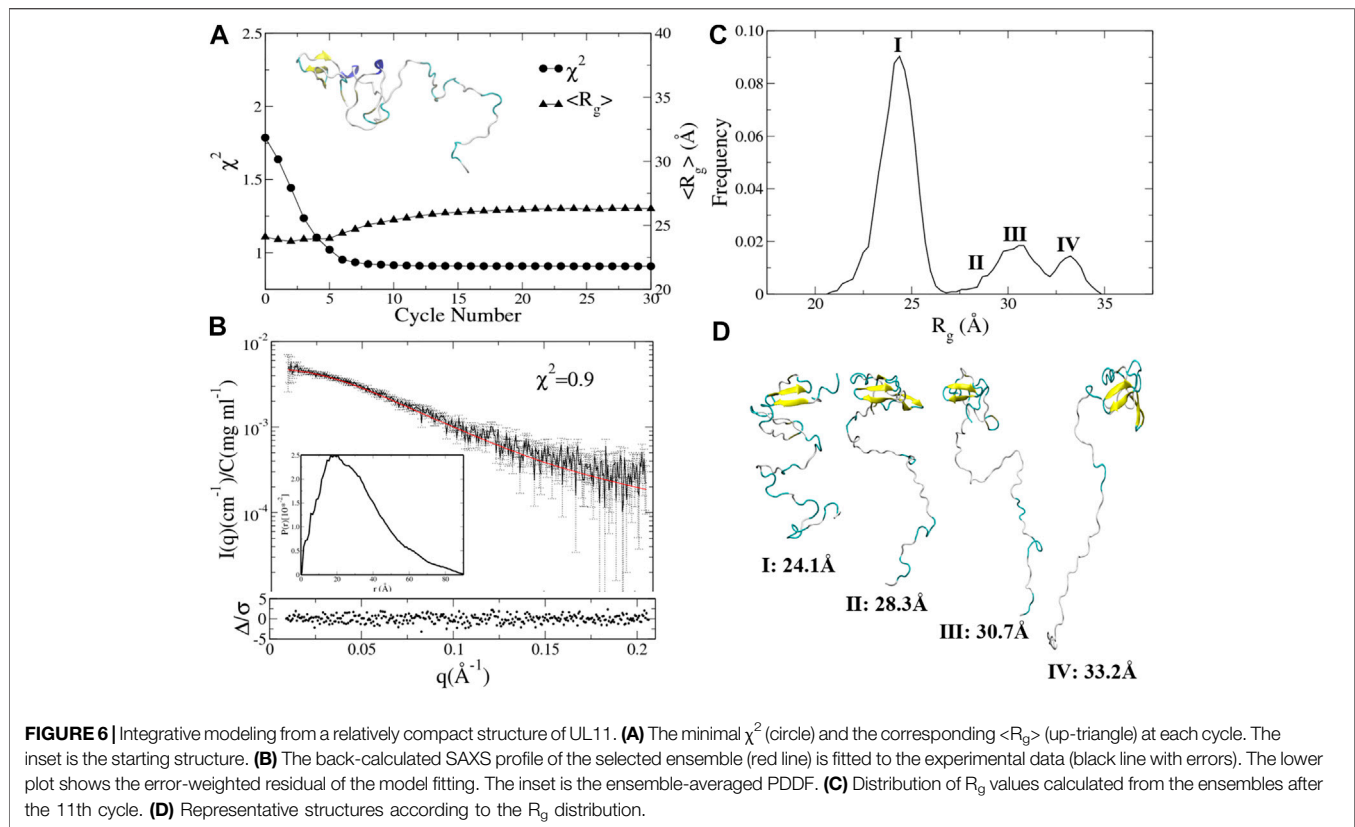
To test the reproducibility of the results, we also conducted the integrative modeling starting from a relatively compact structure of UL11 (inset in Figure 6A) taken from the 150 ns cMD simulation using A99SB/OPC. χ^2 and $\langle R_g \rangle$ of the ensemble are plotted against the cycle number (Figure 6A). χ^2 of the ensemble at cycle 0 is 1.8, and only after seven cycles, it converges to 0.9 (Figure 6A, circle). $\langle R_g \rangle$ of the ensemble at cycle 0 is 23.6 Å, and it converges to 25.8 Å after 11th cycles (Figure 6A, up-triangle). We plotted the calculated SAXS profile of the ensemble at the 12th cycle and its error-weighted residual (Figure 6B). The residual difference plot between the experimental and the computed $I(q)$ is flat, which indicates that the results fit with the data. The normalized ensemble-averaged PDDF is in agreement with the experimental curve (Figure 3B). The R_g distribution of all the ensembles after the 12th cycle also indicates a major peak around 24.1 Å, a minor one between 27.5 and 30.0 Å, and two more peaks with the R_g values larger than 30.0 Å (Figure 6C). The representative structures of the peaks (Figure 6D) correspond to states of UL11 from the relatively compact, the intermediate, and to the extended conformations. It has been found that the two independent integrative models of UL11 starting from the different structures show fairly consistent results.

It is worth noting that the total time scale of the integrative modeling is only 144 ns, but it can achieve a more efficient sampling and better convergence than the 150 ns aMD simulations (Figure 4).

In a previous work (Metrick et al., 2020), the authors ran RANCH, an internal program of EOM2, to generate a coarse-grained structural pool using a simple exclusion energy term. Then EOM was applied to the pool to pick an ensemble by fitting the SAXS data. The ensemble also included states from compact to extended. Our results of integrative modeling support their study. However, our ensembles consist of atomic models generated by fine Amber force field and explicit water model, which should be physically more reasonable than those generated by RANCH. However, more experimental data would be needed to further validate these models.

CONCLUSION

This work integrates an enhanced sampling method and experimental SAXS data to study IDPs. In our strategy, we first need to choose a combination of the force field and water model, such as A99SB/OPC, that is suitable for simulating IDPs, and then an enhanced sampling technique like aMD is taken. After that, integrative modeling is conducted based on iterative multiple independent simulations. Experimental data like SAXS are used to design a scoring function for screening conformations and thus guide the simulations toward an ensemble that fits the experimental data well. Therefore, we think this strategy of integrative modeling is well suited for investigating conformational ensembles of IDPs.



We have carried out the integrative modeling of UL11, which is important for efficient viral replication and tegument assembly. To the best of our knowledge, the understanding of its biochemical structure and mechanism is still limited, except for some coarse-grained structural information (Metrick et al., 2020). In this work, we have predicted an ensemble of atomic structures, which includes both the relatively compact and extended conformations of UL11. This ensemble is in agreement with the available experimental data and may provide information on the functional mechanism of UL11. It has been said that UL11 undergoes LLPS *in vitro* (Metrick et al., 2020). Our study on the monomer and the integrative modeling strategy may be helpful for future research on LLPS.

There are various tools for integrative modeling (Bonomi et al., 2017; Orioli et al., 2020), which use either the refining-while-sampling or the screening-after-sampling strategy. A refining-while-sampling method is efficient, but one needs to modify complicated simulation code to add an energy term for experimental restraints. In a screening-after-sampling method, although there is no need to change the simulation code, the postprocessing reweighting procedure would rely on adequately sampling conformations of the biomolecule, which is, however, a nontrivial issue for IDPs. Our method can be regarded as an iterative screening-after-sampling strategy, so we do not change the MD code. However, the sampling is still efficient because it is guided by the experimental data.

Our integrative modeling method has some other characteristics. The first is that the iterative multiple independent simulations are very suitable for parallel computing. In this work, 24 independent simulations are run

simultaneously, but one can use more CPU/GPU if they are available. The second is the high adaptability. Any sampling methods and ensemble optimization methods can be easily implemented with minor modifications to the scripts. Last but not least, many experimental data may be integrated simultaneously as long as a proper scoring function is designed. One of the future improvements is to input multiple initial models at the beginning of the integrative modeling in order to sample the conformations of IDPs as adequately as possible.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZZ and CD designed the study. CD collected data and carried out the calculation. SW did the structure prediction. ZZ and CD wrote the manuscript.

FUNDING

This work is supported by the National Natural Science Foundation of China (91953101, 21573205), the Strategic

Priority Research Program of the Chinese Academy of Science (XDB37040202), the Hefei National Science Center Pilot Project Funds, and the New Concept

Medical Research Fund of USTC. The Supercomputing Center of USTC provides computer resources for this project.

REFERENCES

- Baird, N. L., Starkey, J. L., Hughes, D. J., and Wills, J. W. (2010). Myristylation and palmitoylation of HSV-1 UL11 are not essential for its function. *Virology* 397 (1), 80–88. doi:10.1016/j.virol.2009.10.046
- Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129 (17), 5656–5664. doi:10.1021/ja069124n
- Bernadó, P., and Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8 (1), 151–167. doi:10.1039/c1mb05275f
- Bhowmick, A., Brookes, D. H., Yost, S. R., Dyson, H. J., Forman-Kay, J. D., Gunter, D., et al. (2016). Finding our way in the dark proteome. *J. Am. Chem. Soc.* 138 (31), 9730–9742. doi:10.1021/jacs.6b06543
- Björling, A., Niebling, S., Marcellini, M., van der Spoel, D., and Westenhoff, S. (2015). Deciphering solution scattering data with experimentally guided molecular dynamics simulations. *J. Chem. Theor. Comput.* 11 (2), 780–787. doi:10.1021/ct5009735
- Bonomi, M., Heller, G. T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* 42, 106–116. doi:10.1016/j.sbi.2016.12.004
- Bottaro, S., Bengtson, T., and Lindorff-Larsen, K. (2020). Integrating molecular simulation and experimental data: a bayesian/maximum entropy reweighting approach. *Methods Mol. Biol.* 2112, 219–240. doi:10.1007/978-1-0716-0270-6_15
- Bowzard, J. B., Visalli, R. J., Wilson, C. B., Loomis, J. S., Callahan, E. M., Courtney, R. J., et al. (2000). Membrane targeting properties of a herpesvirus tegument protein-retrovirus Gag chimera. *J. Virol.* 74 (18), 8692–8699. doi:10.1128/jvi.74.18.8692-8699.2000
- Braitbard, M., Schneidman-Duhovny, D., and Kalisman, N. (2019). Integrative structure modeling: overview and assessment. *Annu. Rev. Biochem.* 88, 113–135. doi:10.1146/annurev-biochem-013118-111429
- Burger, V., Gurry, T., and Stultz, C. (2014). Intrinsically disordered proteins: where computation meets experiment. *Polymers* 6 (10), 2684–2719. doi:10.3390/polym6102684
- Carter, L., Kim, S. J., Schneidman-Duhovny, D., Stöhr, J., Poncet-Montange, G., Weiss, T. M., et al. (2015). Prion protein-antibody complexes characterized by chromatography-coupled small-angle X-ray scattering. *Biophys. J.* 109 (4), 793–805. doi:10.1016/j.bpj.2015.06.065
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290
- Cheng, P., Peng, J., and Zhang, Z. (2017). SAXS-oriented ensemble refinement of flexible biomolecules. *Biophys. J.* 112 (7), 1295–1301. doi:10.1016/j.bpj.2017.02.024
- Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K., and Uversky, V. N. (2006). Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* 45 (35), 10448–10460. doi:10.1021/bi060981d
- Colak, R., Kim, T., Michaut, M., Sun, M., Irimia, M., Bellay, J., et al. (2013). Distinct types of disorder in the human proteome: functional implications for alternative splicing. *Plos Comput. Biol.* 9 (4), e1003030. doi:10.1371/journal.pcbi.1003030
- Curtis, J. E., Raghunandan, S., Nanda, H., and Krueger, S. (2012). SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comp. Phys. Commun.* 183 (2), 382–389. doi:10.1016/j.cpc.2011.09.010
- Du, Z., and Uversky, V. N. (2017). A comprehensive survey of the roles of highly disordered proteins in type 2 diabetes. *Int. J. Mol. Sci.* 18 (10), 2010. doi:10.3390/ijms18102010
- Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272 (20), 5129–5148. doi:10.1111/j.1742-4658.2005.04948.x
- Dunker, A. K., and Oldfield, C. J. (2015). Back to the future: nuclear magnetic resonance and bioinformatics studies on intrinsically disordered proteins. *Adv. Exp. Med. Biol.* 870, 1–34. doi:10.1007/978-3-319-20164-1_1
- Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., et al. (2017). : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 50 (Pt 4), 1212–1225. doi:10.1107/S1600576717007786
- Granata, D., Baftizadeh, F., Habchi, J., Galvagnion, C., De Simone, A., Camilloni, C., et al. (2015). The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* 5, 15449. doi:10.1038/srep15449
- Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120 (24), 11919–11929. doi:10.1063/1.1755656
- Harada, R., and Kitao, A. (2015). Nontargeted parallel cascade selection molecular dynamics for enhancing the conformational sampling of proteins. *J. Chem. Theor. Comput.* 11 (11), 5493–5502. doi:10.1021/acs.jctc.5b00723
- Harada, R., and Kitao, A. (2013). Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* 139 (3), 035103. doi:10.1063/1.4813023
- Harada, R., and Shigeta, Y. (2018). Selection rules on initial structures in parallel cascade selection molecular dynamics affect conformational sampling efficiency. *J. Mol. Graph. Model.* 85, 153–159. doi:10.1016/j.jmkgm.2018.08.014
- Hu, Y., Hong, W., Shi, Y., and Liu, H. (2012). Temperature-accelerated sampling and amplified collective motion with adiabatic reweighting to obtain canonical distributions and ensemble averages. *J. Chem. Theor. Comput.* 8 (10), 3777–3792. doi:10.1021/ct300061g
- Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., and Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323 (3), 573–584. doi:10.1016/s0022-2836(02)00969-5
- Izadi, S., Anandakrishnan, R., and Onufriev, A. V. (2014). Building water models: a different approach. *J. Phys. Chem. Lett.* 5 (21), 3863–3871. doi:10.1021/jz501780a
- Izadi, S., and Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* 145 (7), 074501. doi:10.1063/1.4960175
- Kulkarni, P., and Uversky, V. N. (2019). Intrinsically disordered proteins in chronic diseases. *Biomolecules* 9 (4), 147. doi:10.3390/biom9040147
- Kuzmanic, A., Pritchard, R. B., Hansen, D. F., and Gervasio, F. L. (2019). Importance of the force field choice in capturing functionally relevant dynamics in the von willebrand factor. *J. Phys. Chem. Lett.* 10 (8), 1928–1934. doi:10.1021/acs.jpcclett.9b00517
- LeBlanc, S. J., Kulkarni, P., and Wening, K. R. (2018). Single molecule FRET: a powerful tool to study intrinsically disordered proteins. *Biomolecules* 8 (4), 140. doi:10.3390/biom8040140
- MacLean, C. A., Clark, B., and McGeoch, D. J. (1989). Gene UL11 of herpes simplex virus type 1 encodes a virion protein which is myristylated. *J. Gen. Virol.* 70 (Pt 12), 3147–3157. doi:10.1099/0022-1317-70-12-3147
- MacLean, C. A., Dolan, A., Jamieson, F. E., and McGeoch, D. J. (1992). The myristylated virion proteins of herpes simplex virus type 1: investigation of their role in the virus life cycle. *J. Gen. Virol.* 73 (Pt 3), 539–547. doi:10.1099/0022-1317-73-3-539
- McLauchlan, J., and Rixon, F. J. (1992). Characterization of enveloped tegument structures (L particles) produced by alphaherpesviruses: integrity of the tegument does not depend on the presence of capsid or envelope. *J. Gen. Virol.* 73 (Pt 2), 269–276. doi:10.1099/0022-1317-73-2-269
- Mertens, H. D., and Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* 172 (1), 128–141. doi:10.1016/j.jsb.2010.06.012
- Metrick, C. M., Koenigsberg, A. L., and Heldwein, E. E. (2020). Conserved outer tegument component UL11 from herpes simplex virus 1 is an intrinsically disordered, RNA-binding protein. *mBio* 11 (3). doi:10.1128/mBio.00810-20

- Müller, C. W., Schlauderer, G. J., Reinstein, J., and Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 4 (2), 147–156. doi:10.1016/s0969-2126(96)00018-4
- Orioli, S., Larsen, A. H., Bottaro, S., and Lindorff-Larsen, K. (2020). How to learn from inconsistencies: integrating molecular simulations with experimental data. *Prog. Mol. Biol. Transl. Sci.* 170, 123–176. doi:10.1016/bs.pmbts.2019.12.006
- Owen, D. J., Crump, C. M., and Graham, S. C. (2015). Tegument assembly and secondary envelopment of alphaherpesviruses. *Viruses* 7 (9), 5084–5114. doi:10.3390/v7092861
- Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65(6), 1409–1419. doi:10.1080/00268978800101881
- Potoyan, D. A., and Papoian, G. A. (2011). Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* 133 (19), 7405–7415. doi:10.1021/ja1111964
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23(3), 327–341. doi:10.1016/0021-9991(77)90098-5
- Saltzberg, D., Greenberg, C. H., Viswanath, S., Chemmama, I., Webb, B., Pellarin, R., et al. (2019). Modeling biological complexes using integrative modeling platform. *Methods Mol. Biol.* 2022, 353–377. doi:10.1007/978-1-4939-9608-7_15
- Semenyuk, A. V., and Svergun, D. I. (1991). Gnom—a program package for small-angle scattering data processing. *J. Appl. Cryst.* 24, 537–540. doi:10.1107/S002188989100081x
- Shabane, P. S., Izadi, S., and Onufriev, A. V. (2019). General purpose water model can improve atomistic simulations of intrinsically disordered proteins. *J. Chem. Theor. Comput.* 15 (4), 2620–2634. doi:10.1021/acs.jctc.8b01123
- Shkurti, A., Styliari, I. D., Balasubramanian, V., Bethune, I., Pedebos, C., Jha, S., et al. (2019). CoCo-MD: a simple and effective method for the enhanced sampling of conformational space. *J. Chem. Theor. Comput.* 15 (4), 2587–2596. doi:10.1021/acs.jctc.8b00657
- Svergun, D., Barberato, C., and Koch, M. H. J. (1995). CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28, 768–773. doi:10.1107/s0021889895007047
- Trewhella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., et al. (2017). 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol.* 73 (Pt 9), 710–728. doi:10.1107/S2059798317011597
- Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* 2 (Pt 2), 207–217. doi:10.1107/S205225251500202X
- Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18 (5), 343–384. doi:10.1002/jmr.747
- Uversky, V. N. (2009). Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front. Biosci.* 14, 5188–5238. doi:10.2741/3594
- Uversky, V. N. (2014). The triple power of D³: protein intrinsic disorder in degenerative diseases. *Front. Biosci.* 19, 181–258. doi:10.2741/4204
- Yuan, Y., Zhu, Q., Song, R., Ma, J., and Dong, H. (2020). A two-ended data-driven accelerated sampling method for exploring the transition pathways between two known states of protein. *J. Chem. Theor. Comput.* 16 (7), 4631–4640. doi:10.1021/acs.jctc.9b01184
- Zhang, J., and Gong, H. (2020). Frontier expansion sampling: a method to accelerate conformational search by identifying novel seed structures for restart. *J. Chem. Theor. Comput.* 16 (8), 4813–4821. doi:10.1021/acs.jctc.0c00064
- Zhang, Y.-H., Peng, J.-H., and Zhang, Z.-Y. (2015). Structural modeling of proteins by integrating small-angle x-ray scattering data. *Chin. Phys. B* 24 (12), 126101. doi:10.1088/1674-1056/24/12/126101
- Zhang, Z., Shi, Y., and Liu, H. (2003). Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* 84 (6), 3583–3593. doi:10.1016/S0006-3495(03)75090-5
- Zheng, W., and Tekpinar, M. (2011). Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* 101 (12), 2981–2991. doi:10.1016/j.bpj.2011.11.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ding, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.