

# UCSF

## UC San Francisco Previously Published Works

### Title

OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates.

### Permalink

<https://escholarship.org/uc/item/8gr345kr>

### Journal

PLoS Computational Biology, 19(9)

### Authors

Liu, Shuming

Wang, Cong

Latham, Andrew

et al.

### Publication Date

2023-09-01

### DOI

10.1371/journal.pcbi.1011442

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## RESEARCH ARTICLE

## OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates

Shuming Liu<sup>1</sup>, Cong Wang<sup>1</sup>, Andrew P. Latham<sup>1,2</sup>, Xinqiang Ding<sup>1</sup>, Bin Zhang<sup>1\*</sup>

**1** Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, California, United States of America

\* [binz@mit.edu](mailto:binz@mit.edu)

## OPEN ACCESS

**Citation:** Liu S, Wang C, Latham AP, Ding X, Zhang B (2023) OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates. *PLoS Comput Biol* 19(9): e1011442. <https://doi.org/10.1371/journal.pcbi.1011442>

**Editor:** Guanghong Wei, Fudan University, CHINA

**Received:** April 21, 2023

**Accepted:** August 19, 2023

**Published:** September 11, 2023

**Copyright:** © 2023 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** OpenABC source code is available at <https://github.com/ZhangGroup-MITChemistry/OpenABC>.

**Funding:** This work was supported by the National Institutes of Health (Grant R35GM133580) and the National Science Foundation (Grant MCB-2042362) to BZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Biomolecular condensates are important structures in various cellular processes but are challenging to study using traditional experimental techniques. In silico simulations with residue-level coarse-grained models strike a balance between computational efficiency and chemical accuracy. They could offer valuable insights by connecting the emergent properties of these complex systems with molecular sequences. However, existing coarse-grained models often lack easy-to-follow tutorials and are implemented in software that is not optimal for condensate simulations. To address these issues, we introduce OpenABC, a software package that greatly simplifies the setup and execution of coarse-grained condensate simulations with multiple force fields using Python scripting. OpenABC seamlessly integrates with the OpenMM molecular dynamics engine, enabling efficient simulations with performance on a single GPU that rivals the speed achieved by hundreds of CPUs. We also provide tools that convert coarse-grained configurations to all-atom structures for atomistic simulations. We anticipate that OpenABC will significantly facilitate the adoption of in silico simulations by a broader community to investigate the structural and dynamical properties of condensates.

## Author summary

Biomolecular condensates are essential cellular structures that underpin various cellular processes, including RNA splicing, ribosomal RNA processing, and stress response, etc. These membrane-less organelles exhibit liquid-like behavior, enriched with disordered proteins and RNA that can dynamically exchange with the surrounding media. Computational modeling aids in understanding these condensates' structural and dynamic aspects at high resolution, connecting molecular sequences with condensate emergent properties. Both coarse-grained and atomistic simulations can help offer insights into the nature of the molecular interactions that drive phase separation, the microenvironment of the condensates, and their dynamical relaxation. However, software tools for such simulations

are limited. We introduce OpenABC, a flexible software package for GPU-accelerated simulations of biomolecular condensates. It simplifies setup using Python scripts, integrates with the molecular dynamics engine, OpenMM, for efficient simulations, and facilitates transitioning between coarse-grained and atomistic representations. OpenABC aims to empower researchers by easing simulation setups, aiding force field comparison, and advancing biomolecular condensate research.

## Introduction

Biomolecular condensates underlie the organization of many cellular processes, such as speckles for RNA splicing, nucleoli for ribosomal RNA processes, and P granule for stress response, etc. [1–14]. They are also termed membrane-less organelles due to the lack of enclosure and exhibit liquid-like properties. Intrinsically disordered proteins (IDPs) and RNA molecules are enriched inside the condensates [3, 4, 9, 11]. These molecules promote promiscuous, multivalent interactions, leading to spontaneous phase transition and condensate formation [15]. The nature of the molecular interactions that drive phase separation, the microenvironment of the condensates, and their dynamical relaxation, are under active investigation.

Computational modeling can prove invaluable for studying biomolecular condensates by providing detailed structural and dynamic characterizations [16–42]. Particle-based coarse-grained modeling approaches are promising since their computational efficiency enables long-timescale simulations to promote large-scale reorganization for structural relaxation [43–46]. Such simulations may predict condensate physical properties *de novo*, elucidating the connection between molecular sequences and emergent properties [18, 47]. However, the reduced resolution of these coarse-grained models could be insufficient to describe the complex microenvironment of the condensate interior [48–50]. Atomistic simulations with explicit representation of solvent molecules and counter ions can be necessary to further characterize physicochemical interactions that produce the selective partition of small molecules within condensates [49–53]. Combining the two modeling approaches at different resolutions could be particularly powerful since they enable long-timescale simulations for structural relaxation while preserving the fine-resolution details.

While many computational models and force fields have been introduced for simulations of IDPs and biomolecules, software engineering has yet to catch up. There is an urgent need to build user-friendly tools to set up and execute condensate simulations. Preparing biomolecular simulations can be rather involved. Even creating initial configurations for such simulations is often non-trivial. Much-dedicated software has been introduced to prepare atomistic simulations [54–57], and existing molecular dynamics (MD) simulation packages are highly optimized for computational efficiency [54, 57–59]. However, existing tools are not immediately transferable for setting up coarse-grained condensate simulations. Furthermore, coarse-grained force fields are often implemented into disparate simulation engines not necessarily best suited for condensate simulations, hindering cross-validation and the unleashing of full modeling potential. Further software development can significantly reduce the entry barrier for *in silico* studies, allowing more researchers to experience the usefulness of computational modeling. They could facilitate comparing and benchmarking various force fields, driving continuous improvement.

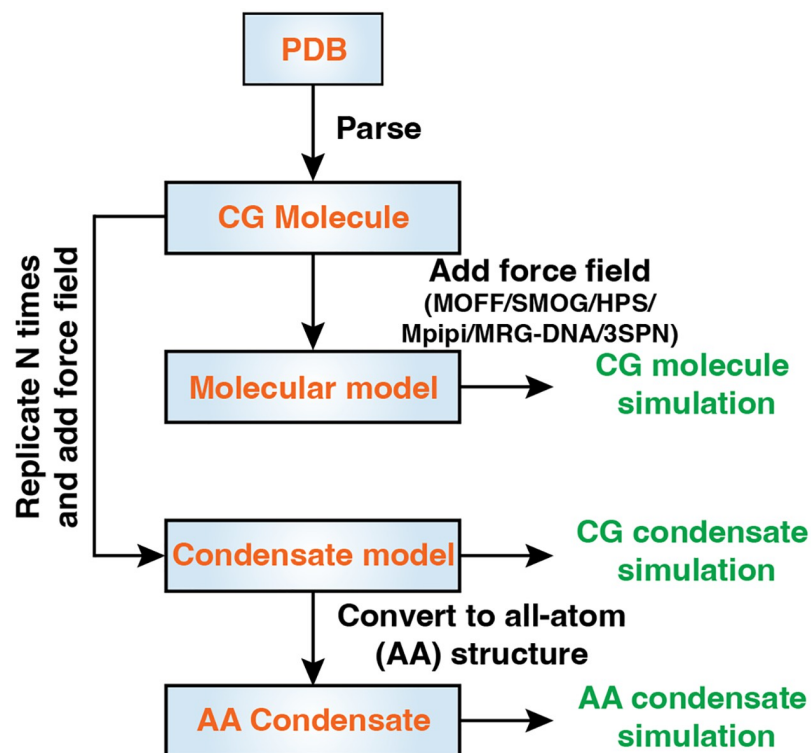
We introduce a software package termed OpenABC for “OpenMM GPU-Accelerated simulations of Biomolecular Condensates”. The package is flexible and implements multiple popular coarse-grained force fields for simulating proteins and nucleic acids. It dramatically

simplifies the simulation setup: only a few lines of Python scripts are needed to carry out condensate simulations starting from initial configurations of a single protein or DNA. The package is integrated with OpenMM, a GPU-accelerated MD engine [60], enabling efficient simulations with advanced sampling techniques. Finally, we include tools that convert coarse-grained configurations to atomistic structures for further condensate modeling with all-atom force fields. Tutorials in Jupyter Notebooks are provided to demonstrate the various capabilities. We anticipate that OpenABC will greatly facilitate the application of existing computer models for simulating biomolecular condensates and the continued force field development.

## Results

### Flexible force field selections for Biomolecular simulations

OpenABC implements several existing force fields for coarse-grained (CG) modeling of protein, RNA, and DNA molecules (Fig 1). Single-bead per amino acid force field for proteins include the hydrophobicity scale (HPS) models [20, 28], the Mpipi force field [29], a generalized structure-based model [41, 61, 62], and the maximum entropy optimized force field (MOFF) [63]. HPS models define interactions between different pairs of amino acids based on various hydrophobicity scales [20, 28]. Recent studies have attempted to improve the accuracy of HPS models with systematic optimizations of the hydrophobicity scale to match experimental



**Fig 1. OpenABC facilitates coarse-grained and atomistic simulations of biomolecular condensates with multiple force fields.** The diagram illustrates the workflow and various functionalities of OpenABC. To set up condensate simulations, the users must provide a configuration file in the PDB format for the molecule of interest. OpenABC parses topological and structural information from the PDB file to build a molecule object. Specifying force field options allows direct simulations of individual molecules. On the other hand, the molecule object can be replicated for condensate simulations. In addition, OpenABC allows the conversion of CG configurations to atomistic structures for simulations with all-atom force fields.

<https://doi.org/10.1371/journal.pcbi.1011442.g001>

observations of IDP monomers [34, 36]. They have been used to study the phase behaviors of numerous proteins [64–66], revealing the contribution of charge distribution patterns, cation- $\pi$  interactions, and the balance between hydrophobic and electrostatic interactions [65, 67] to the stability of condensates.

The Mpipi force field was parameterized using data from all-atom simulations and bioinformatics analysis with a careful calibration of  $\pi$ - $\pi$  and  $\pi$ -cation interactions [29]. These interactions play significant roles in the formation of biomolecular condensates. The force field was shown to accurately capture the radius of gyration and critical temperatures of diverse protein sequences.

SMOG was originally introduced for studying folded proteins using interaction potentials derived from initial input configurations. We generalized the model to describe proteins with disordered domains and leveraged the Miyazawa-Jernigan statistical potential [68] for protein-protein interactions [41, 62].

MOFF was parameterized with the maximum entropy algorithm [69, 70] and the protein folding energy landscape theory [71] to provide consistent descriptions of both folded and disordered proteins [63, 72–74]. It was shown to reproduce the radius of gyration for a collection of proteins, including both ordered and disordered proteins [63, 75]. The balanced interactions among amino acids have proven beneficial in describing complex contacts among phase-separating proteins, including those with both ordered and disordered domains [47, 63, 74].

In addition to protein models, we implemented several force fields for nucleic acids. For example, the molecular renormalization group coarse-graining (MRG-CG) DNA model was initially introduced for simulations with explicit ions to reproduce the salt-dependent DNA persistence length [76]. We adopted it for implicit ion modeling with the Debye-Hückel approximation for electrostatic interactions. We rescaled the strength of bonded interactions to ensure the accuracy of the implicit-ion model in reproducing DNA persistence length at the physiological salt concentration [47]. We further incorporated the DNA model 3SPN [77, 78] into OpenABC for studying sequence specific properties. Unlike MRG-CG DNA that only uses one bead to represent each nucleotide, 3SPN adopts three beads to differentiate sugar, base, and phosphate. Finally, the Mpipi force field can be used to simulate RNA molecules.

While one can in principle combine different force fields for simulating complex systems with both proteins and nucleic acids, care needs to be taken when modeling cross interactions. Previous studies have carried out systematic validations of protein-DNA and protein-RNA interactions and we implemented them into OpenABC, with combinations that include SMOG-3SPN [37, 41, 62, 78–82], MOFF-MRG-CG DNA [47], and Mpipi Protein-RNA [29]. These combinations account for both excluded volume effect and electrostatic interactions. Detailed expressions of all the force field potentials are provided in the [S1 Text—Force Field Definitions](#), with the parameters provided in Tables A-I in [S1 Text](#).

## Simplified setup of condensate simulations

OpenABC leverages the MD simulation engine, OpenMM [60], to offer simulation setup with Python scripting, thus dramatically simplifying the workflow. The software treats each molecule as an object and appends such objects into a container-like class. This class allows the incorporation of various force field options and integration schemes for MD simulations.

An illustration of the typical workflow for condensate simulations is provided in [Fig 1](#). OpenABC first parses a configuration file in the PDB format supplied by users to create a molecule object. The object contains topological and structural information extracted from the input file. Upon introducing interactions defined in various force fields, the molecule object can be used to simulate individual biomolecules. On the other hand, the molecule object can

```

1  from openabc.forcefields.parsers import MOFFParser
2  from openabc.forcefields import MOFFMRGModel
3  from openabc.utils.insert import insert_molecules
4  import simtk.openmm as mm
5  import simtk.openmm.app as app
6  import simtk.unit as unit
7  import os
8
9  # Parse structural and topological information
10 protein = MOFFParser.from_atomistic_pdb('all_atom.pdb', 'Calpha.pdb')
11
12 # Build initial condensate configuration with N = 100 proteins
13 N = 100
14 a, b, c = 100, 100, 100 # box sizes
15 insert_molecules('Calpha.pdb', 'start.pdb', n_mol=N, box=[a, b, c])
16
17 # Create molecule container and OpenMM system
18 condensate = MOFFMRGModel()
19 for i in range(N):
20     condensate.append_mol(protein)
21 top = app.PDBFile('start.pdb').getTopology()
22 condensate.create_system(top, box_a=a, box_b=b, box_c=c)
23 condensate.add_all_default_forces()
24
25 # Initiate MD simulation
26 temperature = 300*unit.kelvin
27 friction_coeff = 1/unit.picosecond
28 timestep = 10*unit.femtosecond
29 integrator = mm.LangevinMiddleIntegrator(temperature, friction_coeff, timestep)
30 init_coord = app.PDBFile('start.pdb').getPositions()
31 condensate.set_simulation(integrator, platform_name='CUDA', init_coord=init_coord)
32 condensate.simulation.minimizeEnergy()
33 condensate.add_reporters(report_interval=10000, output_dcd='output.dcd')
34 condensate.simulation.context.setVelocitiesToTemperature(temperature)
35 condensate.simulation.step(1000000)

```

**Fig 2. OpenABC simplifies simulation setup with Python scripting.** The example code includes all the steps necessary for setting up and performing MD simulations of a protein condensate with MOFF and default settings in a cubic box of length 100 nm. The ten lines included in the highlight box correspond to the creation of the condensate system by parsing topological information from an initial PDB file, building a configuration file by inserting molecules into a box and incorporating the molecular objects, *protein*, into a container class, *condensate*, with appropriate force fields. The rest of the code includes standard simulation setups generic to OpenMM. We chose the Langevin middle integrator to perform simulations at 300 K with a friction coefficient of  $1 \text{ ps}^{-1}$  and a timestep of 10 fs.

<https://doi.org/10.1371/journal.pcbi.1011442.g002>

also be replicated  $N$  times for condensate simulations consisting of  $N$  molecules. As demonstrated in an example code in Fig 2, setting up an entire MD simulation of a protein condensate with default parameters only requires about 20 lines of code.

To enhance conformational sampling of individual molecules and condensates, we provide an implementation of the temperature replica exchange algorithm [83] with PyTorch [84] as



part of the package (see [S1 Text—Implementation of the temperature replica exchange algorithm](#) for details). Furthermore, we introduce utility functions to reconstruct atomistic structures from coarse-grained protein configurations with only  $\alpha$  carbons. This functionality relies on the software “reconstruct atomic model from reduced representation (REMO)” [85] and can facilitate downstream all-atom simulations. More tutorials in Jupyter Notebook format are available online at the OpenABC GitHub repository.

### Efficient simulations with GPU-enabled MD engine

A significant advantage of integrating with OpenMM comes from its native support of GPU acceleration. Simulating implicit solvent coarse-grained condensates on GPUs can be particularly beneficial due to the inhomogeneous distribution of particles arising from implicit solvation [78]. CPU parallelization, which often relies on the spatially-based, domain decomposition strategy, is often less effective because the inhomogeneity in particle density between the condensate and dilute phases produces an imbalanced workload between CPUs.

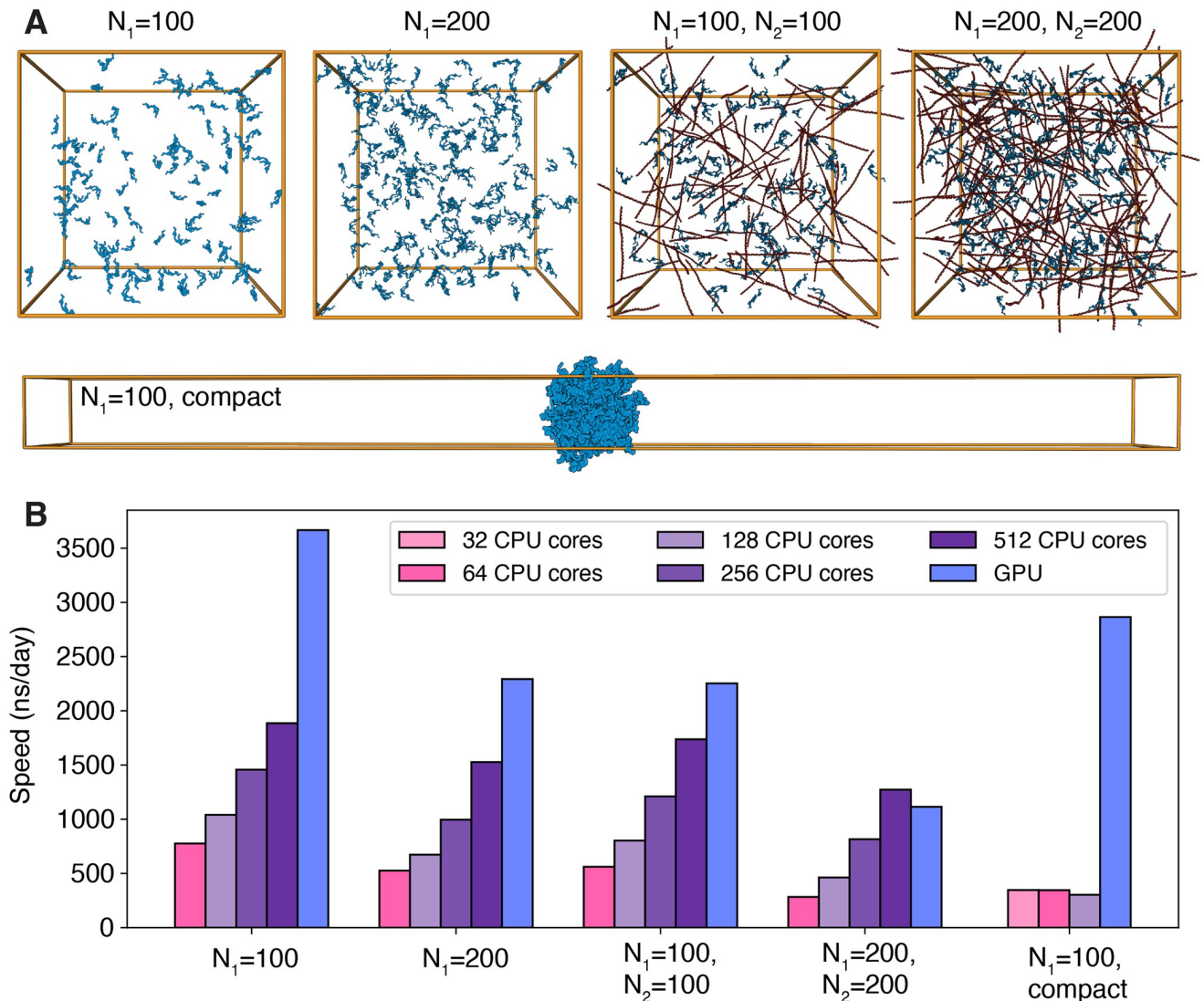
To demonstrate the efficiency of GPU-enabled simulations, we studied five independent condensate systems. The first four systems consist of  $N_1$  HP1 $\alpha$  dimers and  $N_2$  200-bp-long dsDNA randomly distributed in a cubic box of length 200 nm with periodic boundary conditions. In the fifth system, 100 HP1 $\alpha$  dimers in a compact configuration were placed at the center of an elongated box of size  $25 \times 25 \times 400$  nm<sup>3</sup> (Fig 3A). This rectangular setup is typical for the so-called slab simulations to produce a dilute and dense interface along the  $z$ -axis for computing co-existence curves and phase diagrams [20, 86, 87]. MOFF and MRG-CG force fields were used to describe the interactions among coarse-grained particles. We simulated each system for one million steps using the Langevin middle integrator [88] to control the temperature at 300 K, with a friction coefficient of  $1 \text{ ps}^{-1}$  and a time step of 10 fs. For comparison, we simulated the same systems with a closely related integrator using GROMACS, a leading MD engine with state-of-the-art performance on CPUs [54, 57]. More simulation details are provided in the [S1 Text—Benchmarking the performance of condensate simulations](#).

As shown in Fig 3B, OpenMM single GPU performance matches GROMACS with hundreds of CPUs in the first four systems. While GROMACS achieved nearly linear scaling for the first four systems, introducing more CPUs did not lead to any significant speedup in the last system with a dense-dilute interface. As mentioned above, the presence of vacuum regions in slab simulations hinders the efficacy of domain decomposition. On the other hand, OpenMM is less sensitive to the simulation setup and retains superior performance.

The performance of GROMACS depends on our implementation of the CG force fields and may not reflect the theoretical upper limit of the software. In particular, our use of tabulated potentials for the Debye Hückel potential and domain decomposition for parallelization may significantly affect the simulation speed. While performance improvement is possible with additional software engineering, the advantage of CG simulations of condensates on GPUs remains given the differences shown in Fig 3B.

### Application: Validating force field implementations in OpenABC

Before applying the software for extensive simulations, we validated our implementations of various force fields with existing ones. We generated ten configurations for an HP1 $\alpha$  dimer with MOFF and GROMACS through an NVT simulation. As shown in Table J in [S1 Text](#), the potential energies evaluated using MOFF from OpenABC match those reported by GROMACS. Similar comparisons with a protein-DNA complex produce nearly identical energy values as well, as shown in Table K in [S1 Text](#). The protein-DNA complex is formed by an HP1 $\alpha$  dimer with a 200-bp-long dsDNA, and MOFF and MRG-CG DNA were used to



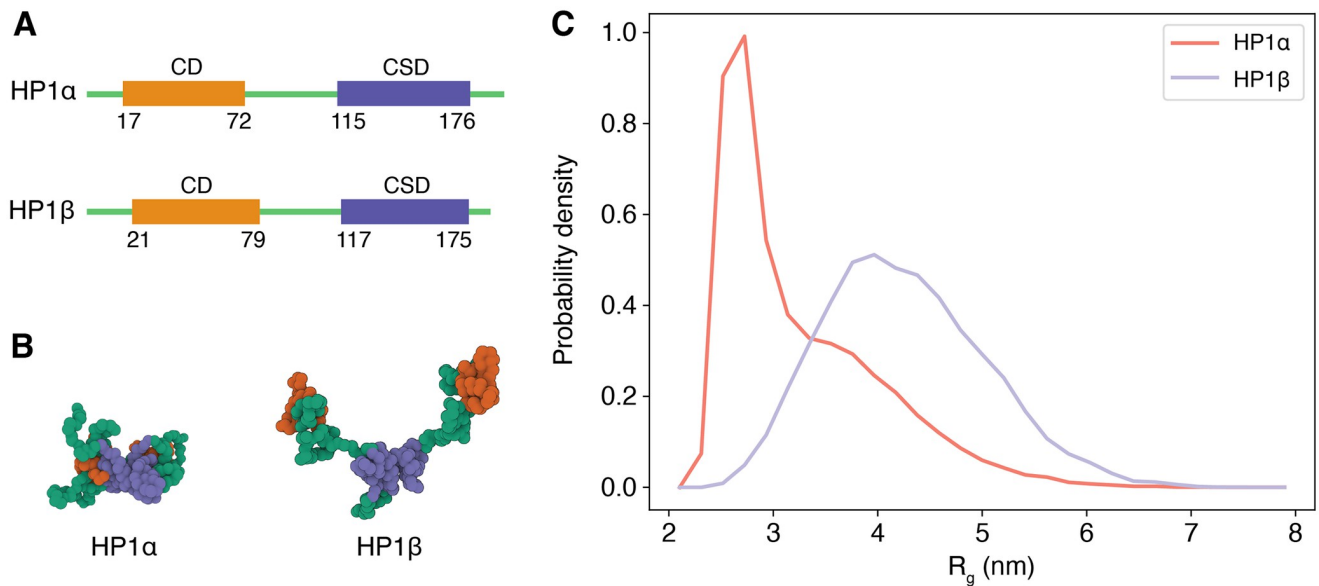
**Fig 3. OpenABC integrates with OpenMM for GPU-accelerated MD simulations.** (A) Snapshots of the five systems used to benchmark simulation performance. The systems consist of  $N_1$  HP1 $\alpha$  dimers (blue) and  $N_2$  200-bp-long dsDNA (red,  $N_2 = 0$  if not specified). The first four systems adopt homogeneous density distributions in cubic boxes of length 200 nm, while the last exhibits a dense-dilute interface in an elongated box of size  $25 \times 25 \times 400 \text{ nm}^3$ . (B) The five data sets compare the performance of CPU simulations using GROMACS with single GPU simulations using OpenMM. The different colors indicate the number of CPUs in GROMACS simulations, as shown in the legends. The benchmarks were performed with Intel Xeon Gold 8260 CPUs and Nvidia Volta V100 GPUs.

<https://doi.org/10.1371/journal.pcbi.1011442.g003>

quantify their interactions. The minor differences between OpenMM and GROMACS energies are mainly caused by using tabulated functions for nonbonded interactions in GROMACS.

We further evaluated the potential energies defined by the HPS model on ten configurations of a disordered protein, DDX4, using both OpenMM and HOOMD-Blue [89]. As shown in Table L in S1 Text, the two sets of energies match exactly, supporting the correctness of our force field implementation. We also validated the Mpipi force field using interaction energies evaluated with OpenMM and LAMMPS [59] for a protein-RNA system, as shown in Table M in S1 Text.





**Fig 4. OpenABC produces consistent results with a previous study, resolving the structural differences between two HP1 homologs.** (A) Secondary structures of HP1 $\alpha$  and HP1 $\beta$  along sequences. (B) Representative structures for HP1 $\alpha$  and HP1 $\beta$  dimer rendered with Mol\* Viewer [90]. The radii of gyration ( $R_g$ ) for the two structures are 2.77 and 4.44 nm, respectively. We colored the chromodomain (CD) in orange, the chromoshadow domain (CSD) in blue, and the rest in green. (C) Probability density distributions of  $R_g$  for HP1 $\alpha$  (red) and HP1 $\beta$  dimer (blue).

<https://doi.org/10.1371/journal.pcbi.1011442.g004>

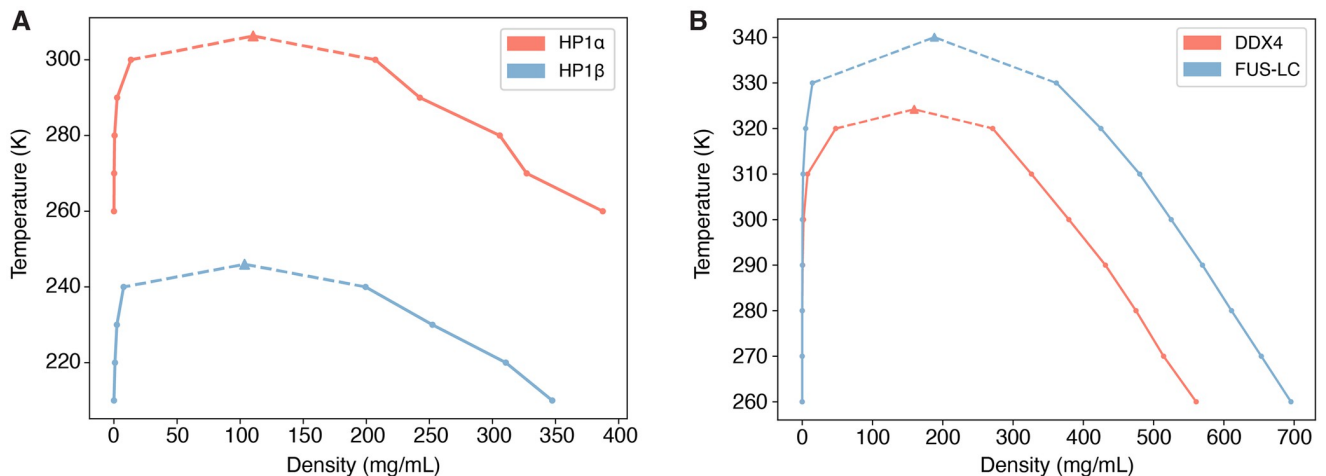
In addition to energy comparisons, we examined the conformational ensembles of HP1 $\alpha$  and HP1 $\beta$  dimers using MOFF with temperature replica exchange simulations [83]. Consistent with our previous study [63], the force field succeeds in resolving the difference in their conformational distribution between the two homologs (Fig 4). The radii of gyration for the two dimers at 300 K are  $3.33 \pm 0.19$  nm, and  $4.27 \pm 0.09$  nm, respectively. These values match the previously reported values computed using GROMACS quantitatively, reproducing experimental trends. Therefore, OpenABC produces consistent results with other software despite differences in integration schemes.

Using the MRG-DNA model, we computed the persistence length of a 200-bp-long DNA segment. The estimated value at a monovalent salt concentration of 100 mM,  $48.83 \pm 2.71$  nm (see Fig A in S1 Text), is consistent with that reported in a previous study using simulations of the same model but with GROMACS [47]. Additional simulation details for estimating the persistence length are provided in the S1 Text—*Estimating the persistence length of MRG-DNA*.

### Application: Coarse-grained simulation of protein condensates

As additional evaluations of force field implementation and to demonstrate the usefulness of OpenABC, we performed slab simulations to determine the phase diagram of four proteins, which are known to form various biomolecular condensates inside the cell. For example, HP1 dimers are involved in chromatin compaction and regulation [91], while DDX4 and FUS are a primary constituent of nuage or germ granules [92] and cytoplasmic RNP granules [93], respectively. The simulations for HP1 $\alpha$  and HP1 $\beta$  were performed with MOFF, while those for FUS LC and DDX4 were modeled with the HPS model using the shifted Urry hydrophobicity scale [28].

The resulting phase diagrams are shown in Fig 5, with the concentrations listed in Tables N-P in S1 Text. The density profiles at different temperatures and the representative snapshots



**Fig 5. OpenABC produces phase diagrams that match previous results.** (A) Phase diagrams for HP1 $\alpha$  (red) and HP1 $\beta$  (blue) dimer condensates computed with MOFF. (B) Phase diagrams for DDX4 (red) and FUS LC (blue) computed with the HPS model parameterized using the Urry hydrophobicity scale. The dots in both plots denote the density values determined from slab simulations, and the triangles represent the critical point obtained from numerical fitting.

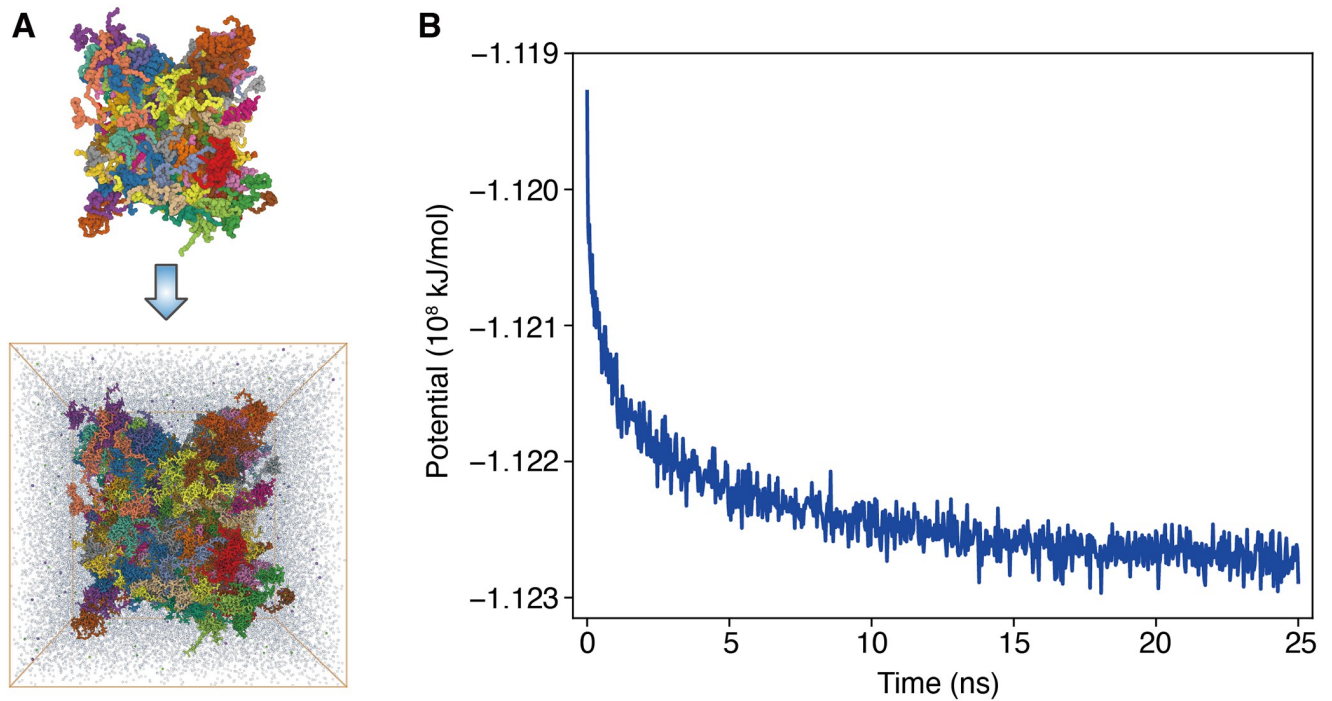
<https://doi.org/10.1371/journal.pcbi.1011442.g005>

at the lowest temperatures are shown in Figs B-C in [S1 Text](#). We fitted the computed phase diagrams with an analytical expression to determine the critical temperature  $T_c$  (see [Methods](#)). The critical temperatures are 306.30 K for HP1 $\alpha$  and 245.99 K for HP1 $\beta$ , consistent with previous results obtained with GROMACS simulations [63]. Similarly, the critical temperatures for DDX4 and FUS LC are 324.21 K and 340.04 K, respectively, matching values reported in a previous study that used the software HOOMD-Blue for simulations [28]. Thus, OpenABC produces statistically indistinguishable results on the phase behavior of protein condensates as in previous studies.

### Application: Atomistic simulation of protein condensates

While residue-level CG models are helpful for long timescale simulations, their limited resolution may prove insufficient to characterize specific properties of condensates, including the solvation environment [48], counter-ion distributions [49], and protein-ligand interactions [50]. Therefore, we implemented functionalities in OpenABC to convert equilibrated CG configurations to atomistic structures. Starting from these structures, well-established tools, such as CHARMM-GUI [55], GROMACS [54, 57], and AMBER [56], can be easily applied to set up explicit solvent simulations with diverse force fields. Furthermore, for explicit solvent simulations, the advantage of OpenMM over other MD packages is less evident. Therefore, we terminate the OpenABC workflow at producing atomistic condensate structures and leave the users with flexibility to choose MD packages and force fields for further studies.

As proof of principle, we converted the final snapshot from the slab simulation of HP1 $\alpha$  dimer at 260 K to an atomistic configuration ([Fig 6](#)). This conversion leverages the software REMO [85] to build atomistic details starting from the  $C\alpha$  positions of each amino acid. We solvated the atomistic HP1 $\alpha$  dimer condensates with water molecules and counter ions. After energy minimization, we carried out an all-atom MD simulation using GROMACS with the CHARMM36m force field [94] and the CHARMM-modified TIP3P water model [95]. More details on simulation preparation can be found in the [S1 Text—Building and relaxing atomistic structures from coarse-grained configurations](#). As shown in [Fig 6](#), the system relaxes with a continuously decreasing potential energy in the first 20 ns and remains stable afterward.



**Fig 6. OpenABC facilitates all-atom simulations by producing equilibrated initial atomistic configurations.** (A) Illustrations of the conversion from a coarse-grained configuration (top) to a fully atomistic model with explicit solvent molecules (bottom). Only 2% of water molecules and counter ions of the atomistic model are shown for clarity. The system consists of 100 HP1 $\alpha$  dimers, and different molecules are shown in one of 25 colors. Both figures are rendered with Mol\* Viewer [90]. (B) The atomistic potential energy evaluated using the CHARMM force field is shown as a function of simulation time.

<https://doi.org/10.1371/journal.pcbi.1011442.g006>

## Conclusion

We introduced a software package, OpenABC, to facilitate coarse-grained and all-atom simulations of biomolecular condensates. The package implements several of the leading coarse-grained force fields for protein and DNA molecules into OpenMM, enabling GPU-accelerated simulations with performances rivaling GROMACS simulations with hundreds of CPUs. New force fields can be quickly introduced within the framework, and we plan to incorporate RNA models into the package as the next step. Comprehensive tutorials are provided to familiarize the users with the various functionalities offered by OpenABC. We anticipate the intuitive Python interface of OpenABC to reduce entry barriers and promote coarse-grained modeling for its adoption by a broader community.

## Materials and methods

### Details of molecular dynamics simulations

We performed temperature replica-exchange simulations [83] with MOFF to determine the conformational ensembles of HP1 $\alpha$  and HP1 $\beta$  dimers. Atomistic protein structures were predicted with RaptorX [96] and used to initialize simulations. Details on modeling HP1 proteins to preserve the tertiary structure of folded domains are provided in the [S1 Text—Setting up MOFF HP1 system](#). Six independent replicas were simulated to maintain temperatures at 300 K, 315.79 K, 333.33 K, 352.94 K, 375.00 K, and 400 K, respectively, with the Langevin middle integrator [88] and a friction coefficient of  $1 \text{ ps}^{-1}$ . Each replica lasted for 200 million steps with a timestep of 10 fs. Exchanges between neighboring replicas were attempted every 1000 steps. More details on the replica exchange simulations are attached in the [S1 Text—Implementation](#)

of the temperature replica exchange algorithm. We discarded the first 100 million steps as equilibration and used the remaining data for analysis.

We carried out slab simulations to evaluate the stability of condensates formed by HP1 $\alpha$  and HP1 $\beta$  dimers. Initial configurations of these simulations were prepared as follows. First, we randomly placed 100 copies of protein dimers into a cubic box of length 75 nm. Then we performed 5-million-step constant pressure and constant temperature (NPT) simulations at one bar and 150 K to compress the system with a timestep of 10 fs. Control of pressure and temperature was achieved by coupling the Monte Carlo barostat with the Langevin middle integrator [88]. The length of the compressed cubic box was about 25 nm. Then we fixed the compressed configuration and extended the box size to 25  $\times$  25  $\times$  400 nm<sup>3</sup>. The rectangular geometry leads to the creation of a dense-dilute interface along the  $z$ -axis. Simulation results are expected to be independent of the exact box lengths and we chose 400 nm to be long enough to support phase coexistence (Fig B in S1 Text). Starting from this initial configuration, we gradually increased the temperature from 150 K to a target value in the first 0.1 million steps. We then performed 200-million-step production simulations at constant volume and constant temperature using the Nosé-Hoover integrator [88] with a collision frequency of 1 ps<sup>-1</sup> and a timestep of 5 fs. Compared to the Langevin thermostat, the Nosé-Hoover integrator allows faster diffusion of protein molecules in the dilute phase to facilitate the equilibration of slab simulations.

Following similar protocols outlined above, we performed slab simulations for disordered regions of protein DDX4 and FUS with the HPS model using parameters derived from the Urry hydrophobicity scale [97]. Detailed amino acid sequences of the two proteins are provided in the S1 Text. For each protein, we first obtained an equilibrium configuration from a 0.1-million-step constant temperature simulation initialized with a straight C $\alpha$  chain. We placed 100 replicas of the equilibrium configurations into a cubic box of length 75 nm. Upon compression by a 5-million-step NPT compression at 1 bar and 150 K with a timestep of 10 fs, the system reaches a cubic box with a size of about 15 nm. We then performed slab simulations with an elongated box of size 15  $\times$  15  $\times$  280 nm<sup>3</sup> and a 10 fs timestep. Nosé-Hoover integrator was again applied with a collision frequency of 1 ps<sup>-1</sup> to maintain the temperature.

## Computing phase diagrams from slab simulations

To determine the concentration of dense and dilute phases from slab simulations, we first identified the largest cluster in a given configuration as the largest connected component of the protein-contact network. Two monomers were defined as in contact if their center-of-mass distance was less than 5 nm, though the computed phase diagrams are rather insensitive to this specific cutoff value (Table O in S1 Text). Subsequently, we translated the system so that the center of mass of the largest cluster coincides with the box center, which was located at  $z = 0$ . We recognized the region with  $|z| < 5$  nm for HPS simulations and  $|z| < 10$  nm for MOFF simulations as the dense phase, while the region with  $|z| > 50$  nm as the dilute phase. The threshold values were chosen to be consistent with prior literature [20, 63] and to roughly follow the size of the condensate as revealed in the density profiles (Figs B-C in S1 Text). The concentrations were determined as the average density value in specified regions using the second half of the simulation trajectories. We fitted the concentration values at various temperatures using the following equation to determine the critical temperature

$$\rho_H - \rho_L = A(T_c - T)^\beta. \quad (1)$$

$\rho_H$  and  $\rho_L$  are the densities at the concentrated and dilute phases. Parameter  $\beta = 0.325$  is the

critical exponent corresponding to the universality class of 3D Ising model [98].  $T_c$  is the critical temperature and  $A$  is the coefficient.

## Supporting information

**S1 Text. Fig A.** The log of the bond vector correlation,  $\log C(n)$ , as a function of the bond separation  $n$ . The dots were obtained from MD simulations, with three colors indicate three independent simulations. The lines are numerical fits to the data. See text *Section: Computing DNA persistence length with the MRG-CG model* for simulation details and computing persistence length from the numerical fitting. **Fig B.** Density profiles obtained from slab simulations of HP1 $\alpha$  (left) and HP1 $\beta$  (right) dimers with the MOFF model. Vertical lines are set at  $z = \pm 10$  and  $\pm 50$  nm. The final snapshots of the slab simulations at the lowest temperatures are shown. CG atoms with  $|z| < 10$  nm are colored in yellow, while the remaining are shown in blue. **Fig C.** Density profiles obtained from slab simulations of DDX4 and FUS LC with the HPS model using the Urry scale optimal parameter set ( $\mu = \mu_{\text{Urry}}^{\text{opt}} = 1$  and  $\Delta = \Delta_{\text{Urry}}^{\text{opt}} = 0.08$ ) at different temperatures. Vertical dashed lines are set at  $z = \pm 5$  nm and  $\pm 50$  nm. The final snapshots of the slab simulations at 260 K are shown. CG atoms with  $|z| < 5$  nm are colored in yellow, while the remaining are shown in blue. This figure shows that the  $|z| < 5$  nm and  $|z| > 50$  nm regimes can represent the concentrated and dilute phases, respectively. **Table A.** The amino acid mass, sizes, and charges used by MOFF and HPS models. Both models share the same amino acid mass and sizes. The charge of HIS differs in the two models, while other amino acids share the same charge. Here  $e$  is the elementary charge. **Table B.** MOFF protein contact  $\epsilon_{ij}$  values as defined in equation S6 in S1 Text. Due to limited space, the numbers are rounded to 3 decimal places. The values are in unit kJ/mol. **Table C.** MRG DNA model bond parameters. All the  $k_{\text{bond},n}$  are in unit kcal/mol/nm<sup>2</sup>, and  $r_0$  unit is nm. **Table D.** MRG DNA model angle parameters. All the  $k_{\text{angle},n}$  are in unit kcal/mol/degree<sup>2</sup>, and  $\theta_0$  unit is degree. **Table E.** MRG DNA model fan bond parameters. All the  $k_{\text{fan bond},n}$  are in kcal/mol/nm<sup>2</sup>, and  $r_{\Delta,0}$  unit is nm.  $\Delta$  means the fan bond between CG nucleotide  $i$  and  $j + \Delta$ , where nucleotide  $i$  and  $j$  are WC-paired. **Table F.** The normalized KR scale and Urry hydrophathy scale values (i.e.  $\lambda_i$  parameters in equation S14 in S1 Text). **Table G.** Mpipi parameter  $\epsilon$  values as defined in equation S25 in S1 Text. Due to limited space, the numbers are rounded to 3 decimal places. The values are in unit kJ/mol. **Table H.** Mpipi parameter  $\sigma$  values as defined in equation S25 in S1 Text. Due to limited space, the numbers are rounded to 3 decimal places. The values are in unit nm. **Table I.** SMOG MJ potential parameter  $\epsilon$  as defined in equation S31 in S1 Text. Due to the limited space, the numbers are rounded to 3 decimal places. The values are in unit kJ/mol. **Table J.** Comparison of potential energies computed with OpenMM and GROMACS using MOFF for ten configurations of HP1 $\alpha$  dimer. **Table K.** Comparison of potential energies computed with OpenMM and GROMACS using MOFF for proteins and MRG for DNA for ten configurations of HP1 $\alpha$  dimer bound to a dsDNA. The energy unit is kJ/mol. See text *Section: Validating the force field implementation in OpenMM* for simulation details. **Table L.** Comparison of potential energies computed with OpenMM and HOOMD-Blue using HPS with Urry or KR scales for ten configurations of protein DDX4. The energy unit is kJ/mol. See text *Section: Validating the force field implementation in OpenMM* for simulation details. **Table M.** Comparison of potential energies computed with OpenMM and LAMMPS using Mpipi force field for a polyR+polyK+polyU system. The system consists of a chain of 10 arginines, a chain of 10 lysines, and 2 individual chains of 10 uracils. The energy unit is kJ/mol. See text *Section: Validating the force field implementation in OpenMM* for simulation details. **Table N.** The coexistence concentrations of HP1 $\alpha$  and HP1 $\beta$  dimers measured by slab simulations with MOFF at different temperatures below the critical temperature. The cutoff distance



for searching the largest cluster is 5 nm. **Table O.** The coexistence concentrations of HP1 $\alpha$  and HP1 $\beta$  dimers measured by slab simulation with MOFF. The concentrations were similarly determined as those shown in Table N but the cutoff distance for searching the largest cluster set as 8 instead of 5 nm. The results are almost identical to the ones shown in Table N, supporting the robustness of phase diagrams with respect to the cutoff distance used for protein clustering. **Table P.** The coexistence concentrations of FUS LC and DDX4 proteins measured by performing slab simulations with HPS model Urry scale and the optimal parameter set ( $\mu = \mu_{\text{Urry}}^{\text{opt}} = 1$  and  $\Delta = \Delta_{\text{Urry}}^{\text{opt}} = 0.08$ ) at different temperatures below the critical temperature. The cutoff distance for searching the largest cluster is 5 nm. (PDF)

## Acknowledgments

We appreciate Azamat Rizuan's help verifying the implementation of HPS models.

## Author Contributions

**Conceptualization:** Shuming Liu, Bin Zhang.

**Data curation:** Shuming Liu, Cong Wang, Bin Zhang.

**Formal analysis:** Shuming Liu, Cong Wang, Bin Zhang.

**Funding acquisition:** Bin Zhang.

**Investigation:** Shuming Liu, Cong Wang, Bin Zhang.

**Methodology:** Shuming Liu, Bin Zhang.

**Project administration:** Bin Zhang.

**Resources:** Bin Zhang.

**Software:** Shuming Liu, Cong Wang, Andrew P. Latham, Xinqiang Ding, Bin Zhang.

**Supervision:** Bin Zhang.

**Validation:** Shuming Liu, Cong Wang, Bin Zhang.

**Visualization:** Shuming Liu, Cong Wang, Bin Zhang.

**Writing – original draft:** Shuming Liu, Bin Zhang.

**Writing – review & editing:** Shuming Liu, Cong Wang, Andrew P. Latham, Xinqiang Ding, Bin Zhang.

## References

1. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proceedings of the National Academy of Sciences*. 2011; 108(11):4334–4339. <https://doi.org/10.1073/pnas.1017150108> PMID: 21368180
2. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoegge C, Gharakhani J, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*. 2009; 324(5935):1729–1732. <https://doi.org/10.1126/science.1172046> PMID: 19460965
3. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews Molecular cell biology*. 2017; 18(5):285–298. <https://doi.org/10.1038/nrm.2017.7> PMID: 28225081
4. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. *Science*. 2017; 357(6357):eaaf4382. <https://doi.org/10.1126/science.aaf4382> PMID: 28935776
5. Narlikar GJ. Phase-separation in chromatin organization. *Journal of biosciences*. 2020; 45:1–5.

6. Choi JM, Holehouse AS, Pappu RV. Physical principles underlying the complex biology of intracellular phase transitions. *Annual review of biophysics*. 2020; 49:107–133. <https://doi.org/10.1146/annurev-biophys-121219-081629> PMID: 32004090
7. Sabari BR, Dall'Agnese A, Young RA. Biomolecular condensates in the nucleus. *Trends in biochemical sciences*. 2020; 45(11):961–977. <https://doi.org/10.1016/j.tibs.2020.06.007> PMID: 32684431
8. Bhat P, Honson D, Guttman M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nature Reviews Molecular Cell Biology*. 2021; 22(10):653–670. <https://doi.org/10.1038/s41580-021-00387-1> PMID: 34341548
9. Lyon AS, Peeples WB, Rosen MK. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology*. 2021; 22(3):215–235. <https://doi.org/10.1038/s41580-020-00303-z> PMID: 33169001
10. Bojja A, Klein IA, Young RA. Biomolecular condensates and cancer. *Cancer cell*. 2021; 39(2):174–192. <https://doi.org/10.1016/j.ccell.2020.12.003> PMID: 33417833
11. Borchers W, Bremer A, Borgia MB, Mittag T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Current opinion in structural biology*. 2021; 67:41–50. <https://doi.org/10.1016/j.sbi.2020.09.004> PMID: 33069007
12. Belmont AS. Nuclear compartments: an incomplete primer to nuclear compartments, bodies, and genome organization relative to nuclear architecture. *Cold Spring Harbor Perspectives in Biology*. 2022; 14(7):a041268. <https://doi.org/10.1101/cshperspect.a041268> PMID: 34400557
13. Latham AP, Zhang B. Molecular determinants for the layering and coarsening of biological condensates. *Aggregate*. 2022; p. e306. <https://doi.org/10.1002/agt2.306> PMID: 37065433
14. Pappu RV, Cohen SR, Dar F, Farag M, Kar M. Phase Transitions of Associative Biomacromolecules. *Chemical Reviews*. 2023;. <https://doi.org/10.1021/acs.chemrev.2c00814> PMID: 36881934
15. Brangwynne CP, Tompa P, Pappu RV. Polymer physics of intracellular phase transitions. *Nature Physics*. 2015; 11(11):899–904. <https://doi.org/10.1038/nphys3532>
16. Hori N, Takada S. Coarse-grained structure-based model for RNA-protein complexes developed by fluctuation matching. *Journal of Chemical Theory and Computation*. 2012; 8(9):3384–3394. <https://doi.org/10.1021/ct300361j> PMID: 26605744
17. Li W, Wang W, Takada S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proceedings of the National Academy of Sciences*. 2014; 111(29):10550–10555. <https://doi.org/10.1073/pnas.1402768111> PMID: 25002491
18. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, et al. Coexisting liquid phases underlie nucleolar subcompartments. *Cell*. 2016; 165(7):1686–1697. <https://doi.org/10.1016/j.cell.2016.04.047> PMID: 27212236
19. Harmon TS, Holehouse AS, Rosen MK, Pappu RV. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *elife*. 2017; 6:e30294. <https://doi.org/10.7554/eLife.30294> PMID: 29091028
20. Dignon GL, Zheng W, Kim YC, Best RB, Mittal J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology*. 2018; 14(1):e1005941. <https://doi.org/10.1371/journal.pcbi.1005941> PMID: 29364893
21. Shi G, Liu L, Hyeon C, Thirumalai D. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature communications*. 2018; 9(1):3161. <https://doi.org/10.1038/s41467-018-05606-6> PMID: 30089831
22. Choi JM, Dar F, Pappu RV. LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLoS computational biology*. 2019; 15(10):e1007028. <https://doi.org/10.1371/journal.pcbi.1007028> PMID: 31634364
23. Dignon GL, Best RB, Mittal J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annual review of physical chemistry*. 2020; 71:53. <https://doi.org/10.1146/annurev-physchem-071819-113553> PMID: 32312191
24. Baul U, Chakraborty D, Mugnai ML, Straub JE, Thirumalai D. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *The Journal of Physical Chemistry B*. 2019; 123(16):3462–3474. <https://doi.org/10.1021/acs.jpcc.9b02575> PMID: 30913885
25. Wang J, Olsson S, Wehmeyer C, Pérez A, Charron NE, De Fabritiis G, et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS central science*. 2019; 5(5):755–767. <https://doi.org/10.1021/acscentsci.8b00913> PMID: 31139712
26. Husic BE, Charron NE, Lemm D, Wang J, Pérez A, Majewski M, et al. Coarse graining molecular dynamics with graph neural networks. *The Journal of chemical physics*. 2020; 153(19):194101. <https://doi.org/10.1063/5.0026133> PMID: 33218238

27. Regy RM, Dignon GL, Zheng W, Kim YC, Mittal J. Sequence dependent phase separation of protein-polynucleotide mixtures elucidated using molecular simulations. *Nucleic acids research*. 2020; 48(22):12593–12603. <https://doi.org/10.1093/nar/gkaa1099> PMID: 33264400
28. Regy RM, Thompson J, Kim YC, Mittal J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science*. 2021; 30(7):1371–1379. <https://doi.org/10.1002/pro.4094> PMID: 33934416
29. Joseph JA, Reinhardt A, Aguirre A, Chew PY, Russell KO, Espinosa JR, et al. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science*. 2021; 1(11):732–743. <https://doi.org/10.1038/s43588-021-00155-3> PMID: 35795820
30. Farr SE, Woods EJ, Joseph JA, Garaizar A, Collepardo-Guevara R. Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions. *Nature communications*. 2021; 12(1):2883. <https://doi.org/10.1038/s41467-021-23090-3> PMID: 34001913
31. Qi Y, Zhang B. Chromatin network retards nucleoli coalescence. *Nature Communications*. 2021; 12(1):6824. <https://doi.org/10.1038/s41467-021-27123-9> PMID: 34819511
32. Kamat K, Lao Z, Qi Y, Wang Y, Ma J, Zhang B. Compartmentalization with nuclear landmarks yields random, yet precise, genome organization. *Biophysical Journal*. 2023; 122(7):1376–1389. <https://doi.org/10.1016/j.bpj.2023.03.003> PMID: 36871158
33. Souza PC, Alessandri R, Barnoud J, Thalmair S, Faustino I, Grünewald F, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods*. 2021; 18(4):382–388. <https://doi.org/10.1038/s41592-021-01098-3> PMID: 33782607
34. Dannenhoffer-Lafage T, Best RB. A data-driven hydrophobicity scale for predicting liquid–liquid phase separation of proteins. *The Journal of Physical Chemistry B*. 2021; 125(16):4046–4056. <https://doi.org/10.1021/acs.jpcc.0c11479> PMID: 33876938
35. Wessén J, Pal T, Das S, Lin YH, Chan HS. A simple explicit-solvent model of polyampholyte phase behaviors and its ramifications for dielectric effects in biomolecular condensates. *The Journal of Physical Chemistry B*. 2021; 125(17):4337–4358. <https://doi.org/10.1021/acs.jpcc.1c00954> PMID: 33890467
36. Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proceedings of the National Academy of Sciences*. 2021; 118(44):e2111696118. <https://doi.org/10.1073/pnas.2111696118>
37. Lin X, Leicher R, Liu S, Zhang B. Cooperative DNA looping by PRC2 complexes. *Nucleic Acids Research*. 2021; 49(11):6238–6248. <https://doi.org/10.1093/nar/gkab441> PMID: 34057467
38. Chen Y, Krämer A, Charron NE, Husic BE, Clementi C, Noé F. Machine learning implicit solvation for molecular dynamics. *The Journal of Chemical Physics*. 2021; 155(8):084101.
39. Nguyen HT, Hori N, Thirumalai D. Condensates in RNA repeat sequences are heterogeneously organized and exhibit reptation dynamics. *Nature chemistry*. 2022; 14(7):775–785. <https://doi.org/10.1038/s41557-022-00934-z> PMID: 35501484
40. Tan C, Jung J, Kobayashi C, Torre DUL, Takada S, Sugita Y. Implementation of residue-level coarse-grained models in GENESIS for large-scale molecular dynamics simulations. *PLOS Computational Biology*. 2022; 18(4):e1009578. <https://doi.org/10.1371/journal.pcbi.1009578> PMID: 35381009
41. Liu S, Lin X, Zhang B. Chromatin fiber breaks into clutches under tension and crowding. *Nucleic Acids Research*. 2022; 50(17):9738–9747. <https://doi.org/10.1093/nar/gkac725> PMID: 36029149
42. Ding X, Zhang B. Contrastive learning of coarse-grained force fields. *Journal of Chemical Theory and Computation*. 2022; 18(10):6334–6344. <https://doi.org/10.1021/acs.jctc.2c00616> PMID: 36112935
43. Di Pierro M, Potoyan DA, Wolynes PG, Onuchic JN. Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proceedings of the National Academy of Sciences*. 2018; 115(30):7753–7758. <https://doi.org/10.1073/pnas.1806297115> PMID: 29987017
44. Kluber A, Burt TA, Clementi C. Size and topology modulate the effects of frustration in protein folding. *Proceedings of the National Academy of Sciences*. 2018; 115(37):9234–9239. <https://doi.org/10.1073/pnas.1801406115> PMID: 30150375
45. Chen X, Chen M, Schafer NP, Wolynes PG. Exploring the interplay between fibrillization and amorphous aggregation channels on the energy landscapes of tau repeat isoforms. *Proceedings of the National Academy of Sciences*. 2020; 117(8):4125–4130. <https://doi.org/10.1073/pnas.1921702117> PMID: 32029593
46. Wu H, Dalal Y, Papoian GA. Binding dynamics of disordered linker histone H1 with a nucleosomal particle. *Journal of molecular biology*. 2021; 433(6):166881. <https://doi.org/10.1016/j.jmb.2021.166881> PMID: 33617899
47. Latham AP, Zhang B. On the stability and layered organization of protein-DNA condensates. *Biophysical Journal*. 2022; 121(9):1727–1737. <https://doi.org/10.1016/j.bpj.2022.03.028> PMID: 35364104

48. Latham AP, Zhu L, Sharon DA, Ye S, Willard AP, Zhang X, et al. Frustrated Microphase Separation Produces Interfacial Environment within Biological Condensates. *bioRxiv*. 2023; p. 2023–03. <https://doi.org/10.1101/2023.03.30.534967> PMID: 37034777
49. Zheng W, Dignon GL, Jovic N, Xu X, Regy RM, Fawzi NL, et al. Molecular details of protein condensates probed by microsecond long atomistic simulations. *The Journal of Physical Chemistry B*. 2020; 124(51):11671–11679. <https://doi.org/10.1021/acs.jpcc.0c10489> PMID: 33302617
50. Thody SA, Clements HD, Baniyadi H, Lyon AS, Sigman MS, Rosen MK. Small Molecule Properties Define Partitioning into Biomolecular Condensates. *bioRxiv*. 2022; p. 2022–12.
51. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*. 2018; 115(21):E4758–E4766. <https://doi.org/10.1073/pnas.1800690115> PMID: 29735687
52. Galvanetto N, Ivanović MT, Chowdhury A, Sottini A, Nüesch M, Nettels D, et al. Ultrafast molecular dynamics observed within a dense protein condensate. *bioRxiv*. 2022; p. 2022–12.
53. Klein IA, Boija A, Afeyan LK, Hawken SW, Fan M, Dall'Agnese A, et al. Partitioning of cancer therapeutics in nuclear condensates. *Science*. 2020; 368(6497):1386–1392. <https://doi.org/10.1126/science.aaz4427> PMID: 32554597
54. Berendsen HJ, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer physics communications*. 1995; 91(1-3):43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
55. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*. 2008; 29(11):1859–1865. <https://doi.org/10.1002/jcc.20945> PMID: 18351591
56. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2013; 3(2):198–210.
57. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
58. Phillips JC, Hardy DJ, Maia JD, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of chemical physics*. 2020; 153(4):044130. <https://doi.org/10.1063/5.0014475> PMID: 32752662
59. Thompson AP, Aktulga HM, Berger R, Bolintineanu DS, Brown WM, Crozier PS, et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*. 2022; 271:108171. <https://doi.org/10.1016/j.cpc.2021.108171>
60. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*. 2017; 13(7):e1005659. <https://doi.org/10.1371/journal.pcbi.1005659> PMID: 28746339
61. Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, et al. SMOG 2: a versatile software package for generating structure-based models. *PLoS computational biology*. 2016; 12(3):e1004794. <https://doi.org/10.1371/journal.pcbi.1004794> PMID: 26963394
62. Ding X, Lin X, Zhang B. Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface. *Nature communications*. 2021; 12(1):1091. <https://doi.org/10.1038/s41467-021-21377-z> PMID: 33597548
63. Latham AP, Zhang B. Consistent force field captures homologue-resolved hp1 phase separation. *Journal of chemical theory and computation*. 2021; 17(5):3134–3144. <https://doi.org/10.1021/acs.jctc.0c01220> PMID: 33826337
64. Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J, et al. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nature structural & molecular biology*. 2019; 26(7):637–648. <https://doi.org/10.1038/s41594-019-0250-x> PMID: 31270472
65. Schuster BS, Dignon GL, Tang WS, Kelley FM, Ranganath AK, Jahnke CN, et al. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences*. 2020; 117(21):11421–11431. <https://doi.org/10.1073/pnas.2000223117>
66. Conicella AE, Dignon GL, Zerze GH, Schmidt HB, D'Ordine AM, Kim YC, et al. TDP-43  $\alpha$ -helical structure tunes liquid-liquid phase separation and function. *Proceedings of the National Academy of Sciences*. 2020; 117(11):5883–5894. <https://doi.org/10.1073/pnas.1912055117> PMID: 32132204
67. Krainer G, Welsh TJ, Joseph JA, Espinosa JR, Wittmann S, de Csilléry E, et al. Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nature communications*. 2021; 12(1):1085. <https://doi.org/10.1038/s41467-021-21181-9> PMID: 33597515

68. Miyazawa S, Jernigan RL. Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of molecular biology*. 1996; 256(3):623–644. <https://doi.org/10.1006/jmbi.1996.0114> PMID: 8604144
69. Pitera JW, Chodera JD. On the use of experimental observations to bias simulated ensembles. *Journal of chemical theory and computation*. 2012; 8(10):3445–3451. <https://doi.org/10.1021/ct300112v> PMID: 26592995
70. Roux B, Weare J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of chemical physics*. 2013; 138(8):02B616. <https://doi.org/10.1063/1.4792208> PMID: 23464140
71. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*. 1995; 21(3):167–195. <https://doi.org/10.1002/prot.340210302> PMID: 7784423
72. Latham AP, Zhang B. Maximum entropy optimized force field for intrinsically disordered proteins. *Journal of chemical theory and computation*. 2019; 16(1):773–781. <https://doi.org/10.1021/acs.jctc.9b00932> PMID: 31756104
73. Latham AP, Zhang B. Improving coarse-grained protein force fields with small-angle X-ray scattering data. *The Journal of Physical Chemistry B*. 2019; 123(5):1026–1034. <https://doi.org/10.1021/acs.jpcc.8b10336> PMID: 30620594
74. Latham AP, Zhang B. Unifying coarse-grained force fields for folded and disordered proteins. *Current opinion in structural biology*. 2022; 72:63–70. <https://doi.org/10.1016/j.sbi.2021.08.006> PMID: 34536913
75. Regmi R, Srinivasan S, Latham AP, Kukshal V, Cui W, Zhang B, et al. Phosphorylation-Dependent Conformations of the Disordered Carboxyl-Terminus Domain in the Epidermal Growth Factor Receptor. *J Phys Chem Lett*. 2020; 11(23):10037–10044. <https://doi.org/10.1021/acs.jpclett.0c02327> PMID: 33179922
76. Savelyev A, Papoian GA. Chemically accurate coarse graining of double-stranded DNA. *Proceedings of the National Academy of Sciences*. 2010; 107(47):20340–20345. <https://doi.org/10.1073/pnas.1001163107> PMID: 21059937
77. Hinckley DM, Freeman GS, Whitmer JK, De Pablo JJ. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *The Journal of chemical physics*. 2013; 139(14):10B604\_1. <https://doi.org/10.1063/1.4822042> PMID: 24116642
78. Lu W, Bueno C, Schafer NP, Moller J, Jin S, Chen X, et al. OpenAWSEM with Open3SPN2: A fast, flexible, and accessible framework for large-scale coarse-grained biomolecular simulations. *PLoS computational biology*. 2021; 17(2):e1008308. <https://doi.org/10.1371/journal.pcbi.1008308> PMID: 33577557
79. Lequieu J, Córdoba A, Schwartz DC, de Pablo JJ. Tension-dependent free energies of nucleosome unwrapping. *ACS central science*. 2016; 2(9):660–666. <https://doi.org/10.1021/acscentsci.6b00201> PMID: 27725965
80. Tan C, Terakawa T, Takada S. Dynamic coupling among protein binding, sliding, and DNA bending revealed by molecular dynamics. *Journal of the American Chemical Society*. 2016; 138(27):8512–8522. <https://doi.org/10.1021/jacs.6b03729> PMID: 27309278
81. Parsons T, Zhang B. Critical role of histone tail entropy in nucleosome unwinding. *J Chem Phys*. 2019; 150(18):185103. <https://doi.org/10.1063/1.5085663> PMID: 31091895
82. Leicher R, Ge EJ, Lin X, Reynolds MJ, Xie W, Walz T, et al. Single-molecule and in silico dissection of the interaction between Polycomb repressive complex 2 and chromatin. *Proc Natl Acad Sci*. 2020; 117(48):30465–30475. <https://doi.org/10.1073/pnas.2003395117> PMID: 33208532
83. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*. 1999; 314(1-2):141–151. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9)
84. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019; 32.
85. Li Y, Zhang Y. REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins: Structure, Function, and Bioinformatics*. 2009; 76(3):665–676. <https://doi.org/10.1002/prot.22380>
86. Dignon GL, Zheng W, Best RB, Kim YC, Mittal J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*. 2018; 115(40):9929–9934. <https://doi.org/10.1073/pnas.1804177115> PMID: 30217894
87. Regy RM, Zheng W, Mittal J. Using a sequence-specific coarse-grained model for studying protein liquid–liquid phase separation. In: *Methods in enzymology*. vol. 646. Elsevier; 2021. p. 1–17.



88. Zhang Z, Liu X, Yan K, Tuckerman ME, Liu J. Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics. *The Journal of Physical Chemistry A*. 2019; 123(28):6056–6079. <https://doi.org/10.1021/acs.jpca.9b02771> PMID: 31117592
89. Anderson JA, Glaser J, Glotzer SC. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science*. 2020; 173:109363. <https://doi.org/10.1016/j.commatsci.2019.109363>
90. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*. 2021; 49(W1):W431–W437. <https://doi.org/10.1093/nar/gkab314> PMID: 33956157
91. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. *Nature reviews Molecular cell biology*. 2004; 5(4):296–305. <https://doi.org/10.1038/nrm1355> PMID: 15071554
92. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell*. 2015; 57(5):936–947. <https://doi.org/10.1016/j.molcel.2015.01.013> PMID: 25747659
93. Burke KA, Janke AM, Rhine CL, Fawzi NL. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Molecular cell*. 2015; 60(2):231–241. <https://doi.org/10.1016/j.molcel.2015.09.006> PMID: 26455390
94. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*. 2017; 14(1):71–73. <https://doi.org/10.1038/nmeth.4067> PMID: 27819658
95. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983; 79(2):926–935. <https://doi.org/10.1063/1.445869>
96. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. *Nature protocols*. 2012; 7(8):1511–1522. <https://doi.org/10.1038/nprot.2012.085> PMID: 22814390
97. Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, et al. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers: Original Research on Biomolecules*. 1992; 32(9):1243–1250. <https://doi.org/10.1002/bip.360320913> PMID: 1420991
98. Rowlinson JS, Widom B. *Molecular theory of capillarity*. Courier Corporation; 2013.