

Toward Predictive Coarse-Grained Simulations of Biomolecular Condensates

Published as part of Biochemistry *special issue* "Computational Biochemistry".

Shuming Liu, Cong Wang, and Bin Zhang*



Cite This: *Biochemistry* 2025, 64, 1750–1761



Read Online

ACCESS |

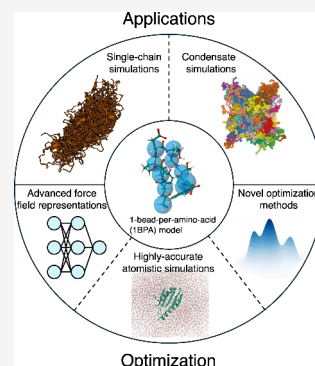
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Phase separation is a fundamental process that enables cellular organization by forming biomolecular condensates. These assemblies regulate diverse functions by creating distinct environments, influencing reaction kinetics, and facilitating processes such as genome organization, signal transduction, and RNA metabolism. Recent studies highlight the complexity of condensate properties, shaped by intrinsic molecular features and external factors such as temperature and pH. Molecular simulations serve as an effective approach to establishing a comprehensive framework for analyzing these influences, offering high-resolution insights into condensate stability, dynamics, and material properties. This review evaluates recent advancements in biomolecular condensate simulations, with a particular focus on coarse-grained 1-bead-per-amino-acid (1BPA) protein models, and emphasizes OpenABC, a tool designed to simplify and streamline condensate simulations. OpenABC supports the implementation of various coarse-grained force fields, enabling their performance evaluation. Our benchmarking identifies inconsistencies in phase behavior predictions across force fields, even though these models accurately capture single-chain statistics. This finding underscores the need for enhanced force field accuracy, achievable through enriched training data sets, many-body potentials, and advanced optimization techniques. Such refinements could significantly improve the predictive capacity of coarse-grained models, bridging molecular details with emergent condensate behaviors.

KEYWORDS: coarse-grained simulations, phase separation, biomolecular condensates, force field accuracy, OpenABC



Phase separation is a fundamental cellular process that drives the formation of biomolecular condensates, such as the nucleolus, stress granules, and P-bodies.^{1–5} The capacity of biomolecules to undergo phase separation and form dynamic, tunable assemblies enables diverse cellular functions by regulating molecular distributions and selective partitioning. This creates distinct environments that influence reaction kinetics and facilitate various cellular processes.^{1,6} These condensates are implicated in numerous biological functions, including genome organization, signal transduction, and RNA metabolism.^{7–9} Dysregulation of phase separation is associated with diseases such as neurodegenerative disorders and cancer.^{2,10} Therefore, a deeper understanding of the mechanisms governing phase separation and the ability to modulate this process hold significant potential for advancing disease treatment strategies.

Over the past decade, experimental studies have revealed the remarkable complexity of biological condensates, particularly in terms of their composition,¹ material properties,^{4,11} and dynamics.^{12,13} These emergent properties are influenced by both intrinsic molecular features, such as interaction patterns,^{4,14,15} and external environmental factors, including temperature, pH, and ionic strength.^{15–17} Developing a theoretical framework that quantitatively links intrinsic and extrinsic factors to these emergent properties is essential for

advancing our understanding of condensate formation and regulation. However, this goal is difficult to achieve through experimental studies alone due to their limited resolution in capturing the structural organization of condensates and inefficiency in systematically exploring the key determinants.

Molecular simulations are valuable complementary tools for investigating the molecular mechanisms underlying condensate formation and regulation.^{18,19} Often described as computational microscopes, they offer high temporal and spatial resolution to characterize systems of interest. These simulations provide a detailed view of the chemical environment within the condensate.^{12,20,21} With advancements in efficiency, they enable the quantification of condensate stability and the identification of thermodynamic and chemical factors, such as temperature, pH, and ionic strength, that influence this stability.^{15,22–25} Moreover, the analysis of simulated phase separation kinetics facilitates the study of condensate dynamics

Received: October 31, 2024

Revised: March 13, 2025

Accepted: March 21, 2025

Published: April 2, 2025



and material properties, linking these characteristics to specific molecular interactions for mechanistic insights.

In this review, we examine recent advancements in molecular simulations of biomolecular condensates with a particular focus on coarse-grained (CG) protein models. Additionally, we highlight the development of OpenABC,²⁶ a software that significantly lowers the entry barrier for condensate simulations and facilitates the cross-examination of various force fields. Our benchmarking revealed inconsistencies among existing force fields in predicting phase behaviors, underscoring the need for further improvement in CG force field accuracy. We emphasize the importance of incorporating diverse training data from atomistic simulations of both folded and disordered proteins, employing many-body potentials to describe interactions with CG particles, and leveraging the latest optimization algorithms for systematic force field parametrization. These advancements could fully realize the potential of CG models to connect molecular sequences with the emergent behaviors of condensates.

■ CHARACTERIZING CONDENSATE CHEMICAL ENVIRONMENT WITH ATOMISTIC SIMULATIONS

Explicit solvent molecular dynamics simulations with atomistic force fields, commonly referred to as atomistic simulations, provide the most robust approach for characterizing biomolecular structure and dynamics.^{27–29} These simulations have also proven to be highly effective in studying biomolecular condensates.

Atomistic simulations have been employed to investigate chemical interactions within biomolecular condensates. Zheng et al.²⁰ conducted simulations on condensates formed by the low-complexity domain of FUS and the N-terminal disordered domain of LAF-1, identifying a range of residue–residue contacts that contribute to the stability of the condensed phase. Latham et al.³⁰ examined the interior of simulated elastin-like polypeptide (ELP)³³ condensates (Figure 1A), revealing an interfacial environment with an incomplete separation of hydrophilic and hydrophobic residues. This interfacial environment supports the prevalence of protein–water hydrogen bonds. Furthermore, Rekhi et al.³¹ demonstrated that phase separation is influenced by a broader array of residue contacts beyond those typically recognized as “stickers” (Figure 1B). Residues previously categorized as “spacers” were shown to provide stabilizing interactions, challenging and extending the traditional sticker-spacer framework.^{4,14}

In addition, all-atom simulations have been utilized to explore the interactions of small molecules with condensates. For example, Wang et al.²¹ examined various condensate–ligand systems, identifying different binding poses of small molecules (Figure 1C). Their findings emphasized the existence of heterogeneous, nonspecific binding environments within protein condensates. This nonspecificity is consistent with reports that the hydrophobicity of small-molecule drugs correlates with their partitioning coefficients into condensates. However, the prevalence of these heterogeneous environments varies among condensates, resulting in distinct preferences for small-molecule partitioning across different condensates.^{34–36}

■ QUANTIFYING EMERGENT PHASE BEHAVIORS USING COARSE-GRAINED MODELS

Although atomistic simulations have been effectively employed to study biomolecular condensates, most research emphasizes

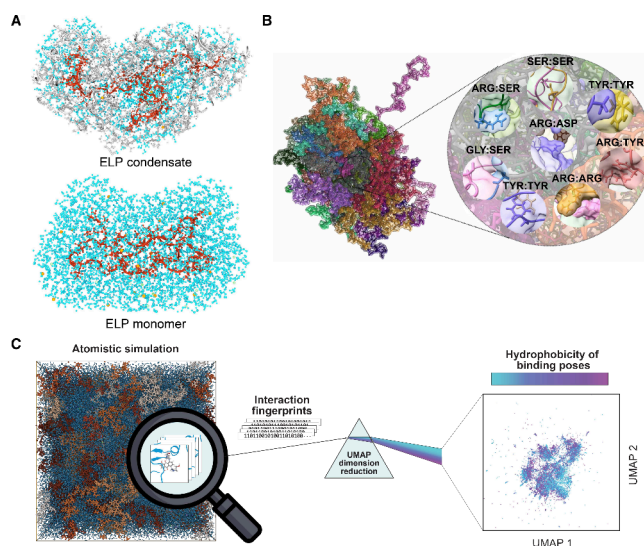


Figure 1. Atomistic simulations allow detailed characterization of the microenvironments inside condensates. (A) Comparison of protein conformation and solvation of a single peptide chain (red) from atomistic simulations of the ELP condensate (top) and monomer (bottom). Atoms within 1 nm of the peptide are shown, including water (cyan), chlorine (green), sodium (orange), and protein atoms from other chains (gray). Figure adapted from ref 30. (B) Snapshot from atomistic simulations reveals the diverse array of contacts that contribute to the phase separation of (GRGDSPYS)₂₅. Figure adapted from ref 31. (C) The condensate formed by HP1α proteins presents a diverse array of binding environments capable of accommodating the small molecule mitoxantrone. These environments are represented by projecting the simulated binding poses onto two variables derived through the UMAP dimensionality reduction technique.³² The binding poses organize into distinct clusters, each characterized by varying levels of hydrophobic interactions with small molecules. Figure adapted from ref 21.

the solvation environment and interactions between proteins and small molecules.^{20,21,30,31} These simulations encounter substantial challenges in adequately sampling the conformational dynamics of proteins within condensate systems, which is necessary to obtain converged statistics of protein contacts. Consequently, it remains difficult to directly link protein–protein interactions and molecular sequence characteristics with condensate stability, composition, and other emergent properties through atomistic simulations.^{37–40}

To enhance simulation efficiency, a range of CG models have been devised, aiming to simplify biomolecular representations. By reduction of the system size, these models significantly accelerate simulations. For instance, the Martini force field groups roughly four heavy atoms into a single CG bead⁴¹ while explicitly representing solvent molecules and ions. By tuning parameters for protein–protein interactions, this force field accurately reproduces the excess protein transfer free energy between dense and dilute phases of the low-complexity domain of FUS.⁴² Additionally, Martini simulations indicate that protein–water interactions significantly influence whether salt addition promotes or inhibits phase separation, highlighting the coupling between distinct interaction patterns in this process.⁴³

Several CG models remove solvent molecules entirely, incorporating solvent effects implicitly through effective interaction energies. In this review, we focus on CG protein models; however, CG models for nucleic acids have also been

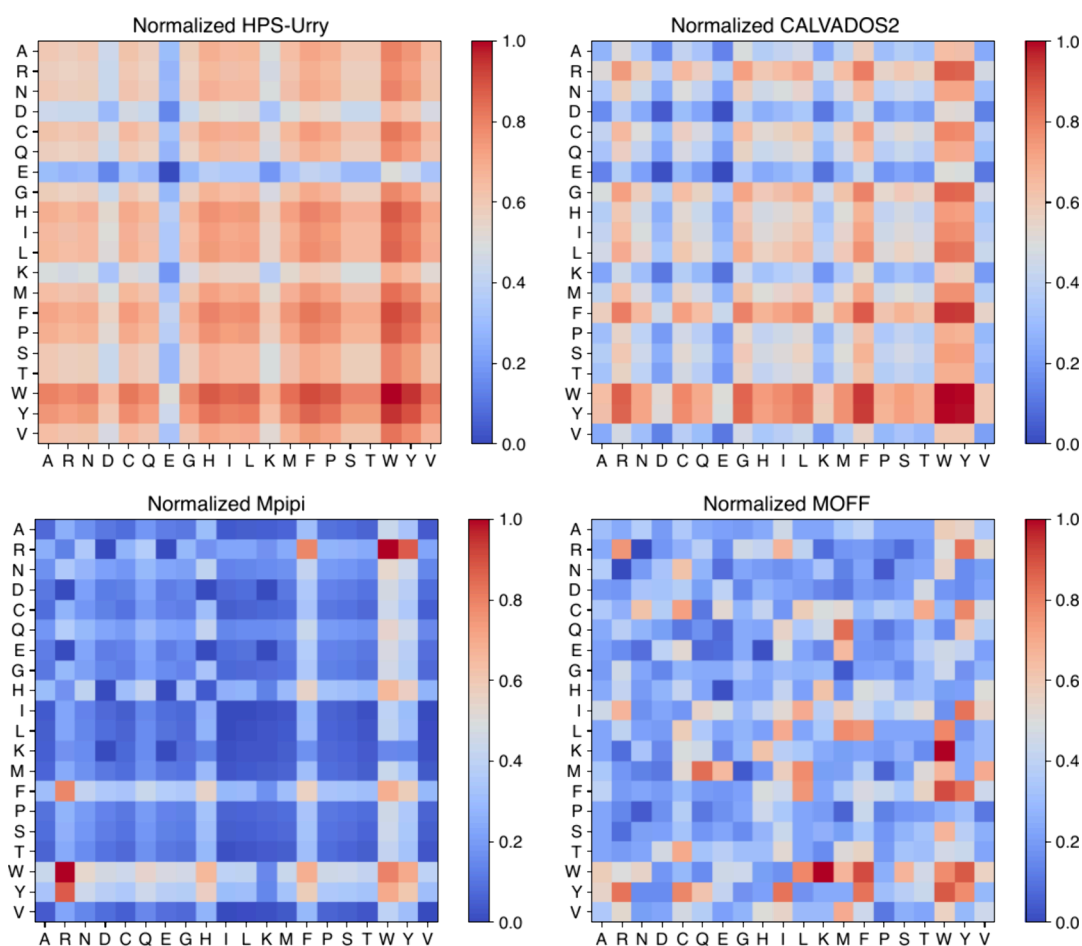


Figure 2. Comparison of nonbonded contact interaction patterns, excluding electrostatic interactions, among amino acids across various 1BPA protein models. Interaction strength parameters (λ for HPS-Urry and CALVADOS2,^{58–61} and ϵ for Mpipi and MOFF)^{62,63} are unified under the notation ϵ for consistency. For each model, ϵ values were normalized to the range [0, 1] to emphasize relative variations within each model. Normalization was performed using the formula $\bar{\epsilon} = \frac{\epsilon - \epsilon_{\min}}{\epsilon_{\max} - \epsilon_{\min}}$, where $\epsilon_{\min}$ and $\epsilon_{\max}$ denote the minimum and maximum interaction parameter values within each model. A larger $\bar{\epsilon}$ corresponds to stronger attraction. This normalization scheme highlights the intrinsic relative differences in contact strengths within each model.

developed and summarized in several comprehensive reviews.^{44–46} A widely employed class of CG protein models used to simulate biomolecular condensates represents each amino acid with a single CG bead. We collectively refer to these models as 1BPA, denoting “1 bead per amino acid”. The bead positions are typically assigned based on the coordinates of α -carbons or the amino acid’s center of mass. Although 1BPA models do not accurately predict secondary structure transitions, requiring explicit backbone atom representation,^{47–57} their computational efficiency makes them suitable for large-scale simulations, particularly condensate simulations.

Multiple 1BPA models have been developed by different research groups.^{22,58–68} While these force fields share covalent bonds for chain connectivity, they differ in nonbonded contact interactions, which describe contact energy between amino acids. These variations stem from distinct optimization algorithms used to calibrate interaction strengths, each tailored to capture specific molecular properties.

The hydrophobicity scale (HPS) models represent one class of 1BPA models.^{23,58–61,64,66,69} In these models, the interaction between two amino acids, i and j , is determined by the hydrophathy scale parameter λ_{ij} , with larger values denoting stronger attractions (Figure 2). These models assign a per-

residue hydrophathy parameter λ_i and apply the mixing rule $\lambda_{ij} = (\lambda_i + \lambda_j)/2$, reducing the total of 210 possible λ_{ij} parameters to 20 independent λ_i values. The λ_i parameters can be derived from existing data, such as atomic partial charges^{23,69} or inverse temperature transition experiments,^{58,70} or they can be systematically optimized to align with experimental measurements such as average radius of gyration (R_g) values.^{59–61,63,66}

Lindoff-Larsen and co-workers significantly extended the HPS model and performed systematic optimization based on a Bayesian algorithm,^{71,72} leading to the development of CALVADOS models. In CALVADOS1 and CALVADOS2, the optimization aimed to reproduce the experimental radius of gyration for a broad set of intrinsically disordered proteins (IDPs).^{59,60} CALVADOS3 introduced further improvements by incorporating multidomain proteins, achieving significant enhancements by adjusting the CG bead position in ordered domains from the $C\alpha$ atom to the center of mass of heavy atoms within each residue.⁶¹

Another 1BPA model called Mpipi has been developed by considering the π – π contact frequencies in PDB structures⁷³ and fitting the potential of mean force between pairs of amino acids, which is computed with all-atom simulations and umbrella sampling.⁶² This force field emphasizes the

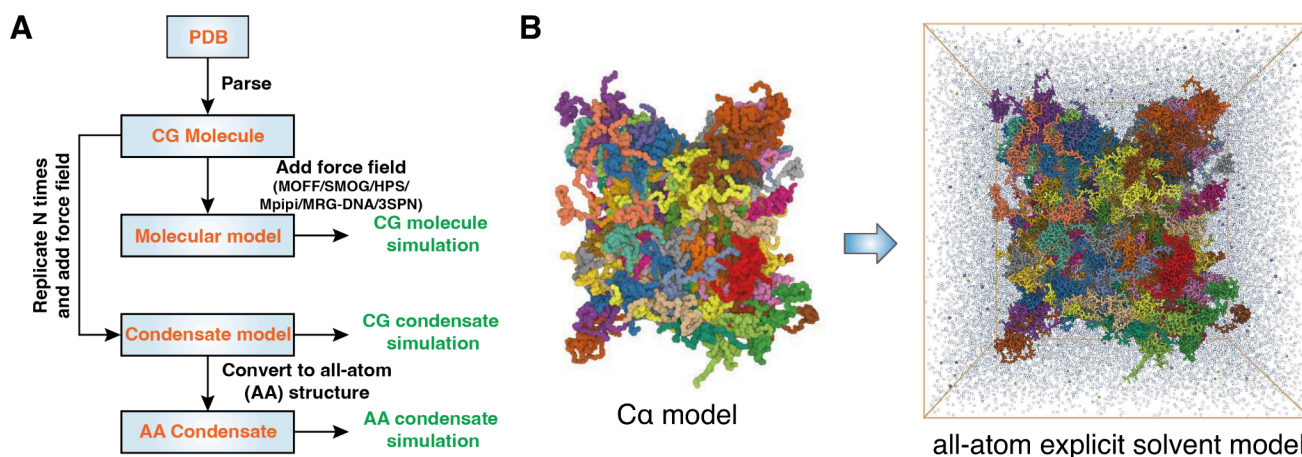


Figure 3. OpenABC supports CG and multiscale simulations of biomolecular condensates using various force fields. (A) Illustration of the OpenABC workflow, featuring Python scripting for both simulation setup and analysis. (B) OpenABC offers tools to convert Ca protein condensates into all-atom models, enabling simulations in explicit solvent environments. Figures adapted from ref 26.

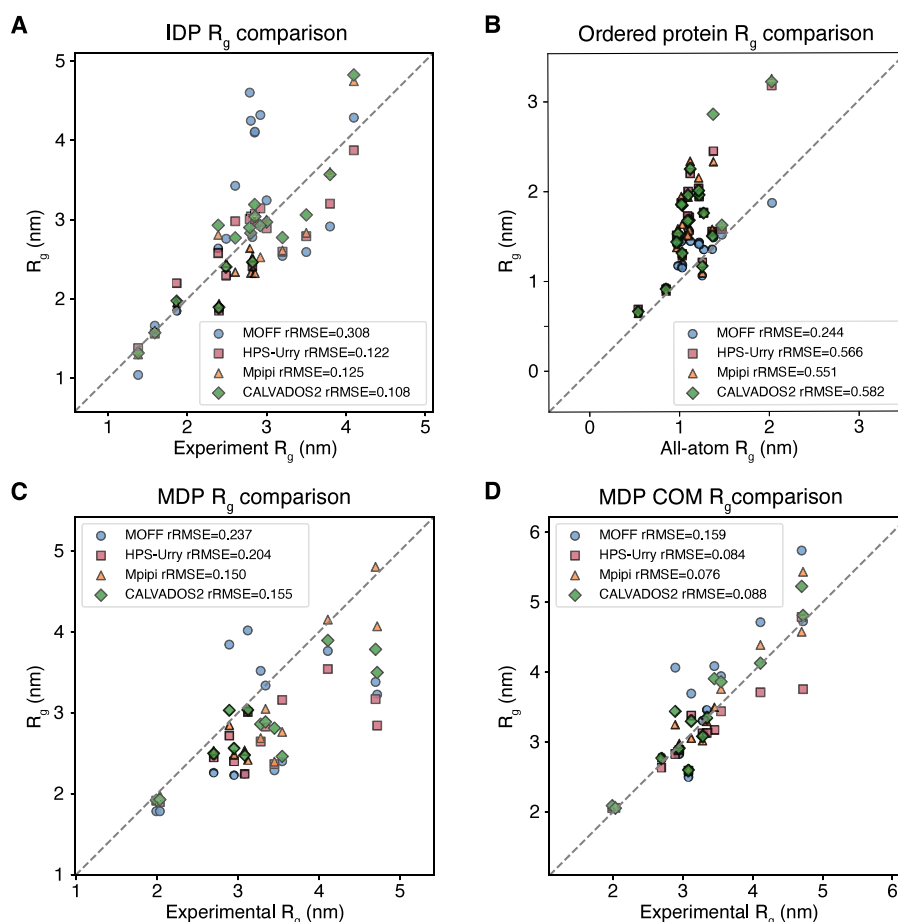


Figure 4. Benchmarking the performance of 1BPA models on single proteins, with a complete list of protein names provided in Table S1. (A) Comparison between simulated and experimental average R_g values for a selection of IDPs, computed using different models. Relative root-mean-square error (rRMSE, eq S1) values for each model are indicated in the legend. (B) Comparison between simulated and all-atom average R_g values for a variety of folded proteins computed with different models and structure-based terms to stabilize ordered domains. rRMSE values for each model are also indicated in the legend. Reference values were obtained from long-time atomistic simulations^{28,92,93} (Table S2). (C) Comparison between simulated and experimental average R_g values for multidomain proteins with CG beads placed at Ca positions. (D) Same comparison as (C), but with CG beads within ordered domains shifted to amino acid's center of mass.

importance of π – π and cation– π interactions in protein condensates, with arginines providing stronger attractions than lysines due to the guanidinium group (Figure 2), consistent

with diverse experiments.^{73–75} Recent advancements in Mpipi have led to the development of the Mpipi-GG⁶⁸ and Mpipi-Recharged⁶⁷ models. Lotthammer et al.⁶⁸ fine-tuned the

interaction parameters to achieve more accurate R_g predictions for a set of IDPs. Meanwhile, R. Tejedor et al.⁶⁷ substituted the Debye–Hückel potential with the Yukawa potential for electrostatic interactions, improving the model's performance for highly charged species by employing parameters derived from atomistic simulations.

Latham and Zhang⁶³ developed a 1BPA model, MOFF, incorporating a novel optimization algorithm. This algorithm integrates the maximum entropy principle^{76,77} to calibrate force field parameters against experimental data and draws on energy landscape theory, which necessitates an energy gap between folded and unfolded structures. Notably, the algorithm supports the inclusion of both folded and disordered proteins in the training set, producing balanced interaction parameters among amino acids. Consequently, MOFF is inherently transferable to both folded and disordered proteins, a capability absent in other models but essential for accurately predicting condensate stability. The simultaneous optimization across folded and disordered proteins yields more nuanced interaction patterns among amino acids compared to alternative models (see Figure 2 and Figure S1). Beyond the novel optimization approach, MOFF distinguishes itself from other 1BPA models by incorporating structure-based energy terms derived from reference structures, which enhance the representation of protein backbone conformations.

The computational efficiency of 1BPA models enables quantification of phase diagrams through slab simulations that investigate the coexistence of both condensed and dilute phases.^{22,23,25} Such quantification links protein sequence features with condensate emergent properties, allowing for the identification of sequence-specific characteristics that influence these properties under perturbative conditions. Simulation studies have identified the stabilizing contributions of nonspecific hydrophobic interactions, as well as hydrogen bonds, salt bridges, and π – π and cation– π interactions, in condensate formation.^{20,31} Amino acid specificity in these models enables differentiation in phase behaviors among homologues,⁶³ post-translational modifications, and various mutants,⁶² providing valuable insights into the regulation of condensate stability in normal cellular processes and the impact of pathogenic mutations.

■ BENCHMARKING COARSE-GRAINED MODELS AND STREAMLINING CONDENSATE SIMULATIONS USING OpenABC

The success of 1BPA models in providing mechanistic insights into condensate organization supports their broader adoption by the scientific community. However, selecting an appropriate model remains challenging due to the availability of multiple variants. These models are often implemented in diverse software platforms, such as HOOMD-Blue,⁷⁸ LAMMPS,⁷⁹ and GROMACS,⁸⁰ complicating performance comparisons. Additionally, many current implementations of CG models do not exploit recent advancements in computational hardware, limiting the efficiency and scalability of CG simulations.

To address these challenges, we developed a CG simulation toolkit, OpenABC²⁶ (Figure 3A), which integrates various protein 1BPA models, including HPS, MOFF, and Mpipi.^{22,62,63} OpenABC also supports the SMOG and 3SPN protein–DNA models,^{81–84} which have proven valuable for elucidating chromatin folding mechanisms.^{85–89}

The integration of diverse models within a single platform allows for direct benchmarking of their performance by

utilizing a consistent numerical integration scheme across different force fields. We evaluated the performance of two HPS models, HPS-Urry and CALVADOS2, alongside MOFF and Mpipi, across 18 disordered proteins. Simulation setups followed the OpenABC tutorials available online, with additional details provided in the Supporting Information. As shown in Figure 4A, all four models accurately reproduce the average R_g values of most disordered proteins. However, some of these proteins were involved in the training set of certain models, which may introduce bias into the evaluation. MOFF exhibits particularly poor performance on A1 variants, primarily due to its reliance on reference structures generated by ColabFold version AlphaFold2^{90,91} for constructing the structure-based potential (see Supporting Information for details). In contrast, other models do not depend on reference structures when simulating IDPs. The R_g values of the reference structures for A1 variants exceed 4.0 nm, substantially larger than the experimental average (<3.0 nm). Excluding A1 variants reduces the average rRMSE to 0.159.

Beyond IDPs, we also simulated a set of folded proteins, with distinct amino acid compositions from IDPs,⁹⁴ to evaluate whether the models accurately capture the collapse transition characteristic of these structures. To enable a consistent comparison, secondary structure potentials were incorporated into all models rather than being restricted to MOFF, as was done for IDPs (see Supporting Information for details). Importantly, no biases favoring long-range native contacts were introduced. The simulations specifically focused on contacts between secondary structure motifs, with their formation being driven primarily by nonbonded interactions in the various 1BPA models.

We found that most 1BPA models overestimate R_g (Figure 4B), while MOFF's results more closely align with values from all-atom simulations.^{28,92,93} We focused exclusively on the R_g and the overall size of the protein conformations rather than the accuracy of their tertiary structures. As illustrated in Figure S2, the root-mean-square deviation (RMSD) relative to the folded structure is large for all four models, although MOFF yields the smallest values. Given the simplicity of these models, it is unreasonable to expect accurate predictions of tertiary conformations.

Finally, we conducted simulations of multidomain proteins, which contain one or more folded domains connected by disordered linkers. In these simulations, the folded domains were restrained to maintain their native structures, in contrast to the simulations of single-domain folded proteins (see Supporting Information for details). The primary objective was to examine interdomain interactions rather than the conformational behavior of individual domains.

Most models in their default settings produced conformations that were more compact than experimental measurements (Figure 4C). Cao et al.⁶¹ reported that shifting the CG bead position in ordered domains from the α -carbon to the amino acid's center of mass improved R_g predictions. Consistent with this finding, we observed that moving the CG bead position to the heavy atom center of mass in ordered domains significantly enhanced the simulation accuracy (Figure 4D). Repositioning the beads likely reduced interdomain contacts,⁶¹ thereby reducing interaction strength.

OpenABC further facilitates CG condensate simulations through Python scripting and optimizes performance via OpenMM and GPU acceleration.⁹⁵ Remarkably, single-GPU performance can rival that of parallel processing on hundreds

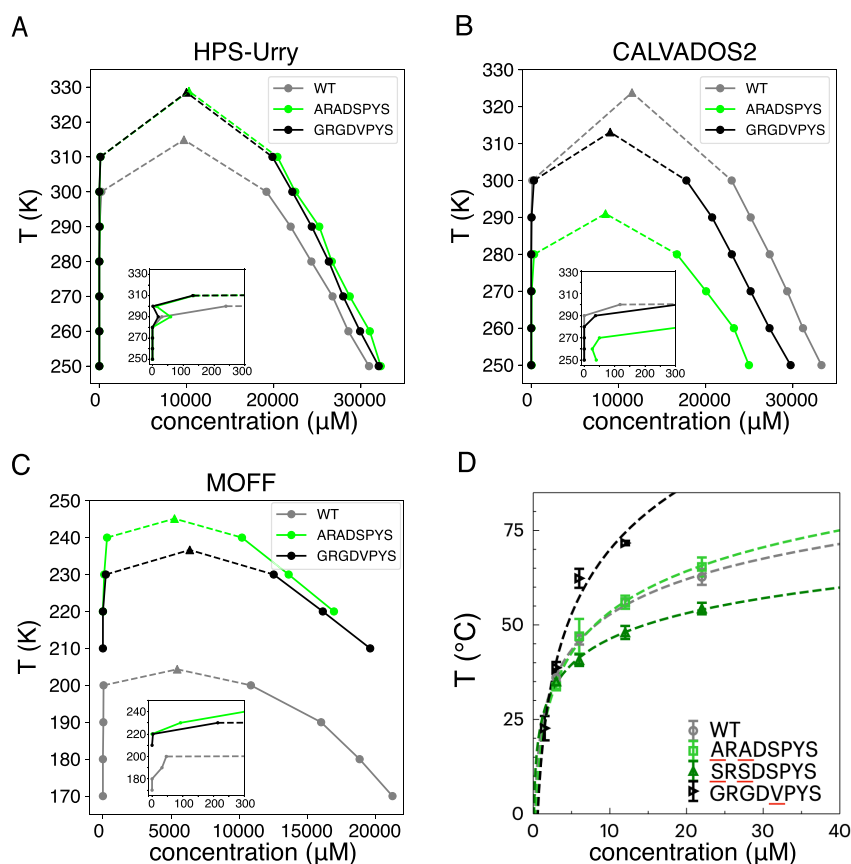


Figure 5. (A–C) Benchmarking the performance of IBPA models on condensates formed by (GRGDSPYS)₂₅ and its variants. The three panels present phase diagrams generated by using HPS-Urry, CALVADOS2, and MOFF, respectively. Legends denote the three condensates, each with its corresponding repeat sequence. (D) Experimental measurements of saturation concentrations from Rekhi et al.³¹ Figure adapted from ref 31.

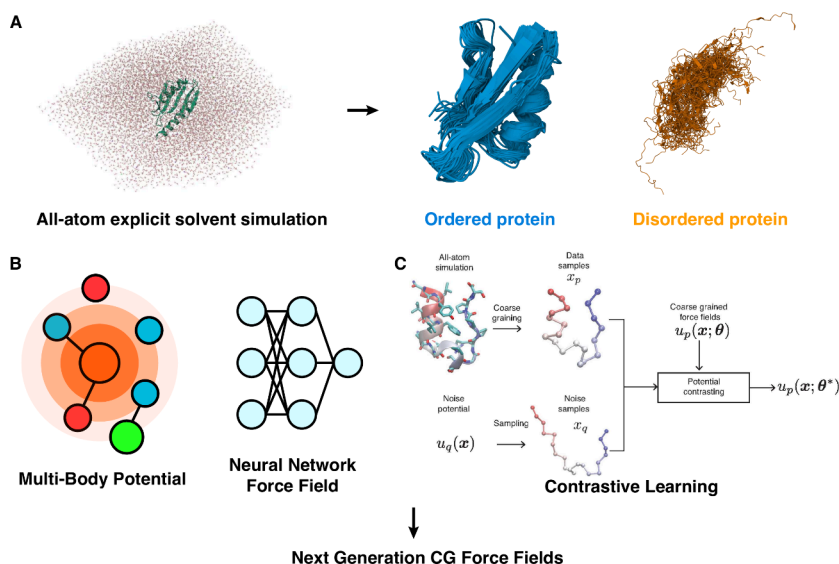


Figure 6. Illustration of the strategy for developing next-generation CG force fields. (A) This approach utilizes conformational ensembles of both folded and disordered proteins derived from atomistic simulations. (B) Multibody potentials, formulated analytically based on local density or employing neural networks, are crucial for enhancing force field accuracy. (C) Parameter optimization in these force fields can benefit from newly developed algorithms, such as potential contrasting, which efficiently exploit extensive training data sets. Figure adapted from ref 65.

of CPU cores.²⁶ Additionally, OpenABC enables the conversion of *Ca* representations to all-atom models (Figure 3B), supporting multiscale computational modeling.

The computational efficiency provided by OpenABC facilitates a systematic evaluation of the protein IBPA model

performance in predicting phase diagrams. While these models demonstrated similar results for single-chain properties of IDPs, they exhibited considerable divergence in condensate simulations. Specifically, the relative stability of condensates formed by the three synthetic IDPs, (GRGDVPYS)₂₅,

(GRGDSPYS)₂₅, and (ARADSPYS)₂₅, varied significantly across the models tested (Figure 5), underscoring the challenges in quantitatively predicting phase behavior.^{96,97} The phase diagram was not calculated using Mpipi, as the OpenABC implementation of the model requires double-precision to maintain simulation stability, which considerably impacts the computational efficiency.

■ ADVANCING THE ACCURACY AND TRANSFERABILITY OF COARSE-GRAINED MODELS

The inconsistent results among the different CG models, as presented in the previous section, underscore the need for further model development. We propose that advancing force fields for the next generation of CG models will require incorporating diverse training data sets, many-body interaction potentials, and novel optimization algorithms. These improvements, we argue, will enable CG models to more accurately predict phase diagrams and establish clearer connections between molecular sequences and the emergent behaviors of condensates. In the following sections, we explore each of these three directions in detail.

Training Data from Atomistic Simulations for Both Folded and Disordered Proteins. The parametrization of existing models frequently relies on low-resolution experimental data. A common example is the use of average radius of gyration (R_g) derived from small-angle X-ray scattering experiments⁹⁸ to fine-tune force field parameters.^{59–61,63,66,99} While such data are valuable, they provide insufficient details to fully characterize the energy landscape of proteins. As we have already shown in Figure 4A, different 1BPA models can all reproduce average R_g values for a broad set of disordered proteins, despite notable differences in their relative interaction strengths (Figure 2).

Obtaining additional high-resolution experimental data that offer greater structural insights can be challenging, if not impossible, particularly for IDPs due to the inherent difficulties in resolving their structures. On the other hand, atomistic force fields have made significant strides in accurately characterizing the structures of both folded and disordered proteins.^{27,28,93} Unlike experimental techniques, simulations with atomistic force fields can generate complete configurational ensembles rather than population averages, offering high-resolution structural information for proteins of interest. These structural ensembles, in principle, could serve as reference data to benchmark existing CG models and provide valuable training sets to further improve their accuracy (Figure 6A).

Besides improving structural resolution, it is equally important to include both ordered and disordered proteins in the training data.¹⁰⁰ Existing models specifically designed for IDPs often assess their performance solely on IDPs. However, when applied to folded proteins, these models frequently fail to collapse the structures, leading instead to extended conformations (Figure 4B).

When applied to folded domains, additional biasing potentials that stabilize tertiary structures are often incorporated into 1BPA models to prevent the expansion observed in Figure 4B. This raises the question of whether it is meaningful to assess whether these models predict collapsed or expanded conformations in the absence of tertiary biasing potentials given that such conditions fall outside their intended application. We argue that improving the accuracy of 1BPA models for folded proteins, particularly in predicting the

collapse transition, is crucial for reliably describing condensate stability. The interactions that stabilize the tertiary structures of folded proteins are chemically analogous to those that stabilize protein–protein interfaces.¹⁰¹ Many proteins that form condensates contain both folded and disordered regions,¹⁰² and interactions between folded domains are essential for maintaining condensate stability. Therefore, improving the accuracy of force fields for folded proteins could also enhance the predictions of protein–protein interactions and condensate stability.

At first glance, the multidomain protein results suggest that nonbonded interactions in 1BPA models are generally reasonable and may even be slightly stronger than necessary for describing protein–protein interactions. This is evident from their tendency to predict overly compact conformations without adjustment (Figure 4C). However, this interpretation contrasts with findings from single-domain folded proteins, where most 1BPA models exhibit interactions that are too weak to drive protein collapse (Figure 4B). One possible explanation is that protein–protein interactions in multidomain proteins are primarily nonspecific and weak, differing in magnitude from the specific interactions that stabilize protein complexes or induce collapse. Further studies are needed to assess the models' accuracy in predicting interaction strengths across a broader range of protein–protein interfaces.

Many-Body Potentials in CG Force Fields. Improving CG force fields requires the incorporation of flexible functional forms for the interaction potentials. Traditionally, the design of CG force fields has been based on the framework of atomistic force fields, which typically assume that nonbonded interactions can be pairwise decomposable. However, many-body interactions often arise as degrees of freedom are reduced.^{103–105} This occurs because configurational entropy is not factor-sensitive and cannot be accurately captured by pairwise interactions alone. Consequently, accounting for many-body interactions is crucial for improving the accuracy of the CG force fields.

A promising approach to modeling many-body interactions involves the use of neural networks or machine learning force fields.^{106–123} Neural networks predict the contribution of individual atoms to the total energy based on the spatial and chemical environment surrounding each atom. Since the numerical representation of these environments requires knowledge of the instantaneous positions of all neighboring atoms, neural network potentials naturally capture many-body effects. Furthermore, their high expressiveness enables neural networks to represent complex interactions without explicitly specifying their functional forms. Despite their promise, neural network-based force fields incur much higher computational costs for force evaluations^{109,124} and typically require reduced integration timesteps to address artificially stiff forces, thus limiting their practicality for simulating condensate systems.

Alternative, local density-dependent multibody terms have been proposed as transferable force fields in coarse-grained (CG) models.^{47,125,126} These analytical functional forms offer a compromise between flexibility and computational efficiency. Their application in CG protein force fields has already been demonstrated,^{47,127} yielding promising outcomes in predicting protein tertiary structures and protein–protein interfaces.^{101,128} Further refinement of these potentials may result in force fields that can be transferred across both folded and disordered proteins.

Algorithms for Optimizing Parameters in CG Force Fields. The parametrization of CG models, which involves a growing number of parameters and flexible functional forms, requires the application of systematic optimization algorithms, particularly when using atomistic simulation data. Several approaches have been developed to address this challenge,^{65,104,105,129,130} focusing on optimizing CG force field parameters to reproduce the potential of mean force, which governs the conformational distributions observed in atomistic simulations.

One such approach is force matching,¹⁰⁴ which seeks to align the mean force—i.e., the negative gradient of the potential of mean force—from all-atom simulations with the gradient of the CG force fields. This method has been applied to train numerous CG force fields, including those based on neural networks.^{107,111,131,132} A significant challenge in these methods is the accurate estimation of forces, particularly due to the limited number of samples available from all-atom simulations. To address this, Kohler et al.¹³³ introduced a novel method utilizing a generative model, specifically normalizing flows, to learn the conformational distribution. This model enables the efficient generation of statistically uncorrelated samples, which supports the accurate and convergent estimation of the mean forces.

Alternatively, maximum likelihood approaches, such as relative entropy minimization,¹²⁹ directly optimize force field parameters to reproduce the conformational distributions obtained from all-atom simulations. These methods aim to align the probability distribution produced by CG force fields with that of the atomistic simulations by minimizing relative entropy or Kullback–Leibler (KL) divergence,¹³⁴ which is equivalent to maximizing the log-likelihood of the model's probability distribution on the training data. For instance, Bottaro et al.¹³⁵ successfully applied the relative entropy method to optimize an implicit solvent model, effectively capturing the structure of disordered peptides. However, these methods tend to be computationally expensive due to the iterative sampling required for force field parameter updates. At each iteration, computational simulations are necessary to generate samples for ensemble averaging, making it difficult to apply these methods to large data sets, such as those involving multiple proteins, where individual simulations must be conducted for each protein.

The inefficiency of maximum likelihood methods, primarily due to the need for iterative sampling, is well documented within the machine learning community. To address this limitation, various approaches have been proposed to enhance their computational efficiency. These methods have also gained traction in the optimization of force fields.^{136,137}

For example, Ding and Zhang⁶⁵ developed the method potential contrasting for CG force field parametrization using noise contrastive learning^{138,139} (Figure 6C). Rather than maximizing the likelihood of the force field directly on the training data, noise contrastive learning incorporates noise samples and optimizes a loss function designed to distinguish true data from these noise samples. In the case of molecular systems, Ding and Zhang⁶⁵ demonstrated that advanced sampling techniques could generate effective noise samples, efficiently covering the relevant configuration space. The potential contrasting approach was shown to be highly effective, successfully parametrizing a CG 1BPA model with a neural network potential that reproduced the free energy

profile of protein folding, as determined from atomistic simulations.

This method's efficiency is particularly noteworthy, as it facilitates the use of large training data sets in the development of transferable models. For instance, Airas et al.¹⁰⁹ demonstrated that potential contrasting could be used to learn a transferable implicit solvent model capable of accurately reproducing the folding landscapes of multiple proteins.

In conclusion, recent advances in atomistic force field accuracy, along with developments in the design and optimization of many-body potentials, provide an ideal foundation for the integration of these elements. This integration can support the development of next-generation CG force fields that achieve a more effective balance between accuracy and computational efficiency for simulating biomolecular condensates.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.4c00737>.

rRMSE definition, simulation details, protein sequences, and supplementary figures illustrating the second Virial coefficients of various CG models and RMSD distributions of selected proteins, along with further tables detailing specific parameters (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Bin Zhang — Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; orcid.org/0000-0002-3685-7503; Email: binz@mit.edu

Authors

Shuming Liu — Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

Cong Wang — Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; orcid.org/0000-0001-5998-3740

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.biochem.4c00737>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant R35GM133580) and the National Science Foundation (Grant MCB-2042362).

■ REFERENCES

- (1) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285–298.
- (2) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.
- (3) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **2019**, *176*, 419–434.

- (4) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annual review of biophysics* **2020**, *49*, 107–133.
- (5) Latham, A. P.; Zhang, B. Molecular determinants for the layering and coarsening of biological condensates: Special Issue: Emerging Investigators. *Aggregate* **2022**, *3*, e306.
- (6) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 215–235.
- (7) Liu, S.; Athreya, A.; Lao, Z.; Zhang, B. From Nucleosomes to Compartments: Physicochemical Interactions Underlying Chromatin Organization. *Annual Review of Biophysics* **2024**, *53*, 221.
- (8) Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **2016**, *352*, 595–599.
- (9) Guo, Q.; Shi, X.; Wang, X. RNA and liquid-liquid phase separation. *Non-coding RNA Research* **2021**, *6*, 92–99.
- (10) Mehta, S.; Zhang, J. Liquid-liquid phase separation drives cellular function and dysfunction in cancer. *Nature Reviews Cancer* **2022**, *22*, 239–252.
- (11) Leicher, R.; Osunsade, A.; Chua, G. N. L.; Faulkner, S. C.; Latham, A. P.; Watters, J. W.; Nguyen, T.; Beckwitt, E. C.; Christodoulou-Rubalcava, S.; Young, P. G.; Zhang, B.; David, Y.; Liu, S. Single-Stranded Nucleic Acid Binding and Coacervation by Linker Histone H1. *Nature Structural & Molecular Biology* **2022**, *29*, 463–471.
- (12) Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. Extreme dynamics in a biomolecular condensate. *Nature* **2023**, *619*, 876–883.
- (13) Jawerth, L.; Fischer-Friedrich, E.; Saha, S.; Wang, J.; Franzmann, T.; Zhang, X.; Sachweh, J.; Ruer, M.; Ijavi, M.; Saha, S.; Mahamid, J.; Hyman, A. A.; Jülicher, F. Protein Condensates as Aging Maxwell Fluids. *Science* **2020**, *370*, 1317–1323.
- (14) Sood, A.; Zhang, B. Preserving condensate structure and composition by lowering sequence complexity. *Biophys. J.* **2024**, *123*, 1815.
- (15) Krainer, G.; Welsh, T. J.; Joseph, J. A.; Espinosa, J. R.; Wittmann, S.; De Csilléry, E.; Sridhar, A.; Toprakcioglu, Z.; Gudiškytė, G.; Czekalska, M. A.; Arter, W. E.; Guillén-Boixet, J.; Franzmann, T. M.; Qamar, S.; George-Hyslop, P. S.; Hyman, A. A.; Collepardo-Guevara, R.; Alberti, S.; Knowles, T. P. J. Reentrant Liquid Condensate Phase of Proteins Is Stabilized by Hydrophobic and Non-Ionic Interactions. *Nat. Commun.* **2021**, *12*, 1085.
- (16) MacKay, J. A.; Callahan, D. J.; FitzGerald, K. N.; Chilkoti, A. Quantitative model of the phase behavior of recombinant pH-responsive elastin-like polypeptides. *Biomacromolecules* **2010**, *11*, 2873–2879.
- (17) Ruff, K. M.; Roberts, S.; Chilkoti, A.; Pappu, R. V. Advances in understanding stimulus-responsive phase behavior of intrinsically disordered protein polymers. *Journal of molecular biology* **2018**, *430*, 4619–4635.
- (18) Saar, K. L.; Qian, D.; Good, L. L.; Morgunov, A. S.; Collepardo-Guevara, R.; Best, R. B.; Knowles, T. P. Theoretical and Data-Driven Approaches for Biomolecular Condensates. *Chem. Rev.* **2023**, *123*, 8988–9009.
- (19) Zhou, H.-X.; Kota, D.; Qin, S.; Prasad, R. Fundamental Aspects of Phase-Separated Biomolecular Condensates. *Chem. Rev.* **2024**, *124*, 8550.
- (20) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular details of protein condensates probed by microsecond long atomistic simulations. *J. Phys. Chem. B* **2020**, *124*, 11671–11679.
- (21) Wang, C.; Kilgore, H. R.; Latham, A. P.; Zhang, B. Nonspecific yet selective interactions contribute to small molecule condensate binding. *J. Chem. Theory Comput.* **2024**, *20*, 10247.
- (22) Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 9929–9934.
- (23) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology* **2018**, *14*, e1005941.
- (24) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. Temperature-controlled liquid-liquid phase separation of disordered proteins. *ACS central science* **2019**, *5*, 821–830.
- (25) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **2020**, *71*, 53–75.
- (26) Liu, S.; Wang, C.; Latham, A. P.; Ding, X.; Zhang, B. OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates. *PLoS Computational Biology* **2023**, *19*, e1011442.
- (27) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (28) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758–E4766.
- (29) Sarthak, K.; Winogradoff, D.; Ge, Y.; Myong, S.; Aksimentiev, A. Benchmarking molecular dynamics force fields for all-atom simulations of biological condensates. *J. Chem. Theory Comput.* **2023**, *19*, 3721–3740.
- (30) Latham, A. P.; Zhu, L.; Sharon, D. A.; Ye, S.; Willard, A. P.; Zhang, X.; Zhang, B. Microphase Separation Produces Interfacial Environment within Diblock Biomolecular Condensates. *eLife* **2025**, *12*, RP90750.
- (31) Rekhi, S.; Garcia, C. G.; Barai, M.; Rizuan, A.; Schuster, B. S.; Kiick, K. L.; Mittal, J. Expanding the molecular language of protein liquid-liquid phase separation. *Nat. Chem.* **2024**, *16*, 1113.
- (32) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, *1*.
- (33) López Barreiro, D.; Minten, I. J.; Thies, J. C.; Sagt, C. M. Structure-property relationships of elastin-like polypeptides: a review of experimental and computational studies. *ACS Biomaterials Science & Engineering* **2023**, *9*, 3796–3809.
- (34) Klein, I. A.; Boija, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall'Agnese, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R.; Sagi, I.; Clark, V. E.; Platt, J. M.; Kar, M.; McCall, P. M.; Zamudio, A. V.; Manteiga, J. C.; Coffey, E. L.; Li, C. H.; Hannett, N. M.; Guo, Y. E.; Decker, T. M.; Lee, T. I.; Zhang, T.; Weng, J. K.; Taatjes, D. J.; Chakraborty, A.; Sharp, P. A.; Chang, Y. T.; Hyman, A. A.; Gray, N. S.; Young, R. A. Partitioning of Cancer Therapeutics in Nuclear Condensates. *Science* **2020**, *368*, 1386–1392.
- (35) Kilgore, H. R.; Young, R. A. Learning the chemical grammar of biomolecular condensates. *Nat. Chem. Biol.* **2022**, *18*, 1298–1306.
- (36) Kilgore, H. R.; Mikhael, P. G.; Overholt, K. J.; Boija, A.; Hannett, N. M.; Van Dongen, C.; Lee, T. I.; Chang, Y.-T.; Barzilay, R.; Young, R. A. Distinct chemical environments in biomolecular condensates. *Nat. Chem. Biol.* **2024**, *20*, 291–301.
- (37) May, A.; Pool, R.; van Dijk, E.; Bijlard, J.; Abeln, S.; Heringa, J.; Feenstra, K. A. Coarse-grained versus atomistic simulations: realistic interaction free energies for real proteins. *Bioinformatics* **2014**, *30*, 326–334.
- (38) Dignon, G. L.; Zheng, W.; Mittal, J. Simulation methods for liquid-liquid phase separation of disordered proteins. *Current opinion in chemical engineering* **2019**, *23*, 92–98.
- (39) Wu, K.; Xu, S.; Wan, B.; Xiu, P.; Zhou, X. A novel multiscale scheme to accelerate atomistic simulations of bio-macromolecules by adaptively driving coarse-grained coordinates. *J. Chem. Phys.* **2020**, *152*, 114115.
- (40) Li, L.; Casalini, T.; Arosio, P.; Salvalaglio, M. Modeling the structure and interactions of intrinsically disordered peptides with multiple replica, metadynamics-based sampling methods and force-field combinations. *J. Chem. Theory Comput.* **2022**, *18*, 1915–1928.

- (41) De Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.
- (42) Benayad, Z.; von Bulow, S.; Stelzl, L. S.; Hummer, G. Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput.* **2021**, *17*, 525–537.
- (43) Zerze, G. H. Optimizing the martini 3 force field reveals the effects of the intricate balance between protein-water interaction strength and salt concentration on biomolecular condensate formation. *J. Chem. Theory Comput.* **2024**, *20*, 1646–1655.
- (44) Dawson, W. K.; Maciejczyk, M.; Jankowska, E. J.; Bujnicki, J. M. Coarse-grained modeling of RNA 3D structure. *Methods* **2016**, *103*, 138–156.
- (45) Korolev, N.; Nordenskiöld, L.; Lyubartsev, A. P. Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome. *Advances in colloid and interface science* **2016**, *232*, 36–48.
- (46) Sun, T.; Minhas, V.; Korolev, N.; Mirzoev, A.; Lyubartsev, A. P.; Nordenskiöld, L. Bottom-up coarse-grained modeling of DNA. *Frontiers in Molecular Biosciences* **2021**, *8*, 645527.
- (47) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.
- (48) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: a coarse-grained force field for intrinsically disordered proteins. *J. Phys. Chem. B* **2018**, *122*, 11115–11125.
- (49) Zhang, Y.; Liu, X.; Chen, J. Toward accurate coarse-grained simulations of disordered proteins and their dynamic interactions. *J. Chem. Inf. Model.* **2022**, *62*, 4523–4536.
- (50) Zhang, Y.; Li, S.; Gong, X.; Chen, J. Toward accurate simulation of coupling between protein secondary structure and phase separation. *J. Am. Chem. Soc.* **2024**, *146*, 342–357.
- (51) Bereau, T.; Deserno, M. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.* **2009**, *130*, 235106.
- (52) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J. F.; Kalimeri, M.; Barducci, A.; Laurin, Y.; Tek, A.; Baaden, M.; Nguyen, P. H.; Derreumaux, P. The OPEP Protein Model: From Single Molecules, Amyloid Formation, Crowding and Hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893.
- (53) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843–1853.
- (54) Das, R.; Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363–382.
- (55) Liwo, A.; Lee, J.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. Protein structure prediction by global optimization of a potential energy function. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 5482–5485.
- (56) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898–7936.
- (57) Chakraborty, D.; Mondal, B.; Thirumalai, D. Brewing COFFEE: ASequence-Specific Coarse-Grained Energy Function for Simulations of DNA- Protein Complexes. *J. Chem. Theory Comput.* **2024**, *20*, 1398–1413.
- (58) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Sci.* **2021**, *30*, 1371–1379.
- (59) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2111696118.
- (60) Tesei, G.; Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe* **2022**, *2*, 94.
- (61) Cao, F.; von Bülow, S.; Tesei, G.; Lindorff-Larsen, K. A coarse-grained model for disordered and multi-domain proteins. *Protein Sci.* **2024**, *33*, e5172.
- (62) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **2021**, *1*, 732–743.
- (63) Latham, A. P.; Zhang, B. Consistent force field captures homologue-resolved HP1 phase separation. *J. Chem. Theory Comput.* **2021**, *17*, 3134–3144.
- (64) Dannenhoffer-Lafage, T.; Best, R. B. A data-driven hydrophobicity scale for predicting liquid-liquid phase separation of proteins. *J. Phys. Chem. B* **2021**, *125*, 4046–4056.
- (65) Ding, X.; Zhang, B. Contrastive learning of coarse-grained force fields. *J. Chem. Theory Comput.* **2022**, *18*, 6334–6344.
- (66) Ding, X. Optimizing Force Fields with Experimental Data Using Ensemble Reweighting and Potential Contrasting. *J. Phys. Chem. B* **2024**, *128*, 6760.
- (67) R Tejedor, A.; Aguirre Gonzalez, A.; Maristany, M. J.; Chew, P. Y.; Russell, K.; Ramirez, J.; Espinosa, J. R.; Collepardo-Guevara, R. Chemically Informed Coarse-Graining of Electrostatic Forces in Charge-Rich Biomolecular Condensates. *ACS Central Science* **2025**, *11*, 302.
- (68) Lotthammer, J. M.; Ginell, G. M.; Griffith, D.; Emenecker, R. J.; Holehouse, A. S. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat. Methods* **2024**, *21*, 465–476.
- (69) Kapcha, L. H.; Rossky, P. J. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *Journal of molecular biology* **2014**, *426*, 484–498.
- (70) Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers: Original Research on Biomolecules* **1992**, *32*, 1243–1250.
- (71) Norgaard, A. B.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Experimental parameterization of an energy function for the simulation of unfolded proteins. *Biophysical journal* **2008**, *94*, 182–192.
- (72) Bottaro, S.; Bengtson, T.; Lindorff-Larsen, K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. *Structural bioinformatics: methods and protocols* **2020**, *2112*, 219–240.
- (73) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **2018**, *7*, e31486.
- (74) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688.
- (75) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699.
- (76) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445–3451.
- (77) Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, 084107.
- (78) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* **2020**, *173*, 109363.
- (79) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In'T Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

- (80) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (81) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. *Journal of molecular biology* **2000**, *298*, 937–953.
- (82) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: a versatile software package for generating structure-based models. *PLoS computational biology* **2016**, *12*, e1004794.
- (83) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013**, *139*, 144903.
- (84) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; De Pablo, J. J. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* **2014**, *141*, 165103.
- (85) Ding, X.; Lin, X.; Zhang, B. Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface. *Nat. Commun.* **2021**, *12*, 1091.
- (86) Liu, S.; Lin, X.; Zhang, B. Chromatin fiber breaks into clutches under tension and crowding. *Nucleic Acids Res.* **2022**, *50*, 9738–9747.
- (87) Qiu, Y.; Liu, S.; Lin, X.; Unarta, I. C.; Huang, X.; Zhang, B. Nucleosome condensate and linker DNA alter chromatin folding pathways and rates. *bioRxiv* **2024**, 1.
- (88) Lin, X.; Zhang, B. Explicit ion modeling predicts physicochemical interactions for chromatin organization. *eLife* **2024**, *12*, RP90073.
- (89) Li, R.; Lin, X. Connected Chromatin Amplifies Acetylation-modulated Nucleosome Interactions. *bioRxiv* **2024**, 1.
- (90) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (91) Mirdita, M.; Schütze, K.; Moriawaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682.
- (92) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.
- (93) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein-protein complexes. *J. Chem. Theory Comput.* **2020**, *16*, 2494–2507.
- (94) Zhang, Y.; Zheng, J.; Zhang, B. Protein Language Model Identifies Disordered, Conserved Motifs Driving Phase Separation. *bioRxiv* **2024**, 1.
- (95) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simonnet, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Computational Biology* **2017**, *13*, e1005659.
- (96) Pal, T.; Wessén, J.; Das, S.; Chan, H. S. Differential Effects of Sequence-Local versus Nonlocal Charge Patterns on Phase Separation and Conformational Dimensions of Polyampholytes as Model Intrinsically Disordered Proteins. *J. Phys. Chem. Lett.* **2024**, *15*, 8248–8256.
- (97) Feito, A.; Sanchez-Burgos, I.; Rey, A.; Collepardo-Guevara, R.; Espinosa, J. R.; Tejedor, A. R. Capturing Single-Molecule Properties Does Not Ensure Accurate Prediction of Biomolecular Phase Diagrams. *Mol. Phys.* **2024**, *122*, e2425757.
- (98) Li, T.; Senesi, A. J.; Lee, B. Small angle X-ray scattering for nanoparticle research. *Chem. Rev.* **2016**, *116*, 11128–11180.
- (99) Latham, A. P.; Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *J. Chem. Theory Comput.* **2020**, *16*, 773–781.
- (100) Latham, A. P.; Zhang, B. Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 63–70.
- (101) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19244–19249.
- (102) Van Der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589–6631.
- (103) Dyson, H. J.; Wright, P. E.; Scheraga, H. A. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13057–13061.
- (104) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114.
- (105) Jin, J.; Pak, A. J.; Durumeric, A. E.; Loose, T. D.; Voth, G. A. Bottom-up coarse-graining: Principles and perspectives. *J. Chem. Theory Comput.* **2022**, *18*, S759–S791.
- (106) Schütt, K.; Kindermans, P.-J.; Saucedo Felix, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in neural information processing systems*; 2017; Vol. 30.
- (107) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; De Fabritiis, G.; Noé, F.; Clementi, C. Machine learning of coarse-grained molecular dynamics force fields. *ACS central science* **2019**, *5*, 755–767.
- (108) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
- (109) Airas, J.; Ding, X.; Zhang, B. Transferable Implicit Solvation via Contrastive Learning of Graph. *Neural Networks. ACS Central Science* **2023**, *9*, 2286–2297.
- (110) Airas, J.; Zhang, B. Scaling Graph Neural Networks to Large Proteins. *J. Chem. Theory Comput.* **2025**, *21*, 2055.
- (111) Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N. E.; Giorgino, T.; Husic, B. E.; Clementi, C.; Noé, F.; De Fabritiis, G. Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **2023**, *14*, 5739.
- (112) Tu, H.; Han, Y.; Wang, Z.; Chen, A.; Tao, K.; Ye, S.; Wang, S.; Wei, Z.; Li, J. RotNet: A Rotationally Invariant Graph Neural Network for Quantum Mechanical Calculations. *Small Methods* **2024**, *8*, 2300534.
- (113) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.
- (114) Zeng, J.; Zhang, D.; Lu, D.; Mo, P.; Li, Z.; Chen, Y.; Rynik, M.; Huang, L.; Li, Z.; Shi, S.; Wang, Y.; Ye, H.; Tuo, P.; Yang, J.; Ding, Y.; Li, Y.; Tisi, D.; Zeng, Q.; Bao, H.; Xia, Y.; Huang, J.; Muraoka, K.; Wang, Y.; Chang, J.; Yuan, F.; Bore, S. L.; Cai, C.; Lin, Y.; Wang, B.; Xu, J.; Zhu, J.-X.; Luo, C.; Zhang, Y.; Goodall, R. E. A.; Liang, W.; Singh, A. K.; Yao, S.; Zhang, J.; Wentzcovitch, R.; Han, J.; Liu, J.; Jia, W.; York, D. M.; E, W.; Car, R.; Zhang, L.; Wang, H. DeePMD-kit v2: A Software Package for Deep Potential Models. *J. Chem. Phys.* **2023**, *159*, 054801.
- (115) Unke, O. T.; Chmiela, S.; Saucedo, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chem. Rev.* **2021**, *121*, 10142–10186.
- (116) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Ketten, S.; Gomez-Bombarelli, R.; Jaakkola, T. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv* **2022**, 1.

- (117) Wang, Y.; Wang, T.; Li, S.; He, X.; Li, M.; Wang, Z.; Zheng, N.; Shao, B.; Liu, T.-Y. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nat. Commun.* **2024**, *15*, 313.
- (118) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **2019**, *31*, 1902765.
- (119) Gkeka, P.; Stoltz, G.; Barati Farimani, A.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H.; Peter, C.; Pietrucci, F.; Silveira, A.; Tkatchenko, A.; Trstanova, Z.; Wiewiora, R.; Lelièvre, T. Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757–4775.
- (120) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **2020**, *153*, 124111.
- (121) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*; 2022; Vol. 35, pp 11423–11436.
- (122) Thölke, P.; De Fabritiis, G. TorchMD-NET: equivariant transformers for neural network based molecular potentials. *arXiv* **2022**, 1.
- (123) Takaba, K.; Friedman, A. J.; Cavender, C. E.; Behara, P. K.; Pulido, I.; Henry, M. M.; MacDermott-Opeskin, H.; Iacovella, C. R.; Nagle, A. M.; Payne, A. M.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Wang, Y. Machine-Learned Molecular Mechanics Force Fields from Large-Scale Quantum Chemical Data. *Chemical Science* **2024**, *15*, 12861–12878.
- (124) Airas, J.; Zhang, B. Scaling Graph Neural Networks to Large Proteins. *J. Chem. Theory Comput.* **2025**, *21*, 2055.
- (125) Sanyal, T.; Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *J. Chem. Phys.* **2016**, *145*, 034109.
- (126) DeLyser, M. R.; Noid, W. Analysis of local density potentials. *J. Chem. Phys.* **2019**, *151*, 224106.
- (127) Liwo, A.; Baranowski, M.; Czaplowski, C.; Gołaś, E.; He, Y.; Jagiela, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; Niazdvedtski, A.; Oldziej, S.; Scheraga, H. A.; Sieradzan, A. K.; Ślusarz, R.; Wirecki, T.; Yin, Y.; Zaborowski, B. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20*, 2306.
- (128) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* **2016**, *138*, 8126–8133.
- (129) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.
- (130) Dhamankar, S.; Webb, M. A. Chemically specific coarse-graining of polymers: Methods and prospects. *J. Polym. Sci.* **2021**, *59*, 2613–2643.
- (131) Peng, Y.; Pak, A. J.; Durumeric, A. E.; Sahrman, P. G.; Mani, S.; Jin, J.; Loose, T. D.; Beiter, J.; Voth, G. A. OpenMSCG: A software tool for bottom-up coarse-graining. *J. Phys. Chem. B* **2023**, *127*, 8537–8550.
- (132) Chen, Y.; Krämer, A.; Charron, N. E.; Husic, B. E.; Clementi, C.; Noé, F. Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys.* **2021**, *155*, 084101.
- (133) Kohler, J.; Chen, Y.; Kramer, A.; Clementi, C.; Noé, F. Flow-matching: Efficient coarse-graining of molecular dynamics without forces. *J. Chem. Theory Comput.* **2023**, *19*, 942–952.
- (134) Kullback, S.; Leibler, R. A. On information and sufficiency. *annals of mathematical statistics* **1951**, *22*, 79–86.
- (135) Bottaro, S.; Lindorff-Larsen, K.; Best, R. B. Variational optimization of an all-atom implicit solvent force field to match explicit solvent simulation data. *J. Chem. Theory Comput.* **2013**, *9*, 5641–5652.
- (136) Varnai, C.; Burkoff, N. S.; Wild, D. L. Efficient parameter estimation of generalizable coarse-grained protein force fields using contrastive divergence: a maximum likelihood approach. *J. Chem. Theory Comput.* **2013**, *9*, 5718–5733.
- (137) Arts, M.; Garcia Satorras, V.; Huang, C.-W.; Zugner, D.; Federici, M.; Clementi, C.; Noé, F.; Pinsler, R.; van den Berg, R. Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2023**, *19*, 6151–6159.
- (138) Gutmann, M.; Hyvärinen, A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*; 2010; pp 297–304.
- (139) Gutmann, M. U.; Hyvärinen, A. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of machine learning research* **2012**, *13*, 307–361.