npj computational materials

www.nature.com/npjcompumats

ARTICLE OPEN

Check for updates

A machine learning enabled hybrid optimization framework for efficient coarse-graining of a model polymer

Zakiya Shireen[1], Hansani Weeratunge[1], Adrian Menzel[2], Andrew W. Phillips[2], Ronald G. Larson[3], Kate Smith-Miles[4] and Elnaz Hajizadeh[1]✉

This work presents a framework governing the development of an efficient, accurate, and transferable coarse-grained (CG) model of a polyether material. The framework combines bottom-up and top-down approaches of coarse-grained model parameters by integrating machine learning (ML) with optimization algorithms. In the bottom-up approach, bonded interactions of the CG model are optimized using deep neural networks (DNN), where atomistic bonded distributions are matched. In the top-down approach, optimization of nonbonded parameters is accomplished by reproducing the temperature-dependent experimental density. We demonstrate that developed framework addresses the thermodynamic consistency and transferability issues associated with the classical coarse-graining approaches. The efficiency and transferability of the CG model is demonstrated through accurate predictions of chain statistics, the limiting behavior of the glass transition temperature, diffusion, and stress relaxation, where none were included in the parametrization process. The accuracy of the predicted properties are evaluated in context of molecular theories and available experimental data.

npj Computational Materials (2022)8:224 ; https://doi.org/10.1038/s41524-022-00914-4

INTRODUCTION

Molecular dynamics (MD) simulation techniques provide a powerful route to establish the structure–property relationships in materials through solving the coupled equations of motions of interacting atoms and molecules in a material system[1–3]. Despite the great success of this computational technique, modeling of the macromolecules, such as proteins and polymers across multiple lengths and time scales is restrained by computational limitations. To overcome the challenges of modeling a macromolecular system such as polymers for longer time scales, coarse-grained models (CG) are required. One of the key features of coarse-graining is the ability to probe polymer behavior over large spatiotemporal scales, which is otherwise difficult to achieve in high-resolution models.

The central problem in coarse-graining polymeric materials is, to retain the chain attributes of interest while constructing the adequate representation of pseudo-atom or CG bead. One of the defining characteristics of polymer materials is the long-time dynamical response that occurs at multiple length scales. Coarse-graining allows the combining of groups of atoms into fewer interaction sites, thus, reducing the degrees of freedom in the system. In addition, the accurate representation requires encoding of finer details, i.e., the chemical specificity should be incorporated into the CG bead descriptor. The CG model can be made to mimic the atomistic structural features by optimizing the interaction parameters by matching the atomistic distributions, the so-called bottom-up approach. Contrary, in the top-down strategy the chemical specificity is subsumed into the descriptor by matching the macroscopic properties such as density (ρ), glass transition temperature (T_g), elastic modulus (E), etc.[4].

The classical methods such as iterative Boltzmann inversion (IBI)[5–8], inverse Monte Carlo[9,10], and relative entropy[11,12] typically aim to map the structural distribution of the atomistic model. However, these structure-based conventional methods have limitations in capturing the correct thermodynamic properties and free energy landscape. While the force-matching methods such as the many-body potential of mean force (PMF) capture the accurate dynamics, they represent the atomistic structural features inaccurately[13–15]. Optimization of parameters defining nonbonded interactions in a bottom-up manner i.e., by optimizing the radial distribution functions (RDF), usually results in larger deviations of the thermal expansion coefficient (α). Hence, a purely bottom-up optimized CG model does not demonstrate thermomechanical consistency, therefore, temperature-dependent density transferability cannot be guaranteed a priori[5,16,17]. Therefore, transferability is achieved when CG potentials are able to accurately predict the structure, thermodynamics properties, and dynamics of the material for temperatures and molecular weights not originally included in the parametrization process. The hybrid approach, i.e., combining the bottom-up and top-down strategies, using conventional coarse-graining methodologies has been developed for a couple of materials systems[4,18]. For instance, the coarse-graining of polystyrene was studied by adopting the top-down strategy, where nonbonded potential parameters were optimized by increasing the pairwise potential energy well depth. The increased well depth reintroduces a rough additive potential energy landscape contrary to the usual smooth potential energy landscape of CG models[4]. Furthermore, using liquid perturbation theory, it has been demonstrated that for common polyethers the nonbonded interactions in the coarse-grained models are temperature-dependent[18]. Therefore, the temperature dependency needs to be accounted for in the coarse-graining strategy.

Although the conventional CG models are straightforward and allow much larger systems with reasonable computation costs[19,20], they still suffer from some important issues in terms of thermomechanical consistency, representability, and

[1]Department of Mechanical Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Melbourne, Australia. [2]Platforms Division, Defence Science and Technology Group, Melbourne, Australia. [3]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA. [4]School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia. ✉email: ellie.hajizadeh@unimelb.edu.au

transferability[13,21]. Despite the fact that the development of fully transferable CG potentials that could capture all material properties of interest across the entire desired range of temperatures remains a formidable challenge, significant efforts have been made to better understand the various coarse-graining mechanisms. Prior research has proposed several strategies to develop temperature-transferable CG models, including a multiscale CG (MS-CG) approach[14,22–24], optimization of nonbonded interactions using multiple macroscopic properties[4], an energy-renormalization approach, etc.[25–27]. Noid and his co-worker's impressive "oeuvre" includes a spectrum of bottom-up strategies of multiscale coarse graining[14,22–24,28,29]. In recent reports, the state-point dependence of the CG potentials is examined for molecular liquids[30] and a dual approach has been considered that combines structure-based and energy-based variational principles to determine effective potentials[31]. Keten and his co-workers have done extensive work, particularly focusing on reproducing the all-atom (AA) dynamics from CG potentials over a wide range of temperatures, using an energy-renormalization approach[25–27] for nonbonded optimization. Therefore, the classical hybrid approach has improved the accuracy of the CG model predictions, but still creates formidable challenges, such as the need for careful fine-tuning of nonbonded *LJ* potential parameters (well depth ϵ) and variations in auxiliary *LJ* potential terms. The fine-tuning of pairwise interaction parameters can be very demanding and time-consuming for complex block copolymers and biological macromolecules[32–35] due to multiple pairwise interaction sites and the inherent iterative process of the optimization.

To address the questions pertaining to generic coarse-graining strategies of complex molecules and accelerate the process, data-driven and machine-learned (ML) potentials are emerging as an alternative approach. The machine-learned CG potentials have been proposed to present the free-energy landscape of the all-atom (AA) molecular model with enhanced efficiency and accuracy[36]. Thus, the ML-enabled CG potentials are becoming a tool to bridge the gap between accurate and computationally expensive all-atom methods and enable the development of approximate but computationally affordable CG methods[37–45]. Since the accuracy of the machine learning model depends intrinsically on the accuracy of the reference model, improved accuracy has been sought by using ab-initio reference models to obtain machine-learned CG potentials that can predict the properties of complex molecules[46–48].

To reduce the computational cost, machine learning and optimization techniques such as particle swarm optimization (PSO), genetic algorithm (GA), and simplex method[39,49,50] were previously integrated to enhance the efficiency of these methods. Therefore, the emerging data-driven technique[51,52] is promising and is credible to achieve accuracy with enhanced efficiency compared to the conventional approaches.

The potential of machine learning algorithms such as neural networks lies in capturing the hidden representations of complex data. They are proven to be capable of formulating complex potentials using AA molecular dynamics simulation data as reference or ground truth[53]. Similar to classical CG potentials, the machine-learned CG (MLCG) methods are also categorized as bottom-up MLCG methods[39,40,49] and top-down MLCG methods[54–58]. Generally, in the bottom-up MLCG methods, a CG potential that is related to representations of CG beads is introduced through multi-layer neural networks[53]. Then the neural network is trained and optimized by utilizing the reference system i.e., data set from the all-atom molecular simulations. In the top-down MLCG method, an empirical relation of the interactions between CG beads (e.g. LJ potential for a simple fluid) is given as a priori[39,40,54]. The interaction parameters in the empirical relation are then optimized by a statistical learning technique to match features such as radial distribution function and temperature-dependent volumetric behavior with corresponding simulations/experimental results.

In this work, we propose a framework by combining bottom-up and top-down MLCG methods that enable thermodynamic consistency and transferability, while maintaining structural representability. Further, we demonstrate the application of the proposed method by coarse-graining poly(tetramethylene oxide) (PTMO), which is one of the major components of elastomeric polyurethanes. The coarse-grained PTMO model represents the soft segment of the polyurethane (PU) chain. The framework involves the integration of MLCG methods with genetic algorithm (GA) and molecular dynamics (MD) simulations. In the developed protocol the bottom-up MLCG approach is used to establish the bonded interactions by training deep neural networks (DNN) and top-down MLCG is used by incorporating a GA optimization scheme to obtain nonbonded interaction parameters by matching temperature-dependent density. One of the key highlights of the proposed framework is that the MD simulations are replaced by a DNN that is trained to predict the temperature-dependent density to further accelerate the optimization component. Therefore, we avoid the need for a large number of simulations, where data are generally not reused. Thus, DNNs are used as surrogate models, which addresses the re-usability issue of the data[59]. Our focus is on the working mechanism of the framework by demonstrating the ability of the CG model to make accurate predictions, and the validation of the trained model. The CG potentials developed in this work provide a statistical mechanical description of the system that is consistent with properties derived directly from a high-resolution united-atom ensemble; however, they may not be completely transferable across different thermodynamic conditions, which we discuss in the conclusion section as a task for future work.

RESULTS

CG model predictions and validations

The performance of DNN accelerated CG potentials is examined by investigating universality and temperature-transferability in two ways. The first way is to compute properties, which have not been included in the training. Second is to demonstrate transferability by the application of the potential parameters for temperatures and molecular weights, which are outside the range of the training set. The machine-learned coarse-grained systems (where N_b is the number of beads/chains and N_c is the number of chains) studied in this work are listed in Table 1 along with the ground truth data.

We first study the structural and conformational properties of the machine-learned coarse-grained model of PTMO melts through MD simulations. The distribution of the radius of gyration R_g of CG and UA models are compared in Fig. 1a for $N_b = 50$. It is to underline that chain size with 50 beads was included in the

Table 1. Machine-learned coarse-grained (MLCG) PTMO systems studied in this work at $T = 453$ K.

Model	M_n (kDa)	N_b (beads/chain)	Chains (N_c)	ρ (g/cm^3)
MLCG	1.8	25	1000	0.868
MLCG	3.6	50	500	0.873
MLCG	7.2	100	250	0.875
MLCG	14.4	200	125	0.876
MLCG	25.2	350	72	0.877
MLCG	36	500	100	0.878
MLCG	72	1000	100	0.879
UA	3.6	50	50	0.875

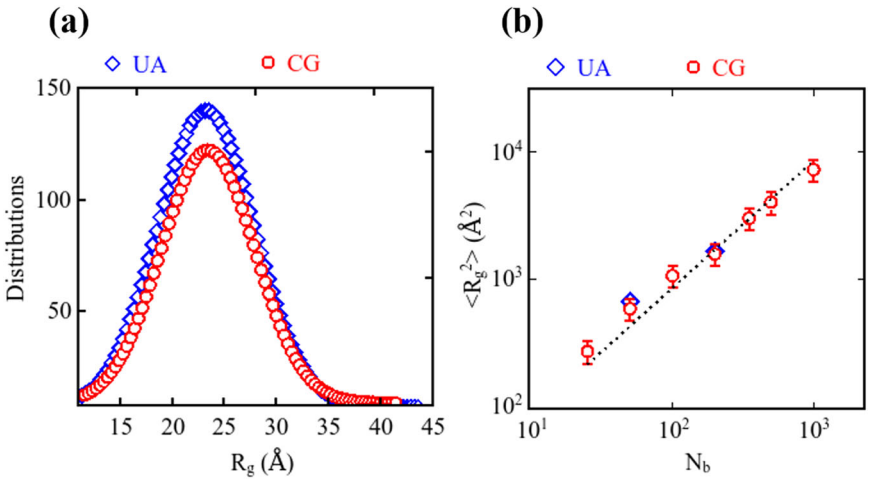

Fig. 1 Resulting chain statistics and structural comparison is shown. a Distribution of radius of gyration R_g is shown for MLCG and UA system for the chain size $N_b = N_m = 50$. **b** The average-mean-squared radius of gyration $\langle R_g^2 \rangle$ as a function of chain length N_b ($T = 453$ K). The diamond symbols represent data from united-atom systems. The dashed line indicates a fit of the power law, $\langle R_g^2 \rangle \propto {N_b}^{2\nu}$ with $\nu = 1/2$ for ideal chains.

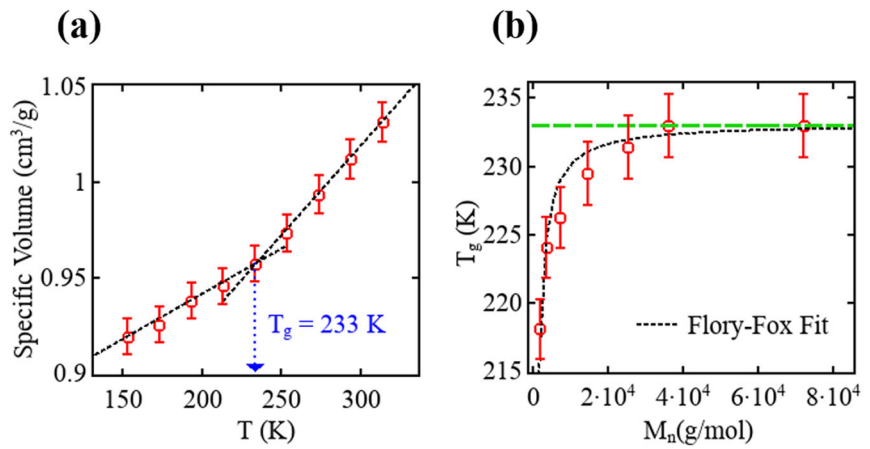

Fig. 2 Prediction of glass transition temperature (T_g) from MLCG model. a The glass transition temperature is estimated from the specific volume at $N_b = 500$. **b** The limiting behavior of the PTMO is shown by fitting the Flory–Fox equation to the predicted T_g from the developed MLCG potential and shown in a dotted line. The dashed line shows the experimental T_g.

training data for potential parametrization. The estimated root mean square of the radius of gyration $\langle R_g^2 \rangle^{1/2} = 24.56 \pm 1.50$ and 25.86 Å from CG and UA systems, respectively. The gyration distributions from the two models are in good agreement. We also analyzed the chain dimensions of the coarse-grained PTMO for all the systems tabulated in Table 1. Figure 1b visualizes the averaged-mean-squared radius of gyration $\langle R_g^2 \rangle$, as a function of chain length N_b. Note that $N_b = N_m$ with N_m being the number of monomers (repeat units) in the UA system. It can be seen that with an increasing chain length of PTMO in the melt, $\langle R_g^2 \rangle$ is approaching random coil hypothesis predictions (linear N_b dependence, shown by dashed line). Figure 1b represents the chain statistics well above and below the chain size ($N_b = N_m = 50$), which was included in training and testing. $\langle R_g^2 \rangle$ for $N_b = 200$ is in good agreement with the chain size of the united atom model at $N_m = 200$, which indicates that the MLCG potential captures the structural features very well.

Figure 2 represents the temperature-dependent volumetric behavior along with predicted limiting behavior in terms of glass transition temperature (T_g) of the PTMO system. It is important to note that the volumetric behavior shown in Fig. 2a is from the temperature range 313–153 K, which was not included in the nonbonded potential parametrization. The dotted arrow in Fig. 2a indicates the predicted $T_g = 233$ K for chain length $N_b = 500$. Figure 2b visualizes the limiting behavior of PTMO melt predicted by MLCG potentials. The predicted glass transition temperatures lie within the range of experimental estimates as tabulated in Table 2. The limiting behavior is identified by fitting the Flory–Fox relationship to the predicted T_g for the range of molecular weight. The dashed line is to show the experimental observation[60] and the dotted line represents the Flory–Fox fit given by the equation:

$$T_g(M_n) = T_{g\infty}(1 - K/M_n) \quad (1)$$

Table 2. Glass transition temperature predicted by MLCG systems at different molecular weights.

M_n (g/mol)	T_g (K)	References
1800−72,000	218.14−233	This work
3500−10,200	198−233	[60]
44,000−44,100	198−207	[73]

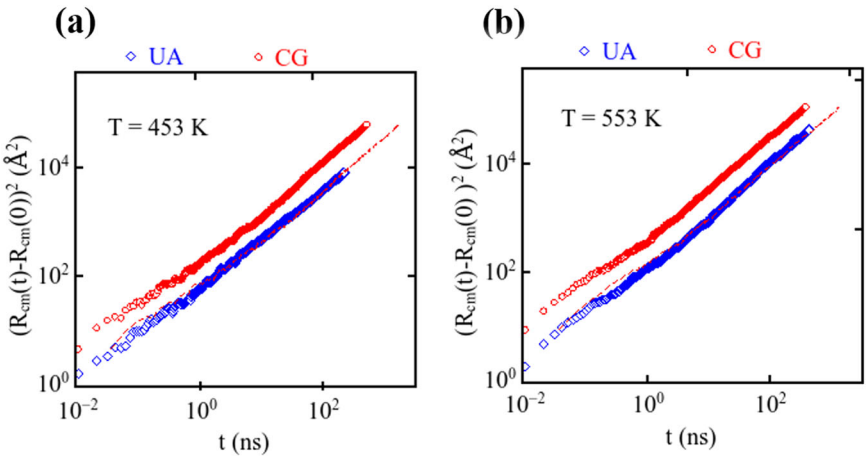

Fig. 3 Dynamic scaling and diffusion comparison between UA and MLCG model is shown. The mean squared displacement (MSD) of the center of mass of CG-PTMO and UA chains are computed at **a** $T = 453$ K and **b** $T = 553$ K. The dashed lines represent the time-scaled MSD of the CG system with scaling factors $S_{UACG} = 3.35$ and 3.55 at 453 and 553 K, respectively.

where $T_{g\infty}$ is the glass transition temperature for PTMO chains approaching infinite length and K is an empirical constant and is equal to 130.0 g/mol from Fig. 2b. We consider the model to be "thermodynamically consistent" if it accurately predicts T_g at higher molecular weights. Therefore, the results shown in Fig. 2 are an indication of the transferability of the developed potentials, using our hybrid approach for different thermodynamic states.

To account for the faster dynamics in CG systems, Fritz et al. proposed a dynamic scaling factor, S_{AACG} to obtain the quantitative agreement between dynamics in atomistic and CG simulations for one component system. The time scaling factor is the ratio of diffusion coefficients of CG and AA system such that $S_{AACG} = D_{CG}/D_{AA}$ and can also be estimated using the mean squared displacement of beads[61]. We estimate the dynamic scaling factor S_{UA-CG} from the linear part of the mean squared displacement of the center of mass ($\langle\langle(R_{cm}(t) - R_{cm}(0))^2\rangle\rangle$) as a function of time using the Einstein equation:

$$D = \lim_{t\to\infty} \frac{1}{6t} \langle (R_{cm}(t) - R_{cm}(0))^2 \rangle \quad (2)$$

The mean squared displacement of CG and UA system is compared and the dynamic scaling factor (S_{UA-CG}) is computed for $N_b = 50$ at $T = 453$ K (chain length and temperature included in training set) and $T = 553$ K. In Fig. 3a the mean squared displacement from CG PTMO system, MSD(t_{CG}) is shifted along the time axis by a feasible time scaling factor $S_{UA-CG} = 3.35$ resulting in an agreement with united-atom data, such that MSD($t_{CG}S_{UA-CG}$) = MSD(t_{UA}). Figure 3b represents the dynamic analysis at a higher temperature ($T = 553$ K). This is to underline that the time-scaled MSD of the CG model is in good agreement with UA, but the scaling factor has increased at a higher temperature. The increased dynamic shift factor with increasing temperature is consistent with earlier findings, where a classical hybrid optimization approach was used to develop the CG model of polymer[4]. Moreover, time-scaled CG curves follow the UA for distances above a few Å and for a time above a few hundred picoseconds. It is reported elsewhere that the agreement between the motion of polymer chains at such a small length scale and a short time scale indicates the inherence of atomistic features in the coarse-grained model[17,62]. It is noteworthy to mention that $T = 553$ K was not included in the parametrization process. Thus, developed potentials show good temperature transferability. However, it is important to note that dynamic scaling factors are empirically derived from the MSD. These scaling factors are effective at reproducing dynamics, particularly at high temperatures, and may require recalibration for different temperatures. Wenjie et al. demonstrated that temperature-transferability of the polymer dynamics over a wide range of temperatures can be achieved by correcting the deviations in activation-free energy[25]. In their energy-renormalization approach, an inference has been drawn from Adam–Gibbs (AG) theory of glass formation, which allows the prediction of dynamic properties over the broad range of temperatures (i.e., Arrhenius regime, the non-Arrhenius regime, and the non-equilibrium glassy regime). The energy-normalization approach achieves this by introducing temperature-dependent rescaling factors for nonbonded parameters ϵ and σ[25,27]. A similar approach could be adopted in a data-driven method, where the ML-enabled framework in the present study provides the basis to further expand its parametrization space which could allow the training of the temperature-dependent LJ parameters to predict dynamics over wider temperature ranges.

Figure 4 shows the relaxation spectrum $G(t)$ for a machine-learned coarse-grained system of PTMO along with united atom simulation. Figure 4a and b show the long and short time behavior of the CG system, respectively. The stress relaxation spectrum is also computed for chain lengths $N_b = 50$ and 500 to demonstrate the predictive power of the CG potentials beyond the entanglement length (experimentally, effective entanglement in polyether diol ≈ 10,500 g/mol corresponding to 145 PTMO repeating units or 181 propylene oxide repeating units[63]). As shown in Fig. 4a the agreement between the UA model and CG PTMO system is excellent from intermediate time scales to long time scales ($10\,ps \le t \le 10^2$ ns) for the chain length of 50 beads. Generally, for AA models sufficient sampling becomes increasingly difficult beyond 1 ns and stress decays rapidly with higher deviations in relaxation time predictions. However, in this study, the multi-tau correlator method[64] is employed to calculate the stress auto-correlation function (SACF) in a canonical ensemble (NVT). The simulation time is increased up to 10^2-10^3 ns, which provides sufficiently accurate estimates of the relaxation spectrum of the UA model. Importantly, the efficiency of the CG model allows for sufficient sampling well beyond 10^3 ns, far from the achievable range of atomistic MD simulations. The relaxation spectrum of chain length $N_b = 500$ is an indication of this with the appearance of a plateau in $G(t)$ up to 10^4 ns. The comparison between two molecular weights also underlines the consistency of the developed potentials in terms of the good agreement from the short time scale (Fig. 4b) to the intermediate time scale (Fig. 4a). The extended plateau for $N_b = 500$ is due to entanglement.

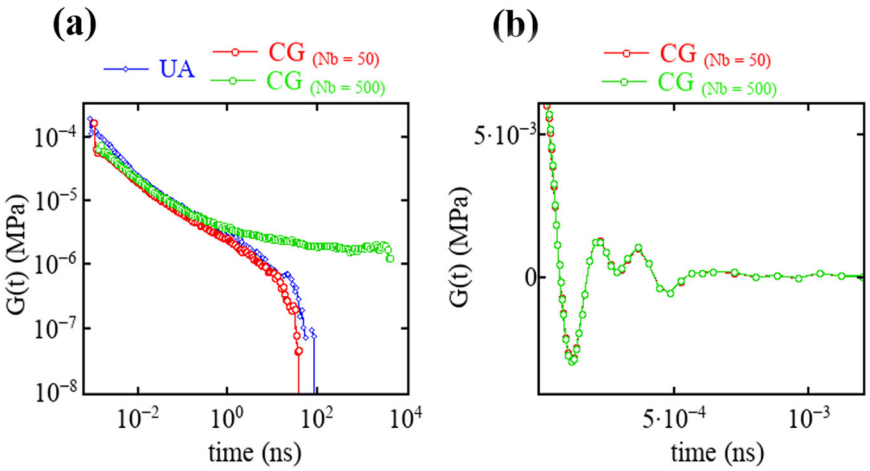

Fig. 4 Molecular-weight dependent relaxation behaviour from MLCG system is visualized. a The relaxation spectrum of CG-PTMO and UA chains are computed at $T = 453$ K. **b** The short timescale relaxation behavior of PTMO chain lengths below and above the entanglement limit.

DISCUSSION

We proposed a framework for the development of efficient, accurate, and transferable coarse-grained (CG) potentials for a model polymer by integrating bottom-up and top-down approaches through machine learning and optimization algorithms. We have demonstrated the versatility of the integrated machine learning method for coarse-graining polyether polymers such as poly(tetramethylene oxide). Using a simple one-bead mapping mechanism, and introducing a deep neural network-trained bottom-up optimization to match the atomistic target bonded distributions, we were able to reproduce the structural features of the polymer chain system. In addition, developing a top-down optimization approach for the nonbonded LJ potential parameters allowed us to match the local structure in terms of the radial distribution function. We also demonstrated that optimizing the nonbonded parameters by matching the specific volume at multiple temperature points in the liquid state provides transferability of potentials to predict the accurate glass transition temperature T_g consistent with experiments. The molecular weight-dependent T_g allows the fitting of the Flory–Fox equation to predict the limiting behavior of PTMO, which is found to be in good agreement with experimental measurements. This indicates that the developed framework may become a versatile approach to enable the transferability of certain thermodynamic state-dependent properties in CG systems. Calculated dynamic scaling factor through comparison of the CG and UA mean squared displacement of the center of mass indicates the accelerated speed of CG systems. By comparing the actual computer wall-time of the UA and CG simulations run for 50 monomer systems, we found out that the performance of the UA system is 9.084 ns/day and for CG system 1735.109 ns/day on a computer of 28 CPUs with an average CPU speed of 98.9%. This means that the CG model of PTMO provides a speed-up of over 200 times compared to the UA model due to the lower degrees of freedom and larger time step. The CG model enables prediction of the long-time behavior, of particular importance for capturing the full stress relaxation behavior of the polymeric systems. This is due to the increased efficiency and faster dynamics of the CG chains.

Therefore, in the present work, we established the advantages of using a hybrid MLCG method for polymer coarse-graining over the classical coarse-graining methods. Using our method, first, we are able to demonstrate simultaneous and accurate prediction of structural, dynamics, and experimentally consistent glass transition along with limiting behavior, which has been elusive to accomplish in traditional coarse-graining methods. Second, our framework enables the transferability of the developed potentials to ensure accurate thermomechanical predictive behavior by including the polymer chain length well beyond the oligomers, in the parametrization (training, testing, and optimization) process. Third, by incorporating experimental data of temperature-dependent density into nonbonded parametrization, the accuracy is enhanced (i.e., experimentally consistent T_g and limiting behavior are achieved) while ensuring the generalizability for thermodynamic state-dependent properties of CG models. Finally, the protocol has been shown to be robust and having been applied to poly(tetramethylene oxide), it could potentially be applied to other polymeric systems.

In closing, we recognize additional directions for future work. In this work, we considered the top-down optimization using temperature-dependent density from experiments, while primarily focusing on establishing an entirely data-driven framework that can capture a variety of material properties. Even though the machine-learned potentials demonstrated that optimized parameters are transferable to predict the thermodynamic properties, further study is necessary to assess the generality of the trends for other properties as well. Of particular interest is to investigate the temperature-transferability of dynamics at temperatures lower than the glass transition by expanding the parametrization space of the framework or amending a physics-informed learning protocol. Future work, therefore, will aim to include temperature-dependent cohesive interaction strength in training and testing for nonbonded optimization, thus building on the framework provided in this work.

METHOD

Characterization of the MLCG framework

We have established a framework through integrating machine learning, optimization, and MD simulations as illustrated in Fig. 5 to develop a temperature-transferable coarse-grained (CG) model that accurately represents poly(tetramethylene oxide). In the present study, target averaged distributions (bonded interaction parameters) are generated from the united-atom (UA) molecular dynamics simulations. A single-bead CG model is utilized, where each bead represents a repeating unit i.e., $-(CH_2CH_2OCH_2CH_2)-$ of PTMO along the polymer chain. The resulting CG bead chains can therefore be denoted by chains of n beads, which corresponds to the number of repeat units in the united-atom resolution of PTMO (see section MD simulation details). Each CG bead is assigned a mass equal to the sum of the atomistic masses of its constituent elements, which is $m = 72$ g/mol. Both UA and

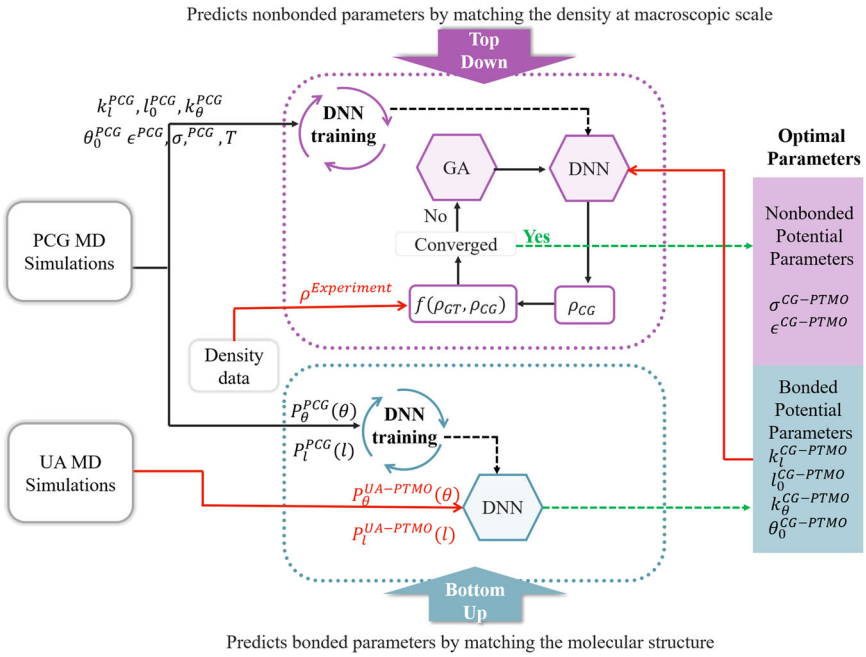

Fig. 5 Schematic representation of the framework is shown. A hybrid optimization approach has been used to develop interaction parameters for coarse-graining polymer molecules using machine learning techniques. The framework includes the bottom-up optimization approach (detail given in MD simulation details section) for predicting bonded interaction parameters ($k_l^{CG-PTMO}, l_0^{CG-PTMO}, k_\theta^{CG-PTMO}, \theta_0^{CG-PTMO}$) by matching bond ($P_l^{UA-PTMO}$) and angle distributions ($P_\theta^{UA-PTMO}$) from a united-atom model of PTMO as ground truth. In the bonded optimization process, the DNN was trained using bond ($P_l^{PCG}(l)$) and angle ($P_\theta^{PCG}(\theta)$) distribution data from prototype-coarse-grained (PCG) simulations. The top-down approach (MD simulation details) was incorporated with a genetic algorithm and deep neural network for predicting the nonbonded interaction parameters ($\sigma^{CG-PTMO}, \epsilon^{CG-PTMO}$) of the CG model. The DNN was trained using parameters ($k_l^{PCG}, l_0^{PCG}, k_\theta^{PCG}, \theta_0^{PCG}, \epsilon^{PCG}, \sigma^{PCG}$, and T) acquired from PCG simulations and was integrated into GA along with parameters obtained in the bottom-up step to predict the density. Consecutively, nonbonded interaction parameters ($\epsilon^{CG-PTMO}, \sigma^{CG-PTMO}$) were optimized by matching temperature-dependent experimental density. The dashed green arrows indicate the optimal parameters i.e., machine-learned CG parameters, obtained using the bottom-up and top-down portions of the strategy.

CG molecular dynamics simulations were performed using the LAMMPS software package[65].

In order to capture the structure-property relation of CG-PTMO, we use the "Prototype" coarse-grained (PCG) systems (Prototype is called PCG throughout this paper) for data acquisition. Thousands of independent MD simulations were run for PCG, to obtain the data for training and testing purposes. These simulations were run using the potential energy (U_{total}) described by Eq. (3).

$$U_{total} = U_{stretch} + U_{bend} + U_{nonb} \quad (3)$$

The total potential energy of the PCG, U_{total} is given by three contributions for which the functional forms are given in Table 4. $U_{stretch}$ is the energy associated with stretching the bonds of equilibrium bond length l_0 between adjacent beads along the polymer chain, U_{bend} is the energy associated with bending, i.e., the angles subtended by consecutive bonds with equilibrium angle θ_0, and U_{nonb} is the energy of nonbonded interactions between beads of inter chains or beads separated by more than three bonds on the same chain. Thus, the target (CG-PTMO) model demands optimization of a total of six potential parameters ($k_l, l_0, \theta, k_\theta, \sigma$, and ϵ) to fully describe the energetics and conformational dynamics of the system. The potential parameters of the target CG system are determined in two steps. In the first step, the bottom-up approach is used to establish the bonded interaction parameters. In the second step, the nonbonded parameters are determined from the top-down approach involving the temperature-dependent volumetric behavior of the system. To explore the parameter space, the bonded and nonbonded potential parameters were randomly sampled over a specific range based on the ground truth information, where ±Δ defines the upper and lower bounds of the explored range from the UA ground truth. The sampling range was defined by refining the region in the search space based on initial 500 (out of 5000) PCG simulations, which were run for wide ±Δ values of bonded parameters by calculating the mean and variance of the bond and angle distributions of the UA-PTMO. For nonbonded parameters, the range for a sampling of ϵ and σ was narrowed down based on the radial distance and magnitude of the first peak of RDF as well as the literature. In particular, the ϵ range was further refined based on an increased well-depth in the top-down optimization of the CG model of various materials[4,14,66].

Bottom-up MLCG: bonded interaction optimization. Multi-layer feed-forward DNNs was developed to learn the relationship between the bonded interaction parameters and local chain structures. The expression of the complex relationship between the underlying potential and structural features is given in Eqs. (4) and (5).

$$(k_l, l_0) = f(P_l(l)) \quad (4)$$

$$(k_\theta, \theta_0) = f(P_\theta(\theta)) \quad (5)$$

The extensive training data set of energies and structural properties of PTMO are taken from 5000 independent PCG systems with random sampling over a broad range of bond and angle potential parameters. The data are normalized and randomly split into two sets. 70% of the data was used to train the model, which included a random split with the ratio 1:4 for validation:training. A 5-fold cross-validation technique was used to reduce the bias in the model as it ensures that every observation

Table 3. Performance of the deep neural networks (DNNs).

DNN	MAE (training)	MAE (5-fold validation)		MAE (resting)	Training time (s)
		Mean	Standard deviation		
Bonded potential[a]	0.025	0.03	0.005	0.028	1539
Angle potential[a]	0.026	0.034	0.004	0.030	963
Density	0.029	0.033	0.002	0.036	1952

[a]The DNNs of bonded and angle potentials present the aggregated mean absolute errors.

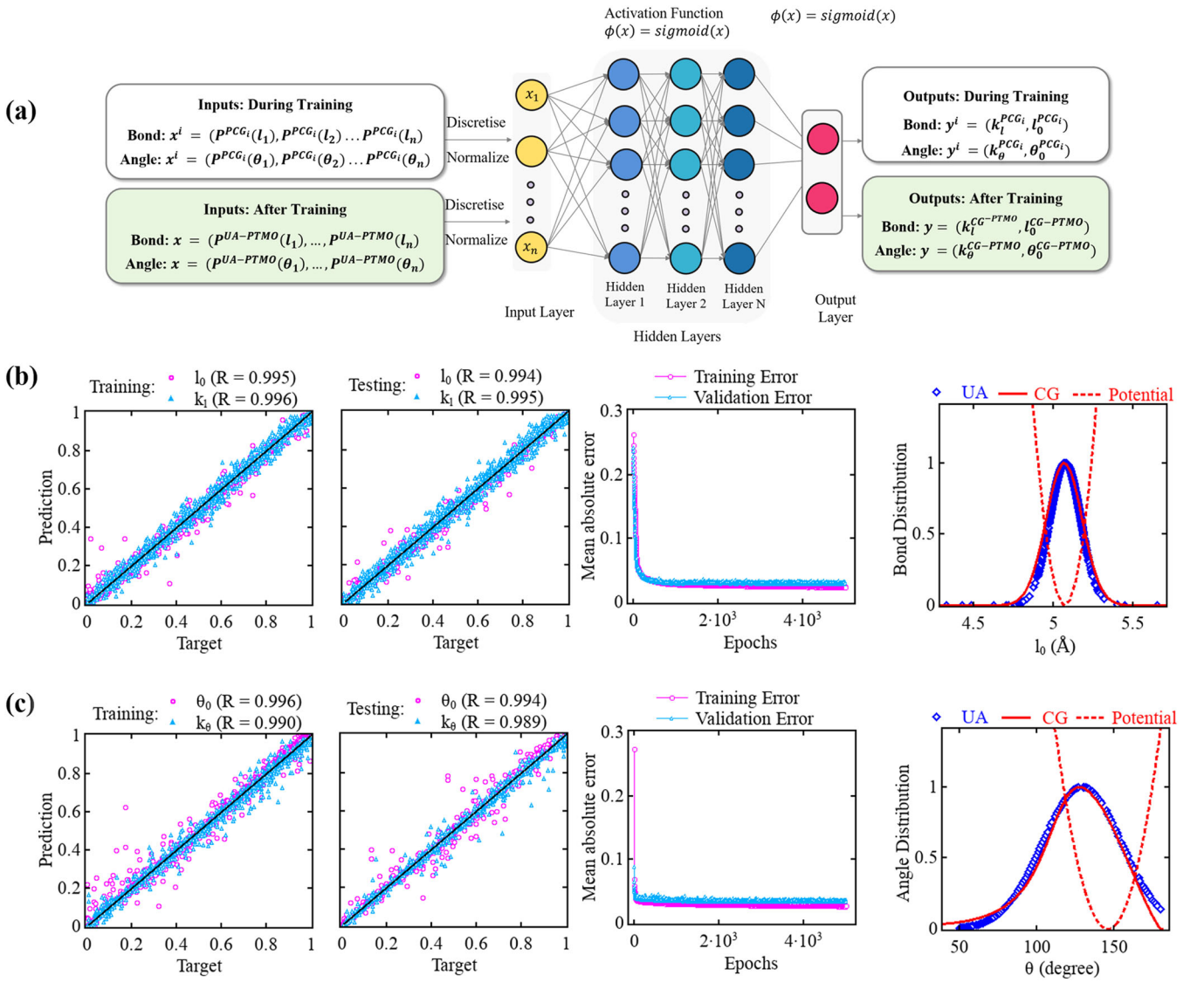

Fig. 6 Bottom-up approach is shown along with performance of the neural networks. a The architecture of the DNN used in obtaining the bonded potential parameters. **b** The scatter plots represent the predictions of the DNN and the target of training and testing data for bond length (l_0) and stretching coefficient (k_l) along with the performance analysis of the neural network during training, and corresponding potential and bond distribution from MLCG predictions. **c** The target and the predictions from the DNN for angle (θ_0) and bending coefficient (k_θ) along with the performance of the DNN during training and corresponding potential and angle distribution from MLCG and ground truth.

from the data set has the chance of appearing in the training and validation set. The remaining 30% of data is used for testing, i.e., to evaluate the generality and the performance of the network for unseen data. The accuracy of the network is assessed using the mean absolute error (MAE) that calculates the deviation between the predicted and target values. The calculated MAE values for testing, training and 5-fold validations are tabulated in Table 3. The network is optimized using the Adam optimizer, which is a stochastic gradient descent algorithm based on adaptive estimates of lower-order moments[67]. Multiple DNN architectures with different activation functions, learning rates, and the number of layers and nodes were evaluated, and the best architecture was

Table 4. Summary of predicted interaction parameters from machine-learned coarse-grained (MLCG) model of PTMO.

	Functional form	CG parameters	Present work (MLCG)	Literature (IBI)
Bonded (bottom-up)	$U_{\text{stretch}}(r) = k_l \cdot (l - l_0)^2$	l_0 (Å)	5.07	4.98
		k_l (kcal/mol Å^2)	35.95	3.03
	$U_{\text{bend}}(\theta) = k_\theta \cdot (\theta - \theta_0)^2$	θ_0 (deg)	146.11	169.5
		k_θ (kcal/mol rad^2)	1.21	0.64
Nonbonded (top-down)	$U_{\text{nonb}}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$	σ	4.82	–
		ϵ	0.73	–

then selected based on the lowest MAE and the least over-fitting.

By adopting the bottom-up machine-learned coarse-graining (MLCG) approach, the data acquired from PCG simulations are fed into the network. The architecture of the developed DNN is shown in Fig. 6a. The input vector x^i is the distribution of the ith PCG system, which is discretized between its minimum and maximum range, i.e., $x_l^i = P(l_{\min} < l < l_{\max})$, and $x_\theta^i = P(\theta)$ where $\theta_{\min} < \theta < \theta_{\max}$. Two distinct DNNs with densely connected layers were trained to predict the bonded potential parameters of the CG-PTMO from the bond and angle distributions as shown in Fig. 6b and c. The DNN for the bond distribution has four hidden layers with 80 nodes for each layer, with a sigmoid activation function at each layer. The output layer consists of two nodes for k_l, and l_0. The DNN developed for the angle potential consists of three hidden layers, with 60 nodes and a sigmoid activation function at each layer. The two nodes in the output layer correspond to k_θ and θ_0.

Once the training of the neural networks is accomplished, the bond and angle distributions from the united-atom MD simulations, i.e., "the ground truth" are fed into the networks to predict the corresponding bonded potential parameters. This way, the DNN provides the bonded interaction parameters of the CG-PTMO model that represents the same bond distributions as the UA-PTMO model.

As shown in Fig. 6b and c the bond and angle distributions of the CG system exhibit a good match with reference values from the UA system, suggesting that the developed stretching and bending potentials, using the deep neural networks successfully represent the PTMO chains. The dashed lines represent the developed potentials obtained via equations listed in Table 4 for bonded contributions. In Table 4, we have also listed the readily available potential parameters for PTMO from the literature. Please note parameters taken from the literature for PTMO are for the soft segment of the polyurea[5]. The CG predictions of the current study are not directly comparable with the CG model by using these potentials due to the unavailability of the nonbonded interaction parameters.

Top-down MLCG: nonbonded interaction optimization. In the conventional CG MD simulations, developing transferable force field parameters have been a long-standing challenge, particularly for complex macromolecules. The traditional approaches involve the optimization of the radial distribution function, which often requires an additional correction term to include pressure fluctuations for accurate prediction of the thermodynamical behavior. In our framework, to predict the accurate thermodynamics across broad temperature ranges, the nonbonded interaction parameters are determined using the density of poly(tetramethylene oxide) from the experiments[68], while bonded potential parameters are obtained in the bottom-up approach using UA data as ground truth.

A genetic algorithm (GA)-based optimization scheme is introduced to map the density of the CG model at different thermodynamic states. The objective function of the algorithm is given as Eq. (6), where T_i is the ith temperature state, ρ_{CG} and ρ_{GT} are the densities of the CG systems and ground truth (GT), respectively. The objective function evaluates the error between the supplemented density data (experimental data) and the predicted density from the CG model.

$$\text{Objective function} \;:\; \min \sum_i \left(\frac{\rho_{\text{CG}}(T_i)}{\rho_{\text{GT}}(T_i)} - 1\right)^2 \tag{6}$$

Since the GA is a meta-heuristic search algorithm based on the natural selection of a population with the process of adaptation for survival, it is a robust and efficient technique to explore complex nonlinear solution space. Compared to other algorithms, such as swarm intelligence-based optimization techniques, GA is less likely to have a premature convergence to a local optimal solution[69].

Due to the iterative process of optimization algorithms, generally, several hundred thousand MD simulations are required to capture the reasonable range in the data. Therefore, approximations or surrogate models with less computational costs can be constructed to replace the MD simulations. In our framework, we have constructed a DNN as a surrogate model to establish the relationship between the CG potential parameters and the temperature-dependent density of the CG model. As a result, the need for further data acquisition for density from computationally expensive MD simulations has been avoided.

To train the deep neural network, bonded potential parameters of the PCG-MD simulations are fed along with temperatures to learn the multiplex relation between density and potential parameters. The multiplex expression is given in Eq. (7), where ρ is the density, and $\sigma, \epsilon, k_l, l_0, k_\theta, \theta_0$ are the potential parameters.

$$\rho(T) = f(\sigma, \epsilon, k_l, l_0, k_\theta, \theta_0) \tag{7}$$

This DNN consists of three hidden layers, each with 40 nodes and rectified linear unit (Relu) activation function. The trained DNN is then integrated into the GA optimization algorithm to expedite the exploration of the optimal LJ potential parameters of the CG system. This ML-based optimization aims to achieve the desired temperature-dependent density, which is consistent with the ground truth (i.e., supplemented experimental data). The average prediction time for the trained DNN to predict the density at one temperature state from a 2.3 GHz core *i*7 processor is ~0.01 s, whereas the PCG-MD simulation takes an average of 60 s. This pinpoints the speed increment by a factor of 6.2×10^3 compared to MD simulations, and therefore, significantly accelerates the optimization process.

From the predicted nonbonded interaction parameters, the resulting local structure i.e., radial distribution function $g(r)$ is shown in Fig. 7c along with RDF from the united atom model and LJ potential (dashed line). The radial distribution functions $g(r)$ of the UA-PTMO and CG-PTMO are calculated at 453 K. The initial peak for both the UA and CG systems is found to correspond to the same radial distance; however, the CG peak value is relatively higher than that of the UA system. This is the result of an increased well-depth as predicted by the DNN in the top-down

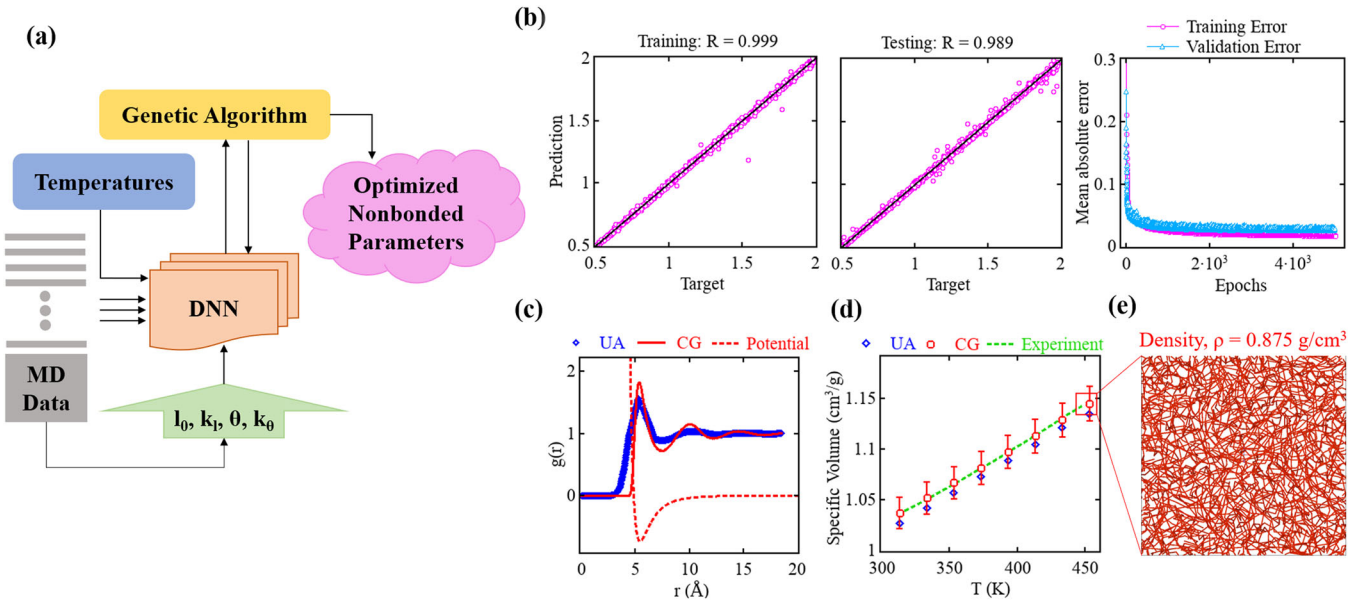

Fig. 7 Top-Down approach is shown. a The schematic of top-down optimization protocol. **b** The scatter plots represent the predictions from the DNN and the target for training and testing data for density (ρ_{CG}), along with the performance of the DNN during training. **c** From the predicted parameters (σ and ϵ) the resulting radial distribution function ($g(r)$) of the CG system is accompanying the corresponding potential and RDF of the UA system. **d** The specific volume of the MLCG system is shown along with ground truth, i.e., experiment (dashed line) and UA PTMO simulations. **e** Snapshot of the CG PTMO system with equilibrated density, $\rho = 0.875$ g/cm^3 at $T = 453$ K.

optimization of the nonbonded parameters. In bottom-up coarse-graining (classical or machine learning), the local structures are optimized, and the resulting RDF almost always matches perfectly with atomistic models[5,40]. However, in top-down approach, where the nonbonded interactions are tuned by incorporating temperature-dependent density (ρ_T), glass transition temperature (T_g), or elastic modulus (E), the resulting radial distribution function has a larger peak value than that from atomistic simulations, as was also found earlier in a top-down optimization of a coarse-grained model of polystyrene[4].

MD SIMULATION DETAILS

We performed united-atom (UA) molecular dynamics simulations to generate reference (the ground truth) structural distributions of PTMO chains in the melt state for the purpose of parametrizing the potential parameters of the CG bead-chain model. Generally, for polymeric materials, the CG potentials are derived from simulations of shorter chains or oligomers with the assumption that the derived potentials are transferable to longer chains[5]. In this study, we use UA chains of PTMO containing 50 repeating units, equivalent to 250 atoms and united atoms along the chain for generating the structural distributions. Considering the computational limitations of the UA-PTMO MD simulations we selected the system of 50 such chains for collecting data as a reference. UA-MD simulations were performed using the LAMMPS software package[65].

The repeat unit $-(CH_2CH_2OCH_2CH_2)-$ of the UA-PTMO chain is defined in such a way that every oxygen and every methylene group is treated as an atom and united atom, respectively. The UA potential used in the reference system is the Transferable Potentials for Phase Equilibria-UA (TraPPE-UA) for the intra (bonded) and inter-molecular (nonbonded) interactions initially developed by Lempesis et al.[3]. Shireen et al. further validated and established the TraPPE-UA potentials to study shear and bulk relaxation behavior of melt PTMO[70]. Before collecting data, the system was equilibrated at $T = 453$ K for 5 ns in a canonical ensemble using the deterministic Nose–Hoover thermostat[71,72]. Subsequently, the equilibrated system was quenched in the NPT ensemble at $P = 1$ atm, from 453 to 313 K in decrements of 20 K). While quenching, isothermal and isobaric conditions were maintained with constants 0.1 and 1 ps, where at each temperature the simulation ran for 5 ns.

We measure two types of structural distributions as well as the temperature-dependent density of the melt PTMO from our UA simulations. First, we collect the bond length l_0 distribution data, i.e., the distances between "virtual" bead centers along the UA-PTMO chain, representing centers of the adjacent beads in the CG-PTMO model. Given that all the beads in the CG chains are of the same type, there is only one type of chemical bond in the CG system, thus, only one set of bond length distribution data is recorded. This distribution is used to parameterize the bond stretching potential (the functional form of the potential is given in Table 4). Second, we collect bond angle θ_0 data and its distribution between three consecutive CG beads, and this distribution is used to parameterize the CG-PTMO bending potential (the functional form of the potential is given in Table 4).

We also measured the temperature-dependent density of melt PTMO at $P = 1$ atm from 453 to 313 K. The measured temperature-dependent density data is shown in Fig. 7d along with MLCG and experimental values. UA-PTMO chains used for ground truth data has repeat units $N_m = 50$, which is below the value at which density becomes insensitive to chain length, therefore, density values are slightly different, and calculated deviations are within 1%.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article. Additional data are available from the corresponding author Dr. Elnaz Hajizadeh upon reasonable request.

CODE AVAILABILITY

Most of the codes of the prediction models are available as open-source code as stated in the cited references. Other codes may be obtained from the corresponding author upon reasonable request.

Received: 19 May 2022; Accepted: 17 October 2022;
Published online: 04 November 2022

REFERENCES

1. Behbahani, A. F. et al. Dynamics and rheology of polymer melts via hierarchical atomistic, coarse-grained, and slip-spring simulations. *Macromolecules* **54**, 2740–2762 (2021).
2. Grest, G. S., Michael Salerno, K., Peters, B. L., Ge, T. & Perahia, D. Resolving properties of entangled polymers melts through atomistic derived coarse-grained models. In *Handbook of Materials Modeling: Methods: Theory and Modeling*, (eds Andreoni, W. & Yip S.) 1397–1410 (Springer, 2020).
3. Lempesis, N., in 't Veld, P. J. & Rutledge, G. C. Atomistic simulation of the structure and mechanics of a semicrystalline polyether. *Macromolecules* **49**, 5714–5726 (2016).
4. Hsu, D. D., Xia, W., Arturo, S. G. & Keten, S. Thermomechanically consistent and temperature transferable coarse-graining of atactic polystyrene. *Macromolecules* **48**, 3057–3068 (2015).
5. Agrawal, V., Arya, G. & Oswald, J. Simultaneous iterative Boltzmann inversion for coarse-graining of polyurea. *Macromolecules* **47**, 3378–3389 (2014).
6. Liu, M. & Oswald, J. Coarse-grained molecular modeling of the microphase structure of polyurea elastomer. *Polymer* **176**, 1–10 (2019).
7. Bayramoglu, B. & Faller, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 9205–9219 (2012).
8. Ohkuma, T. & Kremer, K. A composition transferable and time-scale consistent coarse-grained model for cis-polyisoprene and vinyl-polybutadiene oligomeric blends. *J. Phys.: Mater.* **3**, 034007 (2020).
9. Korolev, N., Luo, D., Lyubartsev, A. P. & Nordenskiöld, L. A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers* **6**, 1655–1675 (2014).
10. Lyubartsev, A. P., Naômé, A., Vercauteren, D. P. & Laaksonen, A. Systematic hierarchical coarse-graining with the inverse Monte Carlo method. *J. Chem. Phys.* **143**, 243120 (2015).
11. Foley, T. T., Shell, M. S. & Noid, W. The impact of resolution upon entropy and information in coarse-grained models. *J. Chem. Phys.* **143**, 12B601_1 (2015).
12. Shell, M. S. Coarse-graining with the relative entropy. *Adv. Chem. Phys.* **10.1002/97**, 395-441, (2016).
13. Dunn, N. J., Foley, T. T. & Noid, W. G. Van der Waals perspective on coarse-graining: progress toward solving representability and transferability problems. *Acc. Chem. Res.* **49**, 2832–2840 (2016).
14. Noid, W. G. Perspective: coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 09B201_1 (2013).
15. Brini, E. et al. Systematic coarse-graining methods for soft matter simulations—a review. *Soft Matter* **9**, 2108–2119 (2013).
16. Louis, A. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* **14**, 9187 (2002).
17. Fritz, D., Harmandaris, V. A., Kremer, K. & van der Vegt, N. F. Coarse-grained polymer melts based on isolated atomistic chains: simulation of polystyrene of different tacticities. *Macromolecules* **42**, 7579–7588 (2009).
18. Huang, H., Wu, L., Xiong, H. & Sun, H. A transferrable coarse-grained force field for simulations of polyethers and polyether blends. *Macromolecules* **52**, 249–261 (2018).
19. Hajizadeh, E., Todd, B. & Daivis, P. Shear rheology and structural properties of chemically identical dendrimer-linear polymer blends through molecular dynamics simulations. *J. Chem. Phys.* **141**, 194905 (2014).
20. Hajizadeh, E., Todd, B. & Daivis, P. A molecular dynamics investigation of the planar elongational rheology of chemically identical dendrimer–linear polymer blends. *J. Chem. Phys.* **142**, 174911 (2015).
21. Li, Y., Abberton, B. C., Kröger, M. & Liu, W. K. Challenges in multiscale modeling of polymer dynamics. *Polymers* **5**, 751–832 (2013).
22. Noid, W. G. et al. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **128**, 244114 (2008).
23. Mullinax, J. & Noid, W. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **131**, 104110 (2009).
24. Rudzinski, J. F. & Noid, W. Coarse-graining entropy, forces, and structures. *J. Chem. Phys.* **135**, 214101 (2011).
25. Xia, W. et al. Energy-renormalization for achieving temperature transferable coarse-graining of polymer dynamics. *Macromolecules* **50**, 8787–8796 (2017).
26. Song, J. et al. Energy renormalization method for the coarse-graining of polymer viscoelasticity. *Macromolecules* **51**, 3818–3827 (2018).
27. Dunbar, M. & Keten, S. Energy renormalization for coarse-graining a biomimetic copolymer, poly (catechol-styrene). *Macromolecules* **53**, 9397–9405 (2020).
28. Dunn, N. J. & Noid, W. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane–toluene mixtures. *J. Chem. Phys.* **144**, 204124 (2016).
29. DeLyser, M. R. & Noid, W. G. Coarse-grained models for local density gradients. *J. Chem. Phys.* **156**, 034106 (2022).
30. Lebold, K. M. & Noid, W. Systematic study of temperature and density variations in effective potentials for coarse-grained models of molecular liquids. *J. Chem. Phys.* **150**, 014104 (2019).
31. Lebold, K. M. & Noid, W. Dual approach for effective potentials that accurately model structure and energetics. *J. Chem. Phys.* **150**, 234107 (2019).
32. Voth, G. A. *Coarse-graining of Condensed Phase and Biomolecular Systems* (CRC Press, 2008).
33. Kmiecik, S. et al. Coarse-grained protein models and their applications. *Chem. Rev.* **116**, 7898–7936 (2016).
34. Singh, N. & Li, W. Recent advances in coarse-grained models for biomolecules and their applications. *Int. J. Mol. Sci.* **20**, 3774 (2019).
35. Shireen, Z. & Babu, S. B. Cage dynamics leads to double relaxation of the intermediate scattering function in a binary colloidal system. *Soft Matter* **14**, 9271–9281 (2018).
36. Ye, H., Xian, W. & Li, Y. Machine learning of coarse-grained models for organic molecules and polymers: progress, opportunities, and challenges. *ACS Omega* **6**, 1758–1772 (2021).
37. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
38. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **3**, 1–13 (2017).
39. Chan, H. et al. Machine learning coarse grained models for water. *Nat. Commun.* **10**, 1–14 (2019).
40. Moradzadeh, A. & Aluru, N. R. Transfer-learning-based coarse-graining method for simple fluids: toward deep inverse liquid-state theory. *J. Phys. Chem. Lett.* **10**, 1242–1250 (2019).
41. Nguyen, D., Tao, L. & Li, Y. Integration of machine learning and coarse-grained molecular simulations for polymer materials: physical understandings and molecular design. *Front. Chem.* **9**, 820417, (2021).
42. Duan, K. et al. Machine-learning assisted coarse-grained model for epoxies over wide ranges of temperatures and cross-linking degrees. *Mater. Des.* **183**, 108130 (2019).
43. McDonagh, J. L., Shkurti, A., Bray, D. J., Anderson, R. L. & Pyzer-Knapp, E. O. Utilizing machine learning for efficient parameterization of coarse grained molecular force fields. *J. Chem. Inf. Model.* **59**, 4278–4288 (2019).
44. Hajizadeh, E., Todd, B. & Daivis, P. Nonequilibrium molecular dynamics simulation of dendrimers and hyperbranched polymer melts undergoing planar elongational flow. *J. Rheol.* **58**, 281–305 (2014).
45. Hajizadeh, E., Yu, S., Wang, S. & Larson, R. G. A novel hybrid population balance-Brownian dynamics method for simulating the dynamics of polymer-bridged colloidal latex particle suspensions. *J. Rheol.* **62**, 235–247 (2018).
46. Ruza, J. et al. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.* **153**, 164501 (2020).
47. Wang, J. et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **5**, 755–767 (2019).
48. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deepcg: constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **149**, 034101 (2018).
49. Bejagam, K. K., Singh, S., An, Y. & Deshmukh, S. A. Machine-learned coarse-grained models. *J. Phys. Chem. Lett.* **9**, 4667–4672 (2018).
50. Reith, D., Meyer, H. & Müller-Plathe, F. Mapping atomistic to coarse-grained polymer models using automatic simplex optimization to fit structural properties. *Macromolecules* **34**, 2335–2345 (2001).
51. Kuenneth, C., Schertzer, W. & Ramprasad, R. Copolymer informatics with multitask deep neural networks. *Macromolecules* **54**, 5957–5961 (2021).
52. Chen, L. et al. Polymer informatics: current status and critical next steps. *Mater. Sci. Eng.* **144**, 100595 (2021).
53. Wang, W. & Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **5**, 1–9 (2019).
54. Jeong, J., Moradzadeh, A. & Aluru, N. Extended deepilst for various thermodynamic states and applications in coarse-graining. *J. Phys. Chem. A* **126**, 1562–1570 (2022).
55. Karuth, A., Alesadi, A., Xia, W. & Rasulev, B. Predicting glass transition of amorphous polymers by application of cheminformatics and molecular dynamics simulations. *Polymer* **218**, 123495 (2021).
56. Scherer, C., Scheid, R., Andrienko, D. & Bereau, T. Kernel-based machine learning for efficient simulations of molecular liquids. *J. Chem. Theory Comput.* **16**, 3194–3204 (2020).
57. Webb, M. A., Delannoy, J.-Y. & De Pablo, J. J. Graph-based approach to systematic molecular coarse-graining. *J. Chem. Theory Comput.* **15**, 1199–1208 (2018).

58. Shireen, Z. & Babu, S. B. Lattice animals in diffusion limited binary colloidal system. *J. Chem. Phys.* **147**, 054904 (2017).
59. Weeratunge, H. et al. A machine learning accelerated inverse design of underwater acoustic polyurethane coatings. *Struct. Multidiscip. Optim.* **65**, 1–13 (2022).
60. Rajendran, G., Mahadevan, V. & Srinivasan, M. Synthesis of some low glass transition temperature polytetrahydrofuran polymers. *Eur. Polym. J.* **25**, 461–463 (1989).
61. Fritz, D., Koschke, K., Harmandaris, V. A., van der Vegt, N. F. & Kremer, K. Multiscale modeling of soft matter: scaling of dynamics. *Phys. Chem. Chem. Phys.* **13**, 10412–10420 (2011).
62. Harmandaris, V. A. & Kremer, K. Dynamics of polystyrene melts through hierarchical multiscale simulations. *Macromolecules* **42**, 791–802 (2009).
63. Pohl, M. et al. Dynamics of polyether polyols and polyether carbonate polyols. *Macromolecules* **49**, 8995–9003 (2016).
64. Ramírez, J., Sukumaran, S. K., Vorselaars, B. & Likhtman, A. E. Efficient on the fly calculation of time correlation functions in computer simulations. *J. Chem. Phys.* **133**, 154103 (2010).
65. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys*. **117**, 1–19 (1995).
66. Joshi, S. Y. & Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **47**, 786–803 (2021).
67. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at arXiv:1412.6980 (2014).
68. Tsujita, Y., Nose, T. & Hata, T. Thermodynamic properties of poly (ethylene glycol) and poly (tetrahydrofuran). I. *P–V–T* relations and internal pressure. *Polym. J.* **5**, 201–207 (1973).
69. Katoch, S., Chauhan, S. S. & Kumar, V. A review on genetic algorithm: past, present, and future. *Multimed. Tools Appl.* **80**, 8091–8126 (2021).
70. Shireen, Z., Hajizadeh, E., Daivis, P. & Brandl, C. Linear viscoelastic shear and bulk relaxation moduli in poly(tetramethylene oxide) (PTMO) using united-atom molecular dynamics. *Comput. Mater. Sci.* **216**, 111824 (2022).
71. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
72. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695 (1985).
73. Ali, S. & Hourston, D. Physical studies of ionically terminated polytetrahydrofuran polymers. IV. Solid-state studies. *J. Appl. Polym. Sci.* **48**, 1915–1992 (1993).

ACKNOWLEDGEMENTS

This research is supported by the Commonwealth of Australia as represented by the Defence Science and Technology Group of the Department of Defence. We acknowledge Spartan-HPC, Argali-HPC, and Melbourne Research Cloud at the Faculty of Engineering and Information Technology, and the University of Melbourne for providing us with the computing facility.

AUTHOR CONTRIBUTIONS

Z.S.: Conceptualization, formal analysis, data visualization, writing, review, and editing. H.W.: Formal analysis, data visualization, review and editing. A.M.: Discussions, data visualization, and review. A.W.P.: Discussions and review. R.G.L.: Discussions and review. K.S.-M.: Discussions and review. E.H.: Supervision, Funding acquisition, project management, review and editing.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Elnaz Hajizadeh.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022