PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d0cp02309d

Is preservation of symmetry necessary for coarse-graining?†

Maghesree Chakraborty, Jinyu Xu and Andrew D. White *

Received 29th April 2020,
Accepted 19th June 2020

DOI: 10.1039/d0cp02309d

rsc.li/pccp

There is a need for theory on how to group atoms in a molecule to define a coarse-grained (CG) mapping. This article investigates the importance of preserving symmetry of the underlying molecular graph of a given molecule when choosing a CG mapping. 26 CG models of seven alkanes with three different CG techniques were examined. We unexpectedly find preserving symmetry has no consistent effect on CG model accuracy regardless of CG method or comparison metric.

1 Introduction

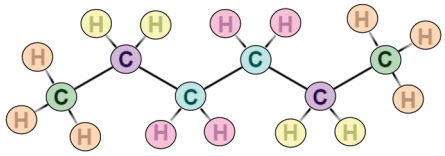
Coarse-grained (CG) simulations have been widely used to study systems to address length-scale challenges in molecular dynamics.[1,2] Selecting a CG mapping and obtaining the corresponding potential energy function are the key steps of defining a CG model. Both of these choices determine how closely a CG simulation reproduces results from the corresponding all-atom (AA) simulation. There are many approaches for fitting the potential,[3,4] but the choice of a CG mapping is still made using chemical intuition. There have been recent efforts to develop more systematic approaches to choose CG mappings,[5–9] including our previous work.[7] Webb *et al.*[5] used spectral grouping iteratively to generate CG representations with successively lower resolutions. Wang and Gómez-Bombarelli[10] recently explored variational auto-encoder CG mappings, which is a promising new data-driven direction. The method, however, has yet to be assessed on more complex molecules and it has yet to be shown if the mappings are optimal. There are pipeline softwares available, like BOCS,[11] VOTCA[12] and Auto-Martini,[13] to facilitate CG system preparation and subsequent simulation. However, these tools either require the user to select the mapping operator or create mapping based on established rules, like Martini CG mappings. Zavadlav *et al.*[14] reported a Bayesian framework to compare different CG mappings of water varying in resolution and number of interaction sites. Kanekal and Bereau[15] have also used a Bayesian framework to investigate the limit of effect of varying the number of CG bead types. Despite the recent attention on systematic selection of CG mappings, there is a lack of studies on which factors influence the quality of CG mappings. In this study we compare different symmetric and asymmetric CG mapping operators of alkanes to understand the importance of preserving symmetry.

Symmetries in molecules have a significant impact on their properties. Previously, molecular symmetry has been exploited to simplify calculation of physical properties[16] (like optical activity,[17,18] dipole moment,[19] melting point,[20,21] solubility,[21] infra-red spectrum[22] and Raman spectrum[23]) and chemical properties.[24,25] Besides the point symmetry groups (spatial coordinates)[26] of molecules, another type of symmetry is called the topological symmetry[27–29] and refers to the symmetry of the underlying molecular graph where atoms are represented as nodes and bonds as edges.[28] Informally, two atoms are symmetric if they are chemically equivalent like the CH3 groups in diethyl ether. These topological symmetry groups can be identified by using graph automorphism on the molecular graph.[27] The symmetry groups consist of chemically equivalent atoms.[29] Molecules that lack global symmetry may have local symmetry groups.[29] Fig. 1 illustrates symmetry groups for hexane, which has global symmetry, and isohexane, which lacks global symmetry. The topological symmetry has recently been used in a recent work by Rosenfeld[30] for molecular synthesis based on topological symmetry. In our previous work,[7] we found that only considering topologically symmetric CG mapping operators reduces the number of unique mappings by an order of magnitude for molecules with heavy atoms between 3 and 9. In this work, we test if considering only symmetric mapping operators is valid on alkanes.

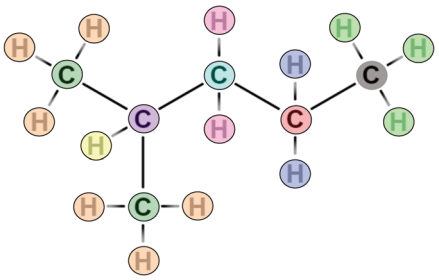
We have considered propane and three isomers of hexane and octane for this study. For each molecule, symmetric and asymmetric CG mapping operators were used to perform bottom-up CG simulations. Symmetric mapping operators refer to those where atoms belonging to the same symmetry groups

Department of Chemical Engineering, University of Rochester, Rochester, New York 14627, USA. E-mail: andrew.white@rochester.edu

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp02309d



Hexane



Isohexane

Fig. 1 Illustration of symmetry groups in hexane and isohexane. The atoms in the same symmetry groups are highlighted with the same color. Even though isohexane lacks global symmetry, it still has symmetry groups.

are kept at the same resolution in the CG representation. In asymmetric mappings, atoms belonging to at least one symmetry group do not have the same resolution in the CG representation. Additionally, we compared how the performances of CG mappings of hexane varied with the choice of different bottom-up approaches to fit the CG potential: force-matching (FM), iterative Boltzmann inversion (IBI) and relative entropy (RE). Our goal is not to compare the accuracies of one method to another, but rather to ensure our conclusions about symmetry are independent of CG potential fitting method. Discussions on comparing these methods, including when they are equivalent, can be found in Kmiecik *et al.*[2] and Noid.[3] Ruhle *et al.*[12] also compared FM and IBI for small organic molecules like water, methanol, propane and hexane.

CG alkane simulations have been studied before and we have summarized the variety of CG mapping operators used for alkanes in previous studies in Table 1. The alkanes in italics are included in our study. While we have tried to include relevant previous work, the list is not exhaustive.

There has been limited study on the effect of symmetry on CG model fidelity. We had considered two asymmetric mappings for methanol in a previous work.[7] Recently, Jin *et al.*,[31] mentioned that symmetry mismatch between the FG and CG representations had resulted in failure of MS-CG models in interfacial systems. They developed the center of symmetry CG in order to preserve the symmetry present in the FG model when it is mapped into a CG model by adding a virtual site. Among the previously studied mapping operators for 16 alkanes listed in Table 1, almost all mappings preserve symmetry except the following: 2-3 mapping for *n*-pentane, 2-3-3 mapping for *n*-octane, 2-2-3-2 mapping for nonane, 2-2-3-3 mapping for *n*-decane, 3-3-3-2 and 2-2-2-3-2 mapping for *n*-undecane, 2-2-3-3-3 and 2-2-2-2-3-2 mappings for *n*-tridecane, 2-2-2-3-3-3 mapping for *n*-pentadecane, 2-2-3-3-3 mapping for *n*-hexadecane, and 2-2-2-2-3-3-3 and 3-2-3-3-3-3 mappings of *n*-heptadecane. These are compared with our results below.

Our work is further motivated by previous CG studies which have yielded results contrary to chemical intuition. Some work has shown that the accuracy of CG mapping with the reference fine-grain (FG) simulation does not monotonically increase with increase in the resolution of the mapping.[32] Foley *et al.*[33] has shown how the information content in CG mapping seems to have an optimum with respect to CG mapping operator resolution. There are other reports,[34–36] including our previous work,[7] that corroborate that higher resolution CG mappings do not always outperform lower resolution mappings. This underlines the need of systematically studying factors which are often deemed trivial while using chemical intuition.

2 Methods

Symmetric and asymmetric mapping operators were considered for seven molecules: *n*-propane, *n*-hexane, isohexane (2-methylpentane), 2,3-dimethylbutane, *n*-octane, 3-ethylhexane, and 4-methylheptane. Three hexane isomers (*n*-hexane, isohexane, 2,3-dimethylbutane) and three octane isomers (*n*-octane, 3-ethylhexane and 4-methylheptane) were chosen since we wanted to study linear and branched isomers of 6-carbon and 8-carbon containing alkanes respectively. FM was used to get the corresponding CG potentials for the selected mappings of the 7 molecules. The illustrations of the mapping operators considered are shown in Fig. 2.

We have also investigated how the choice of method for obtaining CG potentials affects the performance of different mapping operators. This was limited to 6 mappings of hexane, labelled in red in Fig. 2. We compared mapping operators using CG potentials obtained by FM, IBI and RE. The FG simulation for each molecule was performed using GROMACS-2016[37] for 1 ns with the OPLS-AA force field and a 1 fs time step. The densities (in g cm^{-3}) used for FG simulations are as follows: propane – 0.635, *n*-hexane – 0.650, 2-methylpentane – 0.655, 2,3-dimethylbutane – 0.660, *n*-octane – 0.699, 3-ethylhexane – 0.7079 and 4-methylheptane – 0.705.

For each FG simulation, the *NVT* ensemble was maintained at 300 K for all molecules except propane, for which the FG simulation was conducted at a temperature of 200 K.[12] For the FG simulations, particle-mesh Ewald and truncated cut-off were used to handle coulombic and van der Waals interactions respectively. For all the 7 molecules, FM-CG simulations were conducted according to the methods described in our previous work.[7] FM with exclusions were calculated as outlined in the work by Ruhle and Junghans[38] to exclude the contributions from the bonded interactions. The corresponding CG potentials are included in the ESI† as Fig. S1. For the CG simulations, stochastic dynamics integrator (sd) was used with a 2 fs time step. Time constant for temperature coupling was set to 2 ps

Table 1 List of CG mappings used for alkanes in previous studies

Molecule	CG mappings	Metrics of comparison
Neopentane (2,2-dimethylpropane)	Single site mapping at COM[40]	RDF,[40] VACF,[40] self-diffusion coefficient[40]
n-Pentane	CG bead at each carbon-atom,[41] 2-3,[42,43] 4-1,[43] 2-2-1,[43] 1-3-1,[43] single site mapping[43]	Surface tension,[41,42] self-diffusion coefficient,[42,43] compressibility,[42] enthalpy of vaporization[43]
n-Hexane	CG bead at each carbon-atom,[41] 2 bead mapping,[40,42] 2-2-2,[6,42] 2-1-1-2[6]	Surface tension,[41] RDF,[40] VACF,[40] self-diffusion coefficient,[40,42] compressibility[42]
Cyclohexane	Single site mapping at COM[40]	RDF,[40] VACF,[40] self-diffusion coefficient[40]
n-Heptane	2-3-2[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
n-Octane	CG bead at each carbon-atom,[41] 2-2-2-2,[42] 2-3-3[42]	Surface tension,[41] self-diffusion coefficient,[42] compressibility[42]
Nonane	3-3-3,[42] 2-2-3-2[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
n-Decane	CG bead at each carbon-atom,[41] 2-2-2-2-2,[42] 2-2-3-3[42]	Surface tension,[41] self-diffusion coefficient,[42] compressibility[42]
n-Undecane	3-3-3-2,[42] 2-2-2-3-2[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
n-Dodecane	CG bead at each carbon-atom,[41] CG1,[34] CG2,[34] CG3,[34] CG4,[34] 3-3-3-3,[42] 2-2-2-2-2-2[42]	Surface tension,[34,41] temperature-density relationship,[34] self-diffusion coefficient,[42] compressibility[42]
n-Tridecane	2-2-3-3-3,[42] 2-2-2-2-3-2[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
Tetradecane	CG bead at each carbon-atom,[41] 3-3-2-2-2-2,[42] 2-2-2-2-2-2-2[42]	Surface tension,[41,42] self-diffusion coefficient,[42] compressibility[42]
n-Pentadecane	3-3-3-3-3,[42] 2-2-2-3-3-3[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
n-Hexadecane	CG bead at each carbon-atom,[41] 2-2-3-3-3-3,[42] 2-2-2-2-2-2-2-2[42]	Surface tension,[41,42] self-diffusion coefficient,[42] compressibility[42]
n-Heptadecane	2-2-2-2-3-3-3,[42] 3-2-3-3-3-3[42]	Surface tension,[42] self-diffusion coefficient,[42] compressibility[42]
n-Tetracosane	CG1, CG2, CG3, CG4[34]	Surface tension,[34] temperature-density relationship[34]

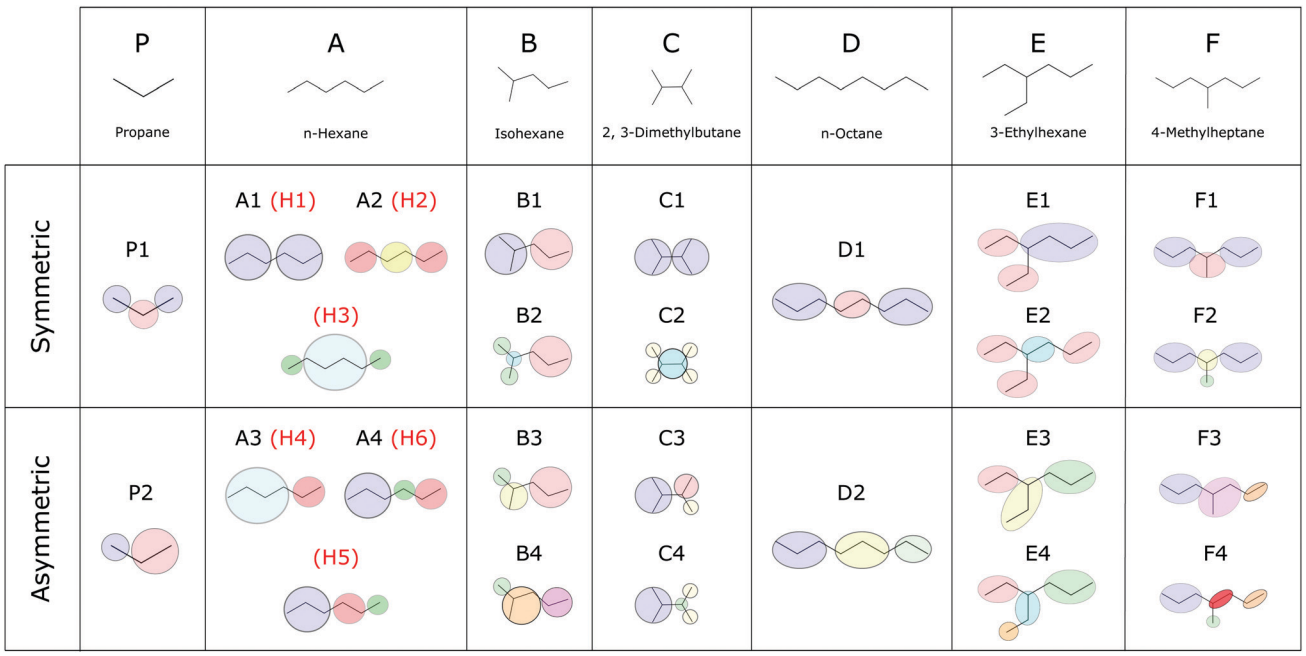


Fig. 2 Illustration of symmetric and asymmetric mapping operators of the seven molecules. We have highlighted in red the alternative labels for hexane CG mapping operators that are used to investigate the effect of different methods (FM, IBI and RE) of obtaining CG potentials on performance of the mappings. All the 24 non-highlighted mappings for the seven molecules were simulated with CG potentials obtained using FM.

and a cut-off of 1.11 nm was used. All bonds and angles were constrained with the SHAKE algorithm.[39] The equilibrium bond and angle values were obtained from the reference FG trajectory mapped using the corresponding mapping operator. The bond and angle values used in the CG simulations are provided in Fig. S3 and S4 (ESI†) respectively. Additional details are also included in the in the ESI.†

The iterative methods, IBI and RE, were also implemented using VOTCA following the procedures reported previously.[12,44] In the RE method, the CG potentials were modeled using cubic B-splines-based piece-wise polynomial functional form.[44] We have included Fig. S2 in the ESI† to compare the potentials from the three bottom-up approaches (FM, IBI and RE) for the 6 hexane mappings. All the CG simulations were run for 1 ns

with a time-step of 2 fs. To evaluate how different mapping operators performed, we compared the center of mass (COM) radial distribution functions (RDFs) and the velocity autocorrelation functions (VACFs) of the CG mapping to those obtained from the corresponding FG simulations. There have been previous studies that used VACFs for analysis in the context of stochastic dynamics.[45–47] For quantitative analysis evaluating the symmetric and asymmetric mappings, we computed the squared error between a CG mapping result and the FG result, normalized over all the CG mappings of a given molecule. For comparing FM, IBI and RE for a particular mapping, we computed the squared error and normalized over the three CG simulation results. The roots of the normalized mean square errors are the final reported values. We also calculated the normalized force error per CG bead[43] for the three methods. Only non-bonded forces were considered for evaluation. FG trajectory, with bonded forces excluded, was mapped into CG coordinates. Mapped forces for each bead were obtained using the eqn (1), where i denotes the atoms constituting the CG bead j.

$$\mathbf{F}_j^{\text{map}} = \sum_{i \in j} \mathbf{F}_i \tag{1}$$

To get the CG forces from FM, IBI and RE for the same trajectory, the mapped trajectory was rerun in GROMACS using the potentials derived from the 3 methods for each mapping. The normalized force error was subsequently evaluated according to eqn (2), where t denotes time-step and j denotes a CG bead.

$$\mathbf{F}^{\text{error}} = \frac{\sum_{t,j} \left\| \mathbf{F}_{t,j}^{\text{CG}} - \mathbf{F}_{t,j}^{\text{map}} \right\|^2}{\sum_{t,j} \left\| \mathbf{F}_{t,j}^{\text{map}} \right\|^2} \tag{2}$$

3 Results and discussion

Fig. 3 and 4 compare the mean square COM-RDF errors and the mean square VACF errors respectively of the asymmetric and symmetric mapping operators of the molecules. The corresponding COM-RDFs and VACFs are included in Fig. S5 of the ESI.† As seen in Fig. 3, preservation of topological symmetry present in the FG model while selecting a CG mapping does not guarantee closer agreement with reference FG COM-RDF. For instance, E1, the symmetric 3-bead CG mapping of 3-ethylhexane, has higher COM-RDF square error than E3, the asymmetric 3-bead CG mapping of 3-ethylhexane, even though both of them have the same degrees of freedom.

Similar results are seen based on COM-RDF square error for other symmetric and asymmetric mapping operator pairs with the same degrees of freedom for hexane (A2–A4), *n*-octane (D1–D2), 3-ethylhexane (E2–E4) and 4-methylheptane (F1–F3, F2–F4). When mean square error for VACF is the metric of comparison, we see in the symmetric and asymmetric CG mapping pair, F2–F4 for 4-methylheptane, that the asymmetric mapping yields lower mean square VACF error than the symmetric one. F2 and F4 mappings have comparable degrees of freedom. Similar results are obtained for symmetric–asymmetric mapping pairs with comparable degrees of freedom for hexane (A2–A4) and 3-ethylhexane (E2–E4). Contradicting results were obtained for other mappings for hexane (A1–A3), *n*-octane (D1–D2), 3-ethylhexane (E1–E3) and 4-methylheptane (F1–F3). Note that the symmetric and asymmetric mapping pairs above have an equal number of beads, and thus equal degrees of freedom, but the asymmetric mappings have more bead types. This gives more trainable parameters for asymmetric mappings as seen in Fig. S1 in the ESI† and could explain the better performance of some of the asymmetric mappings compared to the symmetric ones. Although counter-examples can be found for this hypothesis, like B2–B3 for isohexane, which has better asymmetric performance than symmetric in RDF and have the same number of trainable parameters.

Increasing the degrees of freedom by selecting a higher resolution CG mapping does not guarantee a closer agreement with FG results. As seen in Fig. 3, the 4-bead asymmetric mapping operator for 2,3-dimethylbutane, C4, gives higher COM-RDF square error than the 2-bead mappings, C1 and C3. Two bead propane mapping, P1, gives lower COM-RDF error than three bead P1 mapping. Similar results are seen for isohexane, where 4-bead mapping, B2, gives higher COM-RDF error than 3-bead mappings (B3, B4), and 3-ethylhexane, where 4-bead mappings (E2, E4) give higher COM-RDF errors than 3-bead mappings (E1, E3). Similar unintuitive results are seen even for VACF evaluation parameter. Lower resolution 2-bead mapping for hexane (A1) has lower VACF error than higher resolution mappings. These results corroborate with previously reported works,[33,34] which showed that increasing the resolution of a CG mapping operator does not guarantee better agreement with FG results. Note, for both evaluation metrices, there are instances where higher resolution mappings perform better than lower resolution ones, as expected, for hexane, isohexane, 3-ethylhexane and 4-methylheptane.

Additionally, we note that the performances of CG mappings depend on the evaluation metric. C4, which has higher COM-RDF mean square error than C1 and C3, yields lower VACF mean square error compared to C1 and C3. We also see the reversal of this result where asymmetric mappings which give lower COM-RDF mean square error, give higher VACF mean square error compared to the corresponding symmetric mapping operators. This is seen for propane (P1–P2), *n*-octane (D1–D2) and 3-ethylhexane (E1–E3).

We performed the Wilcoxon signed-rank test on the error values for asymmetric and symmetric mapping pairs to reach a statistical conclusion. For both COM-RDF and VACF, we calculated the difference between normalized error values for asymmetric and symmetric mapping pairs for each of the 7 molecules. This yielded a sample size of 22 paired differences. The Wilcoxon signed-rank test on COM-RDF normalized errors and on VACF normalized errors gave p-values 0.277 and 0.322 respectively. Since the p-values are greater than 0.05 for both the instances, we cannot reject the null hypothesis that there is no difference between the error values for the asymmetric and the symmetric mappings.

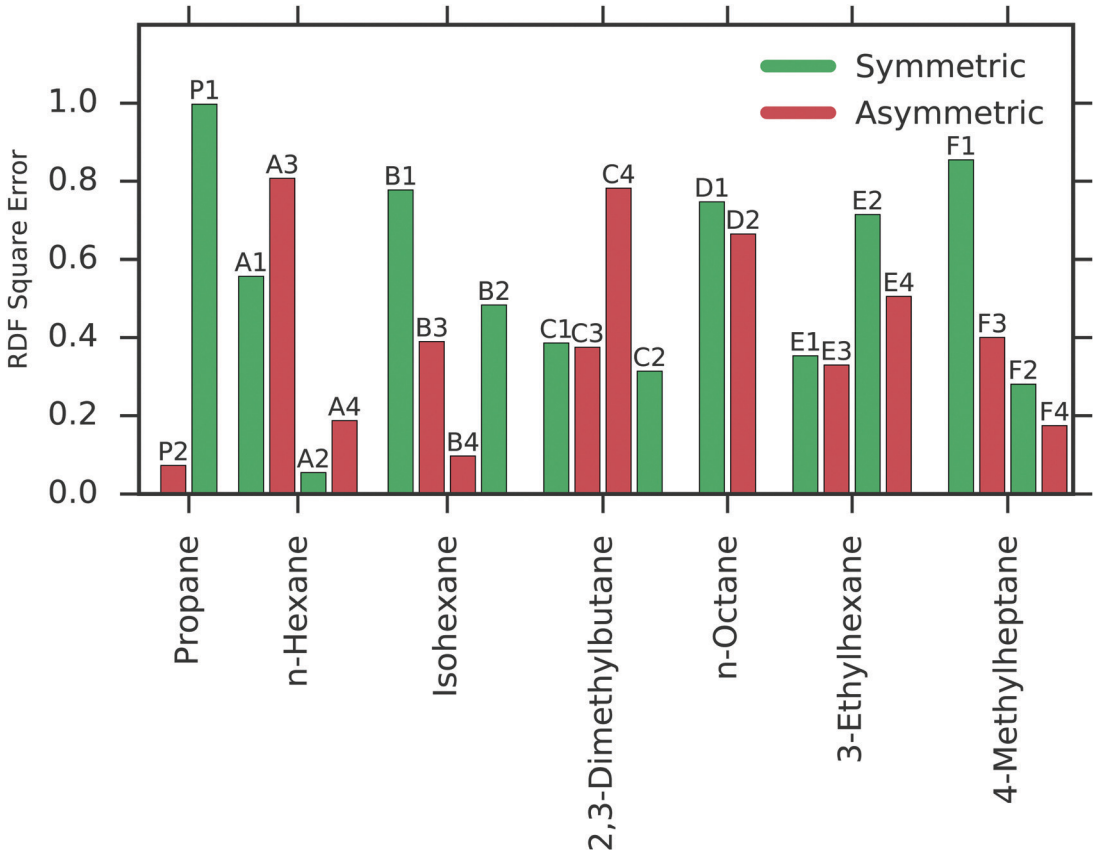


Fig. 3 COM-RDF mean square errors of symmetric and asymmetric mappings of the seven molecules: propane, *n*-hexane, isohexane, 2,3-dimethylbutane, *n*-octane, 3-ethylhexane and 4-methylheptane. For each molecule, the CG mappings have been arranged in the order of increasing resolution. Additionally, for each molecule, the COM-RDF mean square errors have been normalized over all its mappings. CG potentials obtained using FM were used for each of the CG simulation.

Fig. 5 shows the results obtained by comparing the COM-RDF and VACF normalized errors for FM, IBI and RE for six hexane mappings highlighted in red in the Fig. 2 illustration. The corresponding COM-RDFs and VACFs are included in Fig. S6 of the ESI.† The results from the normalized force error evaluation are given in Fig. 6.

Among the 6 hexane mapping operators (H1 through H6), asymmetric mappings yielded the lowest COM-RDF mean square error for all the 3 methods (H5 for FM and IBI, and H6 for RE). The symmetric mapping H2, yields COM-RDF error for FM comparable to H5. Similarly, asymmetric mappings yielded the lowest VACF mean square errors (H6 for IBI and RE, and H5 for FM). Though H3 is a symmetric 3-bead mapping, it has more skewed mass distribution among its beads than H2, a comparable 3-bead symmetric mapping operator. H3 yielded lower mean square error values than H2 for both COM-RDF and VACF for all the methods except FM where H3 has higher COM-RDF mean square error. H3 also yielded the lowest force error among the 6 mappings for FM, IBI and RE. On the contrary, the symmetric H2 mapping gave the highest force errors among the 6 mappings for all the 3 methods. The H2 mapping also gave the highest VACF mean square errors for all the 3 methods and the highest COM-RDF mean square error for IBI and RE among the 3-bead models. Among the 2-bead models, however, the asymmetric H4 mapping yielded the highest VACF and force errors for all the 3 methods and the maximum COM-RDF mean square errors for FM and IBI. Additionally, we observe pairs of symmetric and asymmetric mappings with the same degrees of freedom for all the methods (FM: H2–H6, IBI: H3–H6, RE: H3–H5) where their relative performances vary according to the choice of evaluation metric. Thus for each method, none among the 6 mappings give the best result consistently across the three evaluation parameters.

The performance of symmetric *versus* asymmetric mapping operators varies based on the metric of evaluation regardless of method and molecule. Similar results are found in Table 2 from previous work. An *et al.*,[42] in their work on developing transferable CG models for hydrocarbons, showed that a 3-bead hexane CG mapping better agreed with experimental values of self-diffusion coefficient and expansibility compared to a 2-bead mapping. However, the 2-bead mapping yielded lower error when compressibility and surface tension were considered.[42] In the same work, the asymmetric mapping for *n*-nonane (2-2-3-2) agreed with experimental values better than the symmetric mapping (3-3-3) when self-diffusion coefficient and compressibility were considered. Both the mappings, 2-2-2-3-2 and 3-3-3-2, for undecane studied by An *et al.*[42] are asymmetric. The 5 bead 2-2-2-3-2 mapping yielded compressibility and surface-tension values closer to experimental

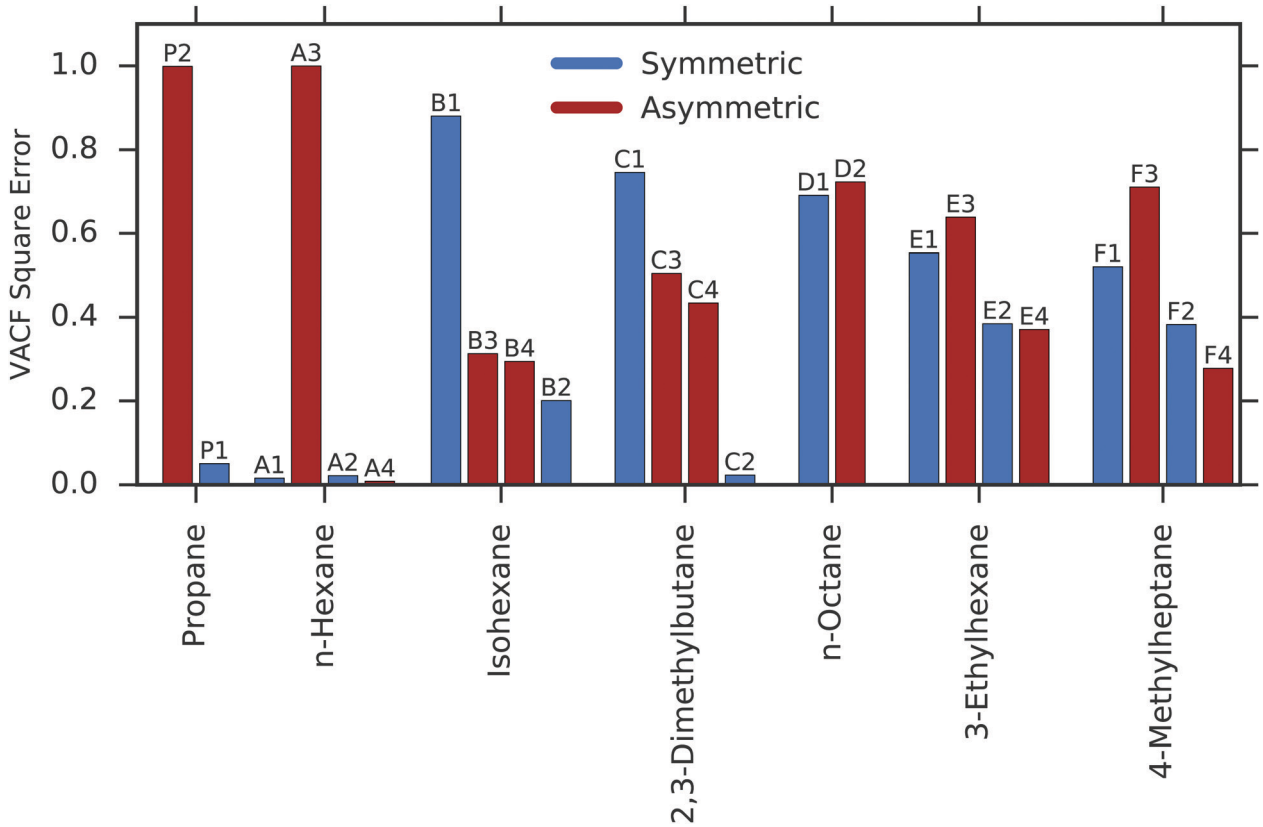


Fig. 4 VACF mean square errors of symmetric and asymmetric mappings of the seven molecules: propane, *n*-hexane, isohexane, 2,3-dimethylbutane, *n*-octanes, 3-ethylhexane and 4-methylheptane. For each molecule, the CG mappings have been arranged in the order of increasing resolution. Additionally for each molecule, the VACF mean square errors have been normalized over all its mappings. CG potentials obtained using FM were used for each of the CG simulation.

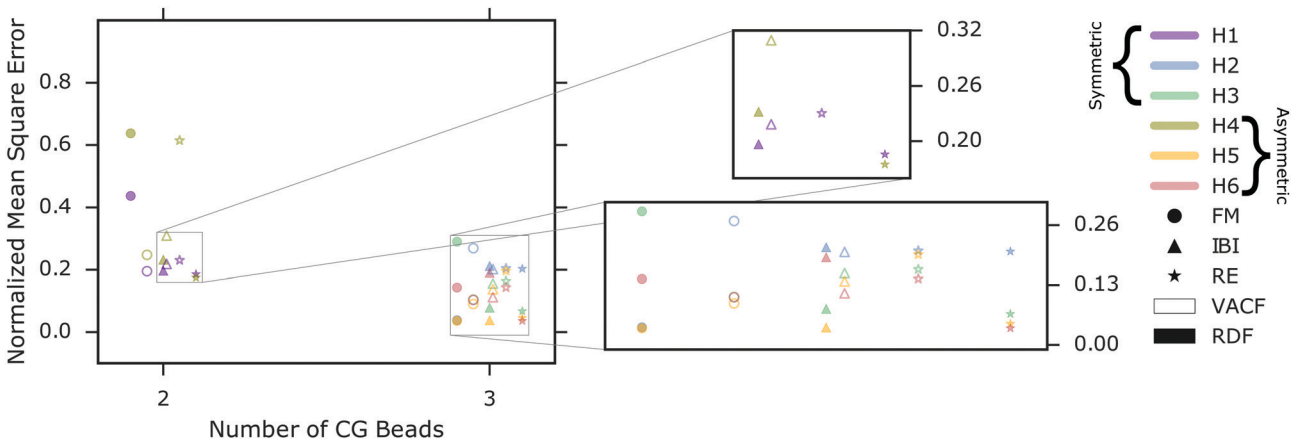


Fig. 5 Comparison of the normalized mean square errors for RDF and VACF for the symmetric and asymmetric mapping operators of hexane. Three different methods (FM, IBI and RE) were also contrasted. The three methods are denoted by different markers and the two evaluation metrics, COM-RDF mean square error and VACF mean square error, are differentiated by filled and unfilled markers, respectively.

results than the 4 bead 3-3-3-2 mapping. On the contrary, the 4 bead mapping gave expansibility and self-diffusion coefficients closer to the experimentally observed values than the 5 bead mapping.

While there are widely used evaluation metrics like the RDF and others as listed in 1, it is still a matter of preference since there is no consensus on the best metric for CG mapping evaluation. One proposed choice is mapping entropy,[44,48] although there are not many studies comparing mapping entropy of mappings and treating different resolutions requires evaluation of partition coefficients. We have chosen COM-RDF since it is not dependent on the number of beads in a CG mapping. This allows us to compare the COM-RDFs of CG

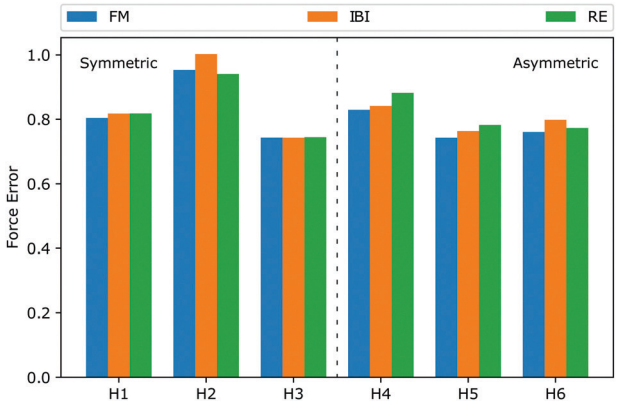


Fig. 6 Normalized square error in non-bonded forces from FM, IBI and RE for the symmetric (H1, H2, H3) and the asymmetric (H4, H5, H6) hexane mappings.

mappings of different resolutions. Our second evaluation metric, VACF, has the same advantage.

4 Conclusions

In this work we show that CG mapping operators which break symmetry sometimes perform better than symmetric CG mapping operators with comparable degrees of freedom. To our knowledge, this is the first attempt to systematically study the effect of CG mapping symmetry on their performance. Further, we provide additional evidence to support previously reported hypothesis that the information content of a CG mapping operators do not monotonically increase with resolution.[33] These two factors can be particularly useful to systematically select multi-scale CG representation of macromolecules like polymers and proteins, where it might be desirable to have specific areas of interest at higher resolutions compared to others. The results reported in this work also warrant further exploration of the possible metrics of comparison between FG and CG simulations.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1764415. We thank the Center for Integrated Research Computing at the University of Rochester for providing the computational resources required to complete this study. We also thank Dr Tristan Bereau for providing helpful feedback. Part of this research was performed while the authors were visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation (Grant No. DMS-1440415).

Notes and references

1 H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole and S. J. Marrink, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 225–248.
2 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.
3 W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.
4 M. G. Saunders and G. A. Voth, *Annu. Rev. Biophys.*, 2013, **42**, 73–93.
5 M. A. Webb, J. Y. Delannoy and J. J. De Pablo, *J. Chem. Theory Comput.*, 2019, **15**, 1199–1208.
6 J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B*, 2014, **118**, 8295–8312.
7 M. Chakraborty, C. Xu and A. D. White, *J. Chem. Phys.*, 2018, **149**, 134106.
8 P. Diggins, C. Liu, M. Deserno and R. Potestio, *J. Chem. Theory Comput.*, 2019, **15**, 648–664.
9 Y. L. Chen and M. Habeck, *PLoS One*, 2017, **12**, 1–17.
10 W. Wang and R. Gómez-Bombarelli, *npj Comput. Mater.*, 2019, **5**, 125.
11 N. J. Dunn, K. M. Lebold, M. R. Delyser, J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B*, 2018, **122**, 3363–3377.
12 V. Ruhle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, **5**, 3211–3223.
13 T. Bereau and K. Kremer, *J. Chem. Theory Comput.*, 2015, **11**, 2783–2791.
14 J. Zavadlav, G. Arampatzis and P. Koumoutsakos, *Sci. Rep.*, 2019, **9**, 99.
15 K. H. Kanekal and T. Bereau, *J. Chem. Phys.*, 2019, **151**, 164106.
16 J. Ivanov, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 596–600.
17 S. F. Mason, *Contemp. Phys.*, 1968, **9**, 239–256.
18 J. L. Carlos, *J. Chem. Educ.*, 1968, **45**, 248–251.
19 R. Obaid and M. Leibscher, *J. Chem. Phys.*, 2015, **142**, 064315.
20 J. Wei, *Ind. Eng. Chem. Res.*, 1999, **38**, 5019–5027.
21 R. Pinal, *Org. Biomol. Chem.*, 2004, **2**, 2692–2699.
22 T. Room, L. Peedu, M. Ge, D. Hüvonen, U. Nagel, S. Ye, M. Xu, Z. Bačic, S. Mamone, M. H. Levitt, M. Carravetta, J. Y. Chen, X. Lei, N. J. Turro, Y. Murata and K. Komatsu, *Philos. Trans. R. Soc., A*, 2013, **371**, 20110631.
23 A. A. Christy, Y. Ozaki and V. G. Gregoriou, *Modern Fourier Transform Infrared Spectroscopy*, Elsevier, Amsterdam, 2001, ch. 4, pp. 41–96.
24 J. D. Dunitz, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 14260–14266.
25 S. C. Lan, P. Raghunath, Y. H. Lu, Y. C. Wang, S. W. Lin, C. M. Liu, J. M. Jiang, M. C. Lin and K. H. Wei, *ACS Appl. Mater. Interfaces*, 2014, **6**, 9298–9306.
26 A. J. Thakkar, *Quantum Chemistry: A concise introduction for students of physics, chemistry, biochemistry and materials science*, IOP Publishing, 2nd edn, 2017, ch. 1, pp. 7–30.
27 D. Chambers and E. Flapan, *Symmetry*, 2014, **6**, 189–209.
28 D. J. Klein and B. Mandal, *MATCH*, 2015, **74**, 247–258.
29 W. Chen, J. Huang and M. K. Gilson, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1301–1313.
30 V. R. Rosenfeld, *J. Math. Chem.*, 2019, **57**, 1850–1867.

31 J. Jin, Y. Han and G. A. Voth, *J. Chem. Phys.*, 2019, **150**, 154103.
32 P. Buslaev and I. Gushchin, *Sci. Rep.*, 2017, **7**, 11476.
33 T. T. Foley, M. S. Shell and W. G. Noid, *J. Chem. Phys.*, 2015, **143**, 243104.
34 M. Dallavalle and N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23034–23042.
35 T. Dannenhoffer-Lafage, A. D. White and G. A. Voth, *J. Chem. Theory Comput.*, 2016, **12**, 2144–2153.
36 S. Markutsya, A. Devarajan, J. Y. Baluyut, T. L. Windus, M. S. Gordon and M. H. Lamm, *J. Chem. Phys.*, 2013, **138**, 214108.
37 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1–2**, 19–25.
38 V. Rühle and C. Junghans, *Macromol. Theory Simul.*, 2011, **20**, 472–477.
39 J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
40 G. Deichmann, V. Marcon and N. F. A. Van Der Vegt, *J. Chem. Phys.*, 2014, **141**, 224109.
41 G. Gyawali, S. Sternfield, R. Kumar and S. W. Rick, *J. Chem. Theory Comput.*, 2017, **13**, 3846–3853.
42 Y. An, K. K. Bejagam and S. A. Deshmukh, *J. Phys. Chem. B*, 2018, **122**, 7143–7153.
43 A. Khot, S. B. Shiring and B. M. Savoie, *J. Chem. Phys.*, 2019, **151**, 244105.
44 S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle and C. Junghans, *PLoS One*, 2015, **10**, 0131754.
45 V. R. Coluci, S. O. Dantas and V. K. Tewary, *Phys. Rev. E*, 2018, **97**, 053310.
46 S. D. Yi and B. J. Kim, *Comput. Phys. Commun.*, 2012, **183**, 1574–1577.
47 S. Izvekov and G. A. Voth, *J. Chem. Phys.*, 2006, **125**, 151101.
48 J. F. Rudzinski and W. G. Noid, *J. Chem. Phys.*, 2011, **135**, 214101.

