Effective Volume Correction for Lennard-Jones Static Potential Matching on Coarse-Graining Small Molecules

Qingdong Zhu,[†] Yuwei Zhang*,[‡] and Fei Xia*,[†]

[†]*School of Chemistry and Molecular Engineering, NYU-ECNU Center for Computational Chemistry at NYU Shanghai, East China Normal University, Shanghai 200062, China.*
[‡]*Jiangsu Key Laboratory of New Power Batteries, Jiangsu Collaborative Innovation Centre of Biomedical Functional Materials, School of Chemistry and Materials Science, Nanjing Normal University, Nanjing 210023, P. R. China.*

Abstract

Coarse-grained (CG) models have been widely employed in simulating the functionality of complex systems due to their lower computational demand, and the accuracy of their simulation outcomes critically depends on their parameters. In the previous study, we developed a general CG potential matching method called the Lennard-Jones Static Potential Matching (LJSPM) method for Lennard-Jones (LJ) interactions, which enables users to derive inter-system transferable LJ parameters efficiently by using only one arbitrary molecular structure. The LJSPM offers a significant means for developing physical-based non-bonded interaction parameters in CG models. However, previous results show that LJSPM generally underestimates LJ repulsive contributions, which leads to mismatching of the CG and AA LJ potential energy surfaces for short-ranged non-bonded pairs. In this work, a physical-based, general, and efficient method named the effective volume correction (EVC) is proposed to fulfill the underestimated short-ranged LJ exclusion effects for CG-LJ parameters matched by the LJSPM method. This new EVC-based LJSPM method can accurately reproduce density, the rational distribution function, and the solvation free energy of small organic molecules including alkanes, alcohols, and amines. This indicates that the EVC-based LJSPM method exhibits good transferability for the development of CG models in small molecules.

1. Introduction

CG models[1-3] offer the advantage of reduced computational demands compared to traditional all-atomic (AA) force fields,[4-6] making them particularly suitable for simulating complex system functionalities. Currently, CG models have been successfully applied to investigate long-time-scale functional studies in complex systems, such as conformational changes in proteins,[7, 8] self-assembly of cell membrane phospholipids and viral capsids,[9] as well as mechanical properties of biomaterials.[10] Due to the immense potential of CG models in studying complex system functions, they have seen rapid development in the last few decades.

CG models can be constructed in the "top-down" or the "bottom-up" ways.[1] The "top-down" way involves directly parameterizing the model using macroscopic experimental data, such as the elastic network model.[11-13] The "bottom-up" approach primarily involves rigorously developing CG models based on statistical theories starting from AA models. It allows the mapping of statistical AA results onto CG models, aiming to ensure that CG models reproduce the statistical outcomes of AA models. Such methods reported in the literature include the iterative Boltzmann inversion,[14] force matching,[15, 16] relative entropy minimization,[17, 18] and the generalized Yvon-Born-Green equations[19, 20] and so on. Although the bottom-up development of CG methods is physically rigorous, a prerequisite for using this method is the availability of an AA trajectory. Additionally, the applicability of this method is greatly influenced by the dynamics sampling of target systems.

To achieve a high transferability of CG models, their energy potential functions of CG models[21-25] are usually defined in a way similar to the AA force fields. The parameters in the CG force fields were developed by combining both of the aforementioned approaches. For example, the Martini protein force fields[26, 27] contain bonded and non-bonded interactions. Its bonding interaction parameters of Martini CG particles were derived using "top-down" fitting of structural data of proteins, while the Lennard-Jones (LJ) parameters of CG particles for vdW interactions were derived using "bottom-up" fitting based on the OPLS[6] AA simulation data.

More recently, Zhang et al. developed a method called the Lennard-Jones Static Potential Matching (LJSPM).[28] With this method, one can directly derive the LJ parameters for estimating the vdW interaction of pairwise CG particles using static structures and AA force field information. The advantage of the LJSPM method lies in that it has explicit analytical expressions for coarse-graining, which allow the derivation of LJ parameters without the need for time-consuming AA dynamic samplings or laborious parameter procedures. The LJ parameters obtained via LJSPM have clear physical meanings and satisfy the Lorentz–Berthelot combining rules[29] when computing LJ interactions. We have employed LJSPM to construct the CG models of DPPC lipids, and the subsequent CG-MD simulations successfully reproduced its self-assembly process.[28] The thermodynamic data such as free energies and enthalpies calculated from the CG-MD trajectories showed good agreement with the corresponding AA results.

The simulated DPPC results suggest that the LJSPM method can serve as an effective means for deriving LJ parameters of CG particles. In the literature, numerous CG models have been developed using a top-down approach. For example, CG models for alkanes[30-32] have been parameterized based on experimental properties like density and enthalpy of vaporization. In this work, we aim to develop accurate CG models based on a bottom-up strategy, with a particular emphasis on utilizing the LJSPM method to determine the LJ parameters of CG models.

Furthermore, we noticed that the previously developed LJSPM tends to underestimate short-range LJ repulsion between particles, leading to a reduction in the effective volume of CG particles and thus substantial discrepancies in solvation energy, density, and radial distribution functions. We develop a novel and efficient correction method called the Effective Volume Correction (EVC) to systematically correct the LJ volume exclusion. Based on the EVC-based LJSPM method, which now possesses clear physical meaning, we achieve significant improvements in accuracy when describing the solvation free energy, radial distribution functions, and density of organic molecules and their mixtures.

2. Theoretical Methods and Simulation Details

2.1 The Lennard-Jones Static Potential Matching Method

In the previous study,[28] we developed a bottom-up coarse-graining method called LJSPM that aims to derive the optimal LJ parameters from a user-defined division. The fundamental of LJSPM is matching the truncated LJ-static potential (LJSP) ψ between AA and CG models by optimizing the CG parameters $\boldsymbol{\varepsilon}^{CG}$ and $\boldsymbol{\sigma}^{CG}$, where the error function is defined in eq.(1).

$$errf(\boldsymbol{\varepsilon}^{CG}, \boldsymbol{\sigma}^{CG}|\sigma_P) = \int_{\mathbb{R}^3} |\psi^{CG}(\boldsymbol{r}, \boldsymbol{\varepsilon}^{CG}, \boldsymbol{\sigma}^{CG}|\sigma_P) - \psi^{AA}(\boldsymbol{r}|\sigma_P)|^2 d\tau \quad (1)$$

ψ is defined as the "elementary LJ interaction potential" between an additional particle probe P and the system, which can be regarded as an electrostatic-potential (ESP)-liked space-depended property for LJ interactions. However, there are several conceptual differences between LJSP and ESP. The first difference is that the concept of LJSP must be based on LJ combining rules, which reduces the original pairwise LJ parameters to atomic parameters. The second difference is that LJSP cannot be completely simplified to a function that only with respect to space. The third difference is that the large exclusion of LJSP should be truncated to avoid overmatching the "high-energy" regions. For N particles whose LJ interactions are deduced by the Lorentz–Berthelot combining rule, the definition of ψ is given in eq.(2), where $\boldsymbol{r}$, σ_P is the position and the atomic sigma parameter of the probe P. $\boldsymbol{r}_i$, σ_i and ε_i is the position, atomic sigma and atomic epsilon parameter of particle i. $min[A, B]$ refers to a truncation function that returns the minimum of A and B.

$$\psi(\boldsymbol{r}|\sigma_P) = \sum_{i \in N} \sqrt{\varepsilon_i} \cdot min\left[\left(\frac{\sigma_P + \sigma_i}{\|\boldsymbol{r}_i - \boldsymbol{r}\|}\right)^{12} - 2\left(\frac{\sigma_P + \sigma_i}{\|\boldsymbol{r}_i - \boldsymbol{r}\|}\right)^{6}, 1.0\right] \quad (2)$$

σ_P can either be defined as the mathematical expectation of all possible σ_i parameters for convenience or canceled out by the adaptive combining rule. Because the optimal atomic CG parameters $\boldsymbol{\varepsilon}^{CG}$ and $\boldsymbol{\sigma}^{CG}$ are a function of σ_p, the optimal atomic LJ parameters is dynamically related to what they coupled, which can be achieved easily by updating the original Lorentz–Berthelot combining rule to an adaptive combining rule in eq.(3), where the function $\varepsilon_i(\sigma)$ and $\sigma_i(\sigma)$ can be

obtained by iteratively matching through eq.(1) with different σ_P. In this work, σ_P is defined as an empirical value of 2.0, which can give similar potential energy predictions with the adaptive combining rule.

$$\begin{cases} \varepsilon_{ij} = \sqrt{\varepsilon_i(\sigma_j)\varepsilon_j(\sigma_i)} \\ \sigma_{ij} = \sigma_i(\sigma_j) + \sigma_j(\sigma_i) \end{cases} \tag{3}$$

2.2 The Effective Volume Correction Method

In the previous study, we found that the bulk LJSPM-based CG biomolecules, such as phospholipids, DNA, and proteins tend to underestimate short-range LJ interactions between CG-participated (CG-CG and CG-AA) pairs. This could lead to excessively strong CG interactions and an overestimation of density. A possible reason for this is that the truncated LJSP in eq.(2) failed to adequately consider the contribution of the repulsive interactions while overemphasizing the attractive contribution in LJ interactions. Therefore, we incorporate a fast-decaying exclusion potential $\varepsilon_{ij}^{exc}\left(\frac{d_{ij}}{r_{ij}}\right)^{12}$ into the interactions of pairwise CG-CG and CG-AA particles, which is shown in eq.(4):

$$\begin{aligned} V_{ij}^{LJ,AA-CG,CG-CG} &= \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \varepsilon_{ij}^{exc}\left(\frac{d_{ij}}{r_{ij}}\right)^{12} \\ &= \varepsilon_{ij}^{\gamma}\left[\left(\frac{\sigma_{ij}^{\gamma}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}^{\gamma}}{r_{ij}}\right)^{6}\right] \end{aligned} \tag{4}$$

Here, d_{ij} is the sum of the "maximum radius" of particles i and j. Maximum radius is a particle-related constant property that is defined as the sum of the maximum distance between its AA members and the particle center, and $\frac{\sigma_{ii}}{2}$. Thus, the pairwise exclusion strength becomes completely related to ε_{ij}^{exc}. It should be noticed that d_{ij} and ε_{ij}^{exc} are explicitly defined for AA particles but only enabled in computing AA-CG exclusion corrections. The greatest advantage of utilizing eq.(4) is that the exclusion-corrected LJ interaction can still be expressed as an LJ function with the corrected parameters $\varepsilon_{ij}^{\gamma}$ and σ_{ij}^{γ}, so that it can be supported by most MD packages without extra coding.

The definition of exclusion correction raises an important question, that is how to determine ε_{ij}^{exc} for all CG-CG and CG-AA interactions by a physical-based and consistent approach. The appending the exclusion term shifts the distance zero point r_{ij}^{0} of the potential, from $2^{-\frac{1}{6}}\sigma_{ij}$ to $2^{-\frac{1}{6}}\sigma_{ij}^{\gamma}$. It is not difficult to understand that the probability of finding $r_{ij} < r_{ij}^{0}$ in an MD simulation is very low because the LJ potential is sharply raised in this region. Thus, the region encircled by LJ zero-point potential surface can be understood as a type of "volume". Appending the exclusion term yields the changing (usually extension) of this "volume". Therefore, we define a property called effective volume to measure the strength of LJ exclusion, which is defined as the volume encircled by the zero-point potential surface. It should be noted that the effective volume is related to the LJ parameters of the observer (probe). e.g. From the perspective of probe P, the effective volume of a CG particle i is denoted as $\mathbb{V}_{i(P)}$ given in eq.(7):

$$\mathbb{V}_{i(P)} = \frac{4\pi}{3}\left(\sigma_{iP}^{\gamma}\right)^{3} \tag{5}$$

The expansion of black like in **Figures 1a** and **1b** shows how the exclusion term influences the size of $\mathbb{V}_{i(P)}$ when $\sigma_P = 1.0$. The colorful stripes indicate how $\mathbb{V}_{i(P)}$ expands when σ_P changes linearly from 1.0 to 2.0.

The effective volume of any CG particle i can be calculated from two representations, CG and AA. For the CG representation, the calculation of effective volume $\mathbb{V}_{i(P)}^{CG}$ is completely identical to eq.(5). For the AA representation, the effective volume $\mathbb{V}_{i(P)}^{AA}$ can be calculated by numerically integrating the elementary volume encircled by the zero-potential surface. A faster approach is to compute the radius of all spherical zero-potential surfaces for all AA members and take the volume encircled by their unions. These two approaches are named as EX1 and EX2, respectively. The repulsive volumes calculated by EX1 and EX2 are given by eq.(6) and eq.(7) respectively, where $V_{P,em}^{LJ}(\boldsymbol{r})$ represents the LJ potential energy interactions between

P and AA members that belong to CG particle i.

$$\mathbb{V}_{i(P)}^{AA,EX1} = \iiint_{|\boldsymbol{r}-\boldsymbol{r}_\mu|<\sigma_P+\sigma_\mu,\forall\mu\in i} d\tau \tag{6}$$

$$\mathbb{V}_{i(P)}^{AA,EX2} = \iiint_{V_{P,em}^{LJ}(\boldsymbol{r})>0} d\tau \tag{7}$$

The effective volume derived from the EX1 method is remarkably similar in shape to the one derived from the EX2 method, but larger, as depicted in **Figures 1c** and **1d**, which show the $\mathbb{V}_{i(P)}^{AA}$ of benzene CG particles calculated using the EX2 and EX1 approach in AA representation. In terms of the liquid density predictions, we found that the EX1 approach usually gives underestimated results while the EX2 approach usually gives overestimated results. Thus, it is appropriate to utilize the average effective volume (AEX) calculated by the EX1 and EX2 approaches to cancel out errors. The formula for calculating this AEX volume is provided in eq.(8) below:

$$\mathbb{V}_{i(P)}^{AA,AEX} = \frac{1}{2}\left(\mathbb{V}_{i(P)}^{AA,EX1} + \mathbb{V}_{i(P)}^{AA,EX2}\right) \tag{8}$$

From the perspective of probe P, the correction of the CG-LJ exclusion yields matching the effective volume of the CG representation from its AA representation $\mathbb{V}_{i(P)}^{AA}$. For the simplest case (one pair of CG particles), the particle j and i serve as the probe particle for particle i and j respectively. In this case, the error function χ^2 for the effective volume matching can be written as eq.(9) (one pair of AA, CG particles can also be handled consistently by considering the AA particle as a one-member CG particle). The condition for minimizing this error function χ^2 is given by eq.(10). By substituting the definition of the effective volume in the CG representation given by eq.(5), the correlation between the corrected σ_{ij}^{γ} and the effective volume in the AA representation is given eq.(11).

$$\chi^2 = \left(\mathbb{V}_{j(i)}^{AA} - \mathbb{V}_{j(i)}^{CG}\right)^2 - \left(\mathbb{V}_{i(j)}^{AA} - \mathbb{V}_{i(j)}^{CG}\right)^2 \tag{9}$$

$$\mathbb{V}_{j(i)}^{CG} = \mathbb{V}_{i(j)}^{CG} = \frac{1}{2}\left(\mathbb{V}_{i(j)}^{AA} + \mathbb{V}_{j(i)}^{AA}\right) \tag{10}$$

$$\sigma_{ij}^{\gamma} = \sqrt[3]{\frac{3}{8\pi}\left(\mathbb{V}_{i(j)}^{AA} + \mathbb{V}_{j(i)}^{AA}\right)} \tag{11}$$

The solution for off-diagonal σ_{ij}^{γ} values can be expressed simply as the coupling of the diagonal value of σ_{ii}^{γ} and σ_{jj}^{γ}, as shown in eq.(12). Therefore, only diagonal values need to be solved and the off-diagonal can be determined by the combining rule, where the diagonal values can be calculated iteratively solving eq.(13). Once all σ_{ij}^{γ} is determined, we can write out ε_{ij}^{exc} by using eq.(14).

$$\left(\sigma_{ij}^{\gamma}\right)^3 = \frac{1}{2}\left[\left(\sigma_{ii}^{\gamma}\right)^3 + \left(\sigma_{jj}^{\gamma}\right)^3\right] \tag{12}$$

$$\sigma_{ii}^{\gamma} = \sqrt[3]{\frac{3}{4\pi}\left(\mathbb{V}_{i(i)}^{AA}\right)} \tag{13}$$

$$\varepsilon_{ij}^{exc} = \frac{\varepsilon_{ij}\sigma_{ij}^{6}\left[\left(\sigma_{ii}^{\gamma}\right)^3 + \left(\sigma_{jj}^{\gamma}\right)^3\right]^2 - \varepsilon_{ij}\sigma_{ij}^{12}}{d_{ij}^{12}} \tag{14}$$

2.3. CG Models

To evaluate the accuracy of LJ interaction parameters obtained through the excluded volume correction algorithm, we selected several types of small molecules and constructed CG models. Our chosen molecules encompassed both straight-chain and branched alkanes, cycloalkanes, alcohol molecules, and amine compounds, as shown in **Table 1**. For the CG representation of these molecules, we adopted a united atom approach, where each carbon atom along with its attached hydrogens in the molecule was represented by a CG particle with its center of mass positioned at the

carbon atom's center. An illustration for the CG divisions of the molecules is listed in **Scheme 1**.

To improve the transferability of CG models, we defined the CG energy potential function contains the general bonded and non-bonded interactions, namely, $V^{CG} = V_{bd}^{CG} + V_{nb}^{CG}$. Bonded interactions V_{BD}^{CG} such as bond lengths, bond angles, and torsional interactions of the CG particles were inherited from the corresponding parameters of the GAFF1 force fields.[33] Non-bonded interactions V_{NB}^{CG} are composed of electrostatic and LJ interactions, namely, $V_{NB}^{CG} = V_{elec}^{CG} + V_{vdW}^{CG}$. Electrostatic interactions of CG models are described by Coulomb's law, with the charges of the CG particles being derived from the numerical summation of the atomic charges in the AA representation. As for the LJ parameters of the CG particles, we employed the EVC-based LJSPM method to derive them based on AA parameters and one arbitrary reference structure (the default structure produced by modeling packages), with the excluded volumes being calculated by either the EX1, EX2 or AEX method.

2.4 CG-MD Simulations

All simulations were carried out using the Gromacs software package.[34] The simulated systems were categorized into two types: pure solvent systems and mixed solvent systems. Pure solvent systems are comprised of 512 identical solvent molecules, while mixed solvent systems contain one solute molecule and 512 solvent molecules. For all systems, energy minimization was initially performed using the steepest descent algorithm, followed by an NPT equilibrium for a period of 1 ns. And finally, sampling NPT simulations of 5 ns were conducted. The temperature of the heat bath was set to 298.15K by utilizing the canonical velocity-rescaling thermostat with a coupling constant of 0.5 ps.[35] And the reference pressure was set to 1 bar by using the Parrinello-Rahman barostat[36] with a coupling constant of 5.0ps. Throughout the entire MD simulations, an integration timestep of 1 fs was employed. The bonds involving hydrogen atoms were constrained using the LINCS algorithm.[37] The Lennard-Jones cut-off was set to 1.2 nm with long-range dispersion corrections applied on energy and pressure. The coulomb interactions were treated with the Particle Mesh Ewald[38] using

an interpolation order of 4 and a grid of 0.12 nm resolution.

2.5 Solvation Free Energy Calculations

The solvation energy was calculated using the free energy perturbation method implemented in the GROMACS packages. The system consisted of one solute and 512 solvents and the LJ and coulomb interactions between the solute and solvents were scaled by a factor λ_{LJ} and λ_{Coul} respectively. Then, $\langle\frac{\partial H}{\partial \lambda_{LJ}}\rangle$ and $\langle\frac{\partial H}{\partial \lambda_{Coul}}\rangle$ were sampled by carrying twenty-five MD simulation replicas successively with different λ_{LJ} and λ_{Coul}. For the first twenty-one replicas, $\lambda_{Coul} = 0.0$ and λ_{LJ} increased linearly from 0.0 to 1.0. Then for the last four replicas, $\lambda_{LJ} = 1.0$ and λ_{Coul} increased to 0.25, 0.50, 0.75, 1.00 respectively.

To circumvent numerical instability in the calculation of LJ interactions, a soft-core LJ potential function was employed between the solute and solvent, with the corresponding parameter alpha and sigma taking values of 0.5 and 0.3, respectively. The overall solvation free energy is calculated by using the Bennett acceptance ratio method.[39]

3. Results and Discussion

3.1 Calculated Density Data for Alkane and Cyclic Molecules

For each molecule labeled as No.1-14 listed in **Table 1** and **Scheme 1**, three different sets of CG parameters were parameterized by the EVC-based LJSPM with the EX1, EX2, and AEX approaches respectively. Then the homogenous solution system of each molecule parameterized by different methods was constructed and relaxed by MD simulations respectively. **Figure 2** shows the comparison of the average density between these CG models and the corresponding AA model respectively.

The comparison of CG and AA results indicates that, for the No.2-10 alkane molecules, the EX1 method significantly underestimates their interaction strength, with a calculated MAD value of 111.6 kg/m³. Specifically, for the BUT molecule, the computed density value of 14.6 kg/m³ is severely underestimated, suggesting that it has undergone vaporization under the given conditions, failing to accurately reproduce

liquid properties. The severe deviation observed for BUT stems from the fact that the EX1 method treats each particle in the AA groups as mutually independent, leading to larger repulsive volumes and weaker interactions. For the No.10-14 cyclic and aromatic molecules, densities calculated using the EX1 method are close to those obtained from AA trajectories, with a calculated MAD value of 37.1 kg/m³. Compared to the EX1 method, the EX2 method tends to overestimate the interaction strength for most simulated systems, thereby leading to higher densities. For the No.1-14 molecules, the calculated MAD of using the EX2 method relative to AA data was 164.4 kg/m³, which is notably greater than the results calculated by the EX1 method.

Among the three methods, the AEX method yields the most accurate density calculations for all molecular systems, with a calculated MAD value of 57.7 kg/m³, which is smaller than the MAD values of EX1 (excluding BUT) and EX2 by 25.2 kg/m³ and 106.7 kg/m³, respectively. The AEX method combines the advantages of both the EX1 and EX2 methods, causing errors to offset each other, hence producing more accurate results. In subsequent discussions, our primary analysis focuses on simulation outcomes derived from the AEX method.

3.2 Radial Distribution Functions of CG Models

The CG models developed based on the AEX method can effectively reproduce the radial distribution functions (RDFs) of AA solvent models. Several representative cases are shown in **Figure 3. Figures 3a** and **3b** illustrate the RDFs of the CG particles in the aromatic BEN and the cyclic CHE molecules, respectively. **Figures 3c** and **3d** separately display the RDF curves for terminal sites and internal sites within OCT molecules. Both CG results closely match those of the ones of AA results.

In the case of the cyclic conjugated BEN, **Figure 3a** demonstrates that the blue RDF curve from the CG model perfectly reproduces the orange curve of the AA model. In **Figure 3b**, the RDF curves from the CG and AA models of CHE exhibit a discernible difference in the peak positions. It is thought that this discrepancy might be caused by the inconsistency between the centers of the CG particles and the spherical LJ potentials. In more detail, the LJ parameter differences between the C and H atoms in the CH_2

group of CHE are not as pronounced as their mass differences, implying that the ideal LJ potential centers should be located between the C atom and its two H atoms. The CG particles are located at the positions of C atoms in the CG models. The distances between two CG interaction centers are shorter than the actual distances between LJ potential centers. This leads to a slight contraction of the LJ interaction range in the cyclic ring compared to the corresponding AA model. Consequently, the cyclic CHE, CHP, and COC molecules possess reduced intermolecular distances during actual CG-MD simulations, resulting in an overall leftward shift of their RDF curves. For the CH groups in the BEN molecules, the center offset is less significant, thus showing good agreement with the AA RDF curve.

3.3 Solvation Free Energies of Alkane Molecules

To further validate the accuracy of the energetic properties of the CG models and their transferability, we conducted two categories of tests. In the first test, we calculated the solvation free energies ΔG for six alkane molecules including BUT, PET, HEX, OCT, CHP, and COC in both BEN and CHE solvents. **Figure 4** presents a comparative chart of the solvation free energies for the CG and AA systems, including 12 molecules in total. The absolute deviations between the solvation free energies of the CG and AA systems were maintained within 3.0 kJ/mol, with a MAD value of 1.5 kJ/mol.

Figure 4 organizes the test results into two halves according to the solvent used. On the left side, the six molecules of BUT, PET, HEX, OCT, CHP, and COC are dissolved in the CHE solvent. The calculated MAD of the solvation free energies for these systems is 1.7 kJ/mol, indicating that their CG models could accurately reproduce the solvation free energies of corresponding AA models. Moreover, the errors in solvation free energy calculated from the CG models haven't significantly increased with the size of the solutes, suggesting that the AEX method maintained high accuracy even for larger molecules. Coupled with the density data calculated in **Figure 2**, it becomes clear that the AEX method not only adequately reproduces the densities of pure alkane solutions but also accurately does the solvation free energies of small alkane molecules in the CHE solvent.

On the right side of **Figure 4**, the same six alkane molecules are dissolved in the BEN solvent. Similarly, the MAD value of the solvation free energies relative to their AA models is 1.3 kJ/mol, demonstrating a high accuracy comparable to that in the CHE solvent. This indicates that the CG models built using the AEX method exhibit good transferability across models and can effectively reproduce key properties such as the density, RDF, and solvation free energies at the AA level, especially for non-polar molecular systems dominated by LJ interactions.

3.4 Solvation Free Energies of Polar Molecules

In the second category of tests, we endeavored to simulate and calculate the solvation free energies of six polar molecules including the MOL, EOL, BOL, TAM, BAM, and NBN in both CHE and BEN solvents individually. To decrease the errors caused by electrostatic interactions, we employed hybrid AA/CG models to perform simulations, where each polar molecule above was represented in the AA model, while other CHE and BEN solvents in each system were described using the CG model.

The solvation free energy data obtained from the AA/CG simulations are illustrated in **Figure 5**. Concerning the solvation free energies of polar molecules in CHE, the MAD value estimated from the AA/CG systems relative to corresponding AA systems is 2.4 kJ/mol. The relatively accurate results are attributed to the small dipole moments of the CH_2 groups in CHE, which lead to weak electrostatic interactions with the polar molecules and allow the AA/CG models to accurately reproduce the solvation free energies of these systems.

By contrast, the right side of **Figure 5** reveals poorer results for the solvation free energies of the six polar molecules in BEN. The MAD value for these cases is 6.5 kJ/mol, which is 4.1 kJ/mol higher than in the CHE case. Notably, for the NBN molecule, the maximum absolute deviation reaches 10.1 kJ/mol. This suggests that in BEN solutions, polar molecules with larger dipoles exert strong polarization effects on BEN, generating dipole-charge interactions that cannot be neglected in the electrostatic interactions. In our future research, we will further develop coarse-grained models for

polar molecules, incorporating accurate descriptions of electrostatic interactions within the models to address these discrepancies.

4. Conclusion

In this study, we introduced an effective volume-based repulsion correction (EVC) method, to refine LJ exclusion contributions that have been underestimated by the LJSPM approach. The EVC method determines exclusion parameters efficiently and systematically by matching the effective volume, which can be ultimately by described by a simple and inter-system transferable combining rule. The reference effective volume can be estimated by three approaches namely EX1, EX2 and AEX, where the AEX approach gives the best predictions. Our simulation results suggest that the accuracy of the EVC-based LJSPM is comprehensively and significantly boosted compared to the original LJSPM for non-polar molecular systems predominantly governed by LJ interactions. Even with all the bottom-up information required by the coarse-graining process taken from one arbitrary reference, the EVC-based LJSPM can still accurately reproduce critical physical properties such as densities and radial distribution functions of several non-polar AA liquids. Furthermore, the EVC-based LJSPM shows promise in accurately describing the solvation free energy information of solution systems, even when the system incorporates polar molecules. This indicates a tremendous potential in developing non-polar CG solvents.

However, a broader application scope is still required for the comprehensive validation of the EVC-based LJSPM method. For example, CG models with coarser representations or stronger electronic interactions. Thus, our future tests or method improvements will focus on describing coarser or polar CG molecules to expand the applicability of the EVC-based LJSPM method.

ASSOCIATED CONTENT

AUTHOR INFORMATION

Corresponding Authors

*Emails: ywzhang@nnu.edu.cn (Y. Z.), and fxia@chem.ecnu.edu.cn (F.X.)

ORCID

Yuwei Zhang: 0009-0002-8741-3889

Fei Xia: 0000-0001-9458-9175

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grants No. 22073029, 21773065). We also acknowledge the support of the NYU-ECNU Center for Computational Chemistry at NYU Shanghai as well as the ECNU Public Platform for Innovation (001) for providing computer time.

References:

1. Noid, W. G. Perspective: Coarse-grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013,** *139* (9), 090901.
2. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016,** *116* (14), 7898-936.
3. Pak, A. J.; Voth, G. A. Advances in Coarse-grained Modeling of Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2018,** *52*, 119-126.
4. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983,** *4* (2).
5. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986,** *7* (2), 230-252.
6. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996** *118* (45), 11225-11236.
7. Zhang, Z.; Sanbonmatsu, K. Y.; Voth, G. A. Key Intermolecular Interactions in the E. coli 70S Ribosome Revealed by Coarse-grained Analysis. *J. Am. Chem. Soc.* **2011,** *133* (42), 16828-38.
8. Zhang, Y.; Cao, Z.; Zhang, J. Z.; Xia, F. Double-Well Ultra-Coarse-Grained Model to Describe Protein Conformational Transitions. *J. Chem. Theory Comput.* **2020,** *16* (10), 6678-6689.
9. Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. Coarse-grained Simulation Reveals Key Features of HIV-1 Capsid Self-assembly. *Nat. Commun.* **2016,** *7*, 11568.
10. Zha, J.; Zhang, Y.; Xia, K.; Grater, F.; Xia, F. Coarse-Grained Simulation of Mechanical Properties of Single Microtubules With Micrometer Length. *Front Mol Biosci* **2020,** *7*, 632122.
11. Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996,** *77* (9), 1905-1908.
12. Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001,** *80* (1), 505-15.
13. Xia, F.; Tong, D.; Lu, L. Robust Heterogeneous Anisotropic Elastic Network

Model Precisely Reproduces the Experimental B-factors of Biomolecules. *J. Chem. Theory Comput.* **2013,** *9* (8), 3704-14.
14. Reith, D.; Meyer, H.; Müller-Plathe, F. Mapping Atomistic to Coarse-Grained Polymer Models Using Automatic Simplex Optimization To Fit Structural Properties. *Macromolecules* **2001,** *34* (7), 2335-2345.
15. Izvekov, S.; Voth, G. A. A Multiscale Coarse-graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005,** *109* (7), 2469-2473.
16. Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The Multiscale Coarse-graining Method. I. A Rigorous Bridge between Atomistic and Coarse-grained Models. *J. Chem. Phys.* **2008,** *128* (24), 244114.
17. Shell, M. S. The Relative Entropy is Fundamental to Multiscale and Inverse Thermodynamic Problems. *J. Chem. Phys.* **2008,** *129* (14), 144108.
18. Chaimovich, A.; Shell, M. S. Coarse-graining Errors and Numerical Optimization Using a Relative Entropy Framework. *J. Chem. Phys.* **2011,** *134* (9), 094112.
19. Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009,** *103* (19), 198104.
20. Mullinax, J. W.; Noid, W. G. A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials. *Journal of Physical Chemistry C* **2010,** *114* (12), 5661-5674.
21. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007,** *111* (27), 7812-24.
22. Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y. D. PACE Force Field for Protein Simulations. 1. Full Parameterization of Parameterization of Version 1 and Verification. *J. Chem. Theory Comput.* **2010,** *6*, 3373–3389.
23. Kar, P.; Gopal, S. M.; Cheng, Y. M.; Predeus, A.; Feig, M. PRIMO: A Transferable Coarse-grained Force Field for Proteins. *J. Chem. Theory Comput.* **2013,** *9* (8), 3769-3788.
24. Shen, H.; Li, Y.; Ren, P.; Zhang, D.; Li, G. An Anisotropic Coarse-Grained Model for Proteins Based On Gay-Berne and Electric Multipole Potentials. *J. Chem. Theory Comput.* **2014,** *10* (2), 731-750.
25. Darre, L.; Machado, M. R.; Brandner, A. F.; Gonzalez, H. C.; Ferreira, S.; Pantano, S. SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **2015,** *11* (2), 723-39.
26. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J.*

Chem. Theory Comput. **2008,** *4* (5), 819-34.
27. Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grunewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domanski, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021,** *18* (4), 382-388.
28. Zhang, Y.; Wang, Y.; Xia, F.; Cao, Z.; Xu, X. Accurate and Efficient Estimation of Lennard-Jones Interactions for Coarse-Grained Particles via a Potential Matching Method. *J. Chem. Theory Comput.* **2022,** *18* (8), 4879-4890.
29. Kong, C. L. Combining Rules for Intermolecular Potential Parameters. II. Rules for the Lennard-Jones (12–6) Potential and the Morse Potential. *J. Chem. Phys.* **1973,** *59* (5), 2464-2467.
30. Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Descriptioni of n-Alkanes. *J. Phys. Chem. B* **1998,** *102* (14), 2569-2577.
31. Avendano, C.; Lafitte, T.; Adjiman, C. S.; Galindo, A.; Muller, E. A.; Jackson, G. SAFT-gamma force field for the simulation of molecular fluids: 2. Coarse-grained models of greenhouse gases, refrigerants, and long alkanes. *J. Phys. Chem. B* **2013,** *117* (9), 2717-33.
32. Papavasileiou, K. D.; Peristeras, L. D.; Bick, A.; Economou, I. G. Molecular Dynamics Simulation of Pure n-Alkanes and Their Mixtures at Elevated Temperatures Using Atomistic and Coarse-Grained Force Fields. *J. Phys. Chem. B* **2019,** *123* (29), 6229-6243.
33. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25* (9), 1157–1174.
34. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.,* **2005,** *26* (16), 1668-88.
35. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101.
36. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
37. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.
38. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C.

Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9* (9), 3878-88.
39. Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.

Table 1. It lists the serial numbers, names, and abbreviations of the test molecules. In this table, No.l-9 represent alkane molecules, No.10-14 denote cyclic molecules, and No. 15-20 pertain to alcohol and amine molecules.

No.	Names	Abbr.	No.	Names	Abbr.
1	Butane	BUT	11	Toluene	TLU
2	Pentane	PET	12	Cyclohexane	CHE
3	Hexane	HEX	13	Cycloheptane	CHP
4	Octane	OCT	14	Cyclooctane	COC
5	Isopentane	IPE	15	Methanol	MOL
6	Isohexane	IHE	16	Ethanol	EOL
7	Neopentane	NPE	17	Butanol	BOL
8	Neohexane	NHE	18	Trimethylamine	TAM
9	Neooctane	NOC	19	Butylamine	BAM
10	Benzene	BEN	20	Benzenamine	NBN

Scheme 1. The CG divisions of the No.1-14 molecules in **Table 1** is presented. In the united-atom scheme, each C atom and its attached H atoms are represented by a single CG particle, whose center is positioned at the location of the C atom. Blue, green, and yellow CG particles denote methyl (CH_3), methylene (CH_2), and methyne (CH) groups in alkanes, respectively. Red CG particles symbolize aromatic CH groups. Carbon atoms without any attached hydrogen atoms are also treated as individual CG particles.

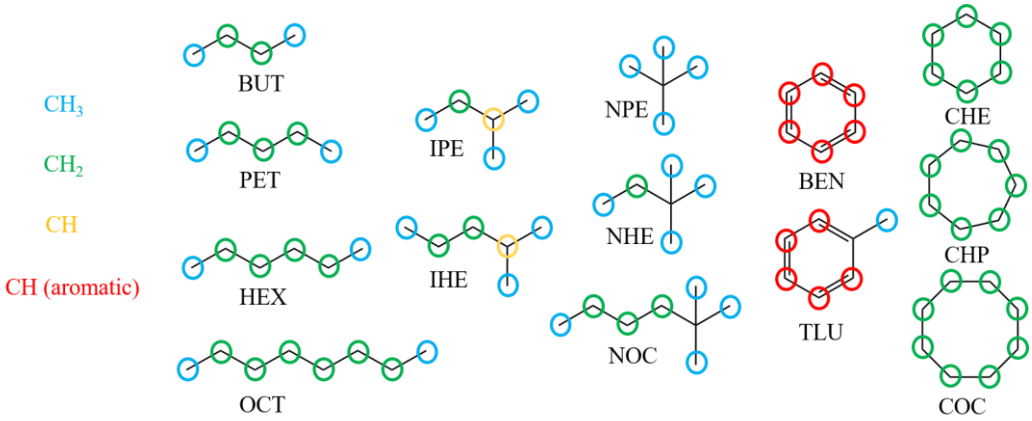

Figure 1. The 2D schematic diagram of effective volumes of a CG particle in a three-site benzene CG model, (a) $V_{i(P)}^{CG}$, exclusion term ignored ($\sigma_{iP}^{\gamma} = \sigma_{iP}$), (b) $V_{i(P)}^{CG}$ exclusion term considered, (c) $V_{i(P)}^{AA,EX2}$, (d) $V_{i(P)}^{AA,EX1}$, with σ_P linearly scaled from 1.0 (black) to 2.0 (yellow).

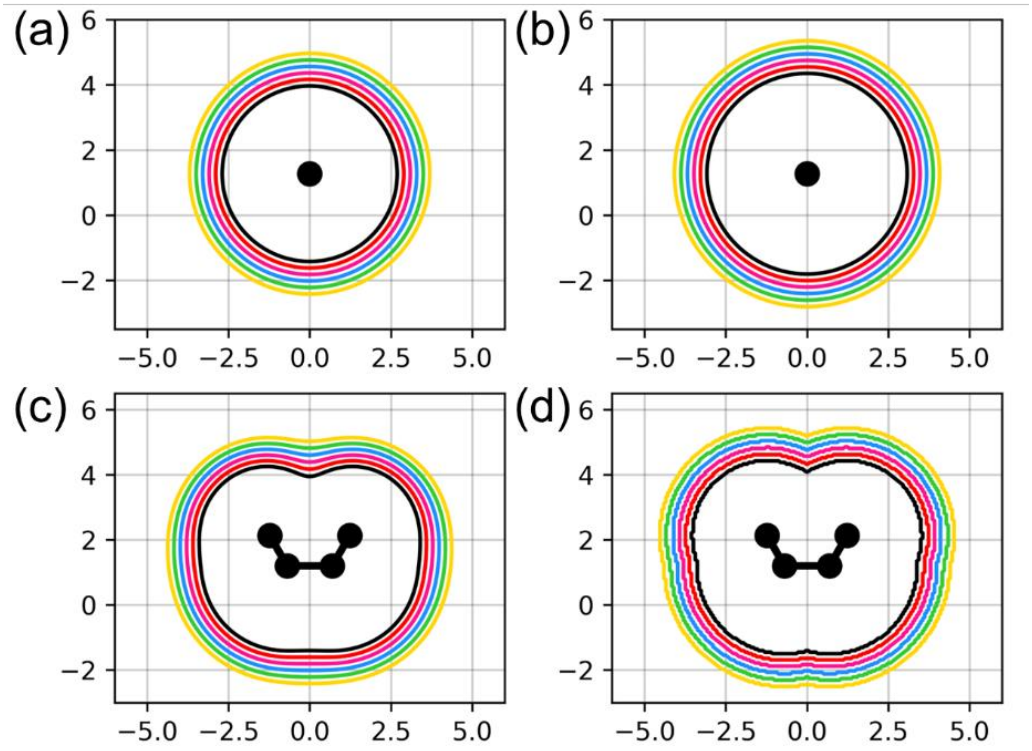

Figure 2. Comparison of the density data calculated from the AA and CG models of the No.1-14 molecules listed in **Table 1** using the LJ parameters derived from EX1, EX2 and AEX methods, respectively.

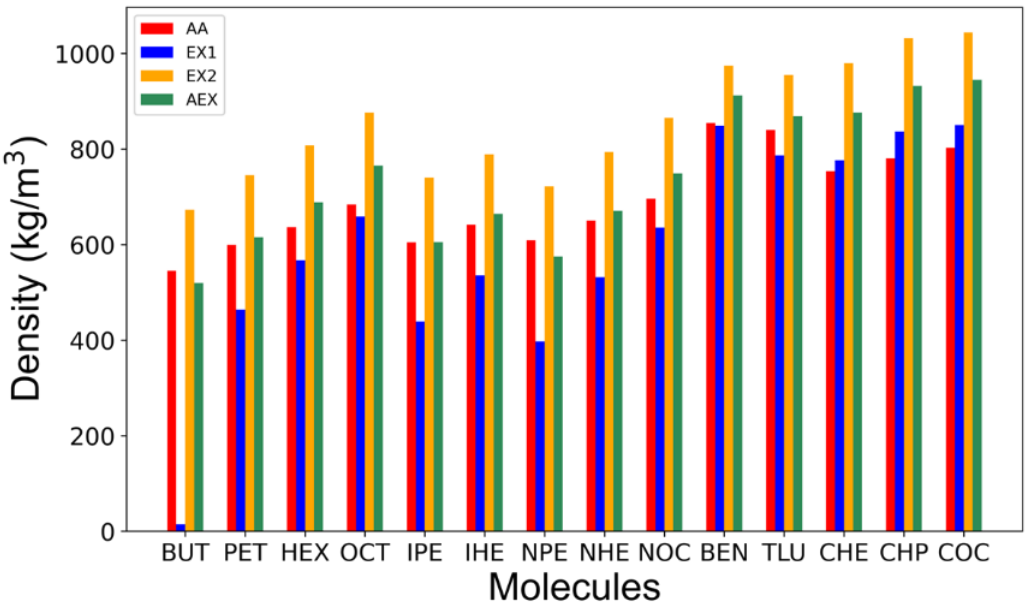

Figure 3. The radial distribution functions (RDFs) calculated from both AA and CG simulations of (a) BEN, (b) CHE, (c) OCT, and (d) OCT are compared with each other. The blue and yellow curves represent the AA and CG RDFs, respectively. The circles marked in the molecules denote the CG sites used for evaluation of RDFs.

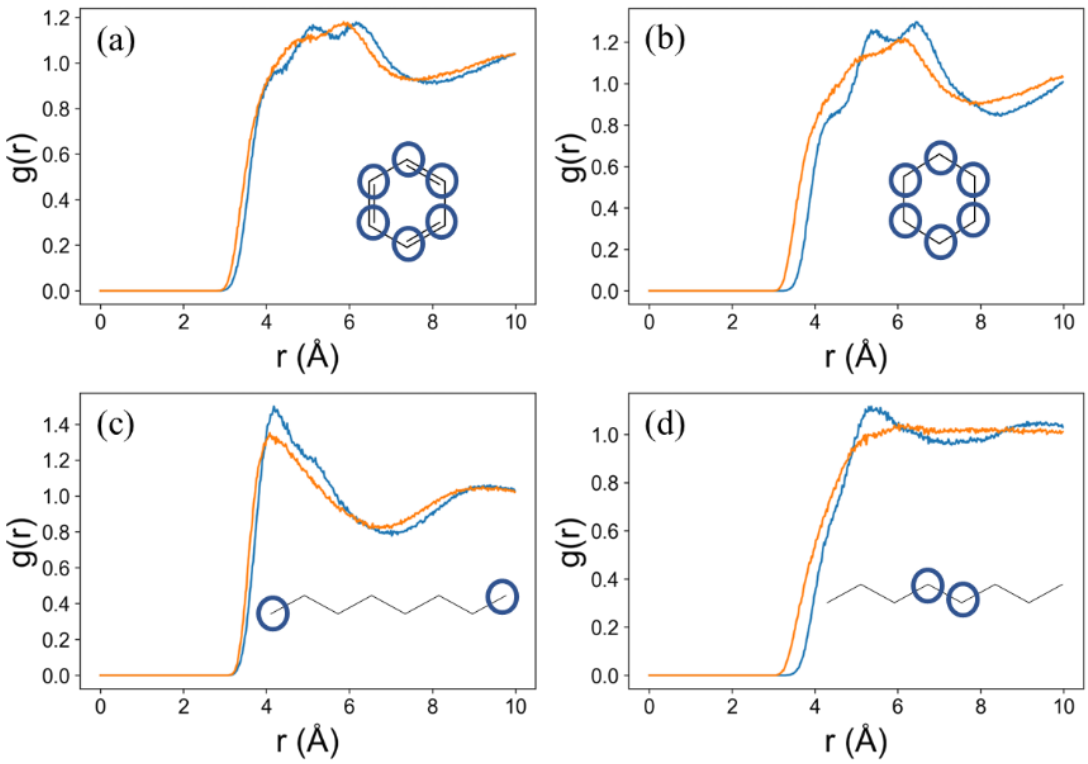

Figure 4. The calculated solvation free energy data for the No. 1-4, 13, and 14 molecules of **Table 1** in solvents CHE and BEN are presented. The energy units are kJ/mol. The red and blue bars denote the results obtained from AA and CG simulations, respectively.

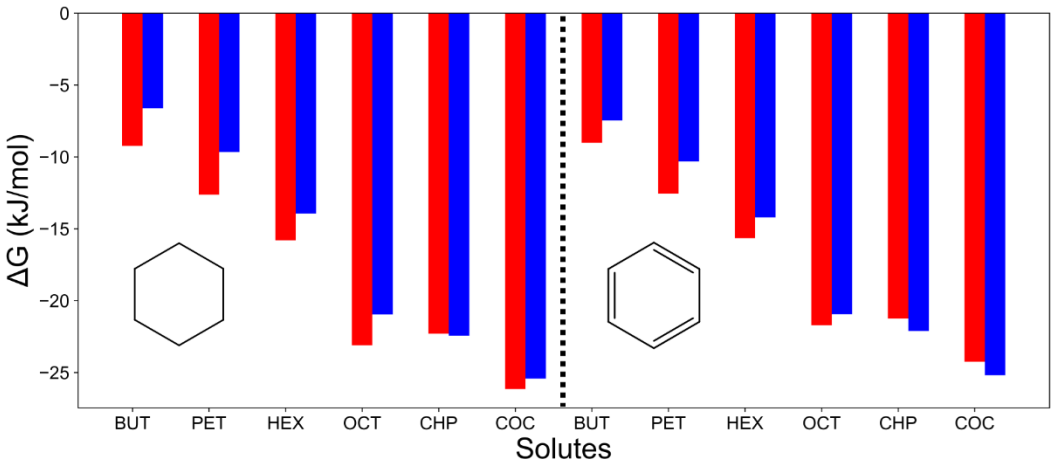

Figure 5. The calculated solvation free energy data for the No. 15-20 molecules of **Table 1** in the CHE and BEN solvents are presented. The energy units are given in kJ/mol. The red bars represent the AA results, while blue bars depict the results from simulations using a hybrid AA/CG model. In the hybrid AA/CG model, individual solute molecules are represented in the AA model, while the remaining solvent molecules are depicted using the CG models.