JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

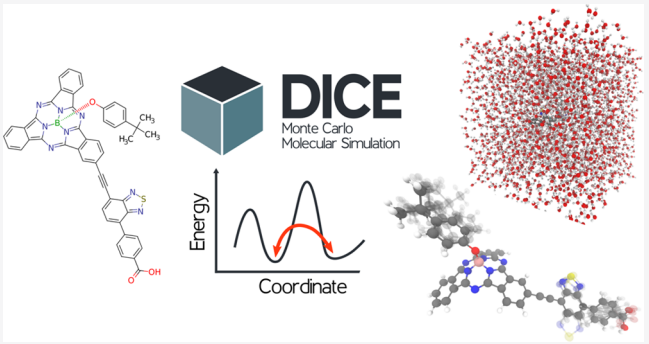
# DICE: A Monte Carlo Code for Molecular Simulation Including the Configurational Bias Monte Carlo Method

Henrique M. Cezar,* Sylvio Canuto,* and Kaline Coutinho*

Cite This: J. Chem. Inf. Model. 2020, 60, 3472−3488 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Solute−solvent systems are an important topic of study, as the effects of the solvent on the solute can drastically change its properties. Theoretical studies of these systems are done with *ab initio* methods, molecular simulations, or a combination of both. The simulations of molecular systems are usually performed with either molecular dynamics (MD) or Monte Carlo (MC) methods. Classical MD has evolved much in the last decades, both in algorithms and implementations, having several stable and efficient codes developed and available. Similarly, MC methods have also evolved, focusing mainly in creating and improving methods and implementations in available codes. In this paper, we provide some enhancements to a configurational bias Monte Carlo (CBMC) methodology to simulate flexible molecules using the molecular fragments concept. In our implementation the acceptance criterion of the CBMC method was simplified and a generalization was proposed to allow the simulation of molecules with any kind of fragments. We also introduce the new version of DICE, an MC code for molecular simulation (available at https://portal.if.usp.br/dice). This code was mainly developed to simulate solute−solvent systems in liquid and gas phases and in interfaces (gas−liquid and solid−liquid) that has been mostly used to generate configurations for a sequential quantum mechanics/molecular mechanics method (S-QM/MM). This new version introduces several improvements over the previous ones, with the ability of simulating flexible molecules with CBMC as one of them. Simulations of well-known molecules, such as *n*-octane and 1,2-dichloroethane in vacuum and in solution, are presented to validate the new implementations compared with MD simulations, experimental data, and other theoretical results. The efficiency of the conformational sampling was analyzed using the acceptance rates of different alkanes: *n*-octane, neopentane, and 4-ethylheptane. Furthermore, a very complex molecule, boron subphtalocyanine, was simulated in vacuum and in aqueous solution showing the versatility of the new implementation. We show that the CBMC is a very good method to perform conformation sampling of complex moderately sized molecules (up to 150 atoms) in solution following the Boltzmann thermodynamic equilibrium distribution.



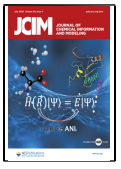



## INTRODUCTION

The efficient sampling of structural and thermodynamic properties of molecules has always been of great interest in molecular science since the advent of computers. Effects of the molecular interactions can change intramolecular properties, such as the relative stability of conformations and intermolecular properties, like specific interactions as hydrogen bonds or $\pi$ stacking. Based on those changes and interactions, other properties such as optical and transport properties or response to electric and magnetic fields may change. For this reason, efficient sampling methods for different types of molecules have always been a topic of interest.

The sampling problem in the thermodynamic equilibrium is usually approached by either Molecular Dynamics (MD) or Monte Carlo (MC) simulations, and some authors compare the good performance of both methods for low rotational barriers[1,2] (around 3−4 kcal/mol) showing the validity of the ergodic theorem. But for large rotational barriers (higher than 10 kcal/mol), it was shown that MD sampling is not ergodic because it becomes trapped in the initial conformation due to the high-energy rotational barrier, while the MC sampling is able to perform a good conformational sampling.[3] Therefore, the MD method is not recommended for conformational sampling of molecules with large barriers, and other techniques should be combined with the MD method to enhance the conformational sampling, such as temperature replica exchange molecular dynamics methods,[4] biased molecular dynamics methods,[5] and several others.[6,7] Great efforts have been made to provide the

Received: January 22, 2020
Published: May 29, 2020







https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

stability and efficiency of molecular simulations with MD methods. These efforts have resulted in the development of several computational packages such as GROMACS,[8] LAMMPS,[9] AMBER,[10] and NAMD,[11] among others. These software packages have been used to tackle different problems in molecular science and have a great user base. On the other hand, the MC software packages are usually designed with a specific class of problems in mind and focus on ensembles that are hardly approached with MD, such as the grand-canonical or Gibbs ensembles, or systems where MD simulation is inefficient, like systems with discontinuity in the potentials or large energy barriers. Some of the MC software suited to molecular simulations are BOSS,[12] Cassandra,[13] Towhee,[14] RASPA,[15] Chameleon,[16] Diadorim,[17] and DICE,[18] among others. From this list, not all of these software packages were developed to sample intramolecular flexibility, and the ones that seek to achieve this purpose use different methods and sometimes have limitations on the systems that can be simulated.

In this paper, we introduce a new version of DICE,[19] which includes an implementation of a configurational bias Monte Carlo (CBMC)[20] method with new features, used to sample the intramolecular flexibility of general molecules with moderate size (up to 150 atoms). DICE is designed to perform molecular simulation of liquids, gases, and solid−liquid and gas−liquid interfaces, with a focus in solute−solvent systems. The software is written in Fortran, supports multithreading with OpenMP and has several algorithms implemented to efficiently sample configurations of molecular systems in the thermodynamic equilibrium, for rigid, flexible, and semiflexible molecules. Details about the code options, algorithms, parallelization efficiency, and input files are given in the Supporting Information (SI). This code was developed in the early 1990s for rigid molecules in the *NVT* or *NPT* ensembles with the molecular interaction described by the Lennard-Jones plus Coulomb potentials, and the first publication featuring simulations performed with DICE was for liquid benzene.[21] Since then, the code has been continuously updated. It has been largely used to study different solvent effect problems[3,22−31] mainly to sample configurations for the sequential hybrid method with quantum mechanics and molecular mechanics approach (S-QM/MM).[22−24,32] In this last version we implemented the CBMC method of Shah and Maginn[33] for flexible molecules, due do its simplicity and possibility of generalization. We also detailed the simplification of the acceptance criterion presented in a previous work[3] and introduced a new scheme that allows fragments of any shape and sample all degrees of freedom including the bond distances that were not sampled in the original method.[33] With this new scheme, noncyclic molecules with any kind of complex topology can be simulated. For cyclic molecules, concerted rotations and other MC methods[34−36] are better suited than the CBMC method.

The choice of CBMC as the method implemented in DICE is based on the simplicity of the method and the ability of extending the implementation to deal with molecules of different types or shapes. Also, the use of rigid fragments as allowed by our implementation permits the sampling of only the most important degrees of freedom, simplifying the parametrization for complex molecules. Historically, most of the algorithms developed to sample the internal degrees of molecules have been developed to sample simple chains. However, there is no consensus in the literature of a single MC method to sample the internal degrees of freedom of a general molecule in an efficient fashion. Those algorithms go from concerted rotations[2,34−36] to changes in the internal coordinates using the Z-matrix representation.[1] Concerted rotations, for example, are particularly useful to sample the dihedral angles of a polymer or peptide backbone with several torsions, i.e., sampling the inner dihedral angles of long chains,[2,34−36] while the use of internal coordinates is useful for simple and small molecules. Besides these successful methods for long chains, CBMC methods have been studied and improved throughout the years for branched molecules.[33,37−42]

In general, the CBMC methods break the molecular structure at some parts and then reconstruct the molecule by reinserting the parts using biased insertion angles favoring the ones with lower energy. The detailed balance is satisfied, and the bias is removed in the acceptance criterion for new configurations. Thus, the thermodynamic equilibrium is guaranteed. For molecules composed by branch and ring groups, it is useful to separate the called hard and soft degrees of freedom. The hard degrees are generally formed by bond distances, angles, and improper angles. Small deformations in the hard degrees cause typically large intramolecular energy variations that are orders of magnitude larger than the thermal energy. Additionally, it is expected that they should have small influence in the molecular conformational distribution. On the other hand, the soft degrees are formed by rotational angles. Small deformations in the soft degrees cause typically intramolecular energy variations competitive to the thermal energy. Then, they are the most important degrees to influence the molecular conformational distribution. This strategy for separating the hard and soft degrees of freedom was proposed by Maginn and coauthors[33,39] in different ways but always considering rigid bond distances. In this method the molecule is divided into fragments and new conformations are generated by reconnecting those fragments with different dihedral angles (soft degrees of freedom). These fragments are parts of the molecule composed by some atoms connected by the hard degrees of freedom containing one common bond that binds the fragment with its neighbors. Therefore, the degrees of freedom within the fragment are considered hard and between the fragments are soft.

An initial stage of the simulation is the sampling of the hard degrees of freedom of each isolated fragment. The sampled geometries of each fragment are stored, creating a library of fragment conformations that is used in the reconstruction of the molecule. Shah and Maginn[33] have proposed fragments of two types: with a central atom, so-called branch, or with simple rings, which were sampled with rigid bonds. In our implementation we used these two fragment types with rigid and flexible bonds and propose an additional fragment type with a general shape. This bond flexibility is important since it can strongly influence the electronic properties of molecules in solution commonly studied by hybrid S-QM/MM method. Therefore, with our implementation, we were able to properly study the solvent effects in electronic properties of chromophores[31] and sample configuration stability, such as *syn:anti* or *cis:trans*, with high energy barriers in solution.[3] The complete CBMC move consists of (i) breaking the molecule in a random fragment connection, (ii) selecting one of the sides and deleting the fragments from this side, keeping the other side unchanged, (iii) selecting one configuration of each deleted fragment library with equal probability, and finally (iv) reconstructing the entire molecule connecting each fragment in the new configuration with a new dihedral angles selected with a probability proportional to the Boltzmann distribution in a small set of trial dihedral angles, (v)


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

calculating the probability of reconstructing the old conformation of the entire molecule with a similar procedure to obtain the acceptance criterion.

We implemented the CBMC method in DICE considering the original method[33] including new ideas: (i) a generalized shape fragment that allows the simulation of noncyclic molecules with any topology and (ii) sampling all hard degrees of freedom, including bond distances and improper dihedral angles that allow the simulation of fully flexible molecules.

To validate our implementation we performed simulations of two well-known systems: *n*-octane in vacuum and in chloroform and 1,2-dichloroethane (DCE) in vacuum, in liquid phase, and in acetonitrile solution. Then, we also tested the influence of the number of CBMC trial insertion angles in the efficiency of the conformational sampling analyzing the acceptance rates of linear or branched alkanes, such as *n*-octane, neopentane (2,2-dimethylpropane), and 4-ethylheptane. Furthermore, we perform simulations with a very complex molecule, a donor−acceptor functionalized boron subphtalocyanine (SubPC). This molecule is composed by a macro-cyclic ring conically shaped of a subphtalocyanine group with boron, i.e., three aza-bridged isoindole units and a central four-coordinate boron atom,[43] with an axial substitution in the boron atom of a 4-(*tert*-butyl)-phenolate and a peripheral substitution in one isoindole unit of a 4-(7-ethynylbenzo[*c*][1,2,5]thiadiazol-4-yl)benzoic acid (EBTBA). This class of molecule is attractive as light harvesting materials in photovoltaic devices, such as dye-sensitized solar cells (DSSC).[44]

First, we perform *n*-octane simulations in vacuum and in solution considering different energy contributions, i.e., just the dihedral energy and the full interacting torsional potential with the bonded and nonbonded terms. The vacuum simulations were made to mimic the ideal gas condition where we could compare directly the torsional potential and the sampled dihedral distribution obtained from the CBMC simulations. We compare these results with those obtained with MD simulations and the populations expected by the rate constants of the ideal gas. These results for *n*-octane are in good agreement with the MD results. Concerning the efficiency of the CBMC implementation, we obtained high acceptance rates (around 60−78%) for moving fragments at the extremes of linear and branched alkanes and good acceptance (around 18−41%) for fragments in the center of the linear alkane (*n*-octane) and moderate acceptance (around 8−23%) for fragments in the center of branched alkane (4-ethylheptane). Therefore, these results show the good efficiency of the CBMC method implemented in the DICE code. Second, we present the simulation results for DCE in vacuum, in liquid phase, and in solution and compared these with the experimental data and other theoretical results, showing a good agreement of the *trans* % populations that is solvent dependent. Finally, we perform simulations of SubPC in vacuum and in aqueous solution showing that the CBMC method can sample well very complex rotational energy profiles in gas phase and in solution, even for voluminous groups. With our new ideas described here, the CBMC method is more general and powerful to simulate flexible noncyclic molecules of any shape and moderate size in solution.

## MONTE CARLO METHODS

**Metropolis Monte Carlo Sampling.** The implementation of the Metropolis MC method in DICE uses standard algorithms for simulations of liquids and gases, such as random selection of one molecule for each MC step, translational displacements in Cartesian coordinates with an auto-adjusted maximum value, rotational movements around a random axis, isotropic volume rescale (in the *NPT* ensemble), image method, periodic boundary condition, neighbor list, cutoff radius, and calculation of the long-range correction for the Lennard-Jones potential assuming a uniform distribution, i.e., $G(r) \approx 1$, and for the Coulomb potential assuming the reaction-field approximation beyond the cutoff radius. A good description of these standard algorithms can be found in many books.[20,45] Additionally, there are some nonstandard procedures, such preferential sampling of the solvent molecules concerning their distance to the solute with the one-over-$r^n$ method,[46,47] and calculations of free energy differences using the free energy perturbation (FEP)[48,49] considering changes in the solute composition, geometry, and the force field that allows the calculation of solvation free energy and reaction paths. DICE is mainly designed to the study of solvent−solute systems. In this section we give a brief overview of the Metropolis MC method to define the notation.

Both the kinetic and potential parts of the energy should in principle be considered for calculating the thermodynamic properties. However, it can be shown using statistical mechanics that the contributions of the kinetic and potential part of the energy can be separated. The kinetic part contribution can be calculated analytically leading to the well-known ideal gas properties. On the other hand, the potential energy contribution is very complex and most of the time cannot be calculated analytically. Then, only the potential energy is considered in the molecular simulations and the kinetic contribution to the thermodynamic properties are added to the final calculated values after the simulation ends.

The Metropolis MC method is an importance sampling method used to sample states in thermodynamic equilibrium that follows the Boltzmann probability distribution

$$P(n) = \frac{\exp[-\beta \mathcal{U}(\mathbf{r}_n^N)]}{\sum_{n=1}^{\infty} \exp[-\beta \mathcal{U}(\mathbf{r}_n^N)]} \tag{1}$$

where $P(n)$ is the probability of a state $n$, $\beta = 1/kT$ is the inverse temperature where $k$ is the Boltzmann constant, and $\mathcal{U}(\mathbf{r}_n^N) = \mathcal{U}(n)$ is the total potential energy of the state $n$ represented by the configuration with the atomic positions of the $N$ atoms system $\mathbf{r}^N$. Sampling from this distribution speeds up the simulation, since only states with a significant probability are sampled and allows the calculation of thermodynamic properties as simple averages.

The sampling of the Boltzmann distribution is done through a Markov chain method, where a new state $n$ is generated based on the previous one, or the old state $o$. This new state may, or may not, be accepted, meaning that the system goes to the new trial configuration or stays in the old one, based on an acceptance criterion that depends only on these two states involved in the trial, $n$ and $o$. The Markov chain can be generated by imposing the detailed balance represented by the following equation

$$\mathcal{K}(o \rightarrow n) = \mathcal{K}(n \rightarrow o) \tag{2}$$

where $\mathcal{K}(o \rightarrow n)$ is the probability of being in the state $o$ and changing to state $n$ and it should be equal to $\mathcal{K}(n \rightarrow o)$ that is the probability of being in state $n$ and changing to state $o$. These probabilities can be rewritten as

$$\mathcal{K}(o \rightarrow n) = P(o)\alpha(o \rightarrow n)\mathrm{acc}(o \rightarrow n) \tag{3}$$

and



https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

$$\mathcal{K}(n \to o) = P(n)\alpha(n \to o)\mathrm{acc}(n \to o) \quad (4)$$

where $P(n)$ is the probability of being in the new state (as showed in eq 1), $\alpha(n \to o)$ is the probability of generating the new state starting from the old state, and $\mathrm{acc}(o \to n)$ is the probability of accepting this trial move.

The Metropolis algorithm[50] was proposed considering a simple case where the procedure of generating any new state is random and generates a uniform probability distribution. Then, the forward and backward trial probability is equal to a constant, i.e., $\alpha(o \to n) = \alpha(n \to o) = \mathrm{const}$. Therefore, the detailed balance (eq 2) can be rewritten as

$$\frac{\mathrm{acc}(o \to n)}{\mathrm{acc}(n \to o)} = \frac{P(n)}{P(o)} \quad (5)$$

Using this condition and the Boltzmann distribution in the *NVT* ensemble (eq 1), the Metropolis acceptance criterion is written as

$$\begin{aligned}\mathrm{acc}(o \to n) &= \min\{1, P(n)/P(o)\} \\ &= \min\{1, \exp[-\beta\Delta\mathcal{U}]\}\end{aligned} \quad (6)$$

where $\Delta\mathcal{U} = \mathcal{U}(n) - \mathcal{U}(o)$ is the energy difference between the new and old states. Since the acceptance criterion satisfies the detailed balance, the generated states lies in the Boltzmann distribution representing the thermodynamic equilibrium.

In the *NPT* ensemble, the Metropolis acceptance criterion is rewritten to take into account the volume trial moves, which an attempt to change the volume is done by changing the box size randomly and rescaling the center of masses of the molecules accordingly. This criterion is given by

$$\begin{aligned}&\mathrm{acc}(o \to n) \\ &\quad = \min\{1, \exp[-\beta(\Delta\mathcal{U} + P\Delta V - N\beta^{-1}\ln(V_n/V_o))]\}\end{aligned} \quad (7)$$

where $P$ is the pressure and $\Delta V = (V_n - V_o)$ is the volume difference between the new and old states.

An important point that should be highlighted here is the simplification that can be applied to the calculation of the total potential energy variation $\Delta\mathcal{U}$ in a new trial configuration when only a single molecule $A$ is selected

$$\Delta\mathcal{U} = \Delta\mathcal{U}^A = \mathcal{U}^A(n) - \mathcal{U}^A(o) \quad (8)$$

where $\mathcal{U}^A(n)$ is the potential energy of the molecule $A$ in a new configuration, considering the intra- and intermolecular interaction with all other molecules. The other interacting terms that not involve the molecule $A$ are canceled in the new and old configurations. Therefore, for the following discussions, we are going to simplify the notation and use the $\mathcal{U}(n)$ and $\mathcal{U}(o)$ to describe the potential energy of the selected molecule in one MC movement trial.

**Configurational Bias Monte Carlo Sampling of Flexible Molecules.** Unlike the Metropolis MC method, the bias MC method uses different probabilities to generate the forward and backward trial configurations, $\alpha(o \to n)$ and $\alpha(n \to o)$, respectively.[20] By proposing moves with probabilities depending on the energy of the new configuration, it is possible to increase the acceptance rates. This bias toward favorable configurations is compensated in the acceptance criterion so the detailed balance is still satisfied and the Boltzmann distribution of states is sampled.

To sample the molecular internal degrees of freedom, we use a CBMC algorithm based in a work of Shah and Maginn.[13,33] In this algorithm, the molecules are divided into fragments. These fragments contain the hard degrees of freedom, i.e., the degrees of freedom that are not very flexible such as bond distances, bond angles and improper dihedral angles. The isolated fragment geometries are sampled based on the Boltzmann distribution in an initial stage of the simulation, where the fragment libraries are created to be used later in the CBMC trial moves. The soft degrees of freedom, on the other hand, are sampled with a CBMC approach, by partially reconstructing the molecules, inserting fragments with a biased movement that takes into account interactions within the molecule and with the solvent.

In our implementation, we propose fragmentation schemes beyond those proposed in the original algorithm, not necessarily looking for the set of smallest fragments. This approach allows fragments that are not branch points or simple rings, so the user can focus in specific sets of degrees of freedom, e.g., in a specific dihedral angle within a molecule. Following our fragmentation scheme, the only class of molecules which, in principle, cannot be studied with the current implementation are the ones where the fragment connections form a cycle, e.g., fragment 1 is connected to fragment 2, which is connected to fragment 3, and so on until the last fragment connects to fragment 1 and closes the cycle.

*Hard and Soft Degrees of Freedom.* The central concept of the algorithm is the separation of hard and soft degrees of freedom. The hard degrees of freedom consist of the bond distances, angles, and improper angles that cause large energy variation due to small changes in the equilibrium positions. However, sometimes some dihedral angles have large rotational barriers, and they may, or may not, be considered also as hard degrees of freedom depending on the user criterion. On the other hand, the soft degrees of freedom consist of dihedral angles that are responsible for large conformational changes such as *anti/syn/gauche* conformations, axial/equatorial conformations, etc.

The total potential energy of a selected molecule $A$ in the state $n$ is written as

$$\mathcal{U}(n) = \mathcal{U}^{\mathrm{hard}}(n) + \mathcal{U}^{\mathrm{soft}}(n) \quad (9)$$

where $\mathcal{U}^{\mathrm{hard}}(n)$ indicates the energy of the hard degrees of freedom, with contributions to the energy given by the bonded and nonbonded interactions within each fragment $i$, $u_i(n)$ over the total number of fragments, $N_{\mathrm{frag}}$

$$\mathcal{U}^{\mathrm{hard}}(n) = \sum_{i=1}^{N_{\mathrm{frag}}} u_i(n) \quad (10)$$

and $\mathcal{U}^{\mathrm{soft}}(n)$ is the energy of the soft degrees of freedom, including the contributions of the dihedral angles, the intramolecular nonbonded energy between fragments $i$ and $j$, $u_{ij}(n)$, and also the intermolecular nonbonded energy between the fragments and other molecules $X$, $u_{i-X}(n)$

$$\mathcal{U}^{\mathrm{soft}}(n) = \sum_{i=2}^{N_{\mathrm{frag}}}\sum_{j=1}^{i-1} u_{ij}(n) + \sum_{i=1}^{N_{\mathrm{frag}}}\sum_{\substack{X=1 \\ X\neq A}}^{N} u_{i-X}(n) = \sum_{i=1}^{N_{\mathrm{frag}}} u_{i\alpha}(n) \quad (11)$$

The first term is the interaction of all $i$th fragment with all the already inserted fragments, $j$, and the second is the interaction of



https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

all *i*th fragments with all the other molecules, *X*. Then, we simplified this notation with the two sum terms by a single sum over $u_{i\alpha}(n)$ where $\alpha$ represents the already inserted fragments and the other molecules. As an example, in Figure 1 we have a

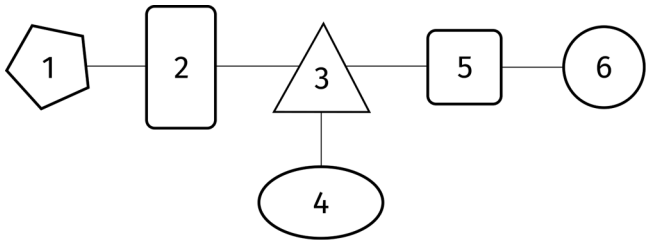



**Figure 1.** Illustration of a pictorial molecule decomposed in six different fragments. These fragments are connected by dihedral angles that are sampled with the CBMC method.

pictorial molecule built of six different fragments. For this molecule

$$\mathcal{U}^{\text{hard}}(n) = u_1(n) + u_2(n) + u_3(n) + u_4(n) + u_5(n) + u_6(n)$$

is the internal energy of each fragment and

$$\mathcal{U}^{\text{soft}}(n) = \{u_{21}(n) + u_{32}(n) + u_{31}(n) + u_{43}(n) + u_{42}(n) + u_{41}(n) + \cdots u_{61}(n)\} + \{u_{1-1}(n) + u_{1-2}(n) + \cdots u_{1-N}(n) + u_{2-1}(n) + u_{2-2}(n) + \cdots u_{2-N}(n) + \cdots u_{6-1}(n) + u_{6-2}(n) + \cdots u_{6-N}(n)\}$$

is the energy interaction between the fragments, $u_{ij}(n)$, and the energy interaction between each fragment and all other $N - 1$ molecules in the simulated system, $u_{i-X}(n)$. Therefore, the sum of the $\mathcal{U}^{\text{hard}}(n)$ and the first part of the $\mathcal{U}^{\text{soft}}(n)$ is the intramolecular potential energy of the selected molecule and the second part of the $\mathcal{U}^{\text{soft}}(n)$ is the intermolecular potential energy of the selected molecule with all others. Details of the possible fragmentation schemes within our algorithm are given in the next section.

*Fragment Definition.* In our implementation, no restrictions are imposed to the fragment definition. Fragments are defined as a set of atoms connected by covalent bonds, and each fragment has two atoms intersected with the neighbor fragment. These two atoms define a rotatable bond between the neighbor fragments and are the central atoms of the dihedral angles between them. Following our scheme, it is possible to fragment a molecule in different ways that should be decided by the user.

Three different types of fragments are supported by our implementation, as illustrated in Figure 2: (a) fragments with a central atom, (b) simple rings, and (c) generic fragments. Each one of the three fragment types may be kept rigid during the simulation, staying in the initial geometry provided by the input geometry, or may be sampled prior to the simulation to generate the called fragment libraries. When the fragments are not rigid, they can be sampled using a semiflexible procedure considering a library of geometries in which only the angles were sampled with rigid bonds or a flexible procedure considering a library of geometries in which all the internal degrees of freedom were sampled, both with the Metropolis sampling technique. A rigid

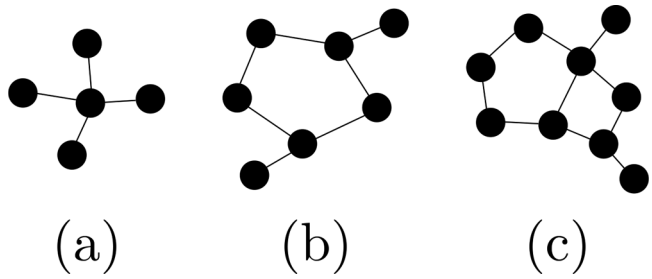



**Figure 2.** Different types of pictorial fragments: (a) fragments with a central atom, (b) simple rings, and (c) generic fragments.

fragment *i* has the advantage of eliminating the need of the force field topology and parameters within the fragment, because the internal energy of the rigid fragments does not change during the entire simulations, i.e., $u_i(n) = u_i(o)$. Then, only three nonbonded parameters are needed, $\{q_i, \epsilon_i, \sigma_i\}$, for each atom within the rigid fragments to calculate its interaction energy with the other molecules, $u_{i-X}(n)$, and additionally the dihedral angle parameters with all the connected fragments to calculate its interacting energy with the previously inserted fragments, $u_{ij}(n)$. The use of rigid fragments will be a good approximation when the small variations of the bond lengths and angles within a fragment are not relevant for the conformational sampling of the entire molecule.

*Metropolis Sampling of the Hard Degrees of Freedom.* In the input each fragment should be identified by its set of atoms and a label that defines its status: *R* for rigid, *F* for semiflexible, and *M* for flexible. If a fragment is rigid, a library is not generated for this fragment. If a fragment is semiflexible, a sampling to create a library for this fragment is performed, using the algorithms for each fragment type, as described below:

(a) Fragments with a central atom: This type of fragment has a collection of atoms that are connected to an atom in the center of the fragment. The sampling of these fragments is done considering fixed bond lengths, with the central atom at the origin and the other atoms having their position described in spherical coordinates. One atom is randomly selected and has its polar and azimuthal angles displaced, with the displacement of the azimuth being done with the cosine of the angle to avoid dealing with Jacobian determinants. This kind of sampling was first introduced in branch point sampling and later generalized.[33,39] The Metropolis MC acceptance criterion is used with this movement to perform the sampling.

(b) Simple rings: This type of fragment has a collection of atoms forming a single ring, which can have additional atoms attached to this ring. The sampling is done combining the sampling of the fragment with the central atom and the crank-shaft move, using the Metropolis MC acceptance criterion.[33] One atom of the ring is randomly selected, and if there are atoms that are not in the ring attached to it, the move on spherical coordinates of the fragments with the central atom may be performed with a 50% chance. If the move in the spherical coordinates is not performed, or if the atom does not have atoms attached to it, the crank-shaft move is done. The crank-shaft move rotates the atom around the axis defined by its neighbor atoms. Atoms attached to the moved atom are also moved as a rigid body. This move also keeps the bond lengths fixed.


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

(c) Generic fragments: This type of fragment has no geometrical restriction. Then, any fragments that are not classified in the two previous types can be considered as generic fragments. However, this type of fragment can only be sampled as a rigid or flexible fragment. It cannot be sampled as a semiflexible fragment.

If a fragment is flexible, a sampling to create a library for this fragment is performed using Cartesian displacements of the atoms and the Metropolis acceptance criterion. One random atom is selected at a time and a random displacement in its coordinates, $\delta x$, $\delta y$, and $\delta z$, is performed. Therefore, the bond lengths, bond angles, and torsional and improper angles can change during the simulation and be sampled, satisfying the Boltzmann distribution.

For all the cases the maximum displacement of the movements is adjusted for each fragment to obtain an acceptance rate of 50%, during the thermalization period before the sampling used to generate the fragment libraries. As all these trial moves have no Jacobian determinant associated with the coordinate change, the Metropolis MC acceptance criterion as presented in eq 6 is used for this sampling, generating fragment libraries that have configurations that satisfy the Boltzmann distribution.

For nonrigid fragments, their libraries are generated with several geometries saved periodically during the initial stage of isolated fragment sampling by the Metropolis technique of the hard degrees of freedom. These libraries are used during the CBMC trial moves as the source of the fragment geometry to be reinserted. Therefore, a fragment selected from its library represents sampling the hard degrees of freedom.

For a fragment $i$, we select a geometry from the library with a probability $1/\kappa_i$, where $\kappa_i$ is the number of fragment geometries in its library. The probability of each geometry to be in the library of fragment $i$ is given by the Boltzmann distribution

$$P_i^{\text{hard}}(n) = \frac{\exp[-\beta u_i(n)]}{\sum_{k=1}^{\kappa_i} \exp[-\beta u_i(k)]} \tag{12}$$

This probability is obtained by sampling the fragment geometries using the Metropolis acceptance criterion as described above and is convenient to simplify the CBMC acceptance criterion.

*CBMC Sampling of the Soft Degrees of Freedom.* Here the molecule can be seen as a chain of fragments (see Figure 1), which are connected to each other by dihedral angles. The CBMC trial move is started by choosing a random flexible molecule from the simulation box with a probability $1/N_{\text{flex}}$, where $N_{\text{flex}}$ is the amount of flexible molecules in the simulated system. It can be the total number of molecules in the system if all of them are flexible or a fraction if some of them are flexible and others are rigid. In this molecule a connection between two fragments is randomly selected with a probability $1/(N_{\text{frag}} - 1)$. From this connection one side is chosen randomly. The user can select between two possibilities for this choice: (i) with 50% of chance for each side or (ii) with a weight that favors smaller parts to be sampled more often, increasing the acceptance rates of the CBMC move. The chosen side of the molecule is deleted and rebuilt by reinserting the $N_{\text{del}}$ deleted fragments with new dihedral angles $\phi_n$ using the CBMC scheme. Nothing is changed on the other side of the molecule. Therefore, with this procedure of changing only the $N_{\text{del}}$ fragments of the molecules, the energy difference between the old and new conformations can be simplified to summing only the energy terms over the deleted fragments due to the cancellation terms of the unchanged fragments. Then, the ratio of the Boltzmann probability between the two conformations should be written as

$$\frac{P(n)}{P(o)} = \prod_{i=1}^{N_{\text{del}}} \frac{\exp[-\beta(u_i(n) + u_{i\alpha}(n))]}{\exp[-\beta(u_i(o) + u_{i\alpha}(o))]} \tag{13}$$

The CBMC trial move scheme is divided into two steps: the forward and reverse moves. These two moves are necessary because, different than the case of Metropolis sampling, the probability of generating a move that goes from old to new state is different from new to old state, i.e, $\alpha(o \to n) \neq \alpha(n \to o)$. To calculate these probabilities in the CBMC trial move, first a new conformation is built by inserting fragment geometries randomly selected from the fragment libraries with new orientations when compared to the conformation of the old state $o$. In this step, the new orientations are selected in a biased scheme selecting one trial dihedral angle from a set $\{\phi_i\}$ with $\kappa_{\phi_i}$ possible angles, each weighted by the Boltzmann distribution. Later, the old conformation is generated by selecting the fragment geometries corresponding to the old conformation from the fragment libraries and by selecting from the new set of dihedral angles the one corresponding to the conformation of the old state. With this approach, both $\alpha(o \to n)$ and $\alpha(n \to o)$ are calculated as a product of probabilities of each independent fragment selection and reinsertion.

The fragments are inserted one at a time, starting with the one attached to the side of molecule that was not deleted in this CBMC trial move. After randomly selecting one geometry of the fragment from the fragment library, a set of $\kappa_{\phi_i}$ random trial orientations for the insertion the fragment are drawn up. Then, the orientations of this $i$th fragment are selected with the biased probability

$$P_i^{\text{soft}}(n) = \frac{\exp[-\beta u_{i\alpha}(n)]}{\sum_{k=1}^{\kappa_{\phi_i}^{(n)}} \exp[-\beta u_{i\alpha}(k)]} \tag{14}$$

This process is repeated until all the deleted fragments are reinserted and the full molecule is rebuilt. For the reverse move, the selected geometries from the fragment libraries are exactly the ones from the old conformation and for the orientations another set of $\kappa_{\phi_i}^{(o)}$ is generated necessarily including the one from the old conformation. The trial orientations in the reverse move are only used to compute the probability, as the chosen orientation is always the one from the old state.

To illustrate the forward move in the pictorial molecule of Figure 1, suppose the connection between fragments 2 and 3 was randomly selected and the right side was chosen also randomly to be rebuilt. Then, the fragments 3, 4, 5, and 6 are deleted and the rebuilding starts with fragment 3. A random geometry for fragment 3 is chosen from the library with a probability $1/\kappa_3$, and $\kappa_\phi$ trial orientations of insertion for this fragment are randomly generated. For each trial orientation, the interaction energy of fragment 3 with the fragments 1 and 2 and with all other molecules is calculated. Using this energy the Boltzmann factors are calculated and one of the trial orientations is chosen with the probability given by eq 14 completing the insertion of fragment 3. Only fragments 4 and 5 are attached to fragment 3, so one of them is randomly selected, e.g., fragment 5. Again, a geometry for fragment 5 is chosen from the fragment library, and the trial orientations are generated. The interaction energy of fragment 5 with fragments 1, 2, and 3 and with all other molecules is


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

calculated to obtain the Boltzmann factors, and one of the trial orientations is chosen. The next fragment to be inserted is fragment 4, because we follow the well-known graph search algorithm breadth first search (BFS). In this scheme, all fragments connected with fragment 3 (first neighbor) will be inserted before the second neighbors fragments and so on. Notice that this way of selecting fragments to be inserted is simply a convention, and another graph search algorithm, depth first search (DFS), could be adopted to insert fragment 6 before fragment 4. The same procedure used to insert fragment 5 is repeated to insert fragment 4, but in this case, the energy also includes the interaction with fragment 5. Finally, fragment 6 is inserted, taking into consideration the interaction with all the fragments from 1 to 5 and the other molecules.

The acceptance criterion of the whole trial move can be found using the super detailed balance[20] that is a restriction of the usual detailed balance condition of eq 2, where the detailed balance condition is satisfied independent of the random sets of insertion trial orientations for the forward $\{\kappa_\phi^{(n)}\}$ and reverse $\{\kappa_\phi^{(o)}\}$ moves.

Now the probability $\alpha(o \to n)$ of generating the new state $n$ from the old state $o$ is written as

$$\alpha(o \to n) = C \prod_{i=1}^{N_{del}} P_i^{hard}(n) P_i^{soft}(n) \tag{15}$$

where $C$ describe the combined probabilities of doing all the random selections in the algorithm, such as selecting one flexible molecule, one fragment connection, one side of molecule to delete the fragments, and one fragment geometry from the library. The probability of generating the old state starting from the new one is found in a similar way, considering that the probabilities in the way back selections are those ones to recover the conformation of the old state. The probabilities in both directions, $\alpha(o \to n)$ and $\alpha(n \to o)$, have the same constant $C$, i.e., the same random probabilities for selections. Therefore, using this two probabilities, the detailed balance (eqs 2, 3, and 4) may be rewritten as

$$\frac{acc(o \to n)}{acc(n \to o)} = \frac{P(n)}{P(o)} \prod_{i=1}^{N_{del}} \frac{P_i^{hard}(o) P_i^{soft}(o)}{P_i^{hard}(n) P_i^{soft}(n)} \tag{16}$$

Thus, substituting eqs 12, 13, and 14 into eq 16, it can be simplified to

$$\frac{acc(o \to n)}{acc(n \to o)} = \frac{W(n)}{W(o)} \tag{17}$$

where $W(n)$ is the Rosenbluth factor of the new conformation defined as

$$W(n) = \prod_{i=1}^{N_{del}} \sum_{k=1}^{\kappa_{\phi_i}^{(n)}} \exp[-\beta u_{i\alpha}(k)] \tag{18}$$

Therefore, the CBMC acceptance criterion is written in a simplified expression as

$$acc(o \to n) = \min\left(1, \frac{W(n)}{W(o)}\right) \tag{19}$$

This simplified expression is well-known in other CBMC schemes,[20] such as the original CBMC algorithm for generating polymer conformations. Now, we have shown that it should be used also in the CBMC proposed by Shah and Maginn[33] and implemented in DICE with new feature such as the generalized fragment type and entirely flexible fragments including the bond length.

*CBMC Acceptance Criterion.* The CBMC acceptance criterion, shown in eq 19 is prone to numerical errors given the characteristic of the Rosenbluth factor (see eq 18). Therefore, to minimize problems with numerical errors, we take the logarithm of the criterion and use the *Log-sum-exp*. We start rewriting the ratio of Rosenbluth factors as

$$\frac{W(n)}{W(o)} = \exp[\ln W(n) - \ln W(o)] \tag{20}$$

which is already an improvement to the numerical stability, since

$$\ln W(n) = \ln\left[\prod_{i=1}^{N_{del}} \sum_{k}^{\kappa_{\phi i}^{n}} \exp(-\beta u_{i\alpha}(k))\right] = \sum_{i=1}^{N_{del}} \ln\left[\sum_{k}^{\kappa_{\phi i}^{n}} \exp(-\beta u_{i\alpha}(k))\right] \tag{21}$$

removes the productory, turning it into a summation.

The *Log-sum-exp* is used in eq 21 to reduce the chance of floating point problems even further. The largest term of the summation inside the logarithm is factored out, making the summation and exponentials numerically safe. Assuming $x_1$ is the largest exponential argument of the summation inside the logarithm, we can do

$$\ln \sum_{i=1}^{N} e^{x_n} = \ln\left(\sum_{i=1}^{N} e^{x_n} \frac{e^{x_1}}{e^{x_1}}\right) = x_1 + \ln\left(1 + \sum_{i=2}^{N} e^{x_n - x_1}\right) \tag{22}$$

where the arguments $x_n - x_1$ are small. We can also use the `log1p(x)` function, present in most computer languages, to calculate $\ln(1 + x)$ with good numerical precision.

Using the *Log-sum-exp*, we end with a numerically more stable version of eq 21

$$\ln W(n) = \sum_{i=1}^{N_{del}} -\beta u_{i\alpha}(\phi_{min}) + \ln\left\{1 + \sum_{\substack{k=1 \\ k \neq \phi_{min}}}^{\kappa_{\phi i}^{n}} \exp[-\beta(u_{i\alpha}(k) - u_{i\alpha}(\phi_{min}))]\right\} \tag{23}$$

with $u_{i\alpha}(\phi_{min})$ being the lowest energy given by $\phi_{min}$, one angle among the $\kappa_\phi$ insertion dihedral angles. Therefore, eq 23 is used in the modified CBMC acceptance criterion implemented in DICE:

$$acc(o \to n) = \exp[\min(0, \ln W(n) - \ln W(o))] \tag{24}$$

Details on the implementation of these criteria in DICE and the expressions for the energy terms in each of the supported force fields are given in the SI.

## RESULTS

We applied the CBMC method for five systems to show the capabilities and correctness of our implementation in the DICE code. Two of them are well-known systems where our results were compared with MD simulations, literature results, and experimental data when available. The other two were used to



https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

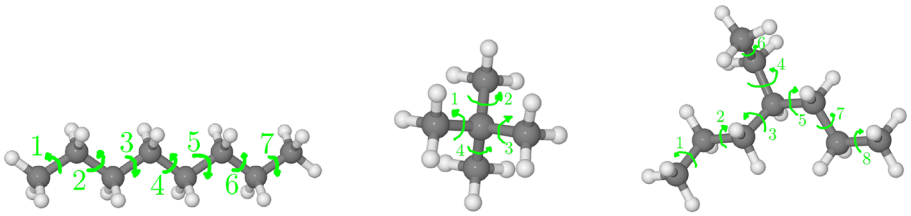



**Figure 3.** Illustration of the rotatable bond of the alkanes simulated with CBMC and its labels.

compare the CBMC sampling efficiency for linear and branched molecules. The last system is a very complex molecule without previous MM simulations, where we compare the distributions from the gas phase simulations with the rotational energy profiles and the sampling in solution with the solute−solvent interactions and dipole interaction stabilization.

The systems studied in this work were: (i) the isolated *n*-octane and in chloroform solution at room conditions, providing comparison with MD results using the same potentials; (ii) the isolated neopentane and 4-ethylheptane at room conditions, comparing the acceptance rates of different rotatable bonds considering the influence of the quantity of trial insertion angles $\kappa_\phi$ in comparison with the linear alkane, *n*-octane; (iii) 1,2-dichloroethane (DCE) in gas and liquid phases and in acetonitrile solution, comparing with experimental data, MD results and other MC simulations found in the literature; (iv) a donor−acceptor functionalized boron subphtalocyanine (SubPC) in the gas phase and water solution.

The software used for the MC simulations was DICE 3.0,[19] and that used for the MD simulations was GROMACS 4.6.7.[8] The input files for the MC and MD simulations are presented in the SI.

**Alkanes.** We have simulated three different alkanes: *n*-octane, neopentane, and 4-ethylheptane (see Figure 3). For *n*-octane in the gas phase and in chloroform solution, we have investigated the consistency of the sampling in comparison with MD. For neopentane, 4-ethylheptane and *n*-octane in gas phase, we analyzed the CBMC sampling efficiency with different quantity of trial insertion angles $\kappa_\phi$ through the acceptance rates of different rotatable bonds, for which labels are shown in Figure 3.

The simulations of the isolated *n*-octane using CBMC was performed in the *NVT* ensemble at 298.15 K in a cubic box of length 57 Å. For the simulation in chloroform solution, the *NPT* ensemble was used, with $N$ = 1001 being one molecule of *n*-octane and 1000 chloroform molecules, $P$ = 1 atm, and the same temperature. During the simulation the average of the cubic box was of 51.7 Å, giving an average density of 1.44 g/cm$^3$ in good agreement with the experimental value of 1.48 g/cm$^3$.[51] We adopt a cutoff radius of approximately 21 Å, and the long-range correction calculated with the continuum model. The *n*-octane molecule was divided with $N_{frag}$ = 8: two fragments relative to its extremities with $CH_3$−C groups and six fragments composing its inner chain with C−$CH_2$−C groups, counting 7 rotatable bonds between the fragments (see Figure 3 and the SI). Bond lengths were kept rigid, and sampling of the hard degrees of freedom was done with the central atom fragment algorithm. We used the OPLS-AA parametrization for the *n*-octane[52] and chloroform[53] molecules. The simulations run for $1 \times 10^7$ MC steps for the isolated *n*-octane and $5 \times 10^8$ MC steps for the *n*-octane in chloroform during the equilibrium stage. Both simulations use $\kappa_i$ = 10 000 (number of configurations in each fragment library) and $\kappa_\phi$ = 18 (number of possible angles for each fragment reinsertion), where the first $\phi$ trial angle was randomly selected in the interval [0, 360°) and the other 17 $\phi_i$ were uniformly distributed with an interval of $\Delta\phi$ = 20°. We called this algorithms to select the trial insertion angles as equidistant $\phi$ sampling (see details in the SI). Additionally, we used a weight option where, after selecting a rotatable bond with a uniform random distribution, the selection of the molecular side to be deleted and regrow favors that one with the smaller amount of atoms to be sampled more often (see details in the SI). In the solution, the molecules are selected with preferential sampling with the one-over-$r^n$ algorithm,[46,47] with $n$ = 2. Here, 30% of the trial moves performed a rigid trial (translation and rotation at the center of mass) and 70% performed a flexible trial (an internal deformation with the CBMC method). The overall CBMC acceptance rates were in average 45% for the isolated *n*-octane and 31% for the *n*-octane in chloroform solution, with $\kappa_\phi$ = 18.

MD simulations were performed for comparison. We used the leapfrog integrator[54] with a time step of 1 fs and the LINCS algorithm[55] to keep the bond lengths fixed. Both the Nosé−Hoover[56,57] and velocity-rescale[58] thermostats, with a coupling constant of $\tau_T$ = 0.1 ps, were tested, but no significant difference was observed between them. Then, the results shown are for the Nosé−Hoover thermostat. In the case of the simulation of *n*-octane in solution, the Berendsen barostat[59] was used with coupling constant of $\tau_T$ = 0.5 ps. We adopt a cutoff radius of 15 Å, and a long-range correction for the energy calculated with reaction-field using the dielectric constant of the chloroform (4.81). The thermodynamic conditions and ensembles are the same used for the CBMC simulations. The simulations run for at least 20 ns during the equilibrium stage.

Knowing that successive configurations obtained from simulations are highly statistically correlated, we used the analysis of the statistical inefficiency of the CBMC and MD simulations to know how many simulation steps are necessary to obtain new statistical information. The statistical inefficiency interval $s$ and the correlation interval (or time) $\tau$ are related by $s \approx 2\tau$.[60] This is relevant in the S-QM/MM context because after the MM simulation it is possible to reduce drastically the amount of configurations that will be used in the QM calculations with the same statistical information. Previously,[32] we have shown that performing QM calculation in 20 000 successive configurations gives the same average and standard deviation values of electron excitations as using only 25 statistically uncorrelated configurations separated by the statistical inefficiency interval. Therefore, here to analyze the statistical inefficiency of the CBMC and MD simulations, we used the energy values of the last $1 \times 10^6$ configurations of both simulations to calculate the autocorrelation function for the energy, $C(t)$, and obtain the statistical inefficiency interval $s$, as performed before.[3,22−31] The calculated values of the $C(t)$ were fitted by a two-exponential decay function, $C(t) = w_1 e^{t/\tau_1} +$


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

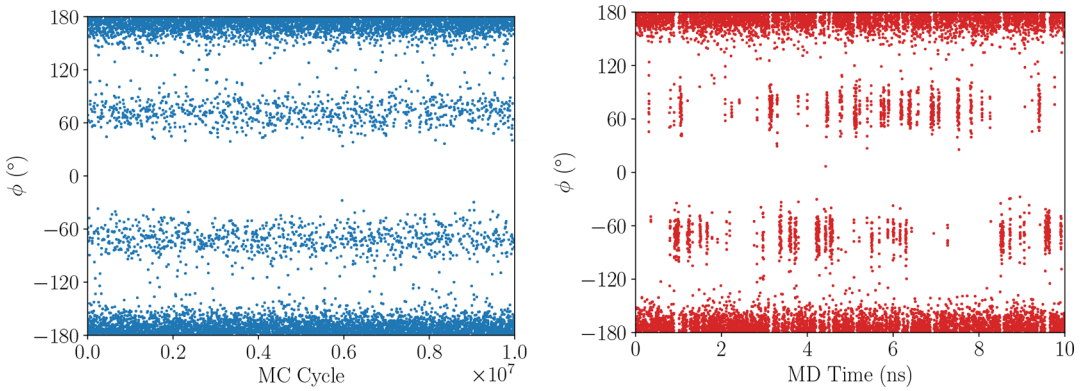



**Figure 4.** Evolution of the central C−C−C−C dihedral angle (C6−C9−C12−C15) of isolated *n*-octane with simulation steps. The same behaviors are observed for *n*-octane in chloroform solution (see the SI).

$w_2 e^{t/\tau_2}$, and the fitted parameters were used to obtain the statistical correlation interval as $s = 2\tau$, where $\tau = w_1\tau_1 + w_2\tau_2$ is the integration of the fitted $C(t)$. Typically, configurations separated by an interval of $s$ give new information with a statistical correlation smaller than 15%. We obtained for the CBMC simulation $s_{CBMC} = 68$ and for the MD simulation $s_{MD} = 3225$ (see details in the SI). Therefore, the CBMC method generates statistically uncorrelated configurations faster than the MD method, in this case around 50 times. For this reason, configurations were saved for further analysis each 100 steps for the CBMC simulation and each 5000 steps, i.e. 5 ps, for the MD simulations ensuring statistically uncorrelated configurations.

We used the saved configurations to analyze the bond and dihedral angles, the population ratio of *trans:gauche* conformations defined by the C−C−C−C dihedral angles and end-to-end C−C length, *R*. The bond angles obtained with CBMC and MD simulations for the isolated *n*-octane and in solution are in very good agreement. Their differences obtained with both simulation methods are smaller than 0.4° (see details in the SI).

The evolution of the central dihedral angle (C6−C9−C12−C15, see the SI for atomic labels) during the simulations is shown in Figure 4, where three population can be easily identified: around −60°, +60°, and ±180° that characterizes the *gauche*(−), *gauche*(+), and *trans* forms, respectively. Additionally in the CBMC simulation, the distribution of this dihedral angle is more uniform, showing a larger frequency of barrier crossing than that obtained in the MD simulation. The vertical strips in the MD distribution describes the time interval where the conformation is trapped in the energy minimum.

Then, the populations were obtained counting the dihedral angle in the interval of [±120°, 0°] for the *gauche*(±) forms and of [±120°, ±180°] for *trans* form. In Table 1, a comparison of the *trans:gauche* population ratio among the results obtained with CBMC and MD simulations are shown for the isolated *n*-octane and in solution. The amount of *gauche*(+) and *gauche*(−) forms for each case are also shown. We estimate a statistical uncertainty of 1% in these populations analyzing independent sets of data. We obtained that (i) the three different C−C−C−C torsional angles of the *n*-octane molecule present a similar *trans:gauche* population ratio around 82:18, indicating a small solvent effect and also small effect of the molecular extremity motion; (ii) comparing the CBMC and MD results, a very good agreement was found with differences smaller than 1%; (iii) comparing the *gauche*(+) and *gauche*(−) populations, the differences obtained with CBMC (up to 1%) are smaller than those obtained with MD (up to 4.5%). This confirm a larger

**Table 1. Comparison of the *trans:gauche* Population of the Dihedral Angles of *n*-Octane for the Isolated Molecule and in Chloroform Solution Obtained with CBMC and MD Simulations[a]**

| | % *trans:gauche* | | % *gauche*(+) | | % *gauche*(−) | |
|---|---|---|---|---|---|---|
| dihedral | CBMC | MD | CBMC | MD | CBMC | MD |
| only torsional interaction | (70:30) | | | | | |
| C1−C2−C6−C9 | 71:29 | 70:30 | 14.3 | 14.3 | 14.3 | 15.9 |
| C2−C6−C9−C12 | 72:28 | 68:32 | 13.8 | 16.8 | 14.0 | 14.9 |
| C6−C9−C12−C15 | 73:27 | 67:33 | 13.6 | 17.4 | 13.7 | 16.0 |
| isolated | (88:12) | | | | | |
| C1−C2−C6−C9 | 80:20 | 75:25 | 9.7 | 11.6 | 10.6 | 13.0 |
| C2−C6−C9−C12 | 83:17 | 82:18 | 9.3 | 9.2 | 8.2 | 8.3 |
| C6−C9−C12−C15 | 82:18 | 82:18 | 8.7 | 8.3 | 8.9 | 9.2 |
| in chloroform | | | | | | |
| C1−C2−C6−C9 | 80:20 | 82:18 | 9.4 | 6.6 | 10.2 | 11.0 |
| C2−C6−C9−C12 | 85:15 | 81:19 | 6.8 | 8.9 | 7.9 | 9.9 |
| C6−C9−C12−C15 | 84:16 | 86:14 | 7.5 | 4.5 | 8.8 | 9.0 |

[a]In parentheses are the theoretical ideal values calculated from the potential energy differences between the *trans* and *gauche*(±) forms (see the SI for details).

statistical inefficiency of the MD method and the necessity of larger amount of configurations for statistical convergence. Furthermore, using only the torsional interaction, it is possible to estimate the theoretical ideal value of 70:30 for the *trans:gauche* population of the C9−C12 rotation based on the equilibrium constant obtained from the energy difference between the *trans* and *gauche* minima of the torsional potential, $\Delta E^{OPLS}_{torsion}(trans \rightarrow gauche) = 0.9$ kcal/mol, having no dependence with the other dihedral angles of the chain (see the details in the SI). For comparison, we performed simulations for this ideal system and obtained a better agreement for the *trans:gauche* population of the CBMC method 73:27 than the MD method, 67:33. Using the total potential energy in the rotation of the C9−C12 bond, it is possible also to estimate the theoretical ideal value of 88:12 for the *trans:gauche* population based on $\Delta E^{OPLS}_{total}(trans \rightarrow gauche) = 1.6$ kcal/mol, obtained with rigid *n*-octane rotation not including the bond angles relaxation that were considered in our simulations. Therefore, comparing this


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

Table 2. Acceptance Rates (%) and Walltime (s) for Simulations of *n*-Octane, Neopentane, and 4-Ethylheptane with Different Quantity of Trial Insertion Angles $\kappa_\phi$[a]

| | *n*-octane | | | neopentane | | | 4-ethylheptane | | |
|---|---|---|---|---|---|---|---|---|---|
| $\kappa_\phi$ = | 12 | 32 | 64 | 12 | 32 | 64 | 12 | 32 | 64 |
| bond 1 | 68.7 | 75.4 | 77.6 | 61.0 | 66.7 | 69.8 | 68.6 | 72.9 | 74.0 |
| bond 2 | 38.1 | 50.8 | 56.2 | 60.3 | 66.8 | 70.0 | 32.8 | 42.0 | 45.4 |
| bond 3 | 22.9 | 37.3 | 44.2 | 60.1 | 66.9 | 69.8 | 8.2 | 13.3 | 15.8 |
| bond 4 | 18.6 | 33.8 | 41.1 | 60.2 | 66.7 | 69.9 | 17.0 | 21.6 | 23.5 |
| bond 5 | 23.3 | 38.0 | 44.1 | | | | 8.3 | 13.4 | 15.9 |
| bond 6 | 38.2 | 50.0 | 55.2 | | | | 64.3 | 68.3 | 69.8 |
| bond 7 | 69.5 | 75.4 | 77.4 | | | | 32.8 | 42.2 | 45.6 |
| bond 8 | | | | | | | 68.5 | 72.7 | 73.9 |
| overall | 39.9 | 51.6 | 56.6 | 60.4 | 66.7 | 69.9 | 37.6 | 43.3 | 45.5 |
| time | 89 | 148 | 222 | 114 | 138 | 174 | 157 | 207 | 284 |

[a]The bonds are labeled according to Figure 3.

estimative with the values obtained in the simulations of isolated *n*-octane (see Table 1), around 82:18, we notice a reduction of 6% in *trans* form due to the angular relaxation considered in the simulation and not considered in the theoretical ideal case.

Additionally we analyzed the size of the *n*-octane using the distribution of the end-to-end *R* length that reveals the combination of the *trans* and *gauche* states of the five dihedral angles in the molecule. These distributions obtained for CBMC and MD simulations of isolated *n*-octane and in solution are also in very good agreement, presenting values between 6.5 and 9.1 Å, and in qualitative agreement with the distribution found in literature for the liquid *n*-octane using a different force field.[61] The end-to-end *R* distributions obtained for CBMC and MD simulations present a well-defined peak at larger lengths (between 8.7 and 9.1 Å) describing the all-*trans* conformation and the broad and asymmetrical peak at low lengths (between 6.5 and 8.7 Å) describing all other conformations mixing the *trans* and *gauche* dihedral angles (see details in SI).

Analyzing the solvation of the *n*-octane in chloroform, we obtained quite similar radial distribution functions $G(r)$ for CBMC and MD simulations. As an example, the two $G(r)$ between the center of mass of octane and the center of mass of the chloroform molecules are presented in the SI.

Therefore, we conclude that CBMC method is capable to reproduce the conformational sampling in good agreement with MD method for systems that present low rotational barriers, in the case of *n*-octane $\Delta E^*(trans \rightarrow gauche)$ = 3.5 kcal/mol. Additionally, we had shown that the CBMC method generates configurations with a smaller statistical inefficiency than MD, resulting in a better CBMC conformational sampling in simulations with the same amount of steps. For systems with large rotational barrier, >9.0 kcal/mol, we had shown before[3] that MD simulations are not able to perform the conformational sampling at room temperature without any enriched sampling technique, retaining the initial conformation for very long time (around 200 ns) even for small molecules. Even for these cases, the CBMC method showed a good performance being able to cross rotational barrier larger than 40.0 kcal/mol and perform a statistical relevance conformational sample.[3]

Now, we are going to analyze the efficiency of the CBMC sampling, as implemented in DICE, with respect to the quantity of trial insertion angles $\kappa_\phi$ in the three alkane molecules: *n*-octane (linear), neopentane (branched), and 4-ethylheptane (branched). Increasing $\kappa_\phi$ generates a better search of the rotational energy profile that can lead to an insertion angle with higher probability of acceptance. Therefore, causing an increase in the overall CBMC acceptance rate and the efficiency of the method. However, the computational cost also increases with $\kappa_\phi$ due to the calculation of the Rosenbluth factor that considers the intra- and intermolecular energy of all trial insertion angles. Then, a balance between the acceptance rate and CPU time must be found. For the MC simulations of isolated neopentane and 4-ethylheptane, we used the same procedures, thermodynamics conditions, and force field as described for *n*-octane. The neopentane molecule was divided with $N_{frag}$ = 5: one central fragment composed of five carbons in a tetrahedral shape and four fragments relative to its extremities with $CH_3-C$ groups, counting four rotatable bonds between the fragments. The 4-ethylheptane molecule was divided with $N_{frag}$ = 9: one central fragment composed of four carbons and one hydrogen in a tetrahedral shape, five fragments composing its inner chain with the $C-CH_2-C$ group, and three fragments relative to its extremities with $CH_3-C$ groups, counting eight rotatable bonds between the fragments. In Table 2 we present the percentage of the acceptance rate for each selected rotatable bond of the molecules (see bond labels in Figure 3), considering $\kappa_\phi$ = 12, 32, and 64. Each rotatable bond divides the molecule in two sides. During the simulations the rotatable bond are randomly selected with equal probability and one side is selected to be deleted and to regrow with insertion of new dihedral angles. In these simulations, we used a weight option that favors the side with smaller amount of atoms to be sampled more often. As an example, for *n*-octane when one of the outer bonds, 1 or 7, is selected the $CH_3$ group (side with one fragment) is selected more often than the $(CH_2)_6-CH_3$ group (side with seven fragments). Therefore, following this procedure, it is expected that the outer rotatable bonds should present higher acceptance rates as the overall probability of accepting the move is composed by the product of the probabilities of inserting each fragment. On the other hand, inner rotatable bonds, such as bond 4 for *n*-octane, should present lower acceptance rates once both sides of this bond have the $(CH_2)_3-CH_3$ group (with four fragments). This tendency is shown in Table 2. For *n*-octane, ~69% of the insertion angle attempts in bonds 1 and 7 were accepted with $\kappa_\phi$ = 12, while for bond 4 this acceptance rate decreases to ~19%. Thus, the acceptance rates for the rotatable bonds closer to the extremities are higher than those in the middle of the chain. Increasing the number of trial insertion angle $\kappa_\phi$ to 64, these acceptance rates increase to ~77% and ~41%, respectively. Then, the benefit of increasing $\kappa_\phi$ is greater for the inner bonds. However, the walltime of the simulation with larger $\kappa_\phi$ increases almost 2.5 times (from 89 to 222 s for



https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

simulations of $1 \times 10^6$ CBMC steps under the same conditions in a desktop PC), making necessary a compromise between the acceptance rate and computational cost (see details about the computational cost in the SI).

For neopentane, all rotatable bonds are equivalent and connect a fragment from extremity, $CH_3$ group, to a central fragment. Then, all acceptance rates are ~70% with $\kappa_\phi = 64$. Comparing with the extremities acceptance rates for *n*-octane, the neopentane presents approximately 7% less probability. Since the force field parameters are the same for both cases, the lower acceptance for neopentane can only be due to the nonbonded interactions, which makes the insertion of the fragment less favorable due to the interaction with the other $CH_3$ groups that are close. For *n*-octane, the nonbonded interactions are not as strong because the other fragment are generally further away. However, due to the smaller number of fragments, the probability of accepting the move that regrows the whole molecule is greater in the case of neopentane. We observe acceptance rates of 6% and 32% for the whole molecule regrow of *n*-octane and neopentane, respectively, with $\kappa_\phi = 12$ and 25% and 49% with $\kappa_\phi = 64$. For 4-ethylheptane, the acceptance rates are around 70% to 74% for rotating the extreme bonds 1, 6, and 8 with $\kappa_\phi = 64$, while for central bonds 3, 4, and 5 the acceptance rate decreases to ~16% up to 24%. For bonds 1 and 8 located in the ends of the heptane chain, the acceptance rates are similar to the extremities of *n*-octane, due to the similarity of the chain length. Even though bond 6 is also in an extremity and with a $CH_3$ group, its acceptance rate is different as the ethyl group is shorter. The shorter chain has different environment due to the nonbonded interactions, changing the acceptance rate even with the same parameters. All other symmetrical bonds have similar acceptance rates: bonds 2 and 7 around 45% and bonds 3 and 5 around 16% with $\kappa_\phi = 64$. Bonds 3 and 5 present the most difficult insertion. Comparing these bonds with *n*-octane, we observe a much higher acceptance in the linear alkane, due to the complexity of inserting the branch with two other close chains. Therefore, the branches indeed lower the acceptance rates, depending on the size of the chains on each branch but with a moderate value with $\kappa_\phi = 32$ or 64. However, even though we consider that our CBMC implementation presents a good performance for linear and branched medium size molecules.

The overall CBMC acceptance rates of our implementation is comparable with other CBMC implementation, such as the coupled-decoupled CBMC (CD-CBMC), reported by Sepehri and coauthors.[40] They obtained overall acceptance rates with 10 insertion trial angles of ~51% for propane and ~21% for 2-methylpropane, while we obtained ~60% for neopentane, ~40% for *n*-octane, and ~38% for 4-ethylheptane with 12 insertion trial angles (see Table 2). However, other insertion methods, such as that one proposed by Sepehri and coauthors[40] can obtain much higher acceptance rates for united-atomic force field models, around 95−99.9%, for linear and two-branched molecules but around 52% for three-branched molecules like neopentane.

**1,2-Dichloroethane.** The solvent effects on the conformational equilibrium for DCE were previously investigated experimentally via IR spectroscopy.[62,63] It was found that the percentage of the *trans* conformations, defined by Cl−C−C−Cl dihedral angle, is 77% in vapor phase, 35% in liquid phase,[62] and 26% in acetonitrile solution.[63] This system was also investigated by MC simulations where the parameters of the OPLS force field was proposed and a good agreement was obtained with the experimental data: 78% in vapor phase, 36% in liquid phase, and 22% in acetonitrile solution.[64]

For the MC and MD simulations of DCE, we used the same procedures and thermodynamics conditions as described for the *n*-octane. We used the OPLS parametrization for the DCE[64] and for acetonitrile.[65] For the simulation in liquid phase and in solution, we used $N = 301$ molecules (one DCE as solute and 300 solvent molecules of DCE or acetonitrile, respectively). The cutoff radius was of $r_{cut} = 0.42L \approx 16$ Å with $L$ being the cubic box size, and the long-range corrections were considered with the continuum model. The DCE molecule was divided with $N_{frag} = 2$: two fragments with $C-CH_2-Cl$ group, counting one rotatable bonds between the fragments. For the liquid simulation with all flexible DCE molecules we obtained an average density of 1.229 g/cm$^3$, in good agreement with the experimental value of 1.245 g/cm$^3$.[66] The molecular flexibility played an important role in the description of the correct density, once in a simulation with rigid *trans*-DCE molecules we obtained a density of 1.194 g/cm$^3$. Our simulations were able to reproduce the radial distribution functions previously reported in the literature.[67]

The implemented CBMC method sample all the internal degrees of freedom accordingly when compared with the MD simulations with the same force field. The evolution of the Cl−C−C−Cl dihedral angle during the simulations in different conditions are shown in the SI. Like the *n*-octane simulations, three populations (*gauche*(−), *gauche*(+), and *trans*) can be easily identify. The density distributions of this dihedral angle of the DCE in gas and liquid phases simulations are shown in Figure 5. For the gas case, we also show the ideal distribution

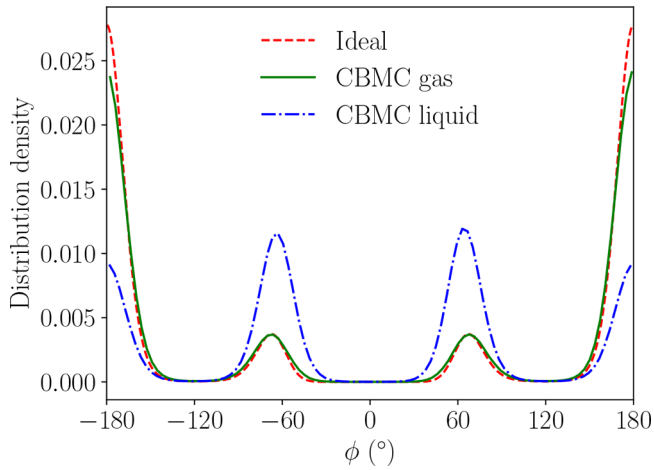



**Figure 5.** Probability density distributions of the (Cl−C−C−Cl) dihedral angle of DCE conformations sampled in gas and liquid phase CBMC simulations.

obtained by plotting the $\exp[-\beta \mathcal{U}_{torsion}(\phi)]$ curve. These distributions (ideal and gas) are similar to the ones reported by Jorgensen et al.[67] with small differences regarding the heights of the peaks, reflecting the small fluctuations in the % *trans* values. The distributions of the gas and liquid simulations clearly show the effect of the liquid environment decreasing the *trans* and increasing the *gauche* conformations of the DCE that is in agreement with the experimental findings that show a decrease of approximately 50% in the *trans* population from the gas to the liquid phase.

To calculate the percentage of *trans* conformations, we performed the same analysis as those performed for *n*-octane and the estimated statistical uncertainty in these populations


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

**Table 3. Percentage of *trans* Configurations of DCE in Different Conditions[a]**

| | % *trans* CBMC | % *trans* MD | % *trans* exp | % *trans* MC[d] |
|---|---|---|---|---|
| gas | 77 ± 3 (80) | 80 ± 15 | 77 ± 2[b] | 78.3 |
| liquid | 30 ± 3 | 34 ± 3 | 35 ± 4[b] | 36.1 |
| liquid (semiflexible) | 33 ± 3 | | | |
| liquid (biased $\phi$) | 32 ± 3 | | | |
| acetonitrile | 21 ± 3 | | 25.6[c] | 22.2 |

[a]In parentheses is the theoretical ideal values calculated from the potential energy differences between the *trans* and *gauche* forms (see the SI for details). Experimental values obtained by [b]Tanabe[62] and [c]Wiberg et al.[63] [d]Simulation values obtained by Jorgensen et al.[64]

were performed analyzing independent sets of data. In Table 3 the values of % *trans* for different conditions are presented and compared with available results. We see that the values of % *trans* obtained with CBMC ((77 ± 3)% in gas, (30 ± 3)% in liquid, and (21 ± 3)% in acetonitrile) are in good agreement with the experimental data, the values of previous simulations[64] using the same force field and the results of our MD simulations. In this case, we also take the opportunity to test the semiflexible conformational sampling and the trial insertion angle algorithms: the random $\phi$ sampling and bias $\phi$ sampling (see details in the SI). Then, three different liquid simulations were performed on the same thermodynamic conditions but with different sampling options implemented in DICE. These simulations were labeled as liquid for simulation performed with flexible fragments and with insertion trial dihedral angles $\phi$ generated randomly between $[0, 360°)$ ((30 ± 3)% of *trans* conformations were obtained); liquid (semiflexible) for simulation performed with rigid fragments and with $\phi$s generated randomly ((33 ± 3)%); and liquid (biased $\phi$) for simulation performed with flexible fragments and with $\phi$s generated using a bias, where more $\phi$s are generated closer to the minima of the torsional potential energy ((32 ± 3)%. These results for the DCE conformations and thermodynamic properties were equivalent in the three simulations within the statistical uncertainty. Then, the rigid fragment sampling does not affect the conformational sampling. Additionally, the biased generation of $\phi$ involves more calculations and causes a larger computational time than the other sampling with the same amount of trial insertion angles ($\kappa_\phi$). Therefore, we used $\kappa_\phi = 32$ for the random $\phi$ sampling and $\kappa_\phi = 16$ for the biases $\phi$ sampling, giving an overall CBMC acceptance rate of 53% and 62%, respectively. We calculated the statistical inefficiency of the CBMC using both $\phi$ samplings and obtained $s_{rand} = 930$ and $s_{bias} = 760$ (around 20% less). Thus, in this case, the biases $\phi$ sampling is more efficient because it generates a greater acceptance ratio of new conformations and statistically uncorrelated configurations faster than the random $\phi$ sampling, even with a smaller $\kappa_\phi$.

**SubPC.** The functionalized boron subphtalocyanine was selected due to its complexity and diversity of functional groups. Some of these groups are voluminous (the subphtalocyanine and the thiadiazol groups); therefore, it is reasonable to assume that MC simulations will find extra difficulty in sampling their rotation in solution or a dense environment. Our main goal here was validate the implementation and test the efficiency of the conformational sampling in solution.

We performed simulations of the isolated SubPC and in aqueous solution using the same procedures and thermodynamic conditions as described for the *n*-octane and DCE. We used the TIP3P model for water.[68] For the simulation in aqueous solution, we used $N = 2501$ being one molecule SubPC and 2500 water molecules. The cutoff radius was of $r_{cut} = 0.42L \approx 19.5$ Å, with corrections for interactions beyond the cutoff given by the continuum model. We obtained an average density of 1.03 g/cm$^3$, in good agreement with the experimental value. We used $\kappa_\phi = 32$ for the equidistant $\phi$ sampling, giving an overall CBMC acceptance rate of 49% for the isolated molecule and 23% for the molecule in solution.

The SubPC molecule was divided into 11 fragments as shown in Figure 6, where fragments 7, 10, and 11 are identical. We

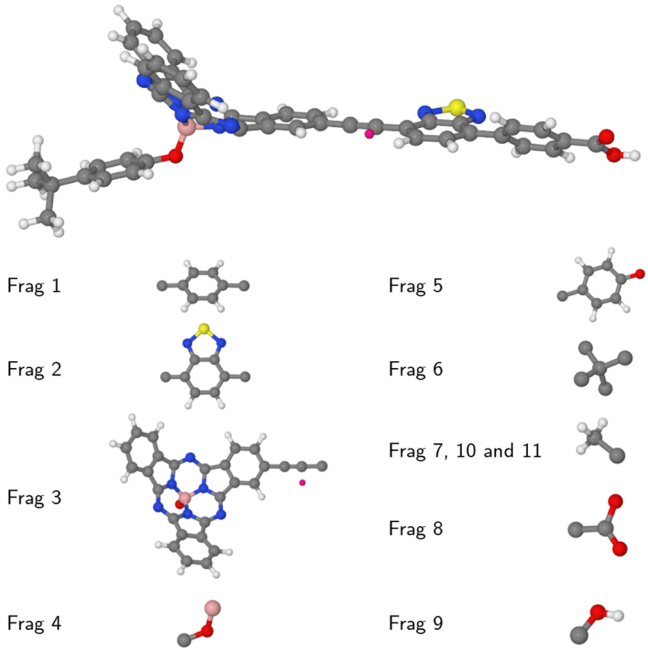



**Figure 6.** Fragmentation scheme for the SubPC molecule. Atoms colors are white for H, gray for C, red for O, blue for N, yellow for S, orange for B, and pink for a dummy site. This dummy site was used to define the dihedral angle between the fragments 3 and 2, due to the linear C−C≡C−C group.

added a dummy atom (Xx) to properly define the rotation between the fragments 3 (subphtalocyanine-ethynyl group) and 2 (thiadiazol group), due to the linear ethynyl group, C−C≡C−C. Bond lengths and bond angles were kept rigid in the optimized geometry obtained by Gotfredsen et al.[44] (using B3LYP/cc-pVDZ), and no sampling of the hard degrees of freedom was performed. Therefore, we treated the SubPC molecule with a semiflexible model with all the fragments rigid. In this case, the use of the semiflexible model simplifies a lot the force field parametrization because only the torsional parameters between the fragments should be parametrized. Usually, the LJ parameters are transferable from other groups and the atomic charges for the Coulomb potential term are obtained with


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

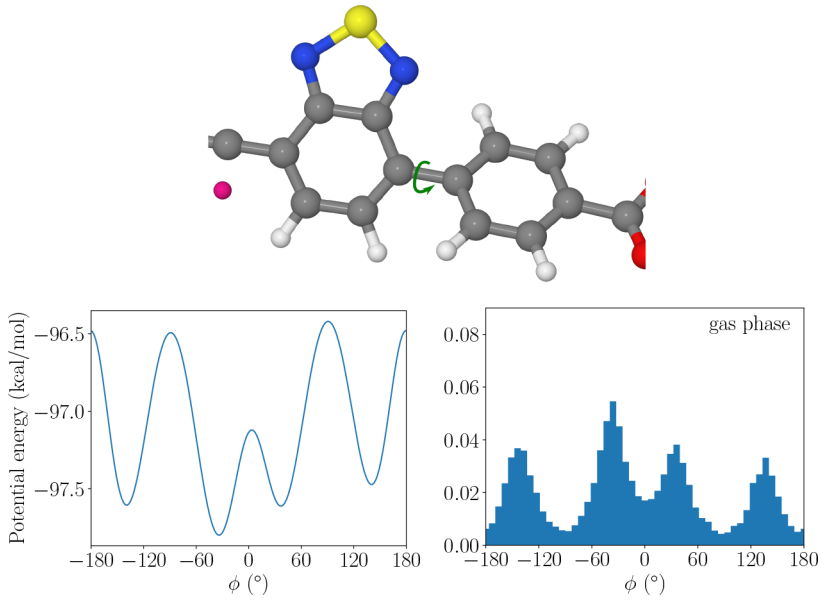



**Figure 7.** (top) Illustration of the dihedral rotation *D*1 between fragments 1 and 2. (bottom left) Total energy profile involved in this rotation (C5−C4−C7−C8) for the isolated SubPC. (bottom right) Dihedral distribution obtained with the CBMC simulation of the isolated SubPC (gas phase).

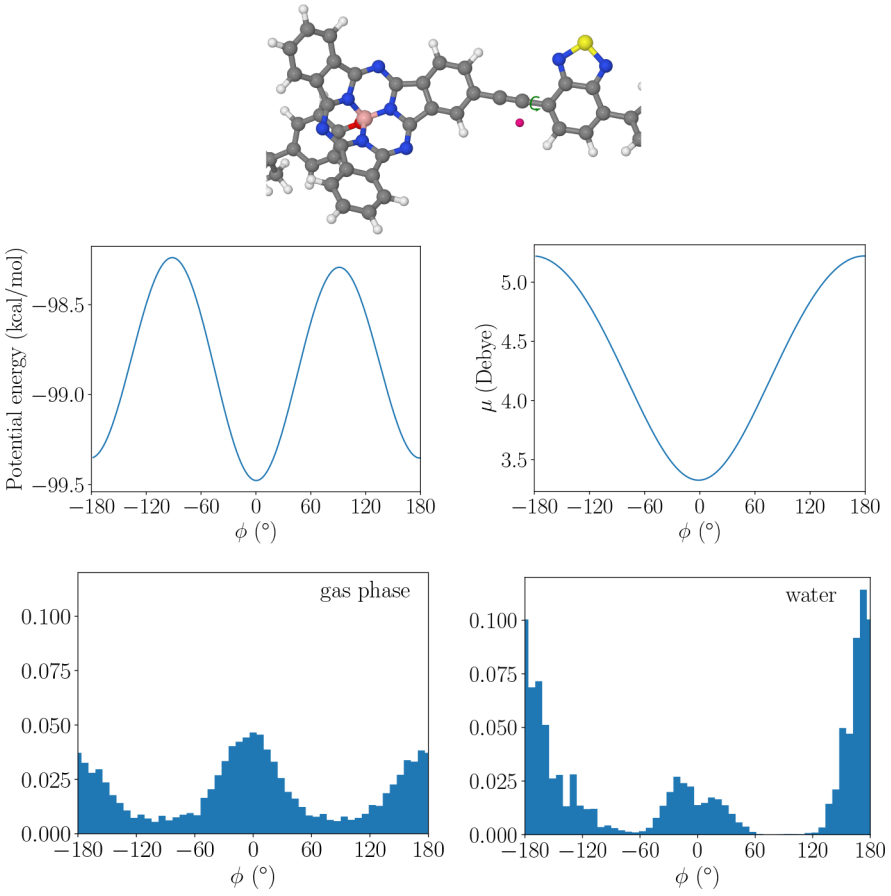



**Figure 8.** (top) Illustration of the dihedral rotation *D*3 between fragments 2 and 3. (middle left) Total energy profile involved in this rotation (C9−C10−C16−Xx94) for the isolated SubPC. (middle right) Dipole moment $\mu$ as a function of the rotation. (bottom left) Dihedral distribution obtained with the CBMC simulation of the isolated SubPC (gas phase). (bottom right) Distribution obtained with the CBMC simulation of SubPC in water.

quantum mechanical (QM) calculation. This simplification is convenient and very useful when the parametrization of some groups is not available or they are very difficult to parametrize, such as in the case of fragment 3, for which parameters are not available in the literature and additionally the parametrization is challenging, once several rings are not planar but in an umbrella-like shape. We used the OPLS-AA force field[52] for the LJ parameters, while the atomic charges were obtained using an



https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

electrostatic fit with the CHELPG method using the CAM-B3LYP/cc-pVDZ level of QM calculations. The torsional parameters of the dihedral angles between fragments were obtained from the QM energy profiles also using CAM-B3LYP/cc-pVDZ.

We analyze the evolution and the distribution of all rotational angles between the fragments, totaling 10 rotations. They are *D*1 C5−C4−C7−C8 between frags 1 and 2; *D*2 C6−C1−C58−O62 between frags 1 and 8; *D*3 C9−C10−C11−Xx94 between frags 2 and 3; *D*4 N25−B27−O28−C29 between frags 3 and 4; *D*5 B27−O28−C29−C30 between frags 4 and 5; *D*6 C31−C32−C56−C61 between frags 5 and 6; *D*7 C32−C56−C57−H82 between frags 6 and 7, C32−C56−C60−H77 between frags 6 and 10, C32−C56−C61−H80 between frags 6 and 11; and *D*8 O62−C58−O59−H91 between frags 8 and 9. All the atomic labels and the distributions of these dihedral angles, from *D*1 to *D*8 are shown in the SI.

Initially we compared the dihedral angles distributions of the isolated SubPC sampled in the CBMC simulations with the total rotational energy profile around the specific bond that connect the two fragments with two rigid parts. We obtained for each dihedral angle that the most frequently sampled conformations have dihedral angles that are in agreement with the minima of the energy profiles and describe peaks in the dihedral angle distributions. As an example, Figure 7 shows the energy profile and the dihedral distribution of the *D*1 rotation in the C−C bond between fragments 1 (benzene group) and fragment 2 (thiadiazol group). In the energy profile four asymmetric minima were identified at around ±140° and ±35°. The minimum at −33° is slightly deeper than the other three, with the largest difference being of about 0.33 kcal/mol for the minimum at 141°. These differences are reflected in the sampled angles in gas phase, as show in the lower panel of Figure 7, where the distribution peaks follows the minima of the potential energy. The same analysis was performed for all other dihedral angles, from *D*1 to *D*8, resulting in the correct distributions (see all graphics in the SI). Therefore, we conclude that in gas phase, even with 11 fragments, the CBMC method can sample the PES correctly.

In solution, we expect that voluminous groups (such as fragments 2 and 3) will present smaller rotational capability caused by the steric effect with the nearby solvent molecules. To investigate if the CBMC method could sample the configurations correctly under these circumstances, we performed simulations of SubPC in aqueous solution and analyzed the sampled distribution of dihedral angles and solute−solvent interaction. Figure 8 shows the dihedral distribution of *D*3 in water and the gas phase, as well as the torsional potential and total dipole moment change during the rotation. As previously shown for *D*1 in Figure 7, the CBMC sampled the minima of the total classical energy profile of *D*3 in gas phase, reflecting the energy difference between the 0° and ±180° minima and resulting in a slightly higher peak around 0°. In water solution, the distribution changes and the ±180° conformation is favored. This change can be explained through the solute−solvent dipolar interaction. The middle right panel of Figure 8 shows that the dipole moment $\mu$ of SubPC increases from 3.3 D when the dihedral is at 0° to 5.2 D at ±180°. Therefore, the solute−solvent energy interaction became stronger due to the dipolar interaction compensating for the internal energy increase. We show this compensation in Figure 9, using the average internal energy and average solute−solvent energy are showed with respect to the *D*3 dihedral angles. The average internal energy

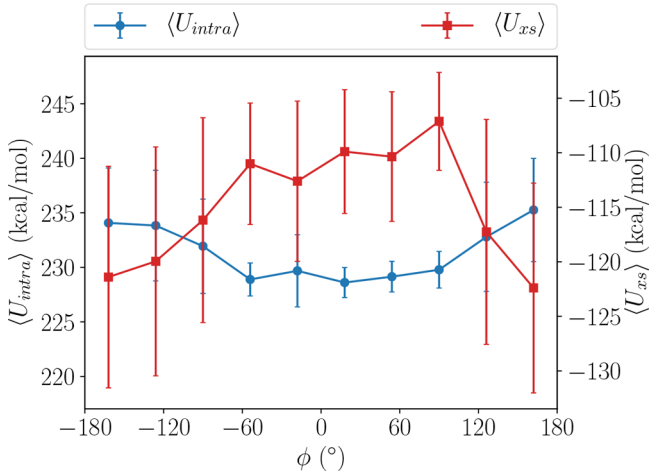



**Figure 9.** Average solute−solvent and average intramolecular energies binned as a function of *D*3, obtained with configurations sampled during the CBMC simulation. The error bars are the standard deviation of the energies in each bin.

increase about 7 kcal/mol when *D*3 goes from 0° to ±180°, while the average solute−solvent energy became stronger by about 12 kcal/mol.

Similarly to this behavior observed for the rotation of *D*3, the distribution of the OH rotation defined by *D*8 is also affected by the solvent interaction, also due to the stabilization of a higher dipole conformation. These results are shown in the SI. Therefore, the CBMC method is able to account for the solute−solvent interaction properly, considering both long-range and specific interactions.

## CONCLUSIONS

In this paper we have presented a new version of DICE, a MC software used for molecular simulation of solute−solvent systems. To tackle the problem of simulating flexible molecules with a MC method, we have implemented and extended the CBMC method of Shah and Maginn.[33] We detailed the simplifications to the acceptance criterion of the method proposed in a previous paper,[3] alongside with extra types of fragments that can be used in the fragmentation scheme needed by the CBMC method. Implementation details were described, as well as the options taken during the implementation of the method.

Simulations for DCE and *n*-octane were run, and the results compared with MD simulations and experimental data when available. For *n*-octane, the gas phase simulations showed that the MD simulations produces configurations with higher statistical inefficiency than the configurations produced in the CBMC configuration. This was shown by the correlation interval (or correlation time), which was 1 order of magnitude larger for MD in gas phase. Based on information of our previous work, this is not surprising, as CBMC easily escapes higher energy barriers compared to MD.[3] An analysis of the conformer population for three cases where different terms of the potential were considered showed that the results of CBMC and MD are consistent. Simulations for *n*-octane in chloroform were also performed, and again, the comparison with MD simulations provided similar results. Additionally, the efficiency of the CBMC sampling was analyzed considering the quantity of trial insertion angles $\kappa_\phi$ in the three alkane molecules: *n*-octane, neopentane, and 4-ethylheptane. The results showed a good


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

efficiency of the CBMC method implemented in the DICE code for linear and branched molecules with moderate size.

The results for DCE show that the CBMC results are consistent with other methodologies and experimental data. We also verify that different options within the algorithm, such as rigid fragments and biased insertion angle generation, have no effect in the sampled geometries, since the conformer population is the same, within the error, independent of the chosen options.

Finally, we have used the large SubPC molecule to show the possibility of using the CBMC method to sample degrees of freedom of moderate size molecules with large fragments in solution. In these simulations, we also used the rigid fragments proposal, simplifying greatly the topology, especially due to the boron subphtalocyanine group. We showed that the CBMC sampling in gas phase reproduced the potential energy minima, with a distribution that reflected the energy differences. In water solution, we observed the solvent effect on the distributions, as conformations with higher dipole moment were favored due to the dipolar interaction. Therefore, the CBMC method is able to sample the dihedral angles efficiently even when large fragments are present in the molecule, showing that the steric effects do not limit the usage of the algorithm.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00077.

More information concerning the details about the DICE options, algorithms, parallelization efficiency, and input files and additional simulation data (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Henrique M. Cezar** − *Instituto de Fisica, Universidade de Sao Paulo, 05508-090 Sao Paulo, SP, Brazil;* orcid.org/0000-0002-7553-0482; Email: henrique.cezar@usp.br

**Sylvio Canuto** − *Instituto de Fisica, Universidade de Sao Paulo, 05508-090 Sao Paulo, SP, Brazil;* orcid.org/0000-0002-9942-8714; Email: canuto@if.usp.br

**Kaline Coutinho** − *Instituto de Fisica, Universidade de Sao Paulo, 05508-090 Sao Paulo, SP, Brazil;* orcid.org/0000-0002-7586-3324; Email: kaline@if.usp.br

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00077


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors thank CAPES for the BioMol project 23038.004630/2014-35; the National Institute of Science and Technology of Complex Fluids (INCT-FCx) with the CNPq grant 141260/2017-3 and FAPESP grant 2014/50983-3; FAPESP and Shell grant 2017/11631-2; and the strategic importance of the support given by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&D levy regulation. The authors also thank all the DICE users, whose feedback reflected in great improvements to the software throughout the years.

## ■ REFERENCES


(1) Jorgensen, W. L.; Tirado-Rives, J. Monte Carlo vs Molecular Dynamics for Conformational Sampling. *J. Phys. Chem.* **1996**, *100*, 14508−14513.

(2) Ulmschneider, J. P.; Ulmschneider, M. B.; Di Nola, A. Monte Carlo vs Molecular Dynamics for All-Atom Polypeptide Folding Simulations. *J. Phys. Chem. B* **2006**, *110*, 16733−16742.

(3) Cezar, H. M.; Canuto, S.; Coutinho, K. Solvent effect on the syn/anti conformational stability: A comparison between conformational bias Monte Carlo and molecular dynamics methods. *Int. J. Quantum Chem.* **2019**, *119*, e25688.

(4) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(5) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919.

(6) Berne, B. J.; Straub, J. E. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181−189.

(7) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E1009−E1018.

(8) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; VAN DER SPOEL, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(9) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.

(10) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comp. Mol. Sci.* **2013**, *3*, 198−210.

(11) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(12) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689−1700.

(13) Shah, J. K.; Marin-Rimoldi, E.; Mullen, R. G.; Keene, B. P.; Khan, S.; Paluch, A. S.; Rai, N.; Romanielo, L. L.; Rosch, T. W.; Yoo, B.; Maginn, E. J. Cassandra: An open source Monte Carlo package for molecular simulation. *J. Comput. Chem.* **2017**, *38*, 1727−1739.

(14) Martin, M. G. MCCCS Towhee: a tool for Monte Carlo molecular simulation. *Mol. Simul.* **2013**, *39*, 1212−1222.

(15) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42*, 81−101.

(16) Alexiadis, O.; Cheimarios, N.; Peristeras, L. D.; Bick, A.; Mavrantzas, V. G.; Theodorou, D. N.; Hill, J. R.; Krokidis, X. Chameleon: A generalized, connectivity altering software for tackling properties of realistic polymer systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, *9*, e1414.

(17) Freitas, L. C. G. DIADORIM: a Monte Carlo Program for liquid simulations including quantum mechanics and molecular mechanics (QM/MM) facilities: applications to liquid ethanol. *J. Braz. Chem. Soc.* **2009**, *20*, 1541−1548.

(18) Coutinho, K.; Canuto, S. *DICE: AMonte Carlo program for molecular liquid simulation*; University of São Paulo, v2.9. 2013.

(19) Cezar, H. M.; Canuto, S.; Coutinho, K. *DICE: A Monte Carlo program for molecular liquid simulation*; University of São Paulo, v. 3.0 (available at https://portal.if.usp.br/dice). 2017.

(20) Frenkel, D.; Smit, B. *Understanding molecular simulation*, 2nd ed.; Academic Press: San Diego, 2002.

(21) Coutinho, K.; Canuto, S. Theoretical description of the absorption spectra of solid and liquid benzene. *J. Mol. Struct.: THEOCHEM* **1993**, *287*, 99−106.

(22) Coutinho, K.; Canuto, S. Solvent Effects from a Sequential Monte Carlo - Quantum Mechanical Approach. *Adv. Quantum Chem.* **1997**, *28*, 89−105.
(23) Coutinho, K.; Canuto, S.; Zerner, M. C. A Monte Carlo-quantum mechanics study of the solvatochromic shifts of the lowest transition of benzene. *J. Chem. Phys.* **2000**, *112*, 9874−9880.
(24) Coutinho, K.; Canuto, S. Solvent effects in emission spectroscopy: A Monte Carlo quantum mechanics study of the n←$\pi^*$ shift of formaldehyde in water. *J. Chem. Phys.* **2000**, *113*, 9132−9139.
(25) Guedes, R. C.; Coutinho, K.; Costa Cabral, B. J.; Canuto, S. Differential Hydration of Phenol and Phenoxy Radical and the Energetics of the Phenol O−H Bond in Solution. *J. Phys. Chem. B* **2003**, *107*, 4304−4310.
(26) Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. An efficient statistically converged average configuration for solvent effects. *Chem. Phys. Lett.* **2007**, *437*, 148−152.
(27) Damasceno, M. V. A.; Cabral, B. J. C.; Coutinho, K. Structure and electronic properties of hydrated mesityl oxide: a sequential quantum mechanics/molecular mechanics approach. *Theor. Chem. Acc.* **2012**, *131*, 1214.
(28) Colherinhas, G.; Fonseca, T.; Castro, M.; Coutinho, K.; Canuto, S. Isotropic magnetic shielding constants of retinal derivatives in aprotic and protic solvents. *J. Chem. Phys.* **2013**, *139*, 094502.
(29) Caputo, M.; Provasi, P.; Benitez, L.; Georg, H.; Canuto, S.; Coutinho, K. A Monte Carlo-Quantum Mechanics Study of Magnetic Properties of Hydrogen Peroxide in Liquid Water. *J. Phys. Chem. A* **2014**, *118*, 6239−6247.
(30) Cornetta, L. M.; Coutinho, K.; Canuto, S.; Varella, M. T. d. N. Free energy barrier for dissociation of the guanosine monophosphate anion in water. *Eur. Phys. J. D* **2016**, *70*, 176.
(31) Cezar, H. M.; Canuto, S.; Coutinho, K. Understanding the absorption spectrum of mesityl oxide dye in solvents of different polarities. *J. Mol. Liq.* **2020**, *307*, 112924.
(32) Coutinho, K.; De Oliveira, M. J.; Canuto, S. Sampling configurations in Monte Carlo simulations for quantum mechanical studies of solvent effects. *Int. J. Quantum Chem.* **1998**, *66*, 249−253.
(33) Shah, J. K.; Maginn, E. J. A general and efficient Monte Carlo method for sampling intramolecular degrees of freedom of branched and cyclic molecules. *J. Chem. Phys.* **2011**, *135*, 134121.
(34) Dodd, L.; Boone, T.; Theodorou, D. A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Mol. Phys.* **1993**, *78*, 961−996.
(35) Tsourtou, F. D.; Peroukidis, S. D.; Peristeras, L. D.; Mavrantzas, V. G. Monte Carlo Algorithm Based on Internal Bridging Moves for the Atomistic Simulation of Thiophene Oligomers and Polymers. *Macromolecules* **2018**, *51*, 8406−8423.
(36) Wu, M. G.; Deem, M. W. Analytical rebridging Monte Carlo: Application to cis/trans isomerization in proline-containing, cyclic peptides. *J. Chem. Phys.* **1999**, *111*, 6625−6632.
(37) Siepmann, J. I.; Frenkel, D. Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* **1992**, *75*, 59−70.
(38) Martin, M. G.; Siepmann, J. I. Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *J. Phys. Chem. B* **1999**, *103*, 4508−4517.
(39) Macedonia, M. D.; Maginn, E. J. A biased grand canonical Monte Carlo method for simulating adsorption using all-atom and branched united atom models. *Mol. Phys.* **1999**, *96*, 1375−1390.
(40) Sepehri, A.; Loeffler, T. D.; Chen, B. Improving the efficiency of configurational-bias Monte Carlo: A density-guided method for generating bending angle trials for linear and branched molecules. *J. Chem. Phys.* **2014**, *141*, 074102.
(41) Sepehri, A.; Loeffler, T. D.; Chen, B. Improving the Efficiency of Configurational-Bias Monte Carlo: A Jacobian-Gaussian Scheme for Generating Bending Angle Trials for Linear and Branched Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 1577−1583.
(42) Sepehri, A.; Loeffler, T. D.; Chen, B. Improving the Efficiency of Configurational-Bias Monte Carlo: Extension of the Jacobian-Gaussian Scheme to Interior Sections of Cyclic and Polymeric Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 4043−4053.
(43) Claessens, C. G.; González-Rodríguez, D.; Rodríguez-Morgade, M. S.; Medina, A.; Torres, T. Subphthalocyanines, Subporphyrazines, and Subporphyrins: Singular Nonplanar Aromatic Systems. *Chem. Rev.* **2014**, *114*, 2192−2277.
(44) Gotfredsen, H.; Neumann, T.; Storm, F. E.; Muñoz, A. V.; Jevric, M.; Hammerich, O.; Mikkelsen, K. V.; Freitag, M.; Boschloo, G.; Nielsen, M. B. Donor-Acceptor-Functionalized Subphthalocyanines for Dye-Sensitized Solar Cells. *ChemPhotoChem.* **2018**, *2*, 976−985.
(45) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 2nd ed.; Oxford University Press: Oxford, 2017.
(46) Owicki, J. C. *Optimization of Sampling Algorithms in Monte Carlo Calculations on Fluids* **1978**, *86*, 159−171.
(47) Bigot, B.; Jorgensen, W. L. Sampling methods for Monte Carlo simulations of n-butane in dilute solution. *J. Chem. Phys.* **1981**, *75*, 1944−1952.
(48) Zwanzig, R. W. High-Temperature Equation of State by a Pertubation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(49) Jorgensen, W.; Buckner, J.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742.
(50) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(51) Ninov, J.; Stefanova, T.; Petrov, P. Vapor-Liquid Equilibria at 101.3 kPa for Diethylamine + Chloroform. *J. Chem. Eng. Data* **1995**, *40*, 199−201.
(52) 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*, 11225−11236.
(53) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.
(54) Hockney, R.; Goel, S.; Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, *14*, 148−158.
(55) 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*, 1463−1472.
(56) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(57) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.
(58) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(59) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(60) Chatfield, C. *The Analysis of Time Series*, 6th ed.; Chapman and Hall, 2003.
(61) Hammonds, K. D.; McDonald, I. R.; Ryckaert, J.-P. Conformer distribution and the kinetics of trans-gauche isomerization in a model of liquid n-octane. *Chem. Phys. Lett.* **1993**, *213*, 27−31.
(62) Tanabe, K. Calculation of infrared band intensities and determination of energy differences of rotational isomers of 1,2-dichloro-, 1,2-dibromo- and 1-chloro-2-bromoethane. *Spectrochim. Acta, Part A* **1972**, *28*, 407−424.
(63) Wiberg, K. B.; Keith, T. A.; Frisch, M. J.; Murcko, M. Solvent Effects on 1,2-Dihaloethane Gauche/Trans Ratios. *J. Phys. Chem.* **1995**, *99*, 9072−9079.
(64) Jorgensen, W. L.; McDonald, N. A.; Selmi, M.; Rablen, P. R. Importance of Polarization for Dipolar Solutes in Low-Dielectric Media: 1,2-Dichloroethane and Water in Cyclohexane. *J. Am. Chem. Soc.* **1995**, *117*, 11809−11810.


https://dx.doi.org/10.1021/acs.jcim.0c00077
J. Chem. Inf. Model. 2020, 60, 3472−3488

(65) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. *J. Comput. Chem.* **2001**, *22*, 1340−1352.
(66) Haynes, W. M. *CRC handbook of chemistry and physics*, 95th ed.; CRC Press, 2014.
(67) Jorgensen, W. L.; Binning, R. C.; Bigot, B. Structures and Properties of Organic Liquids - N-Butane and 1,2-Dichloroethane and Their Conformational Equilibria. *J. Am. Chem. Soc.* **1981**, *103*, 4393−4399.
(68) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.


https://dx.doi.org/10.1021/acs.jcim.0c00077
*J. Chem. Inf. Model.* 2020, 60, 3472−3488