JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

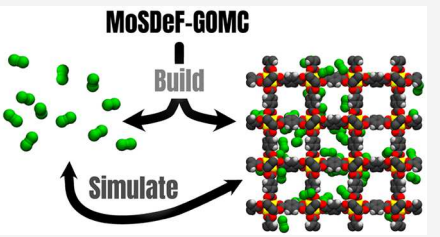
MoSDeF-GOMC: Python Software for the Creation of Scientific Workflows for the Monte Carlo Simulation Engine GOMC

Brad Crawford, Umesh Timalsina, Co D. Quach, Nicholas C. Craven, Justin B. Gilmer, Clare McCabe, Peter T. Cummings, and Jeffrey J. Potoff*

Cite This: J. Chem. Inf. Model. 2023, 63, 1218−1228 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: MoSDeF-GOMC is a python interface for the Monte Carlo software GOMC to the Molecular Simulation Design Framework (MoSDeF) ecosystem. MoSDeF-GOMC automates the process of generating initial coordinates, assigning force field parameters, and writing coordinate (PDB), connectivity (PSF), force field parameter, and simulation control files. The software lowers entry barriers for novice users while allowing advanced users to create complex workflows that encapsulate simulation setup, execution, and data analysis in a single script. All relevant simulation parameters are encoded within the workflow, ensuring reproducible simulations. MoSDeF-GOMC's capabilities are illustrated through a number of examples, including prediction of the adsorption isotherm for CO_2 in IRMOF-1, free energies of hydration for neon and radon over a broad temperature range, and the vapor−liquid coexistence curve of a four-component surrogate for the jet fuel S-8. The MoSDeF-GOMC software is available on GitHub at https://github.com/GOMC-WSU/MoSDeF-GOMC.



INTRODUCTION

The research community has recognized a need for automated and reproducible generation of input files for molecular dynamics (MD) and Monte Carlo (MC) simulations, and a large number of independent efforts have been launched. Some examples include tools for automated force field parameter assignment,[1−4] generation of parameter, topology, and control files for a single simulation,[5−15] to those that support the creation of complex workflows that may involve multiple calculations (potentially with different software) and data analysis.[16−24] Of the input file builders, particularly notable is the well-utilized, web-based CHARMM-GUI.[14,15] CHARMM-GUI is capable of building input for simulations with the CHARMM,[25−28] AMBER,[29,30] OpenFF,[31] and OPLS-AA[32,33] force fields, as well as the coarse-grained MARTINI[34] force field, for all mainstream molecular dynamics codes.[35]

While many tools have been developed to support molecular dynamics codes, software for the generation of input files for Monte Carlo simulations is much more limited.[36] This is likely due to the smaller number of Monte Carlo users, compared to molecular dynamics, and the fact that many Monte Carlo codes use proprietary input files.[37−40] With GPU Optimized Monte Carlo (GOMC),[41,42] we sought to avoid some of these issues by using CHARMM-style topology and parameter files.[43] In principle, the force field, PDB, and PSF files generated by CHARMM-GUI can be used with GOMC. However, CHARMM-GUI only generates one set of input files at a time, precluding its use for studies where simulations must be performed over a wide range of chemical space and/or thermodynamic state points, or for simulations that require two simulation boxes (for example, grand canonical ensemble, and Gibbs ensemble Monte Carlo[44]). Additionally, CHARMM-GUI does not produce control files compatible with GOMC. Despite efforts to borrow as much NAMD[45,46] syntax as possible, there are substantial differences between the control files required for MD and MC simulations. Additionally, while CHARMM-GUI can build a wide range of chemical and biological systems, it does not (currently) support many systems of interest in chemical engineering or chemical physics.

To address the need for streamlined construction of input files and support for complex workflows that utilize GOMC, MoSDeF-GOMC[47,48] was developed as part of the Molecular Simulation Design Framework (MoSDeF) ecosystem.[49,50] MoSDeF serves as the foundation for input file writers for a number of molecular dynamics codes, including HOOMD,[51,52] LAMMPS,[53,54] and GROMACS,[55] and the Monte Carlo software Cassandra[39] (via MoSDeF-Cassandra[36]). MoSDeF-GOMC provides a Python interface to GOMC, greatly simplifying the integration of GOMC with workflow managers, such as signac.[56,57] MosDeF-GOMC adds support for writing force field parameter files in CHARMM format,[43] and coordinate and connectivity files in protein data bank (PDB)

Received: November 29, 2022
Published: February 15, 2023

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

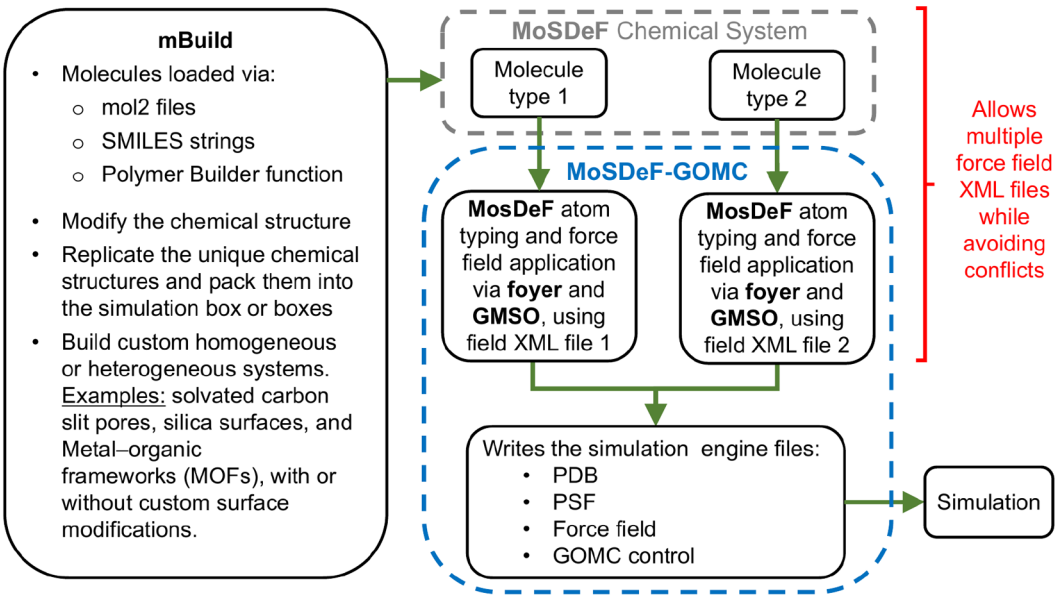


Figure 1. MoSDeF-GOMC Workflow

Table 1. Default Monte Carlo Move Ratios Used by MoSDeF-GOMC

ensemble	translate	rotate	intra-swap[60]	swap[61]	regrowth[62]	crank-shaft[63]	volume
NVT	0.15	0.15	0.3	0	0.3	0.1	0
NPT	0.15	0.15	0.29	0	0.3	0.1	0.01
GEMC-NVT	0.19	0.2	0.1	0.2	0.2	0.1	0.01
GEMC-NPT	0.19	0.2	0.1	0.2	0.2	0.1	0.01
GCMC	0.15	0.15	0.1	0.35	0.15	0.10	0

and protein structure file (PSF) format, respectively. MoSDeF-GOMC also simplifies writing GOMC control files, acting as a semi-expert system, providing reasonable default values for many parameters, while still allowing users full control over any parameter used by GOMC.

The manuscript is organized as follows. In the next section, the design of MoSDeF-GOMC is described. This is followed by a number of use cases, each designed to highlight a specific feature available to users via MoSDeF-GOMC, which include: the calculation of an adsorption isotherm for CO_2 in IRMOF-1, free energies of hydration for Ne and Rn over a broad range of temperatures, and the vapor−liquid coexistence curve of a surrogate for the jet fuel S8. Complete workflows to perform these calculations are available on GitHub.[58]

SOFTWARE ARCHITECTURE

MoSDeF-GOMC Overview. MoSDeF-GOMC version 1.0.0[47,48] is a new package that interfaces with MoSDeF, enabling the generation of coordinate (PDB), connectivity (PSF), CHARMM-style force field parameter files, and the control file required by the Monte Carlo simulation engine GOMC.[41,42] MoSDeF-GOMC abstracts away many details of performing Monte Carlo simulations so that users can focus more of their effort on domain science, and less on debugging input files. All of GOMC's required input files may be generated with a few lines of Python code. MoSDeF-GOMC provides reasonable default values for the GOMC control file writer, while allowing expert users full access to GOMC simulation parameters. When combined with available force field XML files, MoSDeF-GOMC significantly lowers the entry barrier for novice users.

MoSDeF-GOMC has a structured workflow, which is presented in Figure 1. This workflow shows how the chemical system is created, parametrized, and GOMC input files are written. Molecules can be built using a variety of approaches: imported from a mol2 file; using SMILES strings (for all-atom force fields); constructed by removing atoms/beads from other molecules and piecing them together as a monomer or polymer with custom head and tail moieties via MoSDeF's polymer builder function;[59] or with custom build rules defined by the user. Molecules are then dimensionally aligned (if needed), duplicated, and packed into a simulation box (or boxes for simulations in open ensembles). These coordinates are input into MoSDeF-GOMC, where the user can select a force field to apply to each unique molecule type. Applying force fields on a per molecule basis reduces the effort required to create MoSDeF force field XML files, and allows users to validate each force field XML with a generic class of molecules. This reduces the risk of creating incompatible force field XML combinations or modifying an already validated file, introducing possible sources of error. Next, the PDB, PSF, and CHARMM-style[43] force field files are directly output from MoDeF-GOMC's parametrized system. Finally, the GOMC control file is written using the parametrized system, and some required user inputs, such as file names, simulation temperature, and the number of simulation MC steps.

MoSDeF-GOMC supports all of the ensembles available in GOMC (*NVT*, *NPT*, μVT, GEMC-NVT, and GEMC-NPT), and generates the appropriate number of simulation boxes based on the ensemble selected. MoSDeF-GOMC includes keywords to access all of the functionality available in GOMC version 2.76. Most of MoSDeF-GOMC's GOMC control file

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

writer variable names are identical or very close to GOMC variables, so previous knowledge is easily transferred between MoSDeF-GOMC and GOMC.

MoSDeF-GOMC performs a number of preflight checks to ensure users have entered valid information for the control file writer. For example, it checks and ensures that the keywords specified are correct and have been assigned realistic values, input variables are in the correct form or type, and the selected move ratios sum to 1.0. It validates molecule names specified for certain move types, for example, GEMC swap moves need to define molecule/residue names. These checks save users from frustration by preventing MoSDeF-GOMC from generating malformed input files that would produce errors in GOMC.

MoSDeF-GOMC provides reasonable default values for the ratios of various Monte Carlo moves to the GOMC control file writer, which are listed in Table 1. Providing default values enables novice users to run statistically correct simulations with minimal effort. It is expected that expert users will set Monte Carlo move ratios manually to optimize the sampling efficiency of the simulation. Unless explicitly set by the user, the control file writer automatically sets the number of Monte Carlo steps (MCS) for the equilibration period, analysis, and data output to files based on a maximum default step setting (see Table 2).

Table 2. Default Number of Monte Carlo Steps (MCS) Used by MoSDeF-GOMC for Various GOMC Operations

GOMC control variable	maximum default MCS
equilibrium steps (EqSteps)	1×10^6
adjustment steps (AdjSteps)	1×10^3
restart frequency (RestartFreq)	1×10^6
checkpoint frequency (CheckpointFreq)	1×10^6
coordinates frequency (CoordinatesFreq)	1×10^6
DCD frequency (DCDFreq)	1×10^6
console frequency (ConsoleFreq)	1×10^4
pressure frequency (PressureCalc)	1×10^4
block average frequency (BlockAverageFreq)	1×10^4
histogram frequency (HistogramFreq)	1×10^4
histogram sample frequency (SampleFreq)	500

While most keywords are self-explanatory, some benefit from further explanation. The DCDFreq interval is used for outputting trajectory information in DCD format, which is used by a number of molecular dynamics software packages (for example, CHARMM, NAMD, and OpenMM), and has been adopted by GOMC. The DCD format is a single precision binary file that stores atomic coordinates, box vectors, and time information. In GOMC, time information is replaced by the Monte Carlo step number. HistogramFreq is used to set the interval for output of histogram information (number of molecules and energy of the system) observed during grand canonical Monte Carlo simulations. The maximum default MCS is only used when the total number of MCS/10 $\geq$ maximum default MCS, and total simulation MCS/10 $\geq$ 1. If the maximum default number of MCS is not used, then the algorithm checks if the total simulation MCS/10 $\geq$ 1; if true, it sets default values to the integer value of the total simulation MCS/10. If none of the above are true, then the default is set to 1. These defaults ensure reasonable values are provided to GOMC, producing accurate and useful data, while avoiding novice mistakes, such as specifying a very small output interval, which may result in the generation of many gigabytes of useless data.

MoSDeF Ecosystem. The MoSDeF ecosystem[49,50] on which MoSDeF-GOMC is built, is composed of the packages mBuild,[64] Foyer,[65] GMSO,[66,67] and force field-utilities.[68] For this work, the following MoSDeF environment packages were utilized: mBuild version 0.15.1; Foyer version 0.11.3; GMSO version 0.9.1; force field-utilities version 0.2.1.

The generation of atomic coordinates for each molecule is performed by mBuild through various methods.[49,50,65,69] Molecule templates may be loaded from mol2 files for both all-atom and united-atom force fields, while for all-atom representations, molecules may be constructed using SMILES strings.[70−72] mBuild includes a polymer builder,[59] and allows users to write their own custom routines for building molecules. Once molecules are constructed, mBuild uses Packmol[73] to generate initial coordinates for all molecules in the simulation.

The application of force field parameters is performed by the Foyer, while the chemical object information is stored in ParmEd[74] or its replacement GMSO, and either may be called by MoSDeF-GOMC. Force field parameters are stored in XML format, and atom-typing is performed using SMARTS strings. The force field-utilities package transforms the force field XML files to Foyer and GMSO force fields. MoSDeF-GOMC supports a number of potential energy equation forms via GMSO: 12−6 Lennard−Jones and Mie[75] nonbonded potentials; harmonic bonds and angles; OPLS,[32,33] Ryckert-Bellmans, and periodic torsional potentials; and periodic impropers. GMSO supports the use of units, and force field parameters may be entered in their native format, minimizing errors caused during transcription from literature sources. All potential energy forms are automatically converted to a form usable by GOMC and NAMD with appropriate units.

NAMD and GOMC share the same file formats, so their files can be read directly by one another, allowing seamless switching from molecular dynamics (MD) to Monte Carlo (MC) simulations. As highlighted in some of the examples that follow, it can be computationally advantageous to utilize NAMD to quickly equilibrate the system using MD and then perform the desired MC simulations using GOMC. MoSDeF-GOMC can also build PDB, PSF, and parameter files for py-MCMD,[76−78] another Python software that utilizes both NAMD and GOMC to perform hybrid MD/MC simulations. However, unlike standard GOMC simulations, py-MCMD requires custom GOMC and NAMD control files that the user must modify outside of the MoSDeF-GOMC framework.

MoSDeF-GOMC can be coupled with signac[56,57] to create complex, fully automated workflows for performing simulations over a range of thermodynamic and chemical space.[58] These scripts can automate the entire workflow, including building the chemical systems, job submission, data analysis, and figure generation. With an effort similar to what is required to launch a single simulation and perform the resulting data analysis, workflows can be created that build and launch hundreds to thousands of GOMC simulations with minimal user interaction. Users can also expand the range of thermodynamic or chemical space at any point in the study, for example, by adding new temperatures, system compositions, or different molecules. Equally important is that the entire workflow can be shared with the research community, enabling others to reproduce published work exactly.

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

■ ILLUSTRATIVE EXAMPLES

To illustrate the types of simulations that may be created with MoSDeF-GOMC, a number of examples are provided. These include the calculation of an adsorption isotherm for CO_2 in IRMOF-1, the prediction of the free energies of hydration for Ne and Rn, and the vapor−liquid coexistence curve for a four-component surrogate for the jet fuel S8. Some of these examples also highlight unique configurational-bias sampling algorithms that are included in GOMC version 2.76.

CO_2 Adsorption in IRMOF-1. Perhaps the broadest application of grand canonical Monte Carlo simulations in the scientific literature is for the determination of Henry's constants and adsorption isotherms for small molecules such as noble gases,[79,80] light alkanes,[81,82] SO_2,[83] NO_x,[83] H_2S,[84] and CO_2[85] in porous materials. Metal organic frameworks (MOFs) employ a variety of metal ions and organic linkers to create porous materials with high surface area.[86] The modular nature of MOFs presents an extraordinary potential to fine-tune their properties for specific target applications, and grand canonical Monte Carlo simulations have been used to perform high-throughput screening to identify MOFs with exceptional storage or separation capabilities.[79,87]

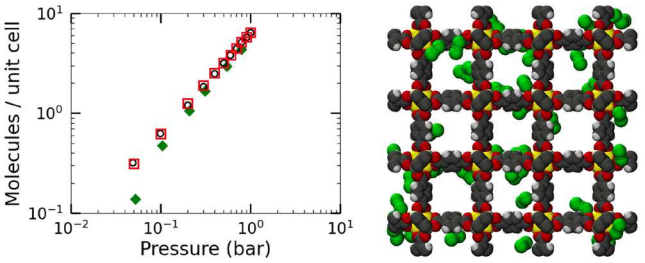


Figure 2. Adsorption isotherm for CO_2 in IRMOF-1 at 298 K (left), and a snapshot of this system (right) illustrating the structure and interactions of CO_2 with the MOF. This work (red squares); Yazaydin et al. (black circles);[92] and experiment (green diamonds).[92] Atoms in the snapshot are color coded as Zinc (yellow), IRMOF-1 oxygen (red), IRMOF-1 carbon (gray), IRMOF-1 hydrogen (white), and CO_2 (green).

Here, an adsorption isotherm for CO_2 in IRMOF-1[88] is determined from grand canonical Monte Carlo simulations. The whole isotherm was generated using a single signac[56,57] workflow, including the MoSDeF-GOMC software, which automated all of the simulations and analysis processes.[58] There are only a few user-defined variables for this workflow, which minimizes the number of changes required to add more data points via simulations for evaluation. The TraPPE force field was used for CO_2,[89] while the DREIDING and UFF force fields were used for the MOF atoms.[90,91] Additional details of the calculations are provided in the Supporting Information. The adsorption isotherm, and a snapshot from one of the simulations, is shown in Figure 2, while numerical data are provided in Table 3. The results produced here are in close agreement with those from Yazaydin et al.,[92] showing only a slight deviation at the lowest pressures. This could be due to the use of a single unit cell ($1 \times 1 \times 1$) by Yazaydin et al., while this work utilized a larger $2 \times 2 \times 2$ unit cell. While the workflow for this example generates data for a single isotherm, expanding it to include additional temperatures is straightforward. In the future, MoSDeF-GOMC with signac could automate high-throughput screening of MOF chemistries to develop new materials with optimal properties for the problem of interest.

Free Energies of Hydration for Noble Gases. Understanding the partitioning of compounds between phases is important for a wide variety of applications,[93,94] including drug design,[95−98] design of separation processes,[99,100] and the prediction of the environmental fate of toxic industrial chemicals.[101−104] The free energy of hydration is an essential quantity needed to understand the partitioning of solutes and may be determined directly from computer simulations. Typically, free energy calculations require a large number of calculations and post-processing of the data; therefore, research groups either use in-house written scripts or one of the numerous automated workflows.[16−18,20,21,105] These workflows were designed primarily to support GROMACS,[55] however, and are not compatible with GOMC.

This section highlights both the ability of GOMC to perform free energy calculations with techniques such as MBAR[106−108] and thermodynamic integration, as well as the ability of MoSDeF-GOMC combined with signac to automate a large number of independent calculations. The free energy data were produced via a single workflow, automating the process of generating and conducting all simulations and data analyses.[58] Analysis of raw free energy data from GOMC was performed with alchemlyb.[106,107,109] A total of 1380 total

Table 3. Adsorption Isotherm for CO_2 in IRMOF-1 at 298 K[a]

this work		Yazaydin et al.[92]			
		simulation		experiment	
P (bar)	molecules per unit cell	P (bar)	molecules per unit cell[b]	P (bar)	molecules per unit cell[b]
0.05	0.3123(3)	0.0490	0.3215	0.0524	0.1374
0.1	0.6256(2)	0.0988	0.6282	0.1046	0.4748
0.2	1.255(1)	0.1994	1.2110	0.2081	1.0577
0.3	1.8879(8)	0.2989	1.8552	0.3121	1.6711
0.4	2.526(3)	0.3990	2.4994	0.5489	2.9595
0.5	3.169(2)	0.4986	3.1742	0.7984	4.3398
0.6	3.814(4)	0.5992	3.8184		
0.7	4.466(2)	0.6983	4.4932		
0.8	5.118(4)	0.7993	5.1374		
0.9	5.779(5)	0.8984	5.8122		
1.0	6.442(7)	0.9990	6.4564		

[a]Standard deviations are represented by () in the last digit. [b]Data extracted from Figure S7 of Yazaydin et al.[92]

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

simulations were performed to produce the data shown in Figure 3. Two molecules (Ne and Rn), 6 temperatures, with 23

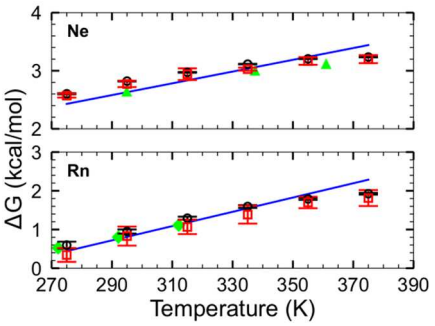


Figure 3. Hydration free energy for neon and radon: this work (red squares), Linnemann et al.[110] (black circles), NIST correlation[115] (blue line), experimental data for neon[113] (green triangles), and radon[114] (green diamonds).

simulations each: 1 NAMD pre-equilibration, 11 GOMC equilibration (one for each λ_i state), and 11 GOMC production free energy simulations. Five replicates were performed for each simulation. This workflow required only a few user-defined variables, including the identity of the molecule to be hydrated, the temperature, and the number of replicates. These free energy calculations could be easily extended to obtain free energies of hydration for any other molecule at any temperature. This workflow could also be modified to use other solvents or force fields.

Free energies of hydration for neon[110] and radon[111] in TIP4P-2005 water[112] from 275 to 375 K are shown in Figure 3, with numerical data listed in Table 4. Data from prior simulations,[110] experiments[113,114] and a correlation from NIST[115] are included for comparison. For neon and radon, the NIST correlation predicts linear behavior for ΔG_{HYD} over the entire temperature range, while both simulations and experiments show nonlinear behavior above 335 K. Small deviations between the two simulation data sets were observed for radon at temperatures below 335 K, although in most cases the data have overlapping error bars. This may be due to differences in how electrostatic interactions were calculated in the two studies. In this work, the Ewald summation was used, while Linnemann et al. used the reaction field method.[110] However, without having access to the simulation code, or workflow for the Linnemann et al. calculations, it is challenging to definitively identify the source of the discrepancies between the two data sets. This situation underscores the need for software like MoSDeF-GOMC, which encodes all simulation parameters (for example, force field parameters, potential cutoffs, long-range corrections for Lennard−Jones and electrostatic interactions, and Monte Carlo move types and ratios) within the workflow. When combined with open-source simulation software such as GOMC, reproducibility is guaranteed.

Vapor−Liquid Equilibria. A common use of GOMC is the calculation of vapor−liquid coexistence with the Gibbs Ensemble Monte Carlo (GEMC) method.[44] In GEMC, the simulation is comprised of two simulation boxes, where one box is typically the vapor phase, and the other is the liquid phase. With the use of specialized sampling algorithms, however, it is also possible to simulate liquid−liquid[116] and gas−solid equilibria.[117] Equilibrium is achieved through exchange of molecules and volume between each phase, while a variety of translations, rotations, and conformational sampling techniques are used to ensure thermal equilibrium. GEMC simulations produce a wealth of information, including saturated liquid and vapor densities, vapor pressure, heat of vaporization, and for mixtures, the composition of each phase.

In this example, vapor−liquid coexistence properties for a jet fuel surrogate for S8 are calculated with GEMC. S8 is a synthetic version of the jet fuel JP-8, which is a complex mixture of linear and branched hydrocarbons with 7−18 carbon atoms.[118] To make computations tractable, it is common to create a "surrogate" fuel with physical properties that are similar to those of the fuel of interest. Typically fuel

Table 4. Hydration Free Energies for Neon and Radon[a]

		ΔG_{HYD} (kcal/mol)			
molecule	T (K)	this work	Linnemann et al.[110]	NIST[115]	experiment[113,114]
Ne	275.0	2.57(3)	2.60(1)	2.430	
	294.9				2.648
	295.0	2.77(5)	2.822(7)	2.632	
	315.0	2.9(1)	2.972(6)	2.834	
	335.0	3.03(3)	3.114(5)	3.037	
	337.4				2.998
	355.0	3.17(7)	3.204(4)	3.239	
	361.0				3.113
	375.0	3.20(7)	3.236(4)	3.442	
Rn	272.15				0.524
	275.0	0.3(2)	0.60(9)	0.444	
	292.15				0.805
	295.0	0.8(2)	0.95(5)	0.812	
	312.15				1.112
	315.0	1.1(2)	1.29(4)	1.181	
	335.0	1.4(2)	1.59(4)	1.550	
	355.0	1.7(1)	1.79(2)	1.918	
	375.0	1.8(2)	1.92(2)	2.287	

[a]Uncertainties are represented by () in the last digit.

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

surrogates are composed of 4−10 compounds, including linear, branched and cyclic alkanes, and aromatics.[119] This particular surrogate for S8 is composed of *n*-dodecane (30.73 mol %), *n*-decane (42.34 mol %), iso-cetane (23.09 mol %), and iso-octane (3.84 mol %) .[120] The molecular structures for the molecules that compose the mixture are given in Figure 4.

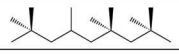
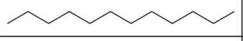
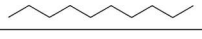
Name	Structure	Formula	Mass
iso−cetane		$C_{16}H_{34}$	226.45 u
iso−octane		C_8H_{18}	114.23 u
n−dodecane		$C_{12}H_{26}$	170.34 u
n−decane		$C_{10}H_{22}$	142.29 u

Figure 4. Chemical structures for molecules in S8-surrogate.

The convergence of a GEMC simulation is tied closely to the acceptance rate for the molecule swap move, which makes the simulation of S8-surrogate particularly challenging, given the size of the molecules and their highly branched topology. Several advanced configurational-bias algorithms have been developed, such as coupled-decoupled configurational-bias Monte Carlo,[61] which is used in GOMC, and the fragment/reservoir method used by Cassandra.[121] Additionally, GOMC has an identity exchange-like move[122] called Molecular Exchange Monte Carlo (MEMC)[123] that greatly increases acceptance rates for the transfer of large, bulky molecules, by exchanging them with a smaller molecule, or a combination of multiple smaller molecules. Without the MEMC move, extraordinarily long simulations would be required to converge even a low-quality solution for this system. While the setup of the MEMC move can be complex, MoSDeF-GOMC substantially simplifies the process for users.

The vapor−liquid coexistence curve and Clausius−Clapeyron plot for S8-surrogate, as predicted by the TraPPE-UA force field,[61,124] are shown in Figure 5 with numerical data listed in

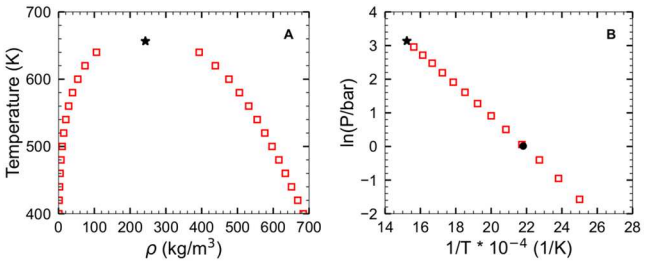


Figure 5. Vapor−liquid coexistence curve (A) and Claussius-Clapayron plot (B) for S8-surrogate. Predictions from simulation (red squares), calculated critical point (black star), and normal boiling point (black circle). Standard deviations are smaller than the symbols.

Table 5. The boiling and critical points are shown in Table 6. GEMC simulations predict a normal boiling point for the four component surrogate of 458.5 K, which compares favorably to the reported experimental boiling point of S8 of 454.15−454.85 K,[118] which is also closely reproduced by both the 4 and 7 component surrogates.[118,120] The compositions of each phase are provided in Table 7, with average standard deviations in the mole fractions for the liquid and vapor phases of 0.0006 and 0.0013, respectively. A snapshot from simulations performed at 640 K is given in Figure 6, which

Table 5. Saturated Liquid and Vapor Densities, Vapor Pressures and Compressibility Factors for S8-Surrogate[a]

T (K)	ρ_{liquid} (kg/m³)	ρ_{vapor} (kg/m³)	P (bar)	ΔH_v (kcal/mol)	Z
400	683.8(6)	0.92(3)	0.208(6)	6.82(3)	0.986(3)
420	668(1)	1.67(7)	0.39(2)	6.60(3)	0.977(3)
440	651.0(9)	2.8(1)	0.67(3)	6.33(5)	0.965(4)
460	633.7(8)	4.4(1)	1.06(3)	6.08(3)	0.954(5)
480	615.7(7)	6.7(2)	1.66(5)	5.79(2)	0.932(5)
500	596.6(7)	10.1(1)	2.49(3)	5.53(2)	0.906(3)
520	576(1)	14.4(3)	3.58(7)	5.21(2)	0.88(1)
540	555(1)	20.4(5)	5.00(9)	4.91(4)	0.85(1)
560	531(1)	28.3(6)	6.8(1)	4.53(4)	0.808(5)
580	505(2)	38.9(9)	8.9(2)	4.13(3)	0.76(1)
600	476(2)	54(1)	11.9(2)	3.67(3)	0.713(7)
620	438(2)	74(1)	15.1(4)	3.09(3)	0.65(2)
640	393(6)	106(6)	19.3(3)	2.38(7)	0.57(2)

[a]Standard deviations are represented by () in the last digit.

Table 6. S8-Surrogate Critical Parameters and Normal Boiling Point[a]

property	this work
T_c (K)	656.5(4)
ρ_c (kg/m³)	243(4)
P_c (bar)	23.0(4)
T_{bp} (K)	458.5(4)

[a]Standard deviations are represented by () in the last digit.

shows the localized aggregation of *n*-decane, while highly branched molecules, such as iso-cetane and iso-octane are dispersed throughout the simulation cell.

CONCLUSIONS

This work has introduced a new software, MoSDeF-GOMC,[47,48] which enables users to quickly and reproducibly create input files for the Monte Carlo software GOMC.[41,42] MoSDeF-GOMC is part of the MoSDeF ecosystem, allowing users to take advantage of all features of MoSDeF, perhaps the most notable of which is the ability to easily move between a variety of simulation engines. For example, for validation purposes, it is helpful to run single-point energy calculations across multiple simulation engines. With this extension to MoSDeF, a user can now create input and run the identical system (if supported) in the Monte Carlo engines GOMC and Cassandra,[36,39] as well as the molecular dynamics engines NAMD,[45,46] LAMMPS,[53,54] and GROMACS.[55]

When used with signac,[56,57] MoSDeF-GOMC allows users to automate large numbers of simulations and data post-processing with a single Python script. Computing free energies of hydration for neon and radon is a good example of this. The workflow for this example spans both chemical and thermodynamic space, uses two different simulation engines (NAMD and GOMC), initiates simulations in different ensembles *NPT* and *NVT*, and performs post-processing on the free energy data with alchemlyb.[106,107,109] In total, 1380 discrete simulations were launched from this workflow. Extending this workflow to cover a much larger chemical space can be done with minimal effort.

Because *all* simulation parameters are encoded in the workflow, any user that has the workflow, and corresponding force field XML file, will be able to run calculations previously

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

Table 7. S8-Surrogate Vapor-Liquid Mole Fractions[a]

	liquid				vapor			
T (K)	$x_{\text{iso-cetane}}$	$x_{\text{iso-octane}}$	$x_{n\text{-dodecane}}$	$x_{n\text{-decane}}$	$y_{\text{iso-cetane}}$	$y_{\text{iso-octane}}$	$y_{n\text{-dodecane}}$	$y_{n\text{-decane}}$
400	0.2642	0.0154	0.3333	0.3871	0.0388	0.1546	0.1555	0.6511
420	0.2653	0.0166	0.3323	0.3858	0.0474	0.1389	0.1721	0.6416
440	0.2614	0.0194	0.3281	0.3911	0.0534	0.1322	0.1845	0.6299
460	0.2648	0.0198	0.3284	0.3870	0.0631	0.1154	0.2014	0.6201
480	0.2636	0.0213	0.3259	0.3892	0.0702	0.1072	0.214	0.6086
500	0.2593	0.0236	0.3226	0.3945	0.0770	0.1023	0.2227	0.5980
520	0.2603	0.0242	0.3216	0.3939	0.0865	0.0927	0.2355	0.5853
540	0.2586	0.0256	0.3197	0.3961	0.0958	0.0857	0.2450	0.5735
560	0.2558	0.0272	0.3175	0.3995	0.1056	0.0792	0.2543	0.5609
580	0.2555	0.0280	0.3163	0.4002	0.1178	0.0718	0.2644	0.5460
600	0.2514	0.0296	0.3141	0.4049	0.1297	0.0663	0.2722	0.5318
620	0.2460	0.0315	0.3116	0.4109	0.1428	0.0608	0.2800	0.5164
620	0.2448	0.0322	0.3106	0.4124	0.1650	0.0526	0.2901	0.4923

[a]Average standard deviations in the mole fractions for the liquid and vapor phases are 0.0006 and 0.0013, respectively.

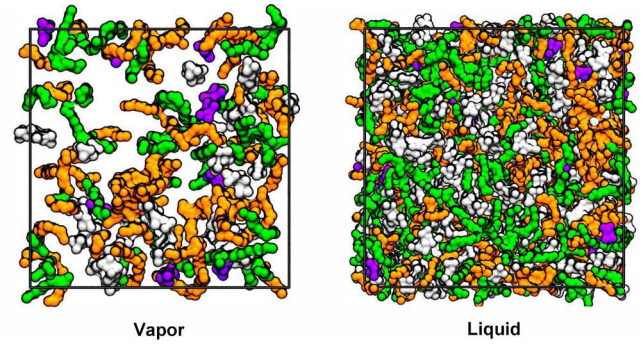


Figure 6. Snapshot from GEMC simulations for S8-surrogate at 640 K. Iso-cetane (white/silver), iso-octane (purple), *n*-dodecane (green), and *n*-decane (orange).

set up with MoSDeF-GOMC and obtain results that are in statistical agreement with published calculations. Given the difficulty experienced by many scientists in reproducing the computational work of others,[125] MoSDeF-GOMC represents a significant advance in the area of reproducible computational science, and embodies the transparent, reproducible, usable by others, and extensible (TRUE) principle advocated for by Cummings and co-workers.[69]

■ ASSOCIATED CONTENT

Data Availability Statement

The open-source signac workflows that use MoSDeF-GOMC are provided via GitHub at https://github.com/GOMC-WSU/Publications/tree/main/2023/Crawford_1. These workflows contain all mol2 files, SMILES strings, force field XML files, and NAMD template control files. These workflows contain all of the information required to perform the simulations presented in this work, including parameters for the automated system building, energy minimization, system equilibration and production runs, and the final data analysis. The open-source MoSDeF-GOMC code can be obtained via GitHub at https://github.com/GOMC-WSU/MoSDeF-GOMC.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01498.

Additional simulation details, parameters, and methods necessary for replication of simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jeffrey J. Potoff − *Department of Chemical Engineering, Wayne State University, Detroit, Michigan 48202-4050, United States;* orcid.org/0000-0002-4421-8787; Phone: +1-313-577-9357; Email: jpotoff@wayne.edu

Authors

Brad Crawford − *Department of Chemical Engineering, Wayne State University, Detroit, Michigan 48202-4050, United States;* orcid.org/0000-0003-0638-7333

Umesh Timalsina − *Institute for Software Integrated Systems (ISIS), Vanderbilt University, Nashville, Tennessee 37212, United States*

Co D. Quach − *Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235-1604, United States; Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37212, United States*

Nicholas C. Craven − *Interdisciplinary Material Science Program, Vanderbilt University, Nashville, Tennessee 37235-0106, United States; Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37212, United States*

Justin B. Gilmer − *Interdisciplinary Material Science Program, Vanderbilt University, Nashville, Tennessee 37235-0106, United States; Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37212, United States*

Clare McCabe − *Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235-1604, United States; Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37212, United States;* orcid.org/0000-0002-8552-9135

Peter T. Cummings − *Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235-1604, United States; Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37212, United States*

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.2c01498

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation (grants OAC-1835713, OAC-1835874 and CBET 2052438). Some of the computational resources used in this work were provided by the Wayne State University Grid computing environment.

■ REFERENCES

(1) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **2012**, *33*, 189−202.
(2) Malde, A.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P.; Oostenbrink, C.; Mark, A. An Automated Force Field Topology Builder (ATB) and Repository Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026−4037.
(3) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: Prediction of hydration free enthalpies. *J. Comput. Aid Mol. Des* **2014**, *28*, 221−233.
(4) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834−5845.
(5) Kota, P. GUIMACS − A Java Based Front End for GROMACS. *In Silico Biol.* **2007**, *7*, 95−99.
(6) Sellis, D.; Vlachakis, D.; Vlassi, M. Gromita: A Fully Integrated Graphical User Interface to Gromacs 4. *Bioinform. Biol.* **2009**, *3*, 99−102.
(7) Roopra, S.; Knapp, B.; Omasits, U.; Schreiner, W. jSimMacs for GROMACS: A Java Application for Advanced Molecular Dynamics Simulations with Remote Access Capability. *J. Chem. Inf. Model.* **2009**, *49*, 2412−2417.
(8) Miller, B. T.; Singh, R. P.; Klauda, J. B.; Hodošček, M.; Brooks, B. R.; Woodcock, H. L. CHARMMing: A New, Flexible Web Portal for CHARMM. *J. Chem. Inf. Model.* **2008**, *48*, 1920−1929.
(9) Hospital, A.; Andrio, P.; Fenollosa, C.; Cicin-Sain, D.; Orozco, M.; Gelpí, J. MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics* **2012**, *28*, 1278−1279.
(10) Ribeiro, J.; Bernardi, R.; Rudack, T.; Stone, J.; Phillips, J.; Freddolino, P.; Schulten, K. QwikMD-Integrative Molecular Dynamics Toolkit for Novices and Experts. *Nature* **2016**, *6*, 26536−26536.
(11) Fortunato, M. E.; Colina, C. M. pysimm: A python package for simulation of molecular systems. *SoftwareX* **2017**, *6*, 7−12.
(12) Jewett, A. I.; Stelter, D.; Lambert, J.; Saladi, S. M.; Roscioni, O. M.; Ricci, M.; Autin, L.; Maritan, M.; Bashusqeh, S. M.; Keyes, T.; Dame, R. T.; Shea, J. E.; Jensen, G. J.; Goodsell, D. S. Moltemplate: A Tool for Coarse-Grained Modeling of Complex Biological Matter and Soft Condensed Matter Physics. *J. Mol. Biol.* **2021**, *433*, 166841.
(13) McCoy, M. D.; Shivakumar, V.; Nimmagadda, S.; Jafri, M. S.; Madhavan, S. SNP2SIM: a modular workflow for standardizing molecular simulation and functional analysis of protein variants. *BMC Bioinform.* **2019**, *20*, 1471−2105.
(14) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(15) Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S. J.; Patel, D. S.; Beaven, A. H.; Lee, K. I.; Rui, H.; Park, S.; Lee, H. S.; Roux, B.; MacKerell, A. D.; Klauda, J. B.; Qi, Y.; Im, W. CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.* **2017**, *38*, 1114−1124.
(16) Homeyer, N.; Gohlke, H. FEW: A workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.* **2013**, *34*, 965−973.
(17) Gapsys, V.; Michielssens, S.; Seeliger, D.; De Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015**, *36*, 348−354.
(18) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485−2490.
(19) Jespers, W.; Esguerra, M.; Aqvist, J.; Gutierrez-de-Teran, H. QligFEP: an automated workflow for small molecule free energy calculations in Q. *J. Cheminform.* **2019**, *11*, 2485−2490.
(20) Petrov, D. Perturbation Free-Energy Toolkit: An Automated Alchemical Topology Builder. *J. Chem. Inf. Model.* **2021**, *61*, 4382−4390.
(21) Carvalho Martins, L.; Cino, E. A.; Ferreira, R. S. PyAutoFEP: An Automated Free Energy Perturbation Workflow for GROMACS Integrating Enhanced Sampling Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4262−4273.
(22) Bayarri, G.; Andrio, P.; Hospital, A.; Orozco, M.; Gelpí, J. BioExcel Building Blocks Workflows (BioBB-Wfs), an integrated web-based platform for biomolecular simulations. *Nucleic Acids Res.* **2022**, *50*, W99−W107.
(23) Atwi, R.; Bliss, M.; Makeev, M.; Rajput, N. MISPR: an open-source package for high-throughput multiscale molecular simulations. *Sci. Rep.* **2022**, *12*, 12.
(24) Shao, Q.; Jiang, Y.; Yang, Z. J. EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. *J. Chem. Inf. Model.* **2022**, *62*, 647−655.
(25) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830−7843.
(26) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671−690.
(27) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162−3180.
(28) Huang, J.; Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(29) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(30) Lee, J.; Hitzenberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W. CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* **2020**, *153*, 035103.
(31) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.
(32) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(33) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(34) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(35) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405−413.
(36) DeFever, R. S.; Matsumoto, R. A.; Dowling, A. W.; Cummings, P. T.; Maginn, E. J. MoSDeF Cassandra: A complete Python interface for the Cassandra Monte Carlo software. *J. Comput. Chem.* **2021**, *42*, 1321−1331.
(37) Martin, M. G. MCCCS Towhee: A tool for monte carlo molecular simulation. *Mol. Simul.* **2013**, *39*, 1212−1222.
(38) 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.
(39) 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.
(40) Hens, R.; Rahbari, A.; Caro-Ortiz, S.; Dawass, N.; Erdös, M.; Poursaeidesfahani, A.; Salehi, H. S.; Celebi, A. T.; Ramdin, M.; Moultos, O. A.; Dubbeldam, D.; Vlugt, T. J. Brick-CFCMC: Open Source Software for Monte Carlo Simulations of Phase and Reaction Equilibria Using the Continuous Fractional Component Method. *J. Chem. Inf. Model.* **2020**, *60*, 2678−2682.
(41) Nejahi, Y.; Soroush Barhaghi, M.; Mick, J.; Jackman, B.; Rushaidat, K.; Li, Y.; Schwiebert, L.; Potoff, J. GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids. *SoftwareX* **2019**, *9*, 20−27.
(42) Nejahi, Y.; Barhaghi, M. S.; Schwing, G.; Schwiebert, L.; Potoff, J. Update 2.70 to "GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids. *SoftwareX* **2021**, *13*, 100627.
(43) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(44) Panagiotopoulos, A. Z.; Quirke, N.; Stapleton, M.; Tildesley, D. J. Phase equilibria by simulation in the Gibbs ensemble alternative derivation, generalization and application to mixture and membrane equilibria. *Mol. Phys.* **1988**, *63*, 527−545.
(45) 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.
(46) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
(47) Crawford, B.; Timalsina, U.; Quach, C. D.; Craven, N.; Gilmer, J.; Cummings, P. T.; Potoff, J. MoSDeF-GOMC: Python software for the creation of scientific workflows for the Monte Carlo simulation engine GOMC. 2022; https://github.com/GOMC-WSU/MoSDeF-GOMC (accessed 2022−10−22).
(48) Crawford, B.; Timalsina, U.; Quach, C. D.; Craven, N.; Gilmer, J. B.; Cummings, P. T.; Potoff, J. MoSDeF-GOMC: Python software for the creation of scientific workflows for the Monte Carlo simulation engine GOMC. **2022** DOI: 10.5281/zenodo.7217632 (accessed 2022−10−22).
(49) Summers, A. Z.; Gilmer, J. B.; Iacovella, C. R.; Cummings, P. T.; McCabe, C. MoSDeF, a Python Framework Enabling Large-Scale Computational Screening of Soft Matter: Application to Chemistry-Property Relationships in Lubricating Monolayer Films. *J. Chem. Theory Comput.* **2020**, *16*, 1779−1793.
(50) Cummings, P.; McCabe, C.; Iacovella, C.; Ledeczi, A.; Jankowski, E.; Jayaraman, A.; Palmer, J.; Maginn, E.; Glotzer, S.; Anderson, J.; Siepmann, J. I.; Potoff, J.; Matsumoto, R.; Gilmer, J.; DeFever, R.; Singh, R.; Crawford, B. Open-Source Molecular Modeling Software in Chemical Engineering, with Focus on the Molecular Simulation Design Framework (MoSDeF). *AIChE J.* **2021**, *67* (3), No. e17206.
(51) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* **2020**, *173*, 109363.
(52) Butler, B. L.; Ramasubramani, V.; Anderson, J. A.; Glotzer, S. C. HOOMD-blue version 3.0 A Modern, Extensible, Flexible, Object-Oriented API for Molecular Simulations. *Proceedings of the 19th Python in Science Conference*. 2020; pp 24−31.
(53) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(54) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
(55) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(56) Ramasubramani, V.; Adorf, C.; Dodd, P.; Dice, B.; Glotzer, S. signac: A Python framework for data and workflow management. In *Proceedings of the 17th Python in Science Conference*. 2018, 152−159.
(57) Adorf, C.; Dodd, P.; Ramasubramani, V.; Glotzer, S. Simple data and workflow management with the signac framework. *Comput. Mater. Sci.* **2018**, *146* (C), 220−229.
(58) Crawford, B.; Potoff, J. signac workflows for MoSDeF-GOMC. 2023; https://github.com/GOMC-WSU/Publications/tree/main/2023/Crawford_1 (accessed 2023−01−03).
(59) Jones, C.; Fothergill, J. mBuild polymer builder. **2022**; DOI: 10.5281/zenodo.7348318 (accessed 2022−10−22).
(60) Chen, B.; Siepmann, J. I. Novel Monte Carlo algorithm for simulating strongly associating fluids: Applications to water, hydrogen fluoride, and acetic acid. *J. Phys. Chem. B* **2000**, *104*, 8725−8734.
(61) 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.
(62) Siepmann, J. I.; Frenkel, D. Conflgurational bias monte carlo: A new sampling scheme for flexible chains. *Mol. Phys.* **1992**, *75*, 59−70.
(63) Baumgärtner, A.; Binder, K. Monte Carlo studies on the freely jointed polymer chain with excluded volume interaction. *J. Chem. Phys.* **1979**, *71*, 2541−2545.
(64) Klein, C.; Sallai, J.; Jones, T. J.; Iacovella, C. R.; McCabe, C.; Cummings, P. T. In *Foundations of Molecular Modeling and Simulation*; Snurr, R. Q., Adjiman, C. S., Kofke, D. A., Eds.; Molecular Modeling and Simulation: Applications and Perspectives; Springer Singapore: Singapore, 2016; pp 79−92.
(65) Klein, C.; Summers, A. Z.; Thompson, M. W.; Gilmer, J. B.; McCabe, C.; Cummings, P. T.; Sallai, J.; Iacovella, C. R. Formalizing atom-typing and the dissemination of force fields with foyer. *Comput. Mater. Sci.* **2019**, *167*, 215−227.
(66) GMSO: General Molecular Simulation Object. 2019; https://github.com/mosdef-hub/gmso (accessed 2022−10−22).
(67) GMSO: General Molecular Simulation Object. 2019; https://gmso.mosdef.org/en/stable/ (accessed 2022−10−22).
(68) forcefield-utilities. 2022; https://github.com/mosdef-hub/forcefield-utilities (accessed 2022−10−22).

(69) Thompson, M. W.; Gilmer, J. B.; Matsumoto, R. A.; Quach, C. D.; Shamaprasad, P.; Yang, A. H.; Iacovella, C. R.; McCabe, C.; Cummings, P. T. Towards molecular simulations that are transparent, reproducible, usable by others, and extensible (TRUE). *Mol. Phys.* **2020**, *118*, No. e1742938.
(70) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(71) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.
(72) Weininger, D. SMILES. 3. DEPICT. Graphical depiction of chemical structures. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 237−243.
(73) Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M. Packmol: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157−2164.
(74) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aid Mol. Des* **2017**, *31*, 147−161.
(75) Mie, G. Zur kinetischen Theorie der einatomigen Körper. *Ann. Phys.* **1903**, *316*, 657−697.
(76) Crawford, B.; Potoff, J. py-MCMD: A Python Library for Performing Hybrid Monte Carlo-Molecular Dynamics Simulations with GOMC and NAMD. 2021; https://github.com/GOMC-WSU/py-MCMD (accessed 2022−08−15).
(77) Crawford, B.; Potoff, J. py-MCMD: Initial Release. **2021**; DOI: 10.5281/zenodo.5142483 (accessed 2022−08−15).
(78) Barhaghi, M. S.; Crawford, B.; Schwing, G.; Hardy, D. J.; Stone, J. E.; Schwiebert, L.; Potoff, J.; Tajkhorshid, E. py-MCMD: Python Software for Performing Hybrid Monte Carlo/Molecular Dynamics Simulations with GOMC and NAMD. *J. Chem. Theory Comput.* **2022**, *18*, 4983−4994.
(79) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Thermodynamic analysis of Xe/Kr selectivity in over 137000 hypothetical metal-organic frameworks. *Chemical Science* **2012**, *3*, 2217−2223.
(80) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal-organic framework with optimally selective xenon adsorption and separation. *Nat. Commun.* **2016**, *7*, ncomms11831.
(81) Vlugt, T. J.; Zhu, W.; Kapteijn, F.; Moulijn, J. A.; Smit, B.; Krishna, R. Adsorption of linear and branched alkanes in the zeolite silicalite-1. *J. Am. Chem. Soc.* **1998**, *120*, 5599−5600.
(82) Kim, J.; Lin, L. C.; Martin, R. L.; Swisher, J. A.; Haranczyk, M.; Smit, B. Large-scale computational screening of zeolites for ethane/ethene separation. *Langmuir* **2012**, *28*, 11914−11919.
(83) Sun, W.; Lin, L.-C.; Peng, X.; Smit, B. Computational screening of porous metal-organic frameworks and zeolites for the removal of SO2 and NOx from flue gases. *AIChE J.* **2014**, *60*, 2314−2323.
(84) Hamon, L.; Leclerc, H.; Ghoufi, A.; Oliviero, L.; Travert, A.; Lavalley, J. C.; Devic, T.; Serre, C.; Férey, G.; De Weireld, G.; Vimont, A.; Maurin, G. Molecular insight into the adsorption of H_2S in the flexible MIL-53(Cr) and rigid MIL-47(V) MOFs: Infrared spectroscopy combined to molecular simulations. *J. Phys. Chem. C* **2011**, *115*, 2047−2056.
(85) Yazaydin, A. Ö.; Snurr, R. Q.; Park, T. H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. Screening of metal-organic frameworks for carbon dioxide capture from flue gas using a combined experimental and modeling approach. *J. Am. Chem. Soc.* **2009**, *131*, 18198−18199.
(86) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341*, 1230444.
(87) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* **2012**, *4*, 83−89.
(88) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal- organic framework. *Nature* **1999**, *402*, 276−279.
(89) Potoff, J.; Siepmann, J. Vapor−liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676−1682.
(90) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(91) Rappe, A.; Casewit, C.; Colwell, K.; Goddard, W. I.; Skiff, W. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(92) Yazaydin, A. Ö.; Snurr, R. Q.; Park, T. H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. Screening of metal-organic frameworks for carbon dioxide capture from flue gas using a combined experimental and modeling approach. *J. Am. Chem. Soc.* **2009**, *131*, 18198−18199.
(93) Hansch, C.; Fujita, T. ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(94) Leo, A.; Hansch, C.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71*, 525−616.
(95) King, E.; Aitchison, E.; Li, H.; Luo, R. Recent Developments in Free Energy Calculations for Drug Discovery. *Front. Mol. Biosci.* **2021**, *8*. DOI: 10.3389/fmolb.2021.712085
(96) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of "Alchemical Perturbation" in Medicinal Chemistry. *J. Med. Chem.* **2018**, *61*, 638−649.
(97) Shirts, M. R.; Mobley, D. L.; Brown, S. P. In *Drug Design: Structure- and Ligand-Based Approaches*; Merz, K. M., Jr, Ringe, D., Reynolds, C. H., Eds.; Cambridge University Press: Cambridge, U.K., 2010; pp 61−86.
(98) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.
(99) Greenblatt, J. B.; Miller, D. J.; Ager, J. W.; Houle, F. A.; Sharp, I. D. *The Technical and Energetic Challenges of Separating (Photo) Electrochemical Carbon Dioxide Reduction Products*. 2018.
(100) Wang, L.; Violet, C.; DuChanois, R. M.; Elimelech, M. Derivation of the Theoretical Minimum Energy of Separation of Desalination Processes. *J. Chem. Educ.* **2020**, *97*, 4361−4369.
(101) Arp, H. P. H.; Niederer, C.; Goss, K. U. Predicting the partitioning behavior of various highly fluorinated compounds. *Environ. Sci. Technol.* **2006**, *40*, 7298−7304.
(102) Bhatnagar, N.; Kamath, G.; Potoff, J. J. Prediction of 1-octanol-water and air-water partition coefficients for nitro-aromatic compounds from molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6467−6474.
(103) Brown, T. N.; Wania, F. Development and Exploration of an Organic Contaminant Fate Model Using Poly-Parameter Linear Free Energy Relationships. *Environ. Sci. Technol.* **2009**, *43*, 6676−6683.
(104) Khawar, M. I.; Nabi, D. Relook on the Linear Free Energy Relationships Describing the Partitioning Behavior of Diverse Chemicals for Polyethylene Water Passive Samplers. *ACS Omega* **2021**, *6*, 5221−5232.
(105) Lundborg, M.; Lindahl, E. Automatic GROMACS Topology Generation and Comparisons of Force Fields for Solvation Free Energy Calculations. *J. Phys. Chem. B* **2015**, *119*, 810−823.
(106) Shirts, M.; Chodera, J. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(107) Chodera, J. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12* (4), 1799−1805.
(108) Bennett, C. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22* (2), 245−268.
(109) Dotson, D.; Beckstein, O.; Wille, D.; Wu, Z.; Kenney, I.; Lee, H.; Lim, V.; Schlaich, A.; Hénin, J.; Barhaghi, M. S.; Joseph, T.; Hsu,

W.-T. alchemistry/alchemlyb. 2021; https://github.com/alchemistry/alchemlyb (accessed 2022−10−22).
(110) Linnemann, M.; Nikolaychuk, P. A.; Muñoz-Muñoz, Y. M.; Baumhögger, E.; Vrabec, J. Henry's Law Constant of Noble Gases in Water, Methanol, Ethanol, and Isopropanol by Experiment and Molecular Simulation. *J. Chem. Eng. Data.* **2020**, *65*, 1180−1188.
(111) Mick, J. R.; Soroush Barhaghi, M.; Potoff, J. J. Prediction of Radon-222 Phase Behavior by Monte Carlo Simulation. *J. Chem. Eng. Data.* **2016**, *61*, 1625−1631.
(112) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(113) Crovetto, R.; Fernández-Prini, R.; Japas, M. L. Solubilities of inert gases and methane in H_2O and in D_2O in the temperature range of 300 to 600 K. *J. Chem. Phys.* **1982**, *76*, 1077−1086.
(114) Lewis, C.; Hopke, P. K.; Stukel, J. J. Solubility of radon in selected perfluorocarbon compounds and water. *Ind. Eng. Chem. Res.* **1987**, *26*, 356−359.
(115) *NIST Chemistry WebBook*. 2021; https://webbook.nist.gov/chemistry/name-ser/ (accessed 2022−07−01).
(116) Zhang, L.; Siepmann, J. I. Pressure dependence of the vapor-liquid-liquid phase behavior in ternary mixtures consisting of n-alkanes, n-perfluoroalkanes, and carbon dioxide. *J. Phys. Chem. B* **2005**, *109*, 2911−2919.
(117) Chen, B.; Siepmann, J. I.; Klein, M. L. Direct Gibbs ensemble Monte Carlo simulations for solid-vapor phase equilibria: Applications to Lennard-Jonesium and carbon dioxide. *J. Chem. Phys. B* **2001**, *105*, 9840−9848.
(118) Bruno, T. J.; Smith, B. L. Improvements in the measurement of distillation curves. 2. Application to aerospace/aviation fuels RP-1 and S-8. *Ind. Eng. Chem. Res.* **2006**, *45*, 4381−4388.
(119) Dooley, S.; Won, S. H.; Chaos, M.; Heyne, J.; Ju, Y.; Dryer, F. L.; Kumar, K.; Sung, C. J.; Wang, H.; Oehlschlaeger, M. A.; Santoro, R. J.; Litzinger, T. A. A jet fuel surrogate formulated by real fuel properties. *Combust. Flame* **2010**, *157*, 2333−2339.
(120) Kim, D.; Martz, J.; Abdul-Nour, A.; Yu, X.; Jansons, M.; Violi, A. A six-component surrogate for emulating the physical and chemical characteristics of conventional and alternative jet fuels and their blends. *Combust. Flame* **2017**, *179*, 86−94.
(121) 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.
(122) Martin, M. G.; Siepmann, J. I. Predicting Multicomponent Phase Equilibria and Free Energies of Transfer for Alkanes by Molecular Simulation. *J. Am. Chem. Soc.* **1997**, *119*, 8921−8924.
(123) Soroush Barhaghi, M.; Potoff, J. J. Prediction of phase equilibria and Gibbs free energies of transfer using molecular exchange Monte Carlo in the Gibbs ensemble. *Fluid Phase Equilib.* **2019**, *486*, 106−118.
(124) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(125) Barba, L. A. The hard road to reproducibility. *Science* **2016**, *354*, 142.

https://doi.org/10.1021/acs.jcim.2c01498
J. Chem. Inf. Model. 2023, 63, 1218−1228