JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

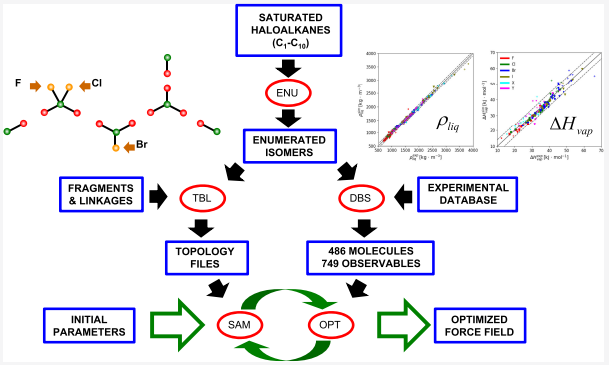
Systematic Optimization of a Fragment-Based Force Field against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Saturated Haloalkanes

Marina P. Oliveira, Maurice Andrey, Salomé R. Rieder, Leyla Kern, David F. Hahn, Sereina Riniker, Bruno A. C. Horta, and Philippe H. Hünenberger*

Cite This: J. Chem. Theory Comput. 2020, 16, 7525−7555 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Direct optimization against experimental condensed-phase properties concerning small organic molecules still represents the most reliable way to calibrate the empirical parameters of a force field. However, compared to a corresponding calibration against quantum-mechanical (QM) calculations concerning isolated molecules, this approach is typically very tedious and time-consuming. The present article describes an integrated scheme for the automated refinement of force-field parameters against experimental condensed-phase data, considering entire classes of organic molecules constructed using a fragment library *via* combinatorial isomer enumeration. The main steps of the scheme, referred to as CombiFF, are as follows: (i) definition of a molecule family; (ii) combinatorial enumeration of all isomers; (iii) query for experimental data; (iv) automatic construction of the molecular topologies by fragment assembly; and (v) iterative refinement of the force-field parameters considering the entire family. As a first application, CombiFF is used here to design a GROMOS-compatible united-atom force field for the saturated acyclic haloalkane family. This force field relies on an electronegativity-equalization scheme for the atomic partial charges and involves no specific terms for σ-holes and halogen bonding. A total of 749 experimental liquid densities ρ_{liq} and vaporization enthalpies ΔH_{vap} concerning 486 haloalkanes are considered for calibration and validation. The resulting root-mean-square deviations from experiment are 49.8 (27.6) kg·m^{-3} for ρ_{liq} and 2.7 (1.8) kJ·mol^{-1} for ΔH_{vap} for the calibration (validation) set. The values are lower for the validation set which contains larger molecules (stronger influence of purely aliphatic interactions). The trends in the optimized parameters along the halogen series and across the compound family are in line with chemical intuition based on considerations related to size, polarizability, softness, electronegativity, induction, and hyperconjugation. This observation is particularly remarkable considering that the force-field calibration did not involve any QM calculation. Once the time-consuming task of target-data selection/curation has been performed, the optimization of a force field only takes a few days. As a result, CombiFF enables an easy assessment of the consequences of functional-form decisions on the accuracy of a force field at an optimal level of parametrization.



1. INTRODUCTION

Classical atomistic simulation[1−3] and, in particular, molecular dynamics[4−8] (MD) has become an established tool complementary to experiment for the investigation of condensed-phase systems relevant in physics, chemistry, and biology. The usefulness of this method results in particular from a favorable trade-off between model resolution and computational cost. Although classical atomistic models represent an approximation to quantum mechanics (QM), they provide a realistic description of atom-based systems at spatial and temporal resolutions on the orders of 0.1 nm and 1 fs, respectively, while their computational cost remains tractable at present for system sizes and time scales on the orders of 10 nm and 1 μs, respectively. The ability of classical MD simulations to represent accurately the properties of a given condensed-phase system depends crucially on the quality of the underlying potential-energy function or force field.[9−15]

The past few years have witnessed an intense renewal of interest for force-field refinement procedures. This is in large part due to a number of computational and methodological advances[16−21] that have enabled an increased accuracy, automation, and throughput of these procedures, including

Received: July 1, 2020
Published: November 24, 2020

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

more powerful processors and coprocessors (graphics processing units (GPUs)[22−27] and field-programmable gate arrays (FPGAs)[28−30]), faster algorithms and improved simulation methodologies (*e.g.*, for nonbonded interactions[31−41] and free-energy calculations[42−49]), more accessible experimental data[50−58] (online literature and databases), and faster QM calculation codes.[59−66] In turn, these progresses have increased the practical usefulness of force-field-based simulations, in particular in the context of material and drug design, where approaches relying on atomistic MD simulations can nowadays keep pace with the fast time scale of industrial and pharmaceutical product development.[45−47,67−75]

One may tentatively distinguish three main strategies in the design of condensed-phase force fields.

1. In fragment-based force fields (FBFFs),[76−80] pioneered in the 1980s, the covalent and nonbonded (NB) interaction parameters, the latter including Lennard-Jones[81] (LJ) coefficients and atomic partial charges (PCs), are specified simultaneously within molecular fragments representative of the relevant chemical functional groups, or the relevant monomers in the case of (bio-)polymers. Molecules are constructed by assembling these fragments or monomers, and invoking an assumption of transferability. The latter suggests that the properties of a given fragment are essentially independent of its covalent surroundings within a specific molecule as well as of the medium surrounding this molecule. The NB parametrization is then performed primarily by fitting against experimental thermodynamic data for small organic compounds, typically considering pure-liquid, solvation, and/or crystal properties.
2. In hybrid force fields (HYFFs),[82,83] introduced in the 1990s, the covalent parameters and LJ coefficients are selected in an FBFF fashion, but the PCs are derived based on a QM calculation involving the target molecule in a given environment (implicit solvent in the QM calculation), typically *via* electrostatic-potential fitting[82−96] or electron-density partitioning.[97−106] The assumption invoked in the HYFF approach is that of a separability between the LJ coefficients and the PCs. Given a selected derivation scheme for the latter, the calibration of the LJ coefficients is performed essentially as in the FBFF case, *i.e.*, against experimental condensed-phase data for small organic compounds.
3. In QM-derived force fields (QDFFs),[107−115] which represent the most recent category, the covalent and NB parameters (LJ coefficients and PCs) are determined simultaneously based on a QM calculation involving the target molecule in a given environment. The derivation schemes for the PCs are the same as in the HYFF case. For the LJ coefficients, the derivation typically involves electron-density partitioning as a starting point.[116−126] The assumption invoked in the QDFF approach is that of a compatibility between single-molecule and condensed-phase NB parameters. Although the fitting of classical force fields against QM data has existed for a long time,[127−131] only recently has the QDFF approach become really competitive for condensed-phase systems.

The HYFF and QDFF schemes are popular nowadays, in particular because they (i) benefit from fast QM calculation methods;[59−66] (ii) promise an exhaustive coverage of the chemical space; (iii) take into account induction effects on the PCs[92,101,132] as well as, possibly, on the LJ coefficients;[112,113,120−122,125,126] and (iv) are easier to automate in terms of topology construction and parameter derivation. The last point is of particular importance if large collections of molecules are to be considered in the calibration and/or application of the force field.

Automation can be applied at three main levels.

1. Automated topology building, which aims at assigning the force-field parameters relevant for a given molecular system based on the tables of a given force-field parameter set (FBFF and HYFF) and/or the results of QM calculations (HYFF and QDFF). Yet another route is to rely on the direct perception of chemical patterns *via* automated substructure search, as done in the OpenFF force field.[133−135] The procedures employed are too diverse and force-field-specific to be listed exhaustively; see, *e.g.*, Antechamber[136] for AMBER/GAFF, ParamChem[137,138] and FFParam[139] for CHARMM/CGenFF, ATB[140−142] for GROMOS, and the SMARTS + SMIRKS perception languages[135] for OpenFF (see also refs 143−149 for other examples). These approaches include in particular automatic typing,[135−137,150−153] torsional scanning,[113,154−160] energy/force/Hessian fitting or matching,[108,161−164] PC derivation schemes[82,83,86−105] (possibly with charge-group partitioning[165,166]), and LJ derivation schemes.[106,112,113,115−126,167−172]
2. Automated force-field optimization, which aims at calibrating the force-field parameters. The target of the refinement is to reproduce experimental condensed-phase observables for a large collection of molecules representative of the chemical space to be covered. This automatic optimization has a long history in the context of target QM data[127−131] (see also refs 109 and 173−177 for more recent work). However, to date, the refinement against experimental data has mainly relied on manual procedures, rendering the task of force-field development extremely laborious and time-consuming. Recent attempts at automating the fitting against condensed-phase observables include in particular the POP scheme[178,179] and the ForceBalance scheme[115,174−176,180−184] (see also refs 185−190 for related semiautomated approaches). However, until now, these have only been applied to parameter optimization for one single compound as opposed to a collection thereof.
3. Automated protocols and workflows for the simulation setup, property calculation, and data analysis, which aim at facilitating and standardizing the calculation procedures. Here also, the approaches are too diverse to be listed exhaustively. They include in particular setup/simulation/analysis workflows,[73,191−197] as well as experimental databases and data-extraction procedures.[50−58]

Owing to advances in automation, it has become feasible to run high-throughput calibration and/or validation protocols[39,40,195,198−205] considering collections of molecules in the hundreds (thousands) and given a computational investment of only a few days (weeks).

In spite of their current popularity, the HYFF and QDFF approaches also present some major shortcomings, namely (i)

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

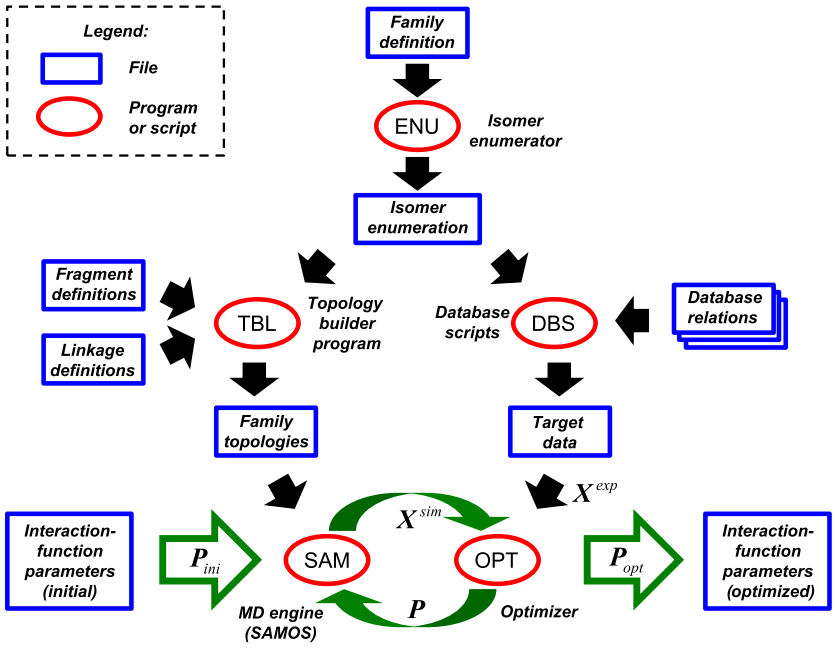


Figure 1. CombiFF workflow for the calibration of a force field based on a given compound family. See Section 2 for explanations and Appendix A for details.

the requirement to specify a reference structure (molecular conformation) and environment (*e.g.*, vacuum or continuum solvent) in the QM calculations leads to an implicit dependence of the topological information (*e.g.*, LJ coefficients and PCs) on configurational information; (ii) the derivation of the PCs and, possibly, LJ coefficients follows recipes rather than physics-based rules, because these parameters are not QM observables and because their calculated values also depend on the choice of a QM level of theory and basis set;[89−93,206−212] and (iii) in practice, some parameters must still be optimized empirically[89,94,95,108,115,211,212] (*e.g.*, LJ coefficients in HYFF, LJ repulsion coefficient[106,112,113,125,126,213,214] in QDFF) and consistently with a specific QM-derivation recipe as well as with other functional-form choices[215] (*e.g.*, treatment of the long-range NB interactions and choice of combination rules).

In contrast, the FBFF scheme acknowledges fully that the NB parameters are effective parameters, all to be determined empirically rather being partly (HYFF) or (nearly) entirely (QDFF) defined using QM-based recipes. These parameters are related to collective condensed-phase properties and correlated with other functional-form and methodological choices made in the force-field design. As a result, the connection between QM-inferred single-molecule properties and effective condensed-phase NB parameters may in fact be rather weak. For this reason, although the substitution of an explicit optimization of the NB parameters against experimental condensed-phase data by the application of QM-based recipes represents a great simplification (potential for automation, fewer explicit parameters, exhaustive chemical-space coverage), it may ultimately place a hard limit on the achievable accuracy.

The goal of the approach presented in this work, referred to as CombiFF, is the design of an integrated scheme for the automated refinement of force-field parameters against experimental condensed-phase data, considering entire classes of organic molecules constructed using a fragment library *via* combinatorial isomer enumeration. The CombiFF scheme is rooted in the FBFF approach and aims at overcoming its major shortcomings. More specifically, it is designed to achieve (i) a very broad (though not exhaustive) coverage of the chemical space; (ii) an appropriate representation of induction effects; and (iii) a complete automation of the topology construction and parameter optimization. In this article, the principles and workflow of the CombiFF approach are described, and its application is illustrated by the design of a GROMOS-compatible united-atom force field for the saturated acyclic haloalkane family.

Haloalkanes have always been of high relevance in chemistry, *e.g.*, as solvents, refrigerants, and reagents. In recent years, they have received renewed attention in the context of supramolecular[216−218] and medicinal[219−225] chemistry. This is in part due to their involvement in noncovalent intermolecular interactions referred to as halogen bonds, which result from a peculiar anisotropy feature in the electron density of the halogen atom known as the σ-hole.[216,217,219,221,222,224−228] Halogen bonds were first observed in crystal structures,[220,226,229−232] and are now well-understood theoretically.[227,233−235] The σ-hole results from a depletion of the electron density at the apex of the halogen atom, an effect that increases in magnitude along the halogen series (*i.e.*, from fluorine to iodine), is stronger for aromatic as opposed to aliphatic halogen atoms, and is exacerbated by substitution with electroattractive groups.[236,237]

Due to the above features, halogenated compounds are often regarded as challenging in terms of force-field design. A number of approaches have been used to account for the σ-hole feature explicitly, namely[227,235] (i) inclusion of an angular dependence in the NB interactions involving halogen atoms;[238−243] (ii) application of an empirical anisotropic (angle-dependent) potential-energy term for the halogen atoms or halogen bonds;[244] (iii) inclusion of a positively charged virtual site at the apex of halogen atoms;[245−254] (iv) use of a multipole expansion[255−258] (possibly mapped to a set of off-center point charges) to represent the charge distribution

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

of halogen atoms; and (v) use of a polarizable force field[259−264] (possibly along with specific anisotropy components). In the present work, however, the CombiFF approach is applied to a force-field functional form exempt of any of these features (as also done, *e.g.*, in refs 109, 188, and 209). This permits in particular to assess whether an explicit representation of σ-holes is necessary to achieve a high force-field accuracy for the molecules and properties considered.

2. METHODOLOGY

The CombiFF workflow for the calibration of a force field based on a given compound family is illustrated in Figure 1. It is explained in the following paragraphs using the example of the saturated haloalkane family considered in this work. This description is only meant to give an overview, and more detailed information can be found in Appendix A. For the ease of reference, a few key numbers (symbols and values) relevant to this discussion are summarized in Table 1.

Table 1. Relevant Numbers Pertaining to the CombiFF Force-Field Calibration and Validation for Saturated Haloalkanes[a]

number	value	meaning
$N_{\mathrm{iso}}^{\mathrm{tot}}$	755 301	total number of constitutional isomers in the 11 families considered (see Table 2)
$N_{\mathrm{iso}}^{\mathrm{sim}}$	486	total number of isomers with available experimental data ($N_{\mathrm{iso}}^{\mathrm{cal}} + N_{\mathrm{iso}}^{\mathrm{val}}$)
$N_{\mathrm{exp}}^{\mathrm{tot}}$	749	total number of experimental data points ($N_{\mathrm{exp}}^{\mathrm{cal}} + N_{\mathrm{exp}}^{\mathrm{val}}$)
$N_{\mathrm{sim}}^{\mathrm{tot}}$	632	total number of distinct P,T-points in this data ($N_{\mathrm{sim}}^{\mathrm{cal}} + N_{\mathrm{sim}}^{\mathrm{val}}$)
$N_{\mathrm{iso}}^{\mathrm{cal}}$	228	compounds in the calibration set
$N_{\mathrm{exp}}^{\mathrm{cal}}$	409	experimental data points for the calibration set ($N_{\rho}^{\mathrm{cal}} + N_{\Delta H}^{\mathrm{cal}}$)
N_{ρ}^{cal}	243	experimental ρ_{liq} data points for the calibration set
$N_{\Delta H}^{\mathrm{cal}}$	166	experimental ΔH_{vap} data points for the calibration set
$N_{\mathrm{sim}}^{\mathrm{cal}}$	332	distinct compounds/P,T-points for this calibration data (*i.e.*, simulations required)
$N_{\mathrm{iso}}^{\mathrm{val}}$	258	compounds in the validation set
$N_{\mathrm{exp}}^{\mathrm{val}}$	340	experimental data points for the validation set ($N_{\rho}^{\mathrm{val}} + N_{\Delta H}^{\mathrm{val}}$)
N_{ρ}^{val}	241	experimental ρ_{liq} data points for the validation set
$N_{\Delta H}^{\mathrm{val}}$	99	experimental ΔH_{vap} data points for the validation set
$N_{\mathrm{sim}}^{\mathrm{val}}$	300	distinct compounds/P,T-points for this validation data (*i.e.*, simulations required)
N_{att}	13	atom types
N_{frg}	15	fragments in the fragment-definition file
$N_{\mathrm{prm}}^{\mathrm{tot}}$	75	parameters of the force field
$N_{\mathrm{prm}}^{\mathrm{att}}$	56	nonbonded parameters (four per atom type, plus four third-neighbor LJ parameters)
$N_{\mathrm{prm}}^{\mathrm{cov}}$	19	covalent parameters
$N_{\mathrm{prm}}^{\mathrm{cal}}$	32	parameters that are actually optimized

[a]See Section 2 for explanations and Appendix A for details.

A family of compounds is defined here as a collection of molecules satisfying specified compositional (*e.g.*, ranges of element counts) and topological (*e.g.*, inclusion or exclusion of specific substructures) constraints. These specifications are gathered in a family-definition file. In the context of saturated haloalkanes, 11 nonoverlapping families A−K are considered, as listed in Table 2. All compounds correspond to the chemical formula $C_nH_{2n+2-m}\tilde{X}_m$ with $1 \le n \le 10$ and $1 \le m \le 4$, where $\tilde{X}_m$ stands for an arbitrary combination of m halogen atoms X ∈ {F,Cl,Br,I}. The definitions of these 11 families and the motivation for their selection are discussed in Appendix A.1.

The molecules belonging to a family are enumerated using an isomer enumeration program (ENU) developed in our group, which delivers a list of all molecules compatible with the constraints specified in the family-definition file. The ENU program is largely inspired from the graph-theoretical approach described in refs 265−267. It relies on fixed valences of four for the carbon atom, and of one for the hydrogen and halogen atoms. No postprocessing filter[268−270] is applied for "molecular realism". Although the program can take stereochemistry into account, this feature was not activated in the present study, *i.e.*, only constitutional isomers are distinguished. This does not represent a serious issue here, since the vast majority (94%) of the molecules considered have at most one stereogenic center (see Section S.1 for details). The isomer list for a given family, including a unique identifier, the chemical formula, and a canonical SMILES-string for each compound, is gathered in an isomer-enumeration file. The numbers of constitutional isomers in each of the 11 nonoverlapping families considered are also reported in Table 2 as N_{iso}, for a total $N_{\mathrm{iso}}^{\mathrm{tot}}$ of 755 301 isomers. This represents only a small subset (13.9%) of the 5 428 090 constitutional isomers of the chemical formula $C_nH_{2n+2-m}\tilde{X}_m$ with $1 \le n \le 10$ and $1 \le m \le 4$, and an even smaller subset of the $\sim 2.8 \times 10^{13}$ constitutional isomers of all saturated haloalkanes with up to 10 carbon atoms (see Section S.2 for details).

For a given family, the experimental-data vector $\mathbf{X}^{\mathrm{exp}}$ consists of a set of experimental observables and associated thermodynamic state points, extracted from a database maintained by our group. In practice, this extraction is only partly automated using database scripts (DBS), and manual curation is still required. For the haloalkanes, the database was queried for the liquid density ρ_{liq} and vaporization enthalpy ΔH_{vap}. The five main data sources accessed were refs 271−275, and only data considered reliable was retained. Illustrative examples of typical situations encountered during the data selection/curation are shown in Figure 2. Considering the union of the 11 families (Table 2), data for ρ_{liq} and/or ΔH_{vap} could be found for $N_{\mathrm{iso}}^{\mathrm{sim}} = 486$ compounds. These were split into a calibration and a validation set, with $N_{\mathrm{iso}}^{\mathrm{cal}} = 228$ and $N_{\mathrm{iso}}^{\mathrm{val}} = 258$ molecules, respectively.

The molecules of these two sets are displayed in Section S.3 (Figures S.2 and S.3, respectively). Acronyms are introduced to avoid the use of lengthy IUPAC names. They involve one letter and four digits. The letter is F, C, B, or I for compounds encompassing carbon, hydrogen, and only halogen atoms of a single type (F, Cl, Br, or I). It is X for hetero-di,tri,tetra-halogenated methanes and Y for larger hetero-di,tri,tetra-halogenated compounds. The first digit stands for the number of carbon atoms (except for X, where it represents the number of different halogen types in the molecule), with the number 10 mapped to the digit zero. The second digit stands for the total number of halogen atoms. The last two digits form a sequential index, further distinguishing compounds for which the first three symbols are identical. For instance, F3201 stands for a fluoroalkane with three carbon and two fluorine atoms, X2301 stands for a halomethane with three halogen atoms of two different types, and Y2301 stands for a haloalkane with two carbon and three halogen atoms of at least two different types.

The $\mathbf{X}^{\mathrm{exp}}$ vectors corresponding to the calibration and validation sets have the dimensions $N_{\mathrm{exp}}^{\mathrm{cal}} = 409$ and $N_{\mathrm{exp}}^{\mathrm{val}} = 340$, respectively. The one used for calibration is stored in a corresponding target-data file. More details on the sources,

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

Table 2. Eleven Families of Saturated Haloalkanes Considered in the Present Work[a]

fam.	n	m	k	N_{iso}	N_{sim}	description
A	1	1−4	1	16	16	C_1 mono- or homo-di,tri,tetra-halogenated
B	1	2−4	2−4	53	26	C_1 hetero-di,tri,tetra-halogenated
C	2−4	1−3	1	172	115	C_2−C_4 mono- or homo-di,tri-halogenated
D	2−4	2−4	2	1764	41	C_2−C_4 hetero-di,tri,tetra-halogenated with exactly two halogen types
E	5	1	1	12	9	C_5 monohalogenated with X on secondary C (R_2−CHX with R ≠ H)
F	5−6	1	1	16	13	C_5−C_6 monohalogenated with X on tertiary C (R_3−CX with R ≠ H)
G	5	1−3	1	12	6	C_5 mono- or homo-di,tri-halogenated of the form ^{t}Bu−CH_iX_{3-i}
H	6	1	1	4	2	C_6 monohalogenated of the form ^{t}Bu-CHX-CH_3
I	5−10	1−3	1	62 964	251	C_5−C_{10} mono- or homo-di,tri-halogenated excluding compounds already in E−H
J	5−9	2−4	2	666 534	6	C_5−C_9 hetero-di,tri,tetra-halogenated with exactly two halogen types
K	10	2	2	23 754	1	C_{10} hetero-di-halogenated with exactly two halogen types

[a]The families, labeled A−K, are nonoverlapping (*i.e.*, no compound belongs to more than one family). All compounds have the chemical formula $C_nH_{2n+2-m}\tilde{X}_m$ with $1 \le n \le 10$ and $1 \le m \le 4$, where $\tilde{X}_m$ stands for an arbitrary combination of m halogen atoms X ∈ {F,Cl,Br,I}. The value of k restricts the number of halogen types allowed in this combination (*e.g.*, $m = 3$ and $k = 2$ indicates that the compound includes three halogen atoms, and that they belong to two different types). The total number N_{iso} of constitutional isomers and the number N_{sim} of isomers actually considered in the simulations are also indicated for each family. The latter number is limited by the availability of experimental data for ρ_{liq} and/or ΔH_{vap}. The 11 families encompass N_{iso}^{tot} = 755 301 isomers, among which N_{iso}^{sim} = 486 are simulated. The union of families A−H defines the calibration set (total N_{iso}^{cal} = 228 compounds with N_{exp}^{cal} = 409 experimental values for ρ_{liq} or ΔH_{vap}). The union of families I−K defines the validation set (total N_{iso}^{val} = 258 compounds with N_{exp}^{val} = 340 experimental values). See Section 2 for explanations and Appendix A.1 for details.

curation, and selection of the experimental data (including a discussion of Figure 2) can be found in Appendix A.2 (see also Section S.4). The reference experimental values retained for ρ_{liq} and/or ΔH_{vap}, along with the associated P,T-points, are listed in Section S.5 (Table S.5). This section also suggests error estimates of about ±10.8 kg·m^{-3} for ρ_{liq} and about ±1.7 kJ·mol^{-1} for ΔH_{vap}, based on standard deviations over multiple experimental values for a subset of molecules. The content of Table S.5 can also be downloaded freely from the internet under ref 276, where the present data is labeled as version 1.0 (further versions will include revisions and/or expansions of the data set).

The principles of the force-field representation employed in the present work are summarized in Appendix A.3. They are essentially compatible with those of the GROMOS force field in its 2016H66 variant,[277] except for one important difference. The partial charges are determined for each molecule based on an electronegativity-equalization[278] (EE) scheme, where charge flows are only allowed along the CX bonds, permitting the definition of neutral charge-groups of at most five atoms, as detailed in Appendix A.4. The two atomic parameters involved in this charge-derivation scheme are the hardness η and the electronegativity χ. Given the use of a geometric-mean combination rule[279,280] for the LJ interactions,[81] the pairwise LJ coefficients are also constructed using two atomic parameters, the collision diameter σ and the well depth ϵ. The atom types of the force field are connected to a unique selection for these four parameters.

For a given compound, the information required to calculate the potential energy associated with a microscopic configuration of the system (*e.g.*, pure liquid or ideal gas) based on the force field includes two distinct components. On the one hand, the molecular topology contains information about a specific molecule in terms of atom properties (list of atoms and their types) and connectivity status (list of bonds, angles, improper dihedrals, torsional dihedrals, and close-neighbor exceptions, along with their types). On the other hand, the interaction-function parameters provide the actual values of the force-field parameters associated with the different types listed in the topology files. For a given family, the topology information pertaining to all molecules it encompasses is gathered in a family-topology file. For a given force field, the parameter information pertaining to all families it is meant to represent is gathered in an interaction-function parameter file. An essential feature of the CombiFF workflow is that these two components are kept strictly separate. The family-topology file solely contains pointers to parameters *via* types, but no explicit parameters. Accordingly, the interaction-function parameter file contains no reference to specific molecules, solely parameters associated with the types.

For a given family, the family-topology file is constructed based on the isomer-enumeration file using a library of fragments, gathered in a fragment-definition file, and a library of linkages, gathered in a linkage-definition file. The family-definition file describes the fragments to be assembled (default types for the set of atoms, bonds, angles, improper dihedrals, torsional dihedrals, and close-neighbor exceptions). The linkage-definition file defines possible modifications operated upon linking two specific fragments in a given connectivity (terms to be added or types to be modified). The family-topology file is constructed automatically by a topology builder program (TBL) developed in our group, which assembles the required fragments using the appropriate linkages based on the SMILES-strings of the compounds from the isomer-enumeration file. In the context of saturated (halo-)alkanes, the N_{frg} = 15 fragments shown in Figure 3 are required. Note that the halogenated fragments involve generic halogen sites that are later substituted by the appropriate halogen type. The principles of the assembly and substitution are illustrated in Figure 4 and detailed in Appendix A.5.

The family-topology file will point *via* types to a subset of parameters in the interaction-function parameter file. We refer to this set as the force-field parameter vector **P**. In the context of saturated (halo-)alkanes, and given the selected force-field representation (Appendices A.3 and A.4), this parameter vector encompasses N_{prm}^{tot} = 75 parameters: four nonbonded interaction parameters (η, χ, σ, and ϵ) for each of the N_{att} = 13 atom types listed in Table 3 and four additional third-neighbor LJ parameters, plus N_{prm}^{cov} = 19 covalent interaction parameters listed in Table 4. The coverage of the different atom types by

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

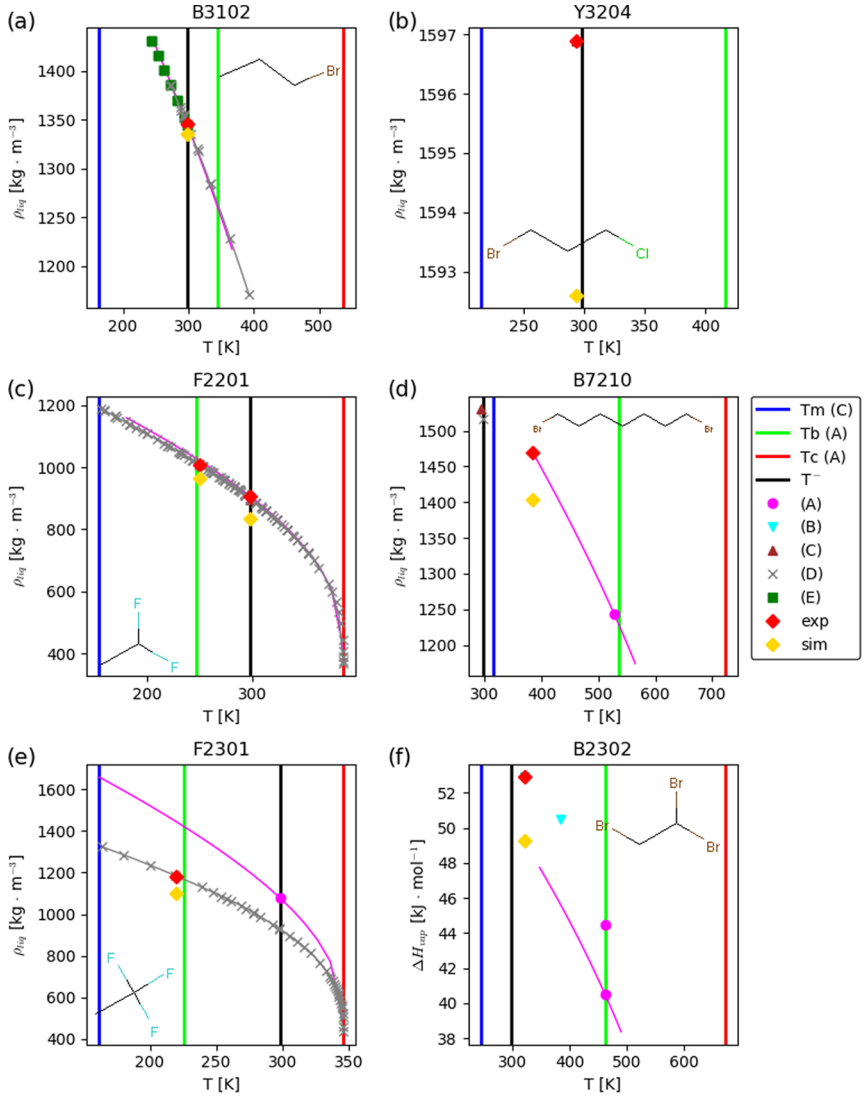


Figure 2. Illustrative examples of typical situations encountered during data selection/curation. For the six graphs (a–f), the indicated property (liquid density ρ_{liq} or vaporization enthalpy ΔH_{vap}) is shown as a function of the temperature T. The four vertical bars indicate the melting point T_m and the boiling point T_b at pressure P° = 1 bar, the critical point T_c (pressure P_c not specified), and the standard temperature T^{--} = 298.15 K. Individual experimental points and equation-based curves are reported from ref 271 (A), ref 272 (B), ref 273 (C), ref 274 (D), and ref 275 (E). The symbol "exp" denotes P,T-points retained for performing simulations, and the symbol "sim" denotes the results obtained with the final (optimized) force field. Note that the pressure may be P° or the vapor pressure P_{vap} of the liquid, depending on the simulation. See Section 2 for explanations and Appendix A.2 for details.

the molecules of the calibration and validation sets, including availability of experimental ρ_{liq} and ΔH_{vap} values, is detailed in Table 5. For the saturated haloalkanes considered here, only a subset of N_{prm}^{cal} = 32 parameters are actually optimized in the force-field calibration. The 43 other parameters are kept fixed at reasonable values. More details on the content of the resulting parameter vector and the selection of appropriate initial values (including a discussion of Tables 3–5) can be found in Appendix A.6 (see also Section S.6).

Given the choice of a parameter vector $\mathbf{P}$, the simulated-result vector $\mathbf{X}^{sim} = \mathbf{X}^{sim}(\mathbf{P})$ can be calculated using MD. In practice, these simulations are performed using a GROMOS-compatible MD engine developed in our group called SAMOS, which stands for Simulation of Atomic and MOlecular Systems. The accuracy of the force field can then be quantified using an objective function $Q = Q(\mathbf{P}; \mathbf{X}^{exp})$ assessing the extent of discrepancy between $\mathbf{X}^{sim}$ and $\mathbf{X}^{exp}$. Considering N_n observables pertaining to N_m molecules, this unitless function is defined here as

$$Q(\mathbf{P}; \mathbf{X}^{exp}) = W^{-1} \sum_{n=1}^{N_n} s_n^{-1} \sum_{m=1}^{N_m} w_{nm} |X_{nm}^{sim}(\mathbf{P}) - X_{nm}^{exp}|$$
$$\text{with} \quad W = \sum_{n=1}^{N_n} \sum_{m=1}^{N_m} w_{nm} \tag{1}$$

Note that Q considers the mean absolute deviation (MAD) rather than the root-mean-square deviation, to limit the impact of outliers during the optimization. The s_n and w_{nm} coefficients adjust the weighting of the different observables/molecules in terms of the relative contributions of corresponding discrepancies to the magnitude of the objective function. The s_n coefficients eliminate the dependence on a unit system and adjust the relative weights of different observables in terms of the perceived (*i.e.*, subjective) extent of "badness". The s_n

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525–7555

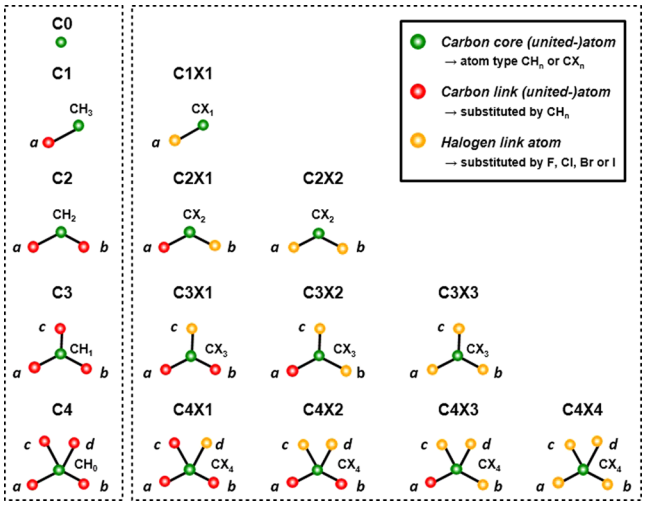


Figure 3. Molecular-topology fragments required for the representation of the saturated (halo-)alkanes. In total, $N_{frg} = 15$ fragments are required. The five fragments on the left are sufficient to construct all the saturated alkanes, and the 10 fragments on the right correspond to the halogenated groups. See Section 2 for explanations and Appendix A.5 for details.

values employed here are 20 kg·m^{-3} for the ρ_{liq} observables and 1 kJ·mol^{-1} for the ΔH_{vap} observables. The unitless w_{nm} coefficients can be used to weigh differently the contribution of specific observable/molecule combinations, *e.g.*, in relation to the presence/absence and, possibly, experimental/simulated error estimates on these numbers. For simplicity, all the observable/molecule combinations included (also considering observables at multiple state points) are associated here with the same weight $w_{nm} = 1$, while all the omitted ones (absence of experimental data) have $w_{nm} = 0$. The prefactor W ensures that the overall magnitude of Q is not affected by the number of data points included. Thus, considering an optimization set with a comparable number of ρ_{liq} and ΔH_{vap} observables, a force field with $Q = 2.5$ would correspond to MADs on the order of 50 kg·m^{-3} for ρ_{liq} and 2.5 kJ·mol^{-1} (about k_BT at room temperature) for ΔH_{vap}, which could represent a reasonable target for the optimization.

The task involved in the force-field calibration corresponds to minimizing Q with respect to $\mathbf{P}$ in the N_k-dimensional parameter space. To guide this optimization, a first-order Taylor-series expansion of $\mathbf{X}^{sim}$ is introduced into eq 1, performed at a reference point $\mathbf{P}^o$ in parameter space. This leads to an approximation $\tilde{Q} = \tilde{Q}(\mathbf{P}; \mathbf{P}^o, \mathbf{X}^{exp})$ of Q given by

$$\tilde{Q}(\mathbf{P};\mathbf{P}^o,\mathbf{X}^{exp}) = W^{-1}\sum_{n=1}^{N_n} s_n^{-1}\sum_{m=1}^{N_m} w_{nm} |X_{nm}^{sim,o} + \mathbf{S}_{nm}^{sim,o}(\mathbf{P}-\mathbf{P}^o) - X_{nm}^{exp}| \quad (2)$$

where $X^{sim,o} = X^{sim}(\mathbf{P}^o)$, and $\mathbf{S}_{nm}^{sim,o}$ is the line of the sensitivity matrix $\underline{\mathbf{S}}^{sim,o} = \underline{\mathbf{S}}^{sim}(\mathbf{P}^o)$ corresponding to the simulated observable n for molecule m. The N_k columns of the latter matrix contain the derivatives of the corresponding quantity with respect to all force-field parameters. Next to the observables themselves, the program SAMOS determines this sensitivity matrix during simulations at $\mathbf{P}^o$, using appropriate statistical−mechanical expressions.[175,178−180,281] Based on eq 2, one can then determine a point $\mathbf{P}^*$ in parameter space that will minimize $\tilde{Q}$ within a specified trust region around $\mathbf{P}^o$. Since the function is convex, this point is unique, and it is selected as a new point $\mathbf{P}^o$ to carry out simulations for the next iteration.

In practice, the optimization algorithm involves the following steps over iterations i starting from zero: (1) select an initial guess $\mathbf{P}_0^o$ for the parameters; (2) run N_{sim}^{cal} simulations to get $X_i^{sim,o}$ and $\underline{\mathbf{S}}_i^{sim,o}$; (3) calculate the real value $Q_i^{real} = Q(\mathbf{P}_i^o; \mathbf{X}^{exp})$ of the objective function at this point in parameter space using eq 1; (4) minimize $\tilde{Q}(\mathbf{P};\mathbf{P}_i^o,\mathbf{X}^{exp})$ in eq 2 with respect to $\mathbf{P}$ starting from $\mathbf{P}_i^o$ and staying within the trust region, leading to $\mathbf{P}_i^*$; (5) calculate the predicted value $Q_{i+1}^{pred} = \tilde{Q}(\mathbf{P}_i^*;\mathbf{P}_i^o,\mathbf{X}^{exp})$ of the objective function; and (6) set $\mathbf{P}_{i+1}^o$ to $\mathbf{P}_i^*$, increment i, and iterate to step (2) until convergence.

Step (2) is the expensive part of the calculation. In contrast, step (4) is inexpensive, and carried out in practice using a simplex minimization. The reason for the introduction of a trust region is that eq 2 involves a linearization, and is thus

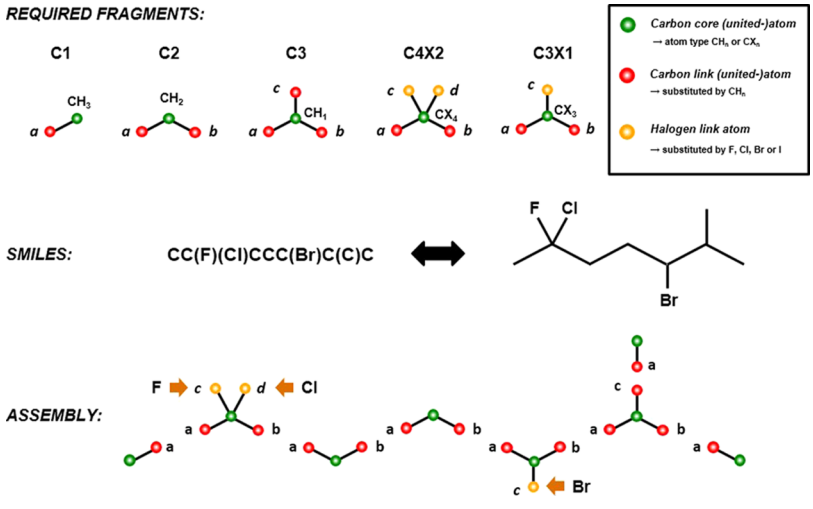


Figure 4. Illustrative example of the SMILES-directed assembly of fragments into a molecular topology. This figure shows the operation of the TBL program for the molecule 2-fluoro-2-chloro-5-bromo-6-methyl-heptane. The SMILES-string (*e.g.*, generated by the ENU program) is used to direct the assembly of the required subset among the fragments shown in Figure 3. The assembly of fragments corresponds to overlapping the selected core−link bonds of the two fragments, with possible adjustments specified in a corresponding linkage entry. See Section 2 for explanations and Appendix A.5 for details.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

Table 3. Atom Types of the Saturated (Halo-)alkane Force Field and Optimized Values of the Associated Force-Field Parameters[a]

atom type	σ (nm)	ϵ (kJ·mol^{-1})	η (e^{-1}·V)	χ (V)	usage
halogen atoms					
F	0.274	0.697	51.132	26.956	fluorine
Cl	0.339	1.437	35.209	20.907	chlorine
Br	0.358	2.026	31.861	18.674	bromine
I	0.382	2.743	26.852	16.339	iodine
aliphatic carbon (united-)atoms (not bonded to a halogen)					
CH0	0.664 (0.336)	0.007			CH_0 carbon atom
CH1	0.502 (0.330)	0.095			CH_1 carbon united-atom (methanetriyl group)
CH2	0.407 (0.316)	0.411			CH_2 carbon united-atom (methylene group)
CH3	0.375 (0.310)	0.867			CH_3 carbon united-atom (methyl group)
CH4	0.371	1.264			CH_4 carbon united-atom (methane molecule)
halogenated carbon (united-)atoms (bonded to at least one halogen)					
CX4	0.590	0.006	9.169	8.593	CH_0 connected to one, two, three, or four halogen atoms
CX3	0.436	0.096	11.410	8.463	CH_1 connected to one, two, or three halogen atoms
CX2	0.390	0.441	14.383	9.125	CH_2 connected to one or two halogen atoms
CX1	0.372	1.098	18.419	11.458	CH_3 connected to one halogen atom

[a]The $N_{att} = 13$ atom types are listed along with their usage and the optimized (final) values of the nonbonded interaction parameters. These are the LJ collision diameter σ and well depth ϵ, along with the EE hardness η and electronegativity χ. The types CH0−CH3 also invole third-neighbor parameters (between parentheses; the correponding ϵ can be deduced since $C_6 = 4\epsilon\sigma^6$ is the same as for the normal parameters). This results in a total of $N_{prm}^{att} = 56$ nonbonded interaction parameters. Note that the parameters are rounded. Exact values can be downloaded from the internet in the GROMOS-compatible files under ref 276. The 14 LJ parameters of the five aliphatic atom types (CH0−CH4) are directly taken from the GROMOS force field and not subject to optimization. The corresponding parameters η and χ are not required as these atoms have zero charge. The values of the $N_{prm}^{cov} = 19$ covalent interaction parameters, which are also excluded from the optimization, can be found in Table 4. Thus, a total of $N_{prm}^{tot} = 75$ parameters fully specifies the force field and a subset of $N_{prm}^{cal} = 32$ parameters (all of them nonbonded) were optimized. The initial values of the nonbonded interaction parameters (at the start of the optimization) are provided in Table S.6. See Section 2 for explanations and Appendices A.3−A.6 for details.

only expected to be accurate when $\mathbf{P}$ is reasonably close to $\mathbf{P}^o$. This limitation is introduced in the form of maximal allowed relative changes in each of the parameters over an iteration, set to 5% for σ, ϵ, and χ, and to 10% for η. Considering steps (3) and (5), from $i = 1$ onward, the real value Q_i^{real} at iteration i can be compared to the predicted value Q_i^{pred} from the previous iteration, giving a hint about the accuracy of the linearization within the imposed trust radius. If the algorithm terminates at iteration i_{max}, the final parameter set is $\mathbf{P}_{i_{max}}^o$ with value $Q_{i_{max}}^{real}$. Although $\mathbf{P}_{i_{max}}^*$ and $Q_{i_{max}+1}^{pred}$ are available, it is preferable to stop at a force field with an explicitly calculated objective function. This procedure is realized in practice by an optimizer (OPT) script. More details on the objective function and the optimization procedure are provided in Appendix A.7.

By minimizing Q in an iterative fashion, one can determine the optimal parameter set (or at least one good parameter choice) for the given family, force-field representation, and experimental data set. For the saturated haloalkanes considered here, $N_{prm}^{cal} = 32$ parameters are optimized considering the calibration set of $N_{iso}^{cal} = 228$ molecules characterized by $N_{exp}^{cal} = 409$ observables. This corresponds to a very favorable observable-to-parameter ratio of 12.8. The parameters of the final force field are given in Tables 3 and 4. The corresponding GROMOS-compatible interaction-function parameter files and molecular topology building-block files for the $N_{iso}^{sim} = 486$ molecules can be downloaded freely from the internet under ref 276. A python script is also provided to construct topologies for other saturated haloalkane molecules based on their SMILES-strings. Finally, the details of the simulation protocols are provided in Appendix A.8.

Given the setup adopted in the present simulations (as described in Appendices A.1−A.8) and access to 332 processors (3 GHz Intel Xeon), *i.e.*, one for each of the N_{sim}^{tot} simulations at different P,T-points to be carried out for each iteration, a typical optimization of the calibration set (10 iterations) requires less than three days of wall-clock time.

The programs/scripts ENU, DBS, TBL, and SAMOS will be documented independently in separate publications. The main focus of the present article is on the basic workflow and parametrization principles.

3. RESULTS AND DISCUSSION

3.1. Parameter Optimization. The evolution of the objective function Q with the iteration number i is illustrated in Figure 5. The graph shows both the real value Q_i^{real} at iteration i calculated according to eq 1 and its predicted value Q_i^{pred} based on the sensitivity analysis of the simulations at the previous iteration $i - 1$ according to eq 2. After two iterations, the predicted and real values agree very well, indicating that the linear approximation is accurate within the chosen trust radius for the parameter variations. During the first three iterations, the objective function also drops sharply. It converges to 1.67 after four iterations, and the three additional iterations provide no further improvement. The final value of Q at iteration seven is 1.63. The corresponding statistics in terms of deviations from experiment over the calibration set is provided in Table 6 (last line). Relative to the experiment, the optimized force field has approximate MADs of 29.6 kg·m^{-3} for ρ_{liq} and 1.9 kJ·mol^{-1} for ΔH_{vap}.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

Table 4. Covalent Types of the Saturated (Halo-)alkane Force Field and Values of the Associated Force-Field Parameters[a]

Bond		
quartic force constant (10^6 kJ·mol^{-1}·nm^{-4})	reference bond length (nm)	usage
15.09	0.133	C−F
8.12	0.176	C−Cl
6.95	0.192	C−Br
5.67	0.216	C−I
7.15	0.153	C−C
Angle		
cosine-harmonic force constant (kJ·mol^{-1})	reference bond angle (deg)	usage
520	109.5	C−C−C
618	109.5	X,C−C−X
Improper Dihedral		
force constant (kJ·mol^{-1}·deg^{-2})	reference improper-dihedral angle (deg)	usage
0.102	35.264	tetrahedral centers
Proper Dihedral		
force constant (kJ·mol^{-1})	phase shift (deg)	multiplicity	usage
5.92	0.0	3	X,C−C−C−C,X

[a]The covalent types are listed along with their usage and the values of the interaction parameters. These are force constants, along with reference bond lengths, bond angles, improper dihedral angles, and, for the torsions, multiplicities and phase-shift angles. Note that the parameters are rounded. Exact values can be downloaded from the internet in the GROMOS-compatible files under ref 276. These $N_{\mathrm{prm}}^{\mathrm{cov}}$ = 19 covalent interaction parameters are either ported from the 2016H66 force field[277] or, whenever unavailable, inferred from QM calculation results. They are not included in the optimization. Values for the $N_{\mathrm{prm}}^{\mathrm{att}}$ = 56 nonbonded interaction parameters of the force field can be found in Table 3. See Section 2 for explanations and Appendices A.3 and A.6 for details.

Table 5. Coverage of the Different Atom Types by the Molecules of the Calibration and Validation Sets, Including Availability of Experimental ρ_{liq} and ΔH_{vap} Values[a]

atom type	N_{cal}	N_{ρ}^{cal}	$N_{\Delta H}^{\mathrm{cal}}$	N^{val}	N_{ρ}^{val}	$N_{\Delta H}^{\mathrm{val}}$
F	71	84	60	32	24	16
Cl	101	106	75	117	112	42
Br	82	83	53	86	82	32
I	45	42	19	30	30	9
CH0	8	8	3	28	28	3
CH1	19	18	11	62	62	13
CH2	73	73	53	231	214	98
CH3	137	143	100	235	218	87
CH4	0	0	0	0	0	0
CX4	80	84	58	60	60	6
CX3	98	100	66	116	100	46
CX2	104	106	74	133	129	61
CX1	4	8	5	0	0	0

[a]The N_{att} = 13 atom types are listed in Table 3. The entries N_{cal}, N_{ρ}^{cal}, and $N_{\Delta H}^{\mathrm{cal}}$ provide numbers of molecules, ρ_{liq} values, and ΔH_{vap} values, respectively, in the calibration set (union of families A−H in Table 2, 228 compounds depicted in Figure S.2) that include at least once the given atom type. The entries N^{val}, N_{ρ}^{val}, and $N_{\Delta H}^{\mathrm{val}}$ provide the corresponding information for the validation set (union of families I−K in Table 2, 258 compounds depicted in Figure S.3).

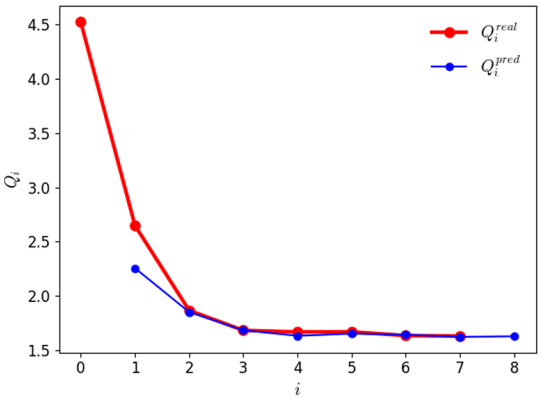


Figure 5. Evolution of the predicted and real values of the objective function against the iteration number along the force-field parameter optimization. The real value Q_i^{real} at iteration i is calculated according to eq 1. The predicted value Q_i^{pred} is calculated based on the sensitivity analysis of the simulations at iteration $i-1$ according to eq 2. The first simulation at $i = 0$ using the initial parameter set (Table S.6) leads to a first real value Q_0^{real} and a first predicted value Q_1^{pred}. The last simulation at $i = 7$ using the final (optimized) parameter set (Table 3) leads to the final real value Q_7^{real} (and a predicted value Q_8^{pred}, which is discarded). See Figure S.8 for an extension of this plot to 25 iterations.

The evolution of the $N_{\mathrm{cal}}^{\mathrm{prm}}$ = 32 force-field parameters subject to calibration (Table 3) as a function of the iteration number i is illustrated in Figure 6. The largest parameter changes typically also occur within the first three iterations. However, some of the parameters still undergo significant variations up to iteration 7. Such a slow relaxation appears to affect primarily the EE parameters and most pronouncedly the hardness of the iodine and of the halogenated carbon atom types. The observation of correlated parameter changes that no longer induce a significant decrease of the objective function suggests the presence of a local quasi-degeneracy in the parameter space. The objective function reaches a flat low-dimensional parameter subspace, where different parameter combinations produce comparable values for the function Q. The final values of the parameters (Table 3) are those obtained at iteration 7. The results of an optimization further extended to 25 iterations are presented in Section S.7. Although the objective function still changes slightly, with a final value of 1.61 at iteration 25, all parameters end up converging.

To investigate the uniqueness of the optimal solution, *i.e.*, whether it converges to a local or a global minimum, the results of seven independent optimizations (25 iterations) starting from different initial parameters are reported in Section S.8. In all cases, the objective function converges to a final value ranging between 1.61 and 1.68. The LJ interaction parameters σ and ϵ systematically converges to similar final values within relatively narrow ranges (Figure S.11). The only exceptions concern the atom type CX1, which is exclusively found in mono-halomethanes, *i.e.*, only four molecules. For this atom type, the final ϵ parameters exhibit a more significant spread. The σ parameter converges to a narrow range, except for a very different value in one of the seven simulations. Compared to the LJ interaction parameters, the converged EE parameters η and χ evidence more variations across the seven optimizations (Figure S.12). However, the final atomic partial charges of the halogen atoms within the different fragments remain essentially consistent across the optimizations, with standard deviations of at most 0.08 e. In agreement with previous findings,[282] this

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

Table 6. Statistics Concerning the Discrepancies between Simulated and Experimental Properties for the Calibration Set Based on the Optimized Force Field[a]

				RMSD		AVED		MAD	
group	N_{mol}	N_{ρ}^{cal}	$N_{\Delta H}^{cal}$	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})
F*1	9	14	8	42.9	1.6	6.7	−1.1	20.0	1.2
F*2	10	15	10	54.8	1.8	−18.3	−1.3	41.1	1.5
F*3	7	7	8	74.5	2.2	−56.1	0.2	58.7	1.8
F*4	1	1	1	312.0	2.8	312.0	2.8	312.0	2.8
C*1	16	18	16	12.8	1.4	2.1	−1.0	9.5	1.2
C*2	17	17	16	17.3	1.4	−6.2	−0.6	13.7	1.2
C*3	16	16	12	16.6	3.7	3.7	1.6	13.2	2.2
C*4	1	1	1	13.6	0.1	−13.6	−0.1	13.6	0.1
B*1	17	18	14	17.1	1.4	5.2	−1.2	13.2	1.2
B*2	17	17	15	37.1	2.6	−9.4	−0.1	26.7	2.0
B*3	14	13	10	36.2	4.8	10.3	−0.0	28.7	3.7
B*4	1	1	1	70.6	3.3	−70.6	−3.3	70.6	3.3
I*1	17	17	9	42.8	1.1	33.9	−1.0	35.6	1.0
I*2	16	16	4	61.9	3.1	23.6	−1.9	43.3	2.2
I*3	1	1	1	271.5	2.8	−271.5	2.8	271.5	2.8
I*4	1	1	0	202.8		−202.8		202.8	
X*2	4	4	2	38.1	2.3	−35.8	−0.5	35.8	2.2
X*3	9	9	5	26.6	4.5	3.6	3.4	21.7	3.4
X*4	13	15	10	57.9	2.3	0.0	2.1	47.3	2.1
Y*2	15	14	6	22.6	3.3	5.7	−1.8	17.5	2.4
Y*3	10	10	7	19.5	2.0	3.0	1.1	12.8	1.7
Y*4	16	14	9	54.1	4.3	7.0	3.2	35.3	3.4
T	228	239	165	49.8	2.7	1.2	0.0	29.6	1.9

[a]For selected groups encompassing N_{mol} molecules, the number N^{cal} of experimental data points, the root-mean-square deviation (RMSD), the average deviation (AVED), and the mean absolute deviation (MAD) are reported for the pure-liquid density ρ_{liq} and the vaporization enthalpy ΔH_{vap}. The letter in the group code stands for the type of halogen atoms present in the molecule: homofluorinated (F), homochlorinated (C), homobrominated (B), homoiodinated (I), hetero-halomethane (X), and hetero-haloalkane with more than one carbon atom (Y). The digit stands for the total number of halogen atoms in the molecule. These two symbols match the first and third symbols employed in the molecule codes used in Section S.3. The line labeled T corresponds to all molecules of the set. The results of five simulations (F1101c, F1301b, F1301c, F2301b, and X2407b), associated with four ρ_{liq} and one ΔH_{vap} values, are omitted from the statistics because they resulted in vaporization. This data is illustrated graphically in Figure 8 (top row).

suggests that there exists a certain extent of degeneracy in how the EE model can account for specific atomic partial charges.

3.2. Results for the Calibration Set. The level of agreement between the optimized force field and experiment in terms of ρ_{liq} and ΔH_{vap} for the calibration set is illustrated in Figure 7 (top row). The corresponding numerical values can be found in Section S.9. The statistics per compound group is provided in Table 6 and illustrated graphically in Figure 8 (top row). For 197 out of the N_{iso}^{cal} = 228 compounds, the deviations in ρ_{liq} and ΔH_{vap} are below 80 kg·m^{-3} and 4 kJ·mol^{-1}, respectively. The 31 molecules (36 different P,T-points) involving larger deviations in either or both values, along with the magnitudes of the discrepancies and the corresponding molecular structures, are listed in Section S.10. Among these, four molecules underwent vaporization. All are fluorinated haloalkanes (F1101, F1301, F2301, and X2407), which are actually gaseous under ambient conditions, and were still simulated at T^{-} = 298.15 K, using either the experimental vapor pressure (three simulations) or P^{o} = 1 bar in the absence of vapor pressure reported in the source (two simulations). These five simulations are omitted from Figure 7, and from the statistics of Table 6 and Figure 8. Considering the other main outliers, most of the involved simulations also correspond to P,T-points which differ from room temperature and atmospheric pressure. This may in part explain the observed discrepancies with the experiment, in addition to possible force-field and experimental errors. Nevertheless, the optimized force field leads to simulated results that agree very well with experiment for the calibration set. The root-mean-square deviation (RMSD) is 49.8 kg·m^{-3} for ρ_{liq}, *i.e.*, less than 6.5% considering that the densities of the compounds in the calibration set are all above 750 kg·m^{-3}. The RMSD value is 2.7 kJ·mol^{-1} for ΔH_{vap}, *i.e.*, only slightly larger than k_BT at room temperature. The corresponding average deviation (AVED) values are close to zero for both properties, suggesting that the RMSD is dominated by statistical as opposed to systematic errors. However, the residual errors are not homogeneously distributed over the different groups of molecules.

In terms of RMSD, two main trends can be distinguished. First, the agreement with experiment is better for chlorinated and brominated compounds compared to fluorinated and iodinated ones. This is probably in part due to force-field description errors, which are likely to affect more adversely the extremes of the halogen chemical series (F, I) compared to its median (Cl, Br), and to the reduced availability of small compounds that are liquid under ambient conditions, with a predominance of gaseous compounds for F and of solid compounds for I. Previous studies have also shown that the fluoroalkanes are challenging compounds in terms of force-field parametrization.[283−285]

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

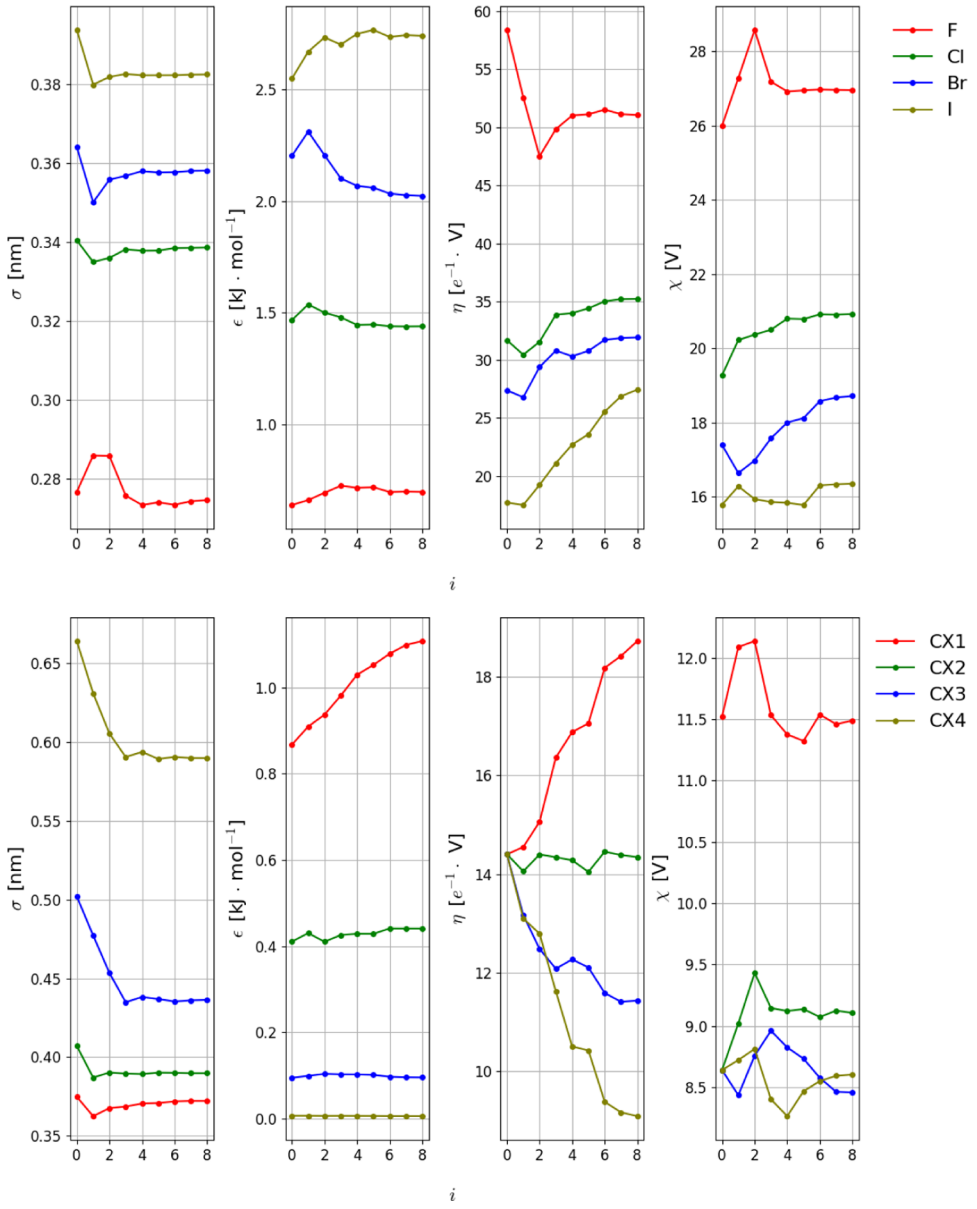


Figure 6. Evolution of the nonbonded interaction parameters against the iteration number along the force-field parameter optimization. The $N_{\text{prm}}^{\text{cal}}$ = 32 parameters considered (Table 3) are the LJ collision diameter σ and well depth ϵ, along with the EE hardness η and electronegativity χ for the atom types F, Cl, Br, and I (top row) and CX1−CX4 (bottom row). The value of each parameter is reported at iteration i. The final force-field parameters are those corresponding to iteration i = 7 (the values at i = 8 correspond to proposed changes for a next iteration and are discarded). See Figure S.9 for an extension of this graph to 25 iterations.

Second, the agreement with experiment tends to slightly worsen when the halogen content of a molecule is increased, *i.e.*, from mono- to di-, tri-, or tetra-halogenated. This is certainly in part related to the fact that increasing halogenation increases the proportion of haloalkane-type intermolecular interactions (*i.e.*, involving halogen and halogenated carbon atoms) as opposed to alkane-type (*i.e.*, aliphatic only) interactions, the latter being comparatively easier to describe accurately and already extremely well captured by the alkane parameters of the GROMOS force field.[286−288]

3.3. Results for the Validation Set. The level of agreement between the optimized force field and experiment in terms of ρ_{liq} and ΔH_{vap} for the validation set is illustrated in Figure 7 (bottom row). The corresponding numerical values can be found in Section S.9. The statistics per compound group is provided in Table 7 and illustrated graphically in Figure 8 (bottom row). For 247 out of the $N_{\text{iso}}^{\text{val}}$ = 258 compounds, the deviations in the values of ρ_{liq} and ΔH_{vap} are below 80 kg·m^{-3} and 4 kJ·mol^{-1}, respectively. The 11 molecules (11 different P,T-points) involving larger deviations in either or both values, along with the magnitudes of the discrepancies and the corresponding molecular structures, are listed in Section S.10. Here, no molecule underwent vaporization.

Just as was the case for the calibration set, the agreement with experiment for the validation set is excellent, with RMSD values of 27.6 kg·m^{-3} for ρ_{liq} and 1.8 kJ·mol^{-1} for ΔH_{vap}, and AVED values that are close to zero. Somewhat counterintuitively, the RMSD values are actually smaller (roughly by a factor 2) for the validation compared to the calibration set. This is likely because the validation set is less challenging. It involves a higher proportion of compounds that are liquid at

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

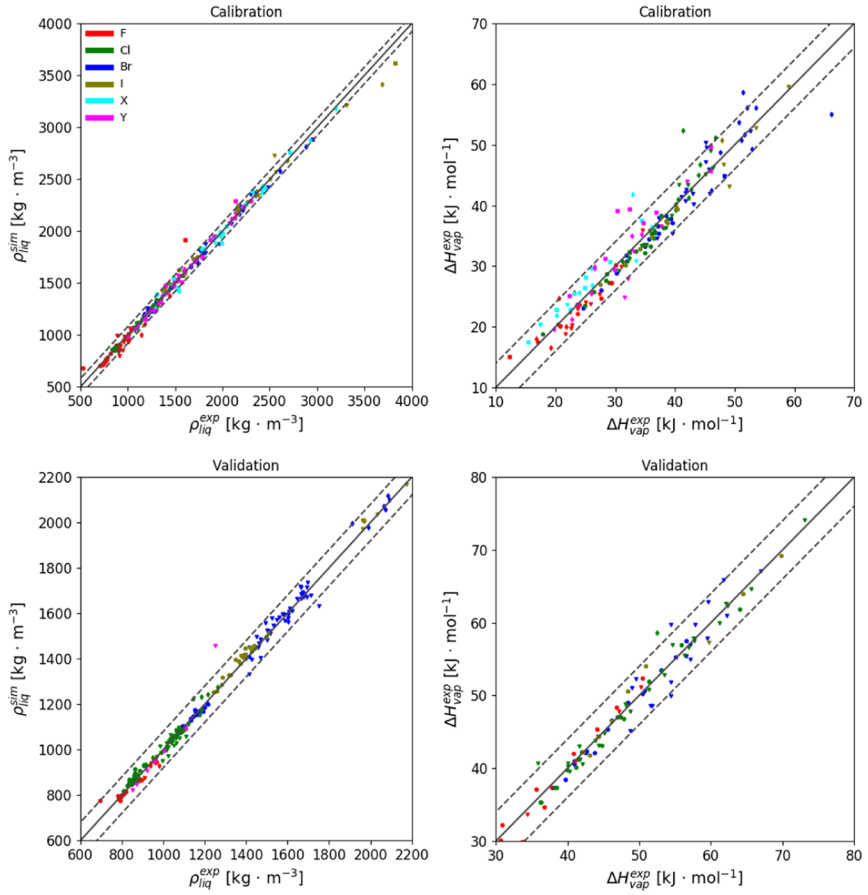


Figure 7. Comparison of simulated and experimental properties based on the optimized force field. The values of ρ_{liq} (left) and ΔH_{vap} (right) are compared for the calibration (top) and validation (bottom) sets. The diagonal solid lines indicate perfect agreement and the range between the two dashed lines indicate agreement within ±80 kg·m^{-3} for ρ_{liq} or ±4 kJ·mol^{-1} for ΔH_{vap}. The symbols refer to the number of halogen atoms in the molecule (bullet, triangle, diamond, and square for 1, 2, 3, or 4 respectively). The corresponding numerical values can be found in Table S.9, the information about outliers in Table S.10, and the statistics per groups of compounds in Tables 6 and 7. In the top graphs, the representative points for five simulations (F1101c, F1301b, F1301c, F2301b, X2407b) are omitted because they resulted in vaporization.

room temperature and atmospheric pressure, and larger molecules that give more weight to the alkane-type (as opposed to the haloalkane-type) interactions. In contrast to the calibration set, the residual errors are also more homogeneously distributed over the different groups of molecules. The families Y*2 and C*3 seem affected by particularly large errors in ρ_{liq} and ΔH_{vap}, respectively, but this is mainly due to two specific outliers (Y6201a and C0301a, respectively). Given that the other molecules in these families show smaller errors, this could indicate problems with the reference experimental data.

3.4. Trends in the Optimized Parameters. As illustrated in Figure 9, the nonbonded interaction parameters for the halogen atoms in the optimized force field show trends along the series that are consistent with chemical intuition. The LJ interaction parameters σ and ϵ both increase along the series, in line with the increase in atomic size and electronic polarizability. The EE parameters χ and η both decrease along the series, in line with a decrease of the electronegativity and an increase of the softness. A comparison of the final χ values with standard electronegativity scales is provided in Section S.11. In view the different definitions adopted, the quantitative comparison is not meaningful.[282] However, the force-field values as well as all the scales considered present a similar relative decrease along the series (Figure S.15). The decrease is similar from Cl to Br and then from Br to I, and slightly larger from F to Cl.

It is also of interest to examine the trends directly in terms of the atomic partial charges. To this purpose, the charges of the halogen atoms in the 10 homohalogenated charge-groups of the force field are compared in Table 8. Considering separately the mono-, di-, and tri-homo-halogenated groups (*i.e.*, the X1, X2, and X3 series), the halogen charge systematically becomes more negative upon increasing the alkyl substitution (*i.e.*, from C1 to C4). Conversely, considering separately the di-, tri-, and tetrasubstituted groups (*i.e.*, the C2, C3, and C4 series), the halogen charge systematically becomes less negative upon increasing the halogenation (*i.e.*, from X1 to X4). Thus, as expected, the substitution of hydrogen atoms by electron-donating (hyperconjugative) alkyl groups tends to enhance the polarization of the carbon−halogen bond, whereas the substitution of hydrogen or carbon atoms by electron-withdrawing halogen atoms inhibits this polarization. Finally, considering each of the 10 charge-groups separately, the halogen charge also systematically becomes less negative along the halogen series, paralleling the evolution of the electronegativity. Similar trends were also observed in a previous QM study,[289] where different charge schemes were used to investigate the effect of the halogen substituents on the charge distribution.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

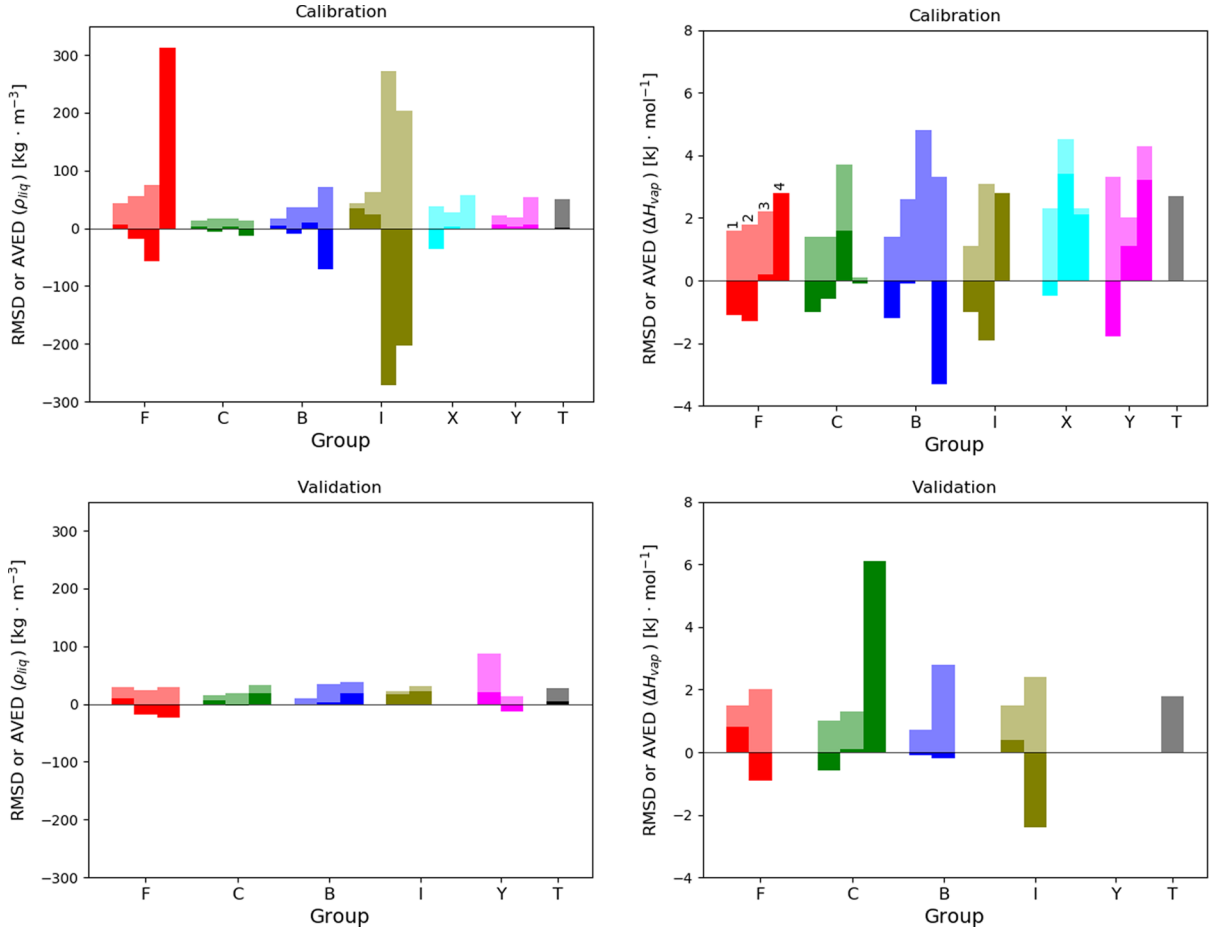


Figure 8. Statistics concerning the discrepancies between simulated and experimental properties based on the optimized force field. The values of the root-mean-square deviation (RMSD, semitransparent bar) and average deviation (AVED, solid bar) in ρ_{liq} (left) and ΔH_{vap} (right) per group of molecules are compared for the calibration (top) and the validation (bottom) set. The labels stand for homofluorinated (F), homochlorinated (C), homobrominated (B), homoiodinated (I), hetero-halomethane (X), hetero-haloalkane with more than one carbon (Y), and all molecules (T). The successive bars in each group correspond to compounds containing one, two, three, or four halogen atoms. The corresponding numerical values can be found in Tables 6 and 7.

The atomic partial charges on the halogen atoms of the different halomethanes are also compared in Figure 10 to corresponding charges obtained from QM calculations at the TPSSh/def2-TZVP level,[290−292] analyzed using the density-derived electrostatic and chemical (DDEC) scheme[293] with a polarizable continuum model[294] (IEFPCM) using permittivities equal or close to the experimental value for each liquid. The corresponding numerical values can be found in Table S.12. The force-field charges range between −0.29 and −0.07 *e*, and the DDEC charges range between −0.26 and 0.00 *e*. The charges of the two sets correlate well in terms of halogen types (between the groups of different colors in the figure), with charge magnitudes decreasing in both cases along the halogen series. However, for each halogen type taken separately (within the colored groups), there is a surprising anticorrelation between the charge sets. Taking the example of fluorine, the mono-, di-, tri-, and tetra-fluoro-methanes (marked as red stars) have comparable charges in the force field and in DDEC. Within the force field, swapping a subset of the fluorine atoms for less electronegative halogen atoms (red bullets) tends to lead to an enhanced negative charge for the remaining fluorine atom(s). This is intuitively reasonable, and the opposite effect is also visible for iodine (violet stars to violet bullets). Somewhat surprisingly, however, the DDEC charges follow opposite trends. This may possibly result from a slight artifactual bias of the electron-density partitioning scheme toward atoms with a larger effective volume, or from a dependency of the DDEC charges on the liquid permittivity *via* the IEFPCM model.

Finally, considering the parameters of the carbon atoms (see Figure 6), the mono- and dihalogenated atom types CX1 and CX2 remain very similar to the aliphatic united-atom types CH3 and CH2, in terms of both σ and ϵ. In contrast, the tri- and tetrahalogenated types CX3 and CX4 have significantly decreased σ and slightly increased ϵ compared to CH1 and CH0 united-atoms. The electronegativities χ are comparable for all halogenated carbon atom types, except for CX1 (higher value). Finally, the hardness shows a pronounced difference, with a regular increase along the sequence from CX4 to CX1.

3.5. Influence of the State Point. Because classical force-field parameters are effective (empirical) quantities, their validity is typically limited to a given phase and to the neighborhood of a selected *P*,*T*-point. The present force field is designed primarily for the liquid phase at temperatures close to T^{-} = 298.15 K and pressures close to P° = 1 bar. The use of somewhat different pressures at T^{-} for compounds that are not liquid at this temperature is probably less problematic than the use of different temperatures. Most of the simulations for the

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

Table 7. Statistics Concerning the Discrepancies between Simulated and Experimental Properties for the Validation Set Based on the Optimized Force Field[a]

				RMSD		AVED		MAD	
group	N_{mol}	N_{ρ}^{val}	$N_{\Delta H}^{val}$	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})	ρ_{liq} (kg·m^{-3})	ΔH_{vap} (kJ·mol^{-1})
F*1	11	9	8	28.7	1.5	9.2	0.8	13.3	1.4
F*2	11	5	8	23.6	2.0	−18.9	−0.9	18.9	1.3
F*3	4	4	0	28.5		−23.7		23.7	
C*1	51	51	14	15.8	1.0	6.6	−0.6	12.1	0.8
C*2	48	43	27	19.4	1.3	−1.7	0.1	14.7	0.8
C*3	11	11	1	32.3	6.1	18.4	6.1	23.8	6.1
B*1	27	27	14	9.8	0.7	0.1	−0.1	8.2	0.5
B*2	52	48	18	33.8	2.8	3.2	−0.2	24.4	2.4
B*3	6	6	0	38.0		18.2		26.3	
I*1	24	24	8	23.0	1.5	17.5	0.4	17.5	1.0
I*2	6	6	1	30.8	2.4	21.8	−2.4	24.0	2.4
Y*2	6	6	0	87.3		19.9		50.0	
Y*3	1	1	0	13.2		−13.2		13.2	
T	258	241	99	27.6	1.8	5.3	0.0	17.6	1.3

[a]For selected groups encompassing N_{mol} molecules, the number N^{val} of experimental data points, the root-mean-square deviation (RMSD), the average deviation (AVED), and the mean absolute deviation (MAD) are reported for the pure-liquid density ρ_{liq} and the vaporization enthalpy ΔH_{vap}. The letter in the group code stands for the type of halogen atoms present in the molecule: homofluorinated (F), homochlorinated (C), homobrominated (B), homoiodinated (I), hetero-halomethane (X), and hetero-haloalkane with more than one carbon atom (Y). The digit stands for the total number of halogen atoms in the molecule. These two symbols match the first and third symbols employed in the molecule codes used in Section S.3. The line labeled T corresponds to all molecules of the set. This data is illustrated graphically in Figure 8 (bottom row).

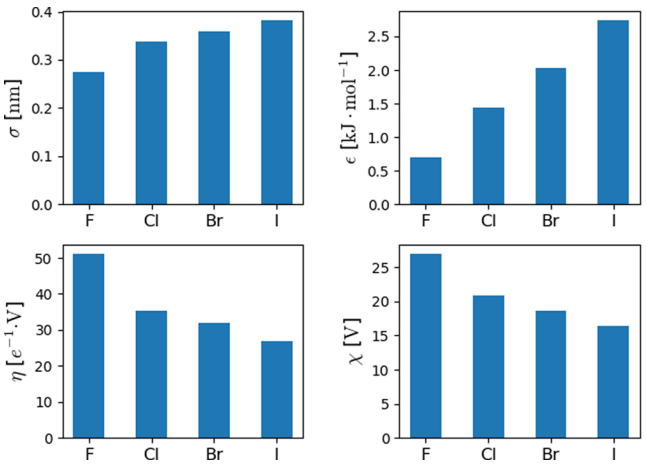


Figure 9. Trends in the nonbonded interaction parameters of the halogen atoms in the optimized force field. The values are shown along the halogen series for the LJ collision diameter σ and well depth ϵ as well as the EE hardness η and electronegativity χ. The values are reported numerically in Table 3. A correlation between the χ values and standard electronegativity scales is provided in Section S.11.

Table 8. Charge of the Halogen Atoms in the 10 Homohalogenated Charge-Groups of the Final Optimized Force-Field[a]

	charge (e)			
type	F	Cl	Br	I
C1X1	−0.247	−0.199	−0.163	−0.123
C2X1	−0.302	−0.270	−0.236	−0.202
C3X1	−0.328	−0.304	−0.270	−0.239
C4X1	−0.331	−0.310	−0.276	−0.245
C2X2	−0.253	−0.214	−0.184	−0.153
C3X2	−0.282	−0.250	−0.219	−0.189
C4X2	−0.289	−0.260	−0.229	−0.199
C3X3	−0.248	−0.213	−0.185	−0.156
C4X3	−0.257	−0.224	−0.196	−0.167
C4X4	−0.231	−0.197	−0.171	−0.144

[a]These charge-groups are identical to the force-field fragments depicted in Figure 3, where all the halogen sites are substituted by the same type of halogen atom.

force-field calibration were performed within 10 K of T^{-}, but still about one-third involved different temperatures, ranging from 145 to 490 K. To investigate the impact of this factor on the accuracy of the force field, additional results are reported in Section S.13. The error in ρ_{liq} and ΔH_{vap} is shown to present no systematic correlation with simulation temperature (Figure S.16). The results of additional simulations are presented for 29 chlorinated and 21 brominated compounds that are liquid at T^{-}, and for which ρ_{liq} and/or ΔH_{vap} are known experimentally for multiple points (or with a fitted curve) between the melting and boiling points. These extra simulations were performed at different temperatures in this range, along with a pressure P°. The error in ρ_{liq} is found to present no systematic correlation with the simulation temperature, while the error in ΔH_{vap} presents a small positive slope of 0.01 kJ·mol·K^{-1} (Figure S.17). These results suggest that the inclusion of liquid-phase data at different state points does not represent a major source of inaccuracy in the present force-field calibration procedure.

4. CONCLUSIONS

The CombiFF approach (Figure 1) is an integrated scheme for the automated refinement of force-field parameters against experimental condensed-phase properties, considering entire classes of organic molecules constructed using a fragment library *via* combinatorial isomer enumeration. The main goal of this article was to describe the workflow of the scheme and to report a first application to the calibration of a GROMOS-based united-atom force field for the saturated acyclic haloalkane family. A key component of the selected force-field representation is that the atomic partial charges are

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

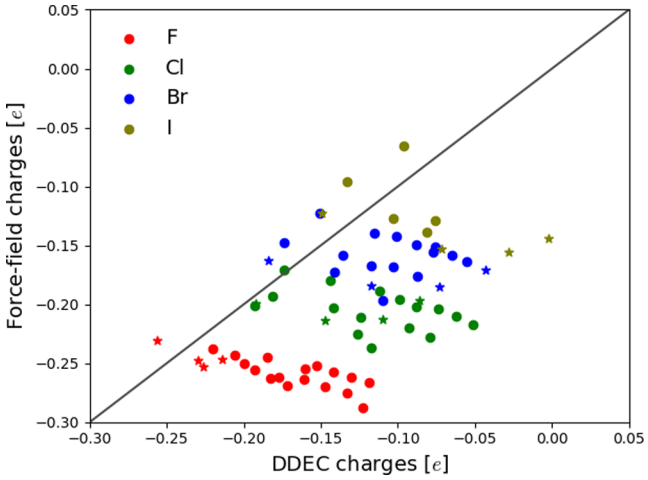


Figure 10. Comparison of atomic partial charges on the halogen atoms considering halomethanes. The charges corresponding to the optimized force field are correlated to those obtained from QM calculations analyzed with the DDEC scheme.[293] The color scheme indicates that the charge is on an atom of the given halogen type. The representative points for the homo-halomethanes (CH_3X, CH_2X_2, CHX_3, and CX_4) are shown as star symbols, and the others as bullets. The corresponding numerical values can be found in Table S.12.

generated using an electronegativity-equalization (EE) scheme, which takes into account induction effects within molecules. As a result, the atom types of the force field are fully characterized by four nonbonded interaction parameters: the LJ collision diameter σ and well depth ϵ, along with the EE hardness η and electronegativity χ.

For the (halo-)alkane force field, 13 atom types are required. Neither the 5 aliphatic atom types nor the 19 covalent interaction parameters were subject to optimization, so that the calibration concerned 32 nonbonded interaction parameters. Their optimization was performed against 409 experimental liquid densities ρ_{liq} and vaporization enthalpies ΔH_{vap} concerning 228 small molecules, corresponding to an observable-to-parameter ratio of 12.8. A collection of 340 additional ρ_{liq} and ΔH_{vap} values concerning 258 other molecules was used for subsequent validation. The optimization of the force field requires less than three days of wall-clock computing time using one processor per simulated system. The solution reached upon optimization is essentially unique for the LJ interaction parameters. However, a larger extent of degeneracy is observed for the EE parameters, with a slower convergence upon iterating and a more pronounced dependence on the starting point of the optimization. This degeneracy leads to a significant variability in the η and χ parameters, whereas the atomic partial charges themselves are less affected.[282] The nonuniqueness of some force-field parameters does not necessarily represent an issue, unless it results from an incompleteness of the target data. As an extreme case of electrostatic degeneracy, it is noted that a force field and its opposite-charge version ($q \rightarrow -q$) are entirely indistinguishable in terms of simulated properties.

The optimized (final) force-field parameters are reported in Tables 3 and 4. They lead to RMSD (AVED) values of 49.8 (1.2) kg·m^{-3} for ρ_{liq} and 2.7 (0.0) kJ·mol^{-1} for ΔH_{vap} over the calibration set. The corresponding values over the validation set are 27.6 (5.3) kg·m^{-3} and 1.8 (0.0) kJ·mol^{-1}, respectively. Thus, in terms of the two properties considered, a remarkable level of agreement with experiment can be achieved using a simple united-atom force field without any specific description of σ-holes and halogen bonding. Of course, an improved representation[238−258] of the anisotropy in the halogen electron density might become important when considering other systems and properties, *e.g.*, mixtures of haloalkanes with more polar liquids, haloarenes or, in the context of drug design, specific interactions between halogenated molecules and proteins.[216,217,219,221,222,224−228,236,237]

The deviations from experiment tend to be smaller for the compounds involving chlorine and bromine compared to those involving fluorine or iodine. This is expected considering that force-field errors are likely to affect more adversely the extremes as opposed to the median of a chemical series (see also refs 283−285 for specific issues related to fluoroalkanes). The discrepancies are also reduced upon increasing the aliphatic content of the molecule. This is also easily understood, as the pairwise interactions between aliphatic groups are comparatively easier to describe and already very well reproduced in the GROMOS force field.[286−288] It also explains why the deviations from experiment are smaller for the validation set (larger molecules) compared to the calibration set (smaller molecules). Finally, the use of about one-third of experimental liquid-phase observables corresponding to *P,T*-points differing from ambient conditions does not seem to represent a major source of inaccuracy in the present force-field calibration. Should it become an issue for other compound families, one could envision the development of temperature-dependent force-field parameters.[295−299]

The trends in the optimized parameters along the halogen series are in line with chemical intuition in terms of size (σ), polarizability/dispersivity (ϵ), hardness (η), and electronegativity (χ). The trends in the atomic partial charges upon alterations of the molecular structure also reproduce well those expected based on considerations related to electronegativity, polarizability, induction, and hyperconjugation. This observation is particularly remarkable considering that the force-field calibration did not involve any QM calculation. Thus, provided that a sufficiently large compound family is considered, the information related to electronic effects within the molecules appears to be unambiguously encoded within the experimental ρ_{liq} and ΔH_{vap} values.

Clearly, some of the force-field representation choices will have to be reconsidered in future work, and the parameter optimization repeated accordingly. These choices include in particular (i) the bond-stretching and bond-angle bending parameters (selected here in a rather *ad hoc* fashion); (ii) the torsional and third-neighbor interaction parameters (relying on a single common type for CCCC, CCCX, and XCCX torsions); (iii) the nonbonded interaction parameters of the aliphatic (united-)atom types (taken directly from GROMOS and not reoptimized); (iv) the use of generic atom types for the halogenated carbon (united-)atoms (possibly in part responsible for the slightly poorer description of the fluorine and iodine compounds); (v) the restriction of EE charge flows to CX bonds only (permitting the definition of neutral charge-groups of at most five atoms); and (vi) the absence of any specific representation for σ-holes and halogen bonding (see above).

Another shortcoming of the CombiFF optimization performed in this work is that stereochemistry has been entirely disregarded. Although the precise control of stereochemistry is straightforward in the simulations, experimental data that is unambiguously resolved in terms of stereo-

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

chemistry is extremely scarce, with the possible exception of cis−trans isomerism for double bonds. Thus, as a first step, simulation may be used to investigate the variability of the thermodynamic properties when considering pure enantiomers, racemic mixtures, or mixtures of diastereomers. This will provide an estimate for the error incurred by neglecting stereochemistry during the force-field calibration.

In the long term, another shortcoming of CombiFF will be that it is only applicable as such to the calibration of fragments present in molecules for which sufficient condensed-phase experimental data is available. In other situations, a combination with other parameter-calibration methods will be required. For example, the missing fragments may be parametrized based on QM data, using procedures (*e.g.*, machine learning) designed to reproduce the optimized CombiFF parameters for the existing (experimentally derived) fragments.

Another important issue in CombiFF is related to the definition of the EE scheme. The most basic EE schemes consider molecules as conductor-like entities. They enable arbitrary charge transfers between atoms irrespective of their topological distance,[278,282,300,301] which results in particular in an incorrect (nonlinear) scaling of the polarizability with molecule size.[302,303] However, EE variants have also been developed that compromise between conductor- and insulant-like behaviors, and restore a correct polarizability scaling. In the present work, a very strong damping has been applied by restriction of EE charge flows to CX bonds only. This permits the definition of neutral charge-groups of at most five atoms, but prevents further charge delocalization to possible neighboring carbon atom(s). In the future, however, CombiFF will be redefined in terms of an EE scheme with damped charge transfers[304] throughout the entire molecule, also involving consideration of bond orders,[304,305] conjugation,[300] and/or mesomery.[306] Rather than readjusting the charges to define neutral charge-groups[165,166] for cutoff truncation, these EE charges will be used within an atomic-cutoff scheme, in combination with either a lattice-sum[307−310] or a reaction-field[311,312] approach to handle the long-range electrostatic interactions. A possible further development would be the design of a polarizable force field of the fluctuating-charge type,[305,313−322] where an on-the-fly EE scheme would incorporate the effect of the configuration-dependent (*i.e.*, local and instantaneous) electric potential on the atomic partial charges during the MD simulation. A particularly appealing feature of such an approach is that it would not require any additional force-field parameters, but a mere CombiFF recalibration of the existing parameters under application of the fluctuating-charge scheme.

The CombiFF workflow is only partly automated. In particular, the data selection and curation, which are of key importance, remain extremely expensive in terms of human time. Work is in progress to try to enhance the exhaustiveness, automation, and reliability (error assessment) of this operation. In contrast, the other steps (compound enumeration, topology creation, optimization procedure) necessitate little human time, and their requirements in terms of computer time remain modest as well (a few days). For this reason, one of the key advantages of CombiFF is that once the time-consuming task of target-data selection has been performed, the optimization can very easily be repeated for any alternative choice of functional-form representation within the force field. In other words, it becomes possible to assess the effect of functional-form decisions on the accuracy at an optimal level of parametrization. For example, one may assess intrinsic accuracy changes related to[215] (i) a change in the cutoff distance; (ii) the use of an alternative functional form for the van der Waals repulsion; (iii) the possible inclusion of a long-range LJ correction; (iv) a change of the combination rules; (v) a change in the model resolution, *e.g.*, united-atom to all-atom; and (vi) the inclusion of off-center sites to represent σ-holes.

Work is also in progress to expand the CombiFF calibration process so as to include other chemical families and liquid properties, as well as properties concerning the real-gas and solid phases.

A. METHODOLOGICAL DETAILS

This appendix complements Section 2 in providing a more detailed description of the CombiFF scheme (Figure 1) and its application to the saturated haloalkane family.

A.1. FAMILY SELECTION

The 11 nonoverlapping families A−K of haloalkanes considered here are listed in Table 2. These molecules are fully saturated (no cycle, no multiple bond) and correspond to linear or branched alkanes with one to four hydrogen atoms substituted by a halogen atom. The chemical formulae and SMILES-strings of the N_{iso}^{tot} = 755 301 constitutional isomers (isomer-enumeration file) were generated automatically (ENU program), and the selection of the N_{iso}^{sim} = 486 compounds with experimental data for ρ_{liq} and/or ΔH_{vap} was made based on our experimental database (DBS scripts). The union of families A−H defines the calibration set of N_{iso}^{cal} = 228 molecules employed to optimize the force-field parameters. The union of families I−K defines the validation set of N_{iso}^{val} = 258 molecules subsequently employed to test the accuracy of the force field. The molecules of these two sets are displayed in Figures S.2 and S.3, respectively.

The calibration set is constructed to cover sufficiently the N_{frg} = 15 fragments (Figure 3) and N_{att} = 13 atom types (Table 3) relevant for the saturated haloalkanes. Families A and B cover the halomethanes, while families C and D cover the haloalkanes with two to four carbon atoms. The four other families E, F, G, and H introduce or improve the coverage of specific halogenated carbon environments, by encompassing secondary mono-haloalkanes, tertiary mono-haloalkanes, and molecules with one or more halogen atoms vicinal to a *tert*-butyl group. The validation set consists of families I−K, which include molecules with up to 10 carbon atoms and 1−3 halogen atoms of the same type, molecules with up to 9 carbon atoms and 2−4 halogen atoms of two different types, and molecules with 10 carbon atoms and 2 halogen atoms of different types.

A.2. EXPERIMENTAL-DATA VECTOR

Considering the 11 families of Table 2, the enumeration (ENU program) delivered a list of N_{iso}^{tot} = 755 301 constitutional isomers with associated chemical formulae and SMILES-strings (isomer-enumeration file). To extract values for ρ_{liq} and ΔH_{vap} from experimental sources (articles, databases), these strings must be mapped to equivalent identifiers such as molecule names (including synonyms), Chemical Abstract Service registry numbers (CAS), IUPAC International Chemical Identifiers (InChI), or associated key identifiers (InChIKey).

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

This mapping was performed by searching for the SMILES-strings in the PubChem database[323] *via* its Application Programming Interface (API). From this source, the molecular formula (for verification), name, CAS number (possibly multiple entries), and InChiKey were collected. This procedure allowed for the identification of 4538 compounds, *i.e.*, only 0.6% of the N_{iso}^{tot} molecules. Note that the SMILES-strings were not canonicalized in this comparison (canonicalization would have led to the identification of 1033 extra compounds).

Based on the identifiers collected, experimental data was queried from our in-house database (DBS scripts). In the context of ρ_{liq} and ΔH_{vap} for saturated haloalkanes, the main sources accessed *via* DBS are (i) the book of Yaws[271] providing fitted equations for ρ_{liq} and ΔH_{vap} as a function of T for 24 268 organic compounds; (ii) the review of Acree[272] providing a compilation of phase-change enthalpies for 6545 organic compounds with 1−10 carbon atoms; (iii) the handbook of Chemistry and Physics[273] (online edition 2018) providing physical constants for 10 867 organic compounds; (iv) the book of Frenkel et al.[274] in the Landolt-Börnstein series providing data points and fitted equations giving ρ_{liq} as a function of T for 937 halohydrocarbons; and (v) the ThermoML database[275] (online) providing experimental thermophysical properties for more than 5500 compounds. Although the measurement temperature is generally unambiguously specified in these sources, this is not always the case for the measurement pressure. In addition, it is not always clear whether ΔH_{vap} refers to a transition from the liquid to the real-gas state or to the ideal-gas state, the latter involving the application of a real-gas (*e.g.*, virial-based) correction. A similar issue arises for ρ_{liq}, where the pressure may be unclear, *e.g.*, 1 bar *vs* vapor pressure of the liquid.

The resulting values were plotted as a function of temperature (single points or empirical curves) using an automated script. Six illustrative examples of such graphs are shown in Figure 2. Panel (a) illustrates the favorable situation where the molecule is liquid at room temperature and atmospheric pressure, there are several experimental data points and/or fitted equations available, and they are in good agreement with each other. In panel (b), the molecule is also liquid under ambient conditions, but there is an experimental data point or a fitted curve from only one source. Panels (c) and (d) illustrate the case where the molecule is gaseous or solid, respectively, at room temperature. Finally, panels (e) and (f) show two cases where the experimental data reported by different sources are conflicting.

Based on the inspection of these graphs, a single value of ρ_{liq} and/or ΔH_{vap} (in some cases two) was then selected manually for each of the N_{iso}^{sim} = 486 molecules. As much as possible, a single value was retained corresponding to a temperature T at (or very close to) T^{-} = 298.15 K and a pressure P at (or very close to) P° = 1 bar. In the absence of such a point, values at lower or higher temperatures had to be selected. This applied in particular to compounds that are not liquid under ambient conditions. For compounds that are gaseous under ambient conditions, two data points at different temperatures were generally selected, one at room temperature, implying a pressure above 1 bar, and the second one at a lower temperature, typically the boiling point at 1 bar. For compounds that are solid under ambient conditions, a data point at a higher temperature was selected, the lowest possible given the available data, which is generally at or just above the melting point (although the value of the latter was not always available). In the above cases, whenever possible, the pressure in the simulation was adjusted to the vapor pressure of the compound at the given temperature, as obtained from Yaws.[324] However, this was not done for the ΔH_{vap} values based on Acree,[272] where no pressure is indicated, and a default value of 1 bar was used in the simulations. In the presence of multiple options (*e.g.*, data available from more than one source, at different state points and/or conflicting across sources), a "most reasonable" choice was made whenever possible. The most dubious experimental data was discarded. More information on these curation choices can be found in Section S.4.

The P,T-points retained, along with the corresponding experimental values of ρ_{liq} and ΔH_{vap}, are listed in Section S.5 (Table S.5). This section also suggests error estimates of about ± 10.8 kg·m^{-3} for ρ_{liq} and about ± 1.7 kJ·mol^{-1} for ΔH_{vap}, based on standard deviations over multiple experimental values for a subset of molecules. For the calibration set, the procedure employed allowed for the identification of N_{ρ}^{cal} = 243 and $N_{\Delta H}^{cal}$ = 166 values for ρ_{liq} and ΔH_{vap}, respectively. The corresponding numbers for the validation set are N_{ρ}^{val} = 241 and $N_{\Delta H}^{val}$ = 99, respectively. As a result, the experimental-data vectors considered in the calibration and validation have N_{exp}^{cal} = 409 and N_{exp}^{val} = 340 components, respectively. The corresponding vector $\mathbf{X}^{exp}$ for the calibration set was stored in the target-data file and used subsequently to optimize the force field. A corresponding file was also made for the validation set and subsequently used to assess the quality of the force field. Because ρ_{liq} and ΔH_{vap} values may be available simultaneously for a given compound at the same P,T-point, the numbers of independent simulations to be carried out are smaller, namely, N_{sim}^{cal} = 332 and N_{sim}^{val} = 300 simulations for the calibration and validation sets, respectively.

The above procedure to select the target experimental data has a number of shortcomings, namely (i) it does not guarantee an exhaustive search (*e.g.*, compounds may be missed if they are absent in PubChem or if an experimental source relies on an unrecognized or erroneous identifier); (ii) it requires a significant extent of manual data curation (selection of appropriate state points and of reliable values); (iii) it involves only limited error checking and uncertainty assessment; and (iv) it can be subject to various sources of ambiguities and errors (*e.g.*, incomplete specification of the thermodynamic state point, approximations involved in a determination method, standard-state definition employed in the source, erroneous compound and/or value). Work is in progress to enhance the exhaustiveness, automation, and reliability of this operation.

A.3. FORCE-FIELD REPRESENTATION

The force-field representation employed here is essentially compatible with that of the GROMOS force field in its 2016H66 variant.[277] The main principles underlying the GROMOS force fields[215,277,325−329] can be summarized as follows.

(A) United-atom representation[330] of aliphatic CH, CH_2, CH_3, and CH_4 groups.[286−288]

(B) Covalent force-field terms including quartic (or harmonic, or constrained) bond-stretching, cosine-harmonic (or angle-harmonic) bond-angle bending,

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

harmonic improper dihedral-angle distortion, and cosine-series torsional dihedral-angle rotation.

(C) LJ representation[81] of the van der Waals interactions.

(D) Nonbonded exclusion of first and second covalent neighbors.

(E) LJ interaction reduction for third covalent neighbors using a special set of parameters, along with unaltered electrostatic interactions.

(F) Application of a geometric-mean combination rule[279,280] for pairwise LJ interaction parameters, distinguishing between non-hydrogen-bonding, uncharged hydrogen-bonding, and oppositely charged interactions.

(G) Freely adjustable atomic partial charges, *i.e.*, not determined by the LJ atom type.

(H) Compatibility with the simple point charge (SPC) water model.[331]

(I) Compatibility with reaction-field electrostatics[332,333] and LJ interactions excluding a long-range correction, based on an effective cutoff distance of 1.4 nm applied to charge-group pairs.

Note that rules D−F represent default rules in GROMOS, which can be overridden in exceptional cases. A more complete description of the functional form of the GROMOS force field can be found in refs 325−328, 334, and 335.

In the present study, rule H is irrelevant, since water is not involved. For simplicity, rule E is also bypassed for the LJ interactions involving a halogen atom or a halogenated carbon (united-)atom. In these cases, a single set of LJ parameters is used for all interactions, including for third neighbors (along with normal electrostatic interactions). Finally, rule G is modified by introducing an electronegativity-equalization scheme to calculate the atomic partial charges as detailed in Appendix A.4. The latter scheme does rely on atom types to determine the hardness and electronegativity of an atom.

In total, N_{att} = 13 atom types are defined in Table 3 for representing all saturated (halo-)alkanes *via* the fragments of Figure 3, namely, halogen atoms (four types), nonhalogenated carbon (united-)atoms (five types), and halogenated carbon (united-)atoms (four types). Each atom type is associated with four nonbonded interaction parameters: the LJ collision diameter σ and well depth ϵ, along with the EE hardness η and electronegativity χ.

Since the main focus of this work is the parametrization of the nonbonded interactions for the halogen functions, neither the purely aliphatic component of the force field nor the covalent interaction parameters are subject to optimization. Consequently, the LJ coefficients associated with the four atom types CH0 to CH4 in Table 3 are taken directly from the GROMOS force field and kept unaltered throughout this work. The covalent interaction parameters of Table 4 are either ported from those of the 2016H66 force field[277] or, whenever unavailable, inferred based on the results of QM calculations as described in Appendix A.6, and kept unaltered as well.

A.4. ATOMIC PARTIAL CHARGES *VIA* ELECTRONEGATIVITY EQUALIZATION

The force fields of the GROMOS family are characterized by freely adjustable atomic partial charges (Appendix A.3, point G). Within the fragment-based approach employed here, one might assign a fix partial charge to each atom in every fragment, and simply add the charges of the overlapped atoms when linking two fragments. However, this approach would account for induction effects in a rather crude way. In practice, it is better to rely on atomic parameters that are less sensitive to the chemical environment of an atom compared to the charge itself. For this reason, the charge-assignment procedure employed in CombiFF is based on an electronegativity-equalization (EE) scheme.

The generic concept of EE encompasses several closely related methods aiming at (i) deriving atomic partial charges taking into account induction effects based on atom and/or bond parameters; (ii) modeling conformation-dependent charge fluxes within molecules; and (iii) modeling environment-induced charge polarization effects within molecules. The EE family of schemes includes in particular the partial equalization of electronegativity[336−343] (PEOE) approach, the CHARGE method,[344] the (full) electronegativity-equalization method[345−347] (EEM), the charge equilibration[313,314] (Qeq) or fluctuating-charge[305,315−322] (FQ) approaches, the chemical potential equalization[303,348−353] (μEq) approach, the atom-bond electronegativity-equalization method[305,354−358] (ABEEM$\sigma\pi$), the Kirchhoff charge model[99,359,360] (KCM), the atom−atom charge transfer[300] (AACT) model, the VC method,[306] the charge transfer with polarization current equilibration[302,361] (QTPIE) scheme, the split-charge equilibration[278,282,301,362−365] (SQE) method, the extended charge equilibrium[366,367] (EQeq) approach, the bond-capacity[304,368,369] (BC) model, and the atom-condensed Kohn−Sham density functional theory (DFT) model approximated to second order[370] (ACKS2). All these schemes rely on Sanderson's EE postulate[371−373] and differ in three main aspects, namely (i) their parameter spaces (*e.g.*, atom and/or bond parameters, consideration of bond types,[304,305] conjugation,[300] and/or mesomery,[306] possible inclusion of off-atom virtual sites[304]); (ii) whether and how the charge redistribution is damped as a function of the topological distance between atoms;[278,282,300−304] and (iii) the way their parameters are calibrated. Generally, EE schemes are used as fitting devices to approximately reproduce QM charge densities or electrostatic potentials, or partial charges derived from these properties.[278,282,301,306,374,375] In contrast, here, the parameters of the EE scheme are optimized to reproduce target condensed-phase properties *via* MD simulations involving the resulting charges.

In practice, an EE type is assigned to each atom within a fragment, which defines two basic properties of the corresponding element in the specific chemical environment. The first one is the hardness η, accounting for the resistance of the neutral atom to have a partial charge deviating from zero. The second one is the electronegativity χ, accounting for the strength with which the neutral atom attracts an extra negative charge. For simplicity, the LJ types (defining σ and ϵ) and EE types (defining χ and η) are jointly assigned to so-called atom types (defining the four parameters simultaneously).

For a molecule of N atoms, the atomic partial charges q_i are calculated based on the η_i and χ_i values of the involved atoms, by minimizing the approximate electronic energy of the isolated molecule in a rigid conformation with respect to possible charge flows between its constituting atoms. This approximate energy is defined as[278]

$$E_{mol}(\mathbf{q}) = \sum_{i=1}^{N}\left(\frac{1}{2}\eta_i q_i^2 + \chi_i q_i\right) + \sum_{i=1}^{N}\sum_{j=i+1}^{N} J_{ij} q_i q_j \quad (A.1)$$

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

The minimization is performed with an electroneutrality constraint on the total charge of the molecule

$$Q_{tot}(\mathbf{q}) = \sum_{i=1}^{N} q_i = 0 \tag{A.2}$$

The pairwise coefficients J_{ij} account for the through-space electrostatic interactions between the excess charges on different atoms. A simple form based on the overlap of Gaussian-shaped clouds[278] is used here

$$J_{ij} = C\frac{1}{\overline{r}_{ij}}\mathrm{erf}\left(\frac{\overline{r}_{ij}}{(\sigma_i^2 + \sigma_j^2)^{1/2}}\right) \tag{A.3}$$

where $C = (4\pi\epsilon_o)^{-1}$, ϵ_o is the permittivity of free space, erf is the error function, $\overline{r}_{ij}$ is an effective interatomic distance, and σ_i and σ_j are the Gaussian widths. For simplicity, these are chosen equal to the corresponding LJ collision diameters. If E_{mol} is given in units of e·V, the natural units of the quantities η and χ are e^{-1}·V and V, respectively, while the value of C in eq A.A.3 is 1.44 e^{-1}·V·nm. It should be noted that the absolute scale of these quantities is only defined by the inclusion of the terms J_{ij}. If these terms were absent, χ and η could be scaled by a common arbitrary factor without affecting the resulting charge set. Since the formulation of these J_{ij} is rather approximate, the relative values of χ and η across atom types are chemically meaningful, but the absolute magnitudes of these quantities should not be overinterpreted.[282]

In practice, the minimization of E_{mol} with the constraint of Q_{tot} according to eqs A.1 and A.2 amounts to solving the matrix equation[374−377]

$$\begin{pmatrix} \eta_1 & J_{12} & \cdots & J_{1N} & -1 \\ J_{21} & \eta_2 & \cdots & J_{2N} & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ J_{N1} & J_{N1} & \cdots & \eta_N & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \overline{\chi} \end{pmatrix} = \begin{pmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_N \\ Q_{tot} \end{pmatrix} \tag{A.4}$$

where $\overline{\chi}$ is the equilibrated electronegativity of the molecule and $Q_{tot} = 0$.

One of the known issues of the EE scheme is that it allows for charge transfers independently of the topological distance between atoms,[278,282,300−304] *i.e.*, molecules behave as perfect conductors. To remedy this problem in an uncomplicated fashion, charge transfers were only allowed here between a halogenated carbon atom (atom types CX1, CX2, CX3, or CX4) and the directly bonded halogen atoms (atom types F, Cl, Br, or I), thereby producing neutral charge-groups of two to five atoms. For the corresponding through-space interactions between charges at a topological distance of one or two bonds, the effective interatomic distances $\overline{r}_{ij}$ in eq A.3 are selected according to the reference distances and angles of the covalent force-field terms in Table 4. The aliphatic united-atoms (types CH0−CH4) are not involved in charge transfers (zero charge) and treated as individual charge-groups of a single atom.

A.5. TOPOLOGY BUILDING

The construction of a molecular topology using the TBL program is illustrated in Figure 4. A fragment corresponds to an atom group in a specific connectivity. It must include one or more core atoms and may include one or more link atoms, with the rule that every link atom has exactly one bond in the fragment, which must connect it to a core atom of the fragment. If a fragment has only core atoms, it is a molecule. Otherwise, it must be linked to other fragments. A linkage between two fragments is specified by selecting a link atom in each of the two fragments and is performed by overlapping the corresponding core−link bonds of the two fragments. A substitution affecting a fragment corresponds to assigning the type of a link atom from a set of allowed choices, specified in its substitution list. For example, it is possible to substitute a link atom labeled X by an arbitrary halogen atom from the substitution list F, Cl, Br, or I.

The fragment-definition file describes the fragments (core atoms, link atoms, and associated substitution lists, along with the types of the covalent term). In the context of saturated haloalkanes, the N_{frg} = 15 fragments shown in Figure 3 are employed. The five fragments on the left are those required for alkanes, and the 10 fragments on the right correspond to halogenated groups (using substitution to handle the different halogen types). The set is complete for all saturated alkanes and haloalkanes although it disregards the stereochemistry. The linkage-definition file defines possible modifications operated upon linking two specific fragments in a given connectivity (terms to be added or types to be modified). This file is created automatically based on default rules but can be edited manually to introduce exceptions if required (not necessary in the present case).

A simple example of fragment-based assembly is shown in Figure 4. The molecule 2-fluoro-2-chloro-5-bromo-6-methylheptane is defined by its SMILES-string CC(F)(Cl)CCC(Br)C(C)C. It can be decomposed using five of the fragments of Figure 3, namely, C1, C2, C3, C4X2, and C3X1, using seven linkages and three substitutions. The corresponding information is stored in an assembly specification file, where all required fragments, linkages, and substitutions are listed. Note that the decomposition is not always unique. Different decompositions may result in topologies that differ in the ordering of the atoms and force-field terms but are otherwise equivalent. However, the SMILES-strings generated by the ENU programs are canonical, and the TBL program can also canonicalize other input SMILES-strings upon request. When this is done, the generated topologies are rigorously unique.

A.6. PARAMETER VECTOR

Given the GROMOS-like united-atom force-field representation described in Appendix A.3, including in particular the use of a geometric-mean combination rule for ϵ and σ (or, equivalently, C_6 and C_{12}), the application of the EE scheme of Appendix A.4 for the charges, involving the specification of χ and η for each atom type, and the N_{att} = 13 atom types listed in Table 3, the nonbonded component of the force field for saturated (halo-)alkanes is entirely specified by N_{prm}^{att} = 56 parameters. To these, one should add the N_{prm}^{cov} = 19 covalent interaction parameters listed in Table 4. This results in N_{prm}^{tot} = 75 parameters for the full force-field specification. In practice, however, the number of parameters to be optimized was reduced taking the following into account.

First, the 24 LJ interaction parameters of the aliphatic united-atoms (CH0−CH4 in Table 3), used for non-halogenated carbon atoms, were not altered relative to the current GROMOS parameters (identical in the 54A7 set[329] and the 2016H66 set[277]). These parameters have been

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

previously optimized[286,288] against experimental data for ρ_{liq} and ΔH_{vap}, as well as solvation free energies in water and cyclohexane. In addition, these united-atoms are kept neutral so that no η and χ values are required.

Second, the 19 covalent interaction parameters were not optimized. They were ported from those of the 2016H66 force field[277] or, whenever unavailable, inferred as follows. The reference bond lengths were set to an average over CCD/6-311G* optimized geometries available in the Computational Chemistry Comparison and Benchmark Database[378] (CCCBDB). The reference bond angles were all set to the ideal tetrahedral value of 109.5°. The corresponding force constants were selected following chemical intuition. For the torsional-dihedral interactions, a generic form was adopted for both the CCCC and the CCCX torsion, involving a threefold potential applied only once for a given central bond, with parameters equal to those of the generic CCCC torsion in GROMOS. Special third-neighbor LJ interaction parameters were only used for the CH0−CH3 atom types.

Following from the above choices, the parameter vector for optimization includes $N_{prm}^{cal} = 32$ parameters. Initial values for these parameters (*i.e.*, values at iteration 0) were either ported from those of the 2016H66 force field[277] or inferred by intuition. The initial values for the nonbonded parameters are provided in Table S.6.

The coverage of the 13 atom types by the molecules of the calibration and validation sets, including availability of experimental ρ_{liq} and ΔH_{vap} values, is detailed in Table 5. Note that the representation of the aliphatic atom types CH0−CH4 is irrelevant here, as these are not subject to parameter optimization, and that the halomethanes are all included in the calibration set. Besides this, both sets appear to appropriately cover all atom types.

A.7. SYSTEMATIC PARAMETER OPTIMIZATION

The force-field optimization problem consists of finding a parameter vector $\mathbf{P}$ leading to a simulated-result vector $\mathbf{X}^{sim}$ that is as close as possible to the experimental-data vector $\mathbf{X}^{exp}$. This can be formulated as the minimization of the objective function $Q(\mathbf{P}; \mathbf{X}^{exp})$ defined by eq 1.

To guide the parameter optimization, the Taylor-series expansion of eq 2 is introduced. In this equation, $\underline{\mathbf{S}}^{sim,o} = \underline{\mathbf{S}}^{sim}(\mathbf{P}^o)$ is the sensitivity matrix, collecting in its lines the derivatives of the simulated observable n for molecule m with respect to the N_k parameters of the force field, measured at point $\mathbf{P}^o$ in parameter space. The corresponding matrix element corresponding to a parameter k is given by

$$S^o_{nm,k} = \left(\frac{\partial X^{sim}_{nm}}{\partial P_k}\right)_{\mathbf{P}=\mathbf{P}^o} \tag{A.5}$$

The evaluation of the sensitivity matrix thus requires determining the derivative of a given ensemble-average observable with respect to all force-field parameters during a simulation, using appropriate statistical−mechanical expressions.[175,178−180,281]

In the NPT ensemble, and dropping the n,m indexes for simplicity, an ensemble-average observable $X^{sim} = X^{sim}(\mathbf{P})$ can be written

$$X^{sim} = \langle \mathcal{X}^{sim} \rangle = Z^{-1} \int d\mathcal{V}\, d\mathbf{r}\, d\mathbf{p}\, \mathcal{X}^{sim} \exp[-\beta(\mathcal{H} + \tilde{P}\mathcal{V})] \tag{A.6}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the Cartesian coordinate and momentum vectors of the system, respectively, $\mathcal{V}$ is the volume, $\mathcal{X}^{sim} = \mathcal{X}^{sim}(\mathbf{r}, \mathbf{p}, \mathcal{V}; \mathbf{P})$ is the instantaneous observable corresponding to X^{sim}, $\mathcal{H} = \mathcal{H}(\mathbf{r}, \mathbf{p}, \mathcal{V}; \mathbf{P})$ is the Hamiltonian, $\tilde{P}$ is the pressure, $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T is the absolute temperature, and Z is the partition function

$$Z = \int d\mathcal{V}\, d\mathbf{r}\, d\mathbf{p} \exp[-\beta(\mathcal{H} + \tilde{P}\mathcal{V})] \tag{A.7}$$

The statistical−mechanical expression for the corresponding derivative is[175,178−180,281]

$$\frac{\partial X^{sim}}{\partial P_k} = \left\langle \frac{\partial \mathcal{X}^{sim}}{\partial P_k} \right\rangle + \beta \left(X^{sim} \left\langle \frac{\partial \mathcal{H}}{\partial P_k} \right\rangle - \left\langle \mathcal{X}^{sim} \frac{\partial \mathcal{H}}{\partial P_k} \right\rangle \right) \tag{A.8}$$

In the case of ρ_{liq}, the instantaneous observable is related to the volume of the pure liquid, and the first term of eq A.8 is zero. For ΔH_{vap}, the instantaneous observable is related to the Hamiltonian itself. It is computed by subtracting the results of eq A.8 for the pure liquid and the ideal gas, and the first term in eq A.8 must be taken into account in both cases. Application of eq A.8 requires the derivative of the potential energy with respect to all force-field parameters. The derivatives with respect to σ, ϵ, and all atomic charges q_i are computed by the MD program SAMOS. At each iteration of the force-field calibration, the derivatives with respect to η and χ are calculated based on the EE model (eq A.1) by propagation of the derivatives with respect to the charges.

A.8. SIMULATION PROTOCOLS

The simulations were performed using a home-developed GROMOS-compatible program in C++ called SAMOS. The calculation of ρ_{liq} requires a simulation of the pure liquid, while that of ΔH_{vap} requires in addition a simulation of the ideal gas.

The pure-liquid simulations relied on molecular dynamics (MD) and were carried out under periodic boundary conditions based on cubic computational boxes containing 512 molecules (box-edge lengths ranging between 3.2 and 6.1 nm). They were performed in the isothermal−isobaric ensemble at the reference pressures P and temperatures T listed in Table S.5. Most of the simulations (70%) in the calibration set were performed at or close to the standard temperature $T^- = 298.15$, within 10 K. The temperature in the remaining simulations ranged from 145 to 490 K. Most of the simulations (66%) were carried out at or close to atmospheric pressure $P^\circ = 1$ bar, within 0.2 bar. The pressure in the remaining simulations ranged from 0.002 to 45.1 bar. The temperature was maintained close to its reference value using a Nosé−Hoover thermostat[379] with a coupling time of 0.1 ps. The pressure was maintained close to its reference value using an Andersen barostat[380] with a coupling time of 0.5 ps. The isothermal compressibility involved in the calculation of the piston mass[381] for the pressure-coupling scheme was set to 4.575×10^{-4} mol·nm^3·kJ^{-1}, for compatibility with the standard GROMOS setup. The use of a weak-coupling[382] thermostat and/or barostat was found inappropriate for the evaluation of eq A.8 (incorrect fluctuations).

The ideal-gas simulations relied on stochastic dynamics[383−387] (SD). They were also carried out under periodic boundary conditions based on cubic computational boxes containing 512 molecules, with all intermolecular interactions

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

turned off. They were performed in the canonical ensemble, at the reference temperatures T listed in Table S.5. Both the box-edge lengths and the pressure are irrelevant. The friction coefficient was set to 1 ps^{-1}. Note that the choice of a specific value for this coefficient affects the dynamics of the system but not the averages and fluctuations of the calculated thermodynamic quantities.

Newton's equations of motion (MD) or Langevin's equations of motion (SD) were integrated using the leap-frog scheme[383] with a time step of 2 fs. Constraints on all bond lengths were enforced by application of the SHAKE procedure[388] with a relative geometric tolerance of 10^{-4}. The nonbonded interactions were calculated using a twin-range scheme[389] based on charge-group distances with short- and long-range cutoff radii set to 0.8 and 1.4 nm, respectively, and an update frequency of five time steps for the short-range pairlist and intermediate-range interactions. For the ideal-gas simulations, the nonbonded interactions are only intramolecular and no long-range correction is applied. For the pure-liquid simulations, a reaction-field correction[332,333] was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance. The corresponding static relative dielectric permittivities ε were set according to the experimental values, whenever available. Otherwise, the calculations relied either on the average of ε over molecules with the same chemical formula or on the value of ε for an analogous molecule. The employed values are reported in Table S.5. The reaction-field self-term and excluded-atom-term contributions[390] to the energy, forces, and virial were included as described in ref 328.

The liquid density ρ_{liq} was calculated by dividing the total mass of the computational box by its volume. The vaporization enthalpy ΔH_{vap} was calculated by subtracting the average potential energy per molecule in the pure-liquid simulation (MD) from the corresponding value for the ideal gas (SD), both simulations being performed at the same temperature T, and increasing the result by RT to account for the volume−pressure enthalpy contribution of the ideal gas. The term $M_{mol}\rho_{liq}^{-1}$ accounting for the volume−pressure enthalpy contribution of the pure liquid, where M_{mol} is the molar mass of the compound, is in all cases very small and was neglected.[391−394]

Each optimization step consisted of 0.3 ns equilibration with the updated parameters, followed by 0.9 ns production. Each iteration started from the final configurations of the previous iteration as initial configurations, unless the system vaporized. The calculation of ρ_{liq} and ΔH_{vap} was performed on the flight, using statistics accummulated every time step. For ΔH_{vap}, the SAMOS program actually simulates the system in the two phases in parallel. No coordinate trajectory was written to file as these would represent massive and unnecessary data.

In view of the long total simulation times and the fast conformational relaxation of the compounds, the choice of initial configurations for the first iteration is not very important. However, given the large number of molecules considered, it has to be automated (homemade scripts). This involved a low-energy conformation of the molecule, replicated 512 times with random positions and orientations to fill the computational box, followed by energy minimization and 0.3 ns equilibration in the appropriate phase.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00683.

Saturated haloalkane family (relevance of stereochemistry, number of isomers, compounds in the calibration and validation sets); experimental data (curation and selection, reference values); optimization protocol (initial values, extended and additional optimizations); comparison with experiment (experimental *vs* simulated properties, outliers, comparison with standard electronegativity scales, comparison with DDEC charges, influence of the state point) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Philippe H. Hünenberger − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland;* orcid.org/0000-0002-9420-7998; Phone: +41 44 632 5503; Email: phil@igc.phys.chem.ethz.ch

Authors

Marina P. Oliveira − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland*

Maurice Andrey − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland*

Salomé R. Rieder − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland*

Leyla Kern − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland*

David F. Hahn − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland;* orcid.org/0000-0003-2830-6880

Sereina Riniker − *Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland;* orcid.org/0000-0003-1893-4031

Bruno A. C. Horta − *Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, Brazil;* orcid.org/0000-0002-3952-1474

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00683

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support by the Swiss National Science Foundation (Grant no. 200021-175944) is gratefully acknowledged. The authors are also grateful to Bill Acree, Sadra Kashef Ol Gheta, Alžbeta Kubincová, and Patrick Bleiziffer for insightful discussion and useful suggestions.

■ REFERENCES

(1) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

(2) Berendsen, H. J. C. *Simulating the Physical World*; Cambridge University Press: Cambridge, U.K., 2007.

(3) Hirst, J. D.; Glowacki, D. R.; Baaden, M. Molecular simulations and visualization: Introduction and overview. *Faraday Discuss.* **2014**, *169*, 9.

(4) Alder, B. J.; Wainwright, T. E. Phase transition for a hard sphere system. *J. Chem. Phys.* **1957**, *27*, 1208.

(5) Alder, B. J.; Wainwright, T. E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459.
(6) van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem., Int. Ed.* **1990**, *29*, 992.
(7) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646.
(8) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular modelling: goals, problems, perspectives. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064.
(9) Halgren, T. A. Potential energy functions. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205.
(10) Hünenberger, P. H.; van Gunsteren, W. F. Empirical Classical Interaction Functions for Molecular Simulations. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Eds.; Kluwer/Escom Science Publishers: Dordrecht, The Netherlands, 1997; Vol. 3, pp 3−82.
(11) Hünenberger, P. H.; van Gunsteren, W. F. Empirical Classical Force Fields for Molecular Systems. In *Lecture Notes in Chemistry*; Sax, A. F., Ed.; Springer Verlag: Berlin, Germany, 1999; pp 177−214.
(12) MacKerell, A. D., Jr. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584.
(13) Monticelli, L.; Tieleman, D. P. Force fields for classical molecular dynamics. *Methods Mol. Biol.* **2013**, *924*, 197.
(14) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129.
(15) Riniker, S. Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview. *J. Chem. Inf. Model.* **2018**, *58*, 565.
(16) Brooks, C. L., III Methodological advances in molecular dynamics simulations of biological systems. *Curr. Opin. Struct. Biol.* **1995**, *5*, 211.
(17) Elber, R. Novel methods for molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **1996**, *6*, 232.
(18) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190.
(19) Norberg, J.; Nilsson, L. Advances in biomolecular simulations: Methodology and recent applications. *Quart. Rev. Biophys.* **2003**, *36*, 257.
(20) van Gunsteren, W. F.; Dolenc, J. Thirty-five years of biomolecular simulation: development of methodology, force fields, and software. *Mol. Simul.* **2012**, *38*, 1271.
(21) Field, M. J. Technical advances in molecular simulation since the 1980s. *Arch. Biochem. Biophys.* **2015**, *582*, 3.
(22) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **2008**, *227*, 5342.
(23) Harvey, M. J.; De Fabritiis, G. An implementation of the smooth particle mesh Ewald method on GPU hardware. *J. Chem. Theory Comput.* **2009**, *5*, 2371.
(24) Hardy, D. J.; Stone, J. E.; Schulten, K. Multilevel summation of electrostatic potentials using graphics processing units. *Parallel Comput.* **2009**, *35*, 164.
(25) Schmid, N.; Bötschi, M.; Van Gunsteren, W. F. A GPU solvent-solvent interaction calculation accelerator for biomolecular simulations using the GROMOS software. *J. Comput. Chem.* **2010**, *31*, 1636.
(26) Taufer, M.; Ganesan, N.; Patel, S. GPU-enabled macromolecular simulation: Challenges and opportunities. *Comput. Sci. Eng.* **2013**, *15*, 56.
(27) Kazachenko, S.; Giovinazzo, M.; Hall, K. W.; Cann, N. M. Algorithms for GPU-based molecular dynamics simulations of complex fluids: Applications to water, mixtures, and liquid crystals. *J. Comput. Chem.* **2015**, *36*, 1787.
(28) Alam, S. R.; Agarwal, P. K.; Smith, M. C.; Vetter, J. S.; Caliga, D. Using FPGA devices to accelerate biomolecular simulations. *Computer* **2007**, *40*, 66.
(29) Gu, Y.; VanCourt, T.; Herbordt, M. C. Explicit design of FPGA-based coprocessors for short-range force computations in molecular dynamics simulations. *Parallel Comput.* **2008**, *34*, 261.
(30) Khan, M. A.; Chiu, M.; Herbordt, M. C. FPGA-Accelerated Molecular Dynamics. In *High-Performance Computing Using FPGAs*; Vanderbauwhede, M.; Benkrid, K., Eds.; Springer, 2013; pp 105−135.
(31) Berendsen, H. J. C. Treatment of Long-Range Forces in Molecular Dynamics. In *Molecular Dynamics and Protein Structure*, Proceedings Workshop 13−18 May 1984, at University of North Carolina; Hermans, J., Eds.; Polycrystal Book Service: Western Springs, Illinois, 1985; pp 18−22.
(32) Tobias, D. J. Electrostatics calculations: recent methodological advances and applications to membranes. *Curr. Opin. Struct. Biol.* **2001**, *11*, 253.
(33) Lagüe, P.; Pastor, R. W.; Brooks, B. R. Pressure-based long-range correction for Lennard-Jones interactions in molecular dynamics simulations: Application to alkanes at interfaces. *J. Phys. Chem. B* **2004**, *108*, 363.
(34) Kastenholz, M. A.; Hünenberger, P. H. Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods. *J. Phys. Chem. B* **2004**, *108*, 774.
(35) Koehl, P. Electrostatics calculations: latest methodological advances. *Curr. Opin. Struct. Biol.* **2006**, *16*, 142.
(36) In't Veld, P. J.; Ismail, A. E.; Grest, G. S. Application of Ewald summation to long-range dispersion forces. *J. Chem. Phys.* **2007**, *127*, No. 144711.
(37) Reif, M. M.; Kräutler, V.; Kastenholz, M. A.; Daura, X.; Hünenberger, P. H. Explicit-solvent molecular dynamics simulations of a reversibly-folding β-heptapeptide in methanol: Influence of the treatment of long-range electrostatic interactions. *J. Phys. Chem. B* **2009**, *113*, 3112.
(38) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **2014**, *114*, 779.
(39) Wennberg, C. L.; Murtola, T.; Páll, S.; Abraham, M. J.; Hess, B.; Lindahl, E. Direct-space corrections enable fast and accurate Lorentz-Berthelot combination rule Lennard-Jones lattice summation. *J. Chem. Theory Comput.* **2015**, *11*, 5737.
(40) Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. Properties of organic liquids when simulated with long-range Lennard-Jones interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2938.
(41) Wang, H.; Nakamura, H.; Fukuda, I. A critical appraisal of the zero-multipole method: Structural, thermodynamic, dielectric, and dynamical properties of a water system. *J. Chem. Phys.* **2016**, *144*, No. 114503.
(42) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395.
(43) Meirovitch, H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181.
(44) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.* **2010**, *31*, 1569.
(45) Michel, J.; Essex, J. W. Predictions of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639.
(46) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: Progresses and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150.
(47) Shirts, M. R. Best practices in free energy calculations for drug design. *Methods Mol. Biol.* **2012**, *819*, 425.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(48) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. Improving the efficiency of free energy calculations in the Amber molecular dynamics package. *J. Chem. Theory Comput.* **2013**, *9*, 4131.
(49) Hansen, N.; van Gunsteren, W. F. Practical aspects of free-energy calculations: A review. *J. Chem. Theory Comput.* **2014**, *10*, 2632.
(50) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database*, version 2009; University of Minnesota: Minneapolis, 2009. http://comp.chem.umn.edu/mnsol.
(51) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Toward a universal model to calculate the solvation thermodynamics of druglike molecules: The importance of new experimental databases. *Mol. Pharmaceutics* **2011**, *8*, 1423.
(52) Barnard, J. M.; Kenny, P. W.; Wallace, P. N. Representing Chemical Structures in Databases for Drug Design. In *Drug Design Strategies. Quantitative Approaches*; Livingstone, D. J.; Davis, A. M., Eds.; RSC Publishing, 2011; pp 164−191.
(53) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule and liquid database. *Bioinformatics* **2012**, *28*, 752.
(54) Brockbank, S. A.; Russon, J. L.; Giles, N. F.; Rowley, R. L.; Wilding, W. V. Critically evaluated database of environmental properties: The importance of thermodynamic relationships, chemical family trends, and prediction methods. *Int. J. Thermophys.* **2013**, *34*, 2027.
(55) Eggimann, B. L.; Sunnarborg, A. J.; Stern, H. D.; Bliss, A. P.; Siepmann, J. I. An online parameter and property database for the TraPPE force field. *Mol. Simul.* **2014**, *40*, 101.
(56) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J Comput.-Aided Mol. Des.* **2014**, *28*, 711.
(57) Beauchamp, K. A.; Behr, J. M.; Rustenburg, A. S.; Bayly, C. I.; Kroenlein, K.; Chodera, J. D. Toward automated benchmarking of atomistic force fields: Neat liquid densities and static dielectric constants from the ThermoML data archive. *J. Phys. Chem. B* **2015**, *119*, 12912.
(58) Frenkel, M. A never-ending search for the truth: Thermodynamics in the uncertain era of the internet. *J. Chem. Thermodyn.* **2015**, *84*, 18.
(59) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005**, *122*, No. 084119.
(60) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge transfer and polarization in solvated proteins from ab initio molecular dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789.
(61) Bowler, D. R.; Miyazaki, T. O(N) methods in electronic structure calculations. *Rep. Prog. Phys.* **2012**, *75*, No. 036503.
(62) Dziedzic, J.; Fox, S. J.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Large-scale DFT calculations in implicit solvent. A case study on the T4 lysozyme L99A/M102Q protein. *Int. J. Quantum Chem.* **2013**, *113*, 771.
(63) Wilkinson, K.; Skylaris, C.-K. Porting ONETEP to graphical processing unit-based coprocessors. 1. FFT box operations. *J. Comput. Chem.* **2013**, *34*, 2446.
(64) Wilkinson, K. A.; Hine, N. D. M.; Skylaris, C.-K. Hybrid MPI-OpenMP parallelism in the ONETEP linear-scaling electronic structure code: Application to the delamination of cellulose nanofibrils. *J. Chem. Theory Comput.* **2014**, *10*, 4782.
(65) Lever, G.; Cole, D. J.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C.-K.; Payne, M. C. Large-scale density functional theory transition state searching in enzymes. *J. Phys. Chem. Lett.* **2014**, *5*, 3614.
(66) Renison, C. A.; Fernandes, K. D.; Naidoo, K. J. Quantum supercharger library: Hyper-parallel integral derivatives algorithms for ab initio QM/MM dynamics. *J. Comput. Chem.* **2015**, *36*, 1410.
(67) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42*, 724.
(68) Maginn, E. J.; Elliott, J. R. Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Ind. Eng. Chem. Res.* **2010**, *49*, 3059.
(69) Guevara-Carrion, G.; Hasse, H.; Vrabec, J. Thermodynamic properties for applications in chemical industry via classical force fields. *Top. Curr. Chem.* **2012**, *307*, 201.
(70) Christ, C. D.; Fox, T. Accuracy assessment and automation of free energy calculations for drug design. *J. Chem. Inf. Model.* **2014**, *54*, 108.
(71) Honarparvar, B.; Govender, T.; Maguire, G. E. M.; Soliman, M. E. S.; Kruger, H. G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.* **2014**, *114*, 493.
(72) Hillisch, A.; Heinrich, N.; Wild, H. Computational chemistry in the pharmaceutical industry: From childhood to adolescence. *Chem. Med. Chem.* **2015**, *10*, 1958.
(73) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695.
(74) Grützner, T.; Schnider, C.; Zollinger, D.; Seyfang, B. C.; Künzle, N. Reducing time to market by innovative development and production strategies. *Chem. Eng. Technol.* **2016**, *39*, 1835.
(75) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153.
(76) Jorgensen, W. L. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335.
(77) Weiner, P. K.; Kollman, P. A. AMBER - Assisted model building with energy refinement - a general program for modeling molecules and their interactions. *J. Comput. Chem.* **1981**, *2*, 287.
(78) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187.
(79) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A consistent empirical potential for water-protein interactions. *Biopolymers* **1984**, *23*, 1513.
(80) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; BIOMOS: Groningen, The Netherlands, 1987.
(81) Lennard-Jones, J. E. The equation of state of gases and critical phenomena. *Physica* **1937**, *4*, 941.
(82) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraint for deriving atomic charges: The RESP model. *J. Phys. Chem. A.* **1993**, *97*, 10269.
(83) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999.
(84) Breneman, C. M.; Wilberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361.
(85) Williams, D. E. Net atomic charge and multipole models for the ab initio molecular electric potential. *Rev. Comput. Chem.* **1991**, *2*, 219.
(86) Henchman, R. H.; Essex, J. W. Generation of OPLS-like charges from molecular electrostatic potential using restraints. *J. Comput. Chem.* **1999**, *20*, 483.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(87) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623.
(88) Chipot, C. Rational determination of charge distributions for free energy calculations. *J. Comput. Chem.* **2003**, *24*, 409.
(89) Udier-Blagović, M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J. Comput. Chem.* **2004**, *25*, 1322.
(90) Stachowicz, A.; Styrcz, A.; Korchowiec, J. Charge sensitivity analysis in force-field-atom resolution. *J. Mol. Model.* **2011**, *17*, 2217.
(91) Stachowicz, A.; Korchowiec, J. Generalized charge sensitivity analysis. *Struct. Chem.* **2012**, *23*, 1449.
(92) Ahmed, A.; Sandler, S. I. Hydration free energies of multifunctional nitroaromatic compounds. *J. Chem. Theory Comput.* **2013**, *9*, 2774.
(93) Jämbeck, J. P. M.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. Partial atomic charges and their impact on the free energy of solvation. *J. Comput. Chem.* **2013**, *34*, 187.
(94) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **2013**, *117*, 2328.
(95) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: A self-consistent force field for condensed-phase simulations of proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515.
(96) Ionescu, C. M.; Sehnal, D.; Falginella, F. L.; Pant, P.; Pravda, L.; Bouchal, T.; Vařeková, S. V.; Geidl, S.; Koča, J. AtomicChargeCalculator: Interactive web-based calculation of atomic charges in large biomolecular complexes and drug-like molecules. *J. Cheminform.* **2015**, *7*, No. 50.
(97) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129.
(98) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, No. 144111.
(99) Yakovenko, O.; Oliferenko, A. A.; Bdzhola, V. G.; Palyulin, V. A.; Zefirov, N. S. Kirchhoff atomic charges fitted to multipole moments: Implementation for a virtual screening system. *J. Comput. Chem.* **2008**, *29*, 1332.
(100) Manz, T. A.; Sholl, D. S. Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455.
(101) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Gomes, J. R. B.; Economou, I. G.; Macedo, E. A. Predicting hydration Gibbs energies of alkyl-aromatics using molecular simulation: A comparison of current force fields and the development of a new parameter set for accurate solvation data. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17384.
(102) Manz, T. A.; Sholl, D. S. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **2012**, *8*, 2844.
(103) Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; orgensen, W. L.; Payne, M. C. Polarized protein-specific charges from atoms-in-molecule electron density partitioning. *J. Chem. Theory Comput.* **2013**, *9*, 2981.
(104) Lee, L. P.; Gabaldon Limas, N.; Cole, D. J.; Pyne, M. C.; Skylaris, C.-K.; Manz, T. A. Expanding the scope of density derived electrostatic and chemical charge partitioning to thousands of atoms. *J. Chem. Theory Comput.* **2014**, *10*, 5377.
(105) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal basis iterative stockholder: Atoms in molecules for force-field development. *J. Chem. Theory Comput.* **2016**, *12*, 3894.
(106) Peérez de la Luz, A.; Aguilar-Pineda, J. A.; Méndez-Bermúdez, J. G.; Alejandre, J. Force field parametrization from the Hirshfeld molecular electronic density. *J. Chem. Theory Comput.* **2018**, *14*, 5949.
(107) Reith, D.; Kirschner, K. N. A modern workflow for force-field development - Bridging quantum mechanics and atomistic computational models. *Comput. Phys. Commun.* **2011**, *182*, 2184.
(108) Grimme, S. A general quantum mechanically derived force field (QMDFF) for molecules and condensed phase sumulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497.
(109) Prampolini, G.; Campetella, M.; De Mitri, N.; Livotto, P. R.; Cacelli, I. Systematic and automated development of quantum mechanically derived force fields: The challenging case of halogenated hydrocarbons. *J. Chem. Theory Comput.* **2016**, *12*, 5525.
(110) Piquemal, J.-P.; Jordan, K. D. Preface: Special topic - From quantum mechanics to force fields. *J. Chem. Phys.* **2017**, *147*, No. 161401.
(111) Xu, P.; Guidez, E. B.; Bertoni, C.; Gordon, M. S. Perspective: *Ab initio* force field methods derived from quantum mechanics. *J. Chem. Phys.* **2018**, *148*, No. 090901.
(112) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the derivation of force field parameters from quantum mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366.
(113) Allen, A. E. A.; Robertson, M. J.; Payne, M. C.; Cole, D. J. Development and validation of the quantum mechanical bespoke protein force field. *ACS Omega* **2019**, *4*, 14537.
(114) Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning. *J. Chem. Inf. Model.* **2019**, *59*, 3485.
(115) Kantonen, S. M.; Muddana, H. S.; Schauperl, M.; Henriksen, N. M.; Wang, L.; Gilson, M. K. Data-driven mapping of gas-phase quantum calculations to general force field Lennard-Jones parameters. *J. Chem. Theory Comput.* **2020**, *16*, 1115.
(116) Chu, X.; Dalgarno, A. Linear response time-dependent density functional theory for van der Waals coefficients. *J. Chem. Phys.* **2004**, *121*, 4083.
(117) Olasz, A.; Vanommeslaeghe, K.; Krishtal, A.; Veszprémi, T.; Van Alsenoy, C.; Geerlings, P. The use of atomic intrinsic polarizabilities in the evaluation of the dispersion energy. *J. Chem. Phys.* **2007**, *127*, No. 224105.
(118) Stone, A. J.; Misquitta, A. J. Atom-atom potentials from ab initio calculations. *Int. Rev. Phys. Chem.* **2007**, *26*, 193.
(119) Stone, A. J. Intermolecular potentials. *Science* **2008**, *321*, 787.
(120) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, No. 073005.
(121) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **2012**, *108*, No. 236402.
(122) Gobre, V. V.; Tkatchenko, A. Scaling laws for van der Waals interactions in nanostructured materials. *Nat. Commun.* **2013**, *4*, No. 2341.
(123) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved density dependent correction for the description of London dispersion forces. *J. Chem. Theory Comput.* **2013**, *9*, 4293.
(124) Bučko, T.; Lebègue, S.; Ángyán, J. G.; Hafner, J. Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. *J. Chem. Phys.* **2014**, *141*, No. 034114.
(125) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular force field parametrization via atoms-in-molecules electron density partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312.
(126) Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating force-field London dispersion coefficients using the exchange-hole dipole moment model. *J. Chem. Theory Comput.* **2017**, *13*, 6146.
(127) Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(128) Hagler, A. T.; Ewig, C. S. On the use of quantum energy surfaces in the derivation of molecular force-fields. *Comput. Phys. Commun.* **1994**, *84*, 131.
(129) Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of class-II force-fields. 1. Methodology and quantum force-field for the alkyl

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

functional-group and alkane molecules. *J. Comput. Chem.* **1994**, *15*, 162.
(130) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490.
(131) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. *J. Comput. Chem.* **2001**, *22*, 1782.
(132) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. *J. Chem. Theory Comput.* **2007**, *3*, 1987.
(133) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. R.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Open force field consortium: Escaping atom types using direct chemical perception with SMIRNOFF v0.1. *bioRxiv* **2018**, *2018*, No. 286542.
(134) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping atom types in force fields using direct chemical perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076.
(135) Zanette, C.; Bannan, C. C.; Bayly, C. I.; Fass, J.; Gilson, M. K.; Shirts, M. R.; Chodera, J. D.; Mobley, D. L. Toward learned chemical perception of force field typing rules. *J. Chem. Theory Comput.* **2019**, *15*, 402.
(136) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247.
(137) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144.
(138) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155.
(139) Kumar, A.; Yoluk, O.; MacKerell, A. D., Jr. FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules. *J Comput Chem.* **2020**, *41*, 958.
(140) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An automated force field topology builder (ATB) and repository: version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026.
(141) 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.-Aided Mol. Des.* **2014**, *28*, 221.
(142) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. The Automated Topology Builder version 3.0 (ATB3.0): Prediction of solvation free enthalpies in water and hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834.
(143) Krieger, E.; Koraimann, G.; Vriend, G. Increasing the precision of comparative models with YASARA NOVA: A self-parameterizing force field. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 393.
(144) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355.
(145) Ribeiro, A. A. S. T.; Horta, B. A. C.; de Alencastro, R. B. MKTOP: a program for automatic construction of molecular topologies. *J. Braz. Chem. Soc.* **2008**, *19*, 1433.
(146) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359.
(147) Margreitter, C.; Petrov, D.; Zagrovic, B. Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* **2013**, *41*, W422.
(148) Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Roux, B.; Im, W. CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues. *Adv. Prot. Chem. Struct. Biol.* **2014**, *96*, 235.
(149) Pevzner, Y.; Frugier, E.; Schalk, V.; Caflisch, A.; Woodcock, H. L. Fragment-based docking: Development of the CHARMMing web user interface as a platform for computer-aided drug design. *J. Chem. Inf. Model.* **2014**, *54*, 2612.
(150) Zhang, Q.; Zhang, W.; Li, Y.; Wang, J.; Zhang, L.; Hou, T. A rule-based algorithm for automatic bond type perception. *J. Cheminform.* **2012**, *4*, No. 26.
(151) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, I. I. I. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **2012**, *33*, 189.
(152) Urbaczek, S.; Kolodzik, A.; Groth, I.; Heuser, S.; Rarey, M. Reading PDB: Perception of molecules from 3D atomic coordinates. *J. Chem. Inf. Model.* **2013**, *53*, 76.
(153) 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.
(154) Guvench, O.; MacKerell, A. D., Jr. Automated conformational energy fitting for force-field development. *J. Mol. Model.* **2008**, *14*, 667.
(155) Dahlgren, M. K.; Schyman, P.; Tirado-Rives, J.; Jorgensen, W. L. Characterization of biaryl torsional energetics and its treatment in OPLS all-atom force fields. *J. Chem. Inf. Model.* **2013**, *53*, 1191.
(156) Burger, S. K.; Ayers, P. W.; Schofield, J. Efficient parameterization of torsional terms for force fields. *J. Comput. Chem.* **2014**, *35*, 1438.
(157) Hopkins, C. W.; Roitberg, A. E. Fitting of dihedral terms in classical force fields as an analytic linear least-squares problem. *J. Chem. Inf. Model.* **2014**, *54*, 1978.
(158) Zgarbovà, M.; Rosnik, A. M.; Luque, F. J.; Curutchet, C.; Jurečka, P. Transferability and additivity of dihedral parameters in polarizable and nonpolarizable empirical force fields. *J. Comput. Chem.* **2015**, *36*, 1874.
(159) 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.
(160) Liu, Z.; Barigye, S. J.; Shahamat, M.; Labute, P.; Moitessier, N. Atom types independent molecular mechanics method for predicting the conformational energy of small molecules. *J. Chem. Inf. Model.* **2018**, *58*, 194.
(161) Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Montiac, S.; Prampolini, G. JOYCE and ULYSSES: Integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736.
(162) Wang, R.; Ozhgibesov, M.; Hirao, H. Partial Hessian fitting for determining force constant parameters in molecular mechanics. *J. Comput. Chem.* **2016**, *37*, 2349.
(163) Vlcek, L.; Sun, W.; Kent, P. R. C. Combining configurational energies and forces for molecular force field optimization. *J. Chem. Phys.* **2017**, *147*, No. 161713.
(164) Zahariev, F.; De Silva, N.; Gordon, M. S.; Windus, T. L.; Dick-Perez, M. ParFit: APython-based object-oriented program for fitting molecular mechanics parameters to ab initio data. *J. Chem. Inf. Model.* **2017**, *57*, 391.
(165) Canzar, S.; El-Kebir, M.; Rool, R.; Elbassioni, K.; Malde, A. K.; Mark, A. E.; Geerke, D. P.; Stougie, L.; Klau, G. W. Charge Group Partitioning in Biomolecular Simulation. In *RECOMB 2012*; Chor, B., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2012; p 29.
(166) Canzar, S.; El-Kebir, M.; Pool, R.; Elbassioni, K.; Malde, A. K.; Mark, A. E.; Geerke, D. P.; Stougie, L.; Klau, G. W. Charge group

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

partitioning in biomolecular simulation. *J. Comput. Biol.* **2013**, *20*, 188.
(167) Hill, J.-R. Use of test particle calculations for the derivation of van der Waals parameters used in force fields. *J. Comput. Chem.* **1997**, *18*, 211.
(168) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Direct derivation of van der Waals force field parameters from quantum mechanical interaction energies. *J. Phys. Chem. B* **2003**, *107*, 9601.
(169) Tafipolsky, M.; Engels, B. Accurate intermolecular potentials with physically grounded electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 1791.
(170) Ansorg, K.; Tafipolsky, M.; Engels, B. Cation-π interactions: Accurate intermolecular potential from symmetry-adapted perturbation theory. *J. Phys. Chem. B* **2013**, *117*, 10093.
(171) Tafipolsky, M.; Ansorg, K. Toward a physically motivated force field: Hydrogen bond directionality from a symmetry-adapted perturbation theory perspective. *J. Chem. Theory Comput.* **2016**, *12*, 1267.
(172) Tafipolsky, M. Challenging dogmas: Hydrogen bond revisited. *J. Phys. Chem. A* **2016**, *120*, 4550.
(173) Pinnick, E. R.; Erramilli, S.; Wang, F. Predicting the melting temperature of ice-Ih with only electronic structure information as input. *J. Chem. Phys.* **2012**, *137*, No. 014510.
(174) Wang, L.-P.; Chen, J.; van Voorhis, T. Systematic parametrization of polarizable force fields from quantum chemistry data. *J. Chem. Theory Comput.* **2013**, *9*, 452.
(175) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885.
(176) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised parameters for the AMOEBA polarizable atomic multipole water model. *J. Phys. Chem. B* **2015**, *119*, 9423.
(177) Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of quantum mechanically derived force-fields parameterized from dispersion-corrected DFT data: The benzene dimer as a prototype for aromatic interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182.
(178) Di Pierro, M.; Elber, R. Automated optimization of potential parameters. *J. Chem. Theory Comput.* **2013**, *9*, 3311.
(179) Di Pierro, M.; Mugnai, M. L.; Elber, R. Optimizing potentials for liquid mixture: A new force field for a *tert*-butanol and water solution. *J. Phys. Chem. B* **2015**, *119*, 836.
(180) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117*, 9956.
(181) Qi, R.; Wang, L.-P.; Wang, Q.; Pande, V. S.; Ren, P. United polarizable multipole water model for molecular mechanics simulation. *J. Chem. Phys.* **2015**, *143*, No. 014504.
(182) McKiernan, K. A.; Wang, L.-P.; Pande, V. S. Training and validation of a liquid-crystalline phospholipid bilayer force field. *J. Chem. Theory Comput.* **2016**, *12*, 5960.
(183) Wade, A. D.; Wang, L.-P.; Huggins, D. J. Assimilating radial distribution functions to build water models with improved structural properties. *J. Chem. Inf. Model.* **2018**, *58*, 1766.
(184) Qiu, Y.; Nerenberg, P. S.; Head-Gordon, T.; Wang, L.-P. Systematic optimization of water models using liquid/vapor surface tension data. *J. Phys. Chem. B* **2019**, *123*, 7061.
(185) Yin, J.; Fenley, A. T.; Henriksen, N. M.; Gilson, M. K. Toward improved force-field accuracy through sensitivity analysis of host-guest binding thermodynamics. *J. Phys. Chem. B* **2015**, *119*, 10145.
(186) Yin, J.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bind3P: Optimization of a water model based on host-guest binding data. *J. Chem. Theory Comput.* **2018**, *14*, 3621.
(187) Naden, L. N.; Shirts, M. R. Rapid computation of thermodynamic properties over multidimensional nonbonded parameter spaces using adaptive multistate reweighting. *J. Chem. Theory Comput.* **2016**, *12*, 1806.
(188) Stroet, M.; Koziara, K. B.; Malde, A. K.; Mark, A. E. Optimization of empirical force fields by parameter space mapping: A single-step perturbation approach. *J. Chem. Theory Comput.* **2017**, *13*, 6201.
(189) Messerly, R. A.; Razavi, S. M.; Shirts, M. R. Configuration-sampling-based surrogate models for rapid parameterization of non-bonded interactions. *J. Chem. Theory Comput.* **2018**, *14*, 3144.
(190) Messerly, R. A.; Barhaghi, M. S.; Potoff, J. J.; Shirts, M. R. Histogram-free reweighting with grand canonical Monte Carlo: Post-simulation optimization of non-bonded potentials for phase equilibria. *J. Chem. Eng. Data* **2019**, *64*, 3701.
(191) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead optimization mapper: Automatic free energy calculations for lead optimization. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755.
(192) Krämer-Fuhrmann, O.; Neisius, J.; Gehlen, N.; Reith, D.; Kirschner, K. N. Wolf$_2$Pack - Portal based atomistic force-field development. *J. Chem. Inf. Model.* **2013**, *53*, 802.
(193) 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.
(194) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating setup for alchemical free energy simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485.
(195) 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.
(196) Pronk, S.; Pouya, I.; Lundborg, M.; Rotskoff, G.; Wesén, B.; Kasson, P. M.; Lindahl, E. Molecular simulation workflows as parallel algorithms: The execution engine of Copernicus, a distributed high-performance computing platform. *J. Chem. Theory Comput.* **2015**, *11*, 2600.
(197) Jespers, W.; Esguerra, M.; Åqvist, J.; Gutiírez-de-Terán, T. QligFEP: An automated workflow for small molecule free energy calculations in Q. *J. Cheminform.* **2019**, *11*, No. 26.
(198) Shivakumar, D.; Deng, Y.; Roux, B. Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theory Comput.* **2009**, *5*, 919.
(199) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **2010**, *6*, 1509.
(200) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **2012**, *8*, 61.
(201) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing solute-water van der Waals interactions to reproduce solvation free energies. *J. Phys. Chem. B* **2012**, *116*, 4524.
(202) Knight, J. L.; Yesselman, J. D.; Brooks, C. L., III Assessing the quality of absolute hydration free energies among CHARMM-compatible ligand parameterization schemes. *J. Comput. Chem.* **2013**, *34*, 893.
(203) Chapman, D. E.; Steck, J. K.; Nernberg, P. S. Optimizing protein-protein van der Waals interactions for the AMBER ff9x/ff12 force field. *J. Chem. Theory Comput.* **2014**, *10*, 273.
(204) Jämbeck, J. P. M.; Lyubartsev, A. P. Update of the general Amber force field for small solutes with an emphasis on free energies of hydration. *J. Phys. Chem. B* **2014**, *118*, 3793.
(205) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force field benchmark of organic liquids. 2. Gibbs energy of solvation. *J. Chem. Inf. Model.* **2015**, *55*, 1192.
(206) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128.
(207) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, *111*, 2242.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(208) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007**, *444*, 205.
(209) Beckstein, O.; Iorga, B. I. Prediction of hydration free energies for aliphatic and aromatic chloro derivatives using molecular dynamics simulations with the OPLS-AA force field. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 635.
(210) Beckstein, O.; Fourrier, A.; Iorga, B. I. Prediction of hydration free energies for the SAMPL4 diverse set of compounds using molecular dynamics simulations with the OPLS-AA force field. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 265.
(211) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation of CM5 charges for condensed-phase modeling. *J. Chem. Theory Comput.* **2014**, *10*, 2802.
(212) Dodda, L. S.; Vilseck, J. Z.; Cutrona, K. J.; Jorgensen, W. L. Evaluation of CM5 charges for nonaqueous condensed-phase modeling. *J. Chem. Theory Comput.* **2015**, *11*, 4273.
(213) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D., Jr.; Roux, B. Optimized Lennard-Jones parameters for druglike small molecules. *J. Chem. Theory Comput.* **2018**, *14*, 3121.
(214) Visscher, K. M.; Geerke, D. P. Deriving force-field parameters from first principles using a polarizable and higher order dispersion model. *J. Chem. Theory Comput.* **2019**, *15*, 1875.
(215) Gonçalves, Y. M. H.; Senac, C.; Fuchs, P. F. J.; Hünenberger, P. H.; Horta, B. A. C. Influence of the treatment of non-bonded interactions on the thermodynamic and transport properties of pure liquids calculated using the 2016H66 force field. *J. Chem. Theory Comput.* **2019**, *15*, 1806.
(216) Gilday, L. C.; Robinson, S. W.; Barendt, T. A.; Langton, M. J.; Mullaney, B. R.; Beer, P. D. Halogen bonding in supramolecular chemistry. *Chem. Rev.* **2015**, *115*, 7118.
(217) Bauzá, A.; Mooibroek, T. J.; Frontera, A. The bright future of unconventional σ/π-hole interactions. *Chem. Phys. Chem.* **2015**, *16*, 2496.
(218) von der Heiden, D.; Vanderkooy, A.; Erdélyi, M. Halogen bonding in solution: NMR spectroscopic approaches. *Coord. Chem. Rev.* **2020**, *407*, No. 213147.
(219) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J. Med. Chem.* **2013**, *56*, 1363.
(220) Sirimulla, S.; Bailey, J. B.; Vegesna, R.; Narayan, M. Halogen interactions in protein-ligand complexes: Implications of halogen bonding for rational drug design. *J. Chem. Inf. Model.* **2013**, *53*, 2781.
(221) Scholfield, M. R.; Vander Zanden, C. M.; Carter, M.; Ho, P. S. Halogen bonding (X-bonding): A biological perspective. *Protein Sci.* **2013**, *22*, 139.
(222) Politzer, P.; Murray, J. S. Halogen bonding: An interim discussion. *Chem. Phys. Chem.* **2013**, *14*, 278.
(223) Xu, Z.; Yang, Z.; Liu, Y.; Lu, Y.; Chen, K.; Zhu, W. Halogen bond: Its role beyond drug-target binding affinity for drug discovery and development. *J. Chem. Inf. Model.* **2014**, *54*, 69.
(224) Ho, P. S. Biomolecular halogen bonds. *Top. Curr. Chem.* **2015**, *358*, 241.
(225) Montaña, A. M. The σ and π holes. The halogen and tetrel bondings: Their nature, importance and chemical, biological and medicinal implications. *ChemistrySelect* **2017**, *2*, 9094.
(226) Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J. Am. Chem. Soc.* **1996**, *118*, 3108.
(227) Wolters, L. P.; Schyman, P.; Pavan, M. J.; Jorgensen, W. L.; Bickelhaupt, F. M.; Kozuch, S. The many faces of halogen bonding: A review of theoretical models and methods. *WIREs Comput. Mol. Sci.* **2014**, *4*, 523.
(228) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G. The halogen bond. *Chem. Rev.* **2016**, *116*, 2478.
(229) Zhang, Q.; Xu, Z.; Zhu, W. The underestimated halogen bonds forming with protein side chains in drug discovery and design. *J. Chem. Inf. Model.* **2017**, *57*, 22.
(230) Zhang, Q.; Xu, Z.; Shi, J.; Zhu, W. Underestimated halogen bonds forming with protein backbone in protein data bank. *J. Chem. Inf. Model.* **2017**, *57*, 1529.
(231) Kuhn, B.; Gilberg, E.; Taylor, R.; Cole, J.; Korb, O. How significant are unusual protein-ligand interactions? Insights from database mining. *J. Med. Chem.* **2019**, *62*, 10441.
(232) Shinada, N. K.; de Brevern, A. G.; Schmidtke, P. Halogens in protein-ligand binding mechanism: A structural perspective. *J. Med. Chem.* **2019**, *62*, 9341.
(233) Riley, K. E.; Hobza, P. The relative roles of electrostatics and dispersion in the stabilization of halogen bonds. *Phys. Chem. Chem. Phys.* **2013**, *15*, 17742.
(234) Politzer, P.; Murray, J. S.; Clark, T. σ-Hole bonding: A physical interpretation. *Top. Curr. Chem.* **2015**, *358*, 19.
(235) Kolář, M. H.; Hobza, P. Computer modeling of halogen bonds and other σ-hole interactions. *Chem. Rev.* **2016**, *116*, 5155.
(236) Riley, K. E.; Murray, J. S.; Fanfrlík, J.; Řezáč, J.; Solá, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. Halogen bond tunability I: The effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine. *J. Mol. Model.* **2011**, *17*, 3309.
(237) Riley, K. E.; Murray, J. S.; Fanfrlík, J.; Řezáč, J.; Solá, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. Halogen bond tunability II: the varying roles of electrostatic and dispersion contributions to attraction in halogen bonds. *J. Mol. Model.* **2013**, *19*, 4651.
(238) Carter, M.; Rappé, A. K.; Ho, S. Scalable anisotropic shape and electrostatic models for biological bromine halogen bonds. *J. Chem. Theory Comput.* **2012**, *8*, 2461.
(239) Scholfield, M. R.; Coates Ford, M.; Vander Zanden, C. M.; Billman, M. M.; Ho, P. S.; Rappé, A. K. Force field model of periodic trends in biomolecular halogen bonds. *J. Phys. Chem. B* **2015**, *119*, 9140.
(240) de Azevedo Santos, L.; Prandi, I. G.; Ramalho, T. C. Could quantum mechanical properties be reflected on classical molecular dynamics? The case of halogenated organic compounds of biological interest. *Front. Chem.* **2019**, *7*, 848/1.
(241) Wang, L.; Gao, J.; Bi, F.; Song, B.; Liu, C. Toward the development of the potential with angular distortion for halogen bond: A comparison of potential energy surfaces between halogen bond and hydrogen bond. *J. Phys. Chem. A* **2014**, *118*, 9140.
(242) Koebel, M. R.; Schmadeke, G.; Posner, R. G.; Sirimulla, S. AutoDock VinaXB: Implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. *J. Cheminf.* **2016**, *8*, No. 27.
(243) Shaw, R.; Hill, J. A simple model for halogen bond interaction energies. *Inorganics* **2019**, *7*, No. 19.
(244) Liu, Y.; Xu, Z.; Yang, Z.; Chen, K.; Zhu, W. A knowledge-based halogen bonding scoring function for predicting protein-ligand interactions. *J. Mol. Model.* **2013**, *19*, 5015.
(245) Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564.
(246) Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M. Halogen bonding in ligand-receptor systems in the framework of classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508.
(247) Ibrahim, M. A. A. AMBER empirical potential describes the geometry and energy of noncovalent halogen interactions better than advanced semiempirical quantum mechanical method PM6-DH2X. *J. Phys. Chem. B* **2012**, *116*, 3659.
(248) Kolář, M.; Hobza, P. On extension of the current biomolecular empirical force field for the description of halogen bonds. *J. Chem. Theory Comput.* **2012**, *8*, 1325.
(249) Jorgensen, W. L.; Schyman, P. Treatment of halogen bonding in the OPLS-AA force field: Application to potent anti-HIV agents. *J. Chem. Theory Comput.* **2012**, *8*, 3895.
(250) Kolář, M.; Hobza, P.; Bronowska, A. K. Plugging the explicit σ-holes in molecular docking. *Chem. Commun.* **2013**, *49*, No. 981.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(251) Gutiérrez, I. S.; Lin, F. Y.; Vanommeslaeghe, K.; Lemkul, J. A.; Armacost, K. A.; Brooks, C. L., III; MacKerell, A. D., Jr. Parametrization of halogen bonds in the CHARMM general force field: Improved treatment of ligand-protein interactions. *Bioorg. Med. Chem.* **2016**, *24*, 4812.
(252) Nunes, R.; Vila-Viçosa, D.; Machuqueiro, M.; Costa, P. J. Biomolecular simulations of halogen bonds with a GROMOS force field. *J. Chem. Theory Comput.* **2018**, *14*, 5383.
(253) Nunes, R.; Vila-Viçosa, D.; Costa, P. J. Tackling halogenated species with PBSA: Effect of emulating the σ-hole. *J. Chem. Theory Comput.* **2019**, *15*, 4241.
(254) Campetella, M.; De Mitri, N.; Prampolini, G. Automated parameterization of quantum-mechanically derived force-fields including explicit sigma holes: A pathway to energetic and structural features of halogen bonds in gas and condensed phase. *J. Chem. Phys.* **2020**, *153*, No. 044106.
(255) Titov, O. I.; Shulga, D. A.; Palyulin, V. A.; Zefirov, N. S. Perspectives of halogen bonding description in scoring functions and QSAR/QSPR: Substituent effects in aromatic core. *Mol. Inf.* **2015**, *34*, 404.
(256) El Hage, K.; Bereau, T.; Jakobsen, S.; Meuwly, M. Impact of quadrupolar electrostatics on atoms adjacent to the sigma-hole in condensed-phase simulations. *J. Chem. Theory Comput.* **2016**, *12*, 3008.
(257) Franchini, D.; Dapiaggi, F.; Pieraccini, S.; Forni, A.; Sironi, M. Halogen bonding in the framework of classical force fields: The case of chlorine. *Chem. Phys. Lett.* **2018**, *712*, 89.
(258) Titov, O. I.; Shulga, D. A.; Palyulin, V. A. Quadrupole correction: From molecular electrostatic potential to free energies of halogen bonding. *J. Chem. Theory Comput.* **2019**, *15*, 1159.
(259) Du, L.; Gao, J.; Bi, F.; Wang, L.; Liu, C. A polarizable ellipsoidal force field for halogen bonds. *J. Comput. Chem.* **2013**, *34*, 2032.
(260) Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Modeling organochlorine compounds and the σ-hole effect using a polarizable multipole force field. *J. Phys. Chem. B* **2014**, *118*, 6456.
(261) Adluri, A. N. S.; Murphy, J. N.; Tozer, T.; Rowley, C. N. Polarizable force field with a σ-hole for liquid and aqueous bromomethane. *J. Phys. Chem. B* **2015**, *119*, 13422.
(262) Lin, F.-Y.; MacKerell, Jr. Polarizable empirical force field for halogen-containing compounds based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2018**, *14*, 1083.
(263) Lin, F.-Y.; MacKerell, A. D., Jr. Improved modeling of halogenated ligand-protein interactions using the Drude polarizable and CHARMM additive empirical force fields. *J. Chem. Inf. Model.* **2019**, *59*, 215.
(264) Rackers, J. A.; Ponder, J. W. Classical Pauli repulsion: An anisotropic, atomic multipole model. *J. Chem. Phys.* **2019**, *150*, No. 084104.
(265) Ullmann, J. R. An algorithm for subgraph isomorphism. *J. ACM* **1976**, *23*, 31.
(266) Grund, R. *Konstruktion molekularer Graphen mit gegebenen Hybridisierungen und überlappungsfreien Fragmenten*; Lehrstuhl II für Mathematik der Universität Bayreuth, 1994.
(267) Schneider, N.; Sayle, R. A.; Landrum, G. A. Get your atoms in order. An open-source implementation of a novel and robust molecular canonicalization algorithm. *J. Chem. Inf. Model.* **2015**, *55*, 2111.
(268) Rishton, G. M. Reactive compounds and in vitro fake positives in HTS. *Drug Discovery Today* **1997**, *2*, 382.
(269) Rishton, G. M. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today* **2003**, *8*, 86.
(270) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864.
(271) Yaws, C. L. *Thermophysica Properties of Chemicals and Hydrocarbons*, 2nd ed.; Gulf Professional Publishing (Elsevier): Oxford, U.K., 2014.
(272) Acree, W., Jr.; Chickos, J. S. Phase transition enthalpy measurements of organic and organometallic compounds. Sublimation, vaporization and fusion enthalpies from 1880 to 2015. Part 1. C_1-C_{10}. *J. Phys. Chem. Ref. Data* **2016**, *45*, No. 033101.
(273) Rumble, J. R. *CRC Handbook of Chemistry and Physics*, 98th ed.; CRC Press/Taylor and Francis: Boca Raton, 2018.
(274) Frenkel, M.; Hong, X.; Dong, Q.; Yan, X.; Chirico, R. D. In *Thermodynamic Properties of Organic Compounds and Their Mixtures. Densities of Halohydrocarbons*; Frenkel, M.; Marsh, K. N., Eds.; Landolt-Börnstein Series; Springer-Verlag: Berlin/Heidelberg, 2003; Vol. *IV/8J*.
(275) Frenkel, M.; Chirico, R. D.; Diky, V.; Dong, Q.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A.; Stein, S. E.; Königsberger, E.; Goodwin, A. R. H. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML). *Pure Appl. Chem.* **2006**, *78*, 541.
(276) Hünenberger, P. H. *CombiFF Data Collection in the ETHZ Research Collection* (tar-file CombiFF_saturated_haloalkanes, version 1.0 corresponds to the published article) DOI: 10.3929/ethz-b-000445271 (accessed 2020).
(277) Horta, B. A. C.; Merz, P. T.; Fuchs, P.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-compatible force field for small organic molecules in the condensed phase: The 2016H66 parameter set. *J. Chem. Theory Comput.* **2016**, *12*, 3825.
(278) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. The electronegativity equalization method and the split charge equilibration applied to organic systems: Parametrization, validation, and comparison. *J. Chem. Phys.* **2009**, *131*, No. 044127.
(279) Hagler, A. T.; Huler, E.; Lifson, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **1974**, *96*, 5319.
(280) Lifson, S.; Hagler, A. T.; Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O···H hydrogen bonds. *J. Am. Chem. Soc.* **1979**, *101*, 5111.
(281) Bourasseau, E.; Haboudou, M.; Boutin, A.; Fuchs, A. H.; Ungerer, P. New optimization method for intermolecular potentials: Optimization of a new anisotropic united atoms potential for olefins: Prediction of equilibrium properties. *J. Chem. Phys.* **2003**, *118*, 3020.
(282) Verstraelen, T.; Bultinck, P.; Van Speybroeck, V.; Ayers, P. W.; Van Neck, D.; Waroquier, M. The significance of parameters in charge equilibration models. *J. Chem. Theory Comput.* **2011**, *7*, 1750.
(283) Song, W.; Rossky, P. J.; Maroncelli, M. Modeling alkane-perfluoroalkane interactions using all-atom potentials: Failure of the usual combining rules. *J. Chem. Phys.* **2003**, *119*, 9145.
(284) von Rudorff, G. F.; Watermann, T.; Sebastiani, D. Perfluoroalkane force field for lipid membrane environments. *J. Phys. Chem. B* **2014**, *118*, 12531.
(285) Nikitin, A. M.; Milchevskiy, Y. V.; Lyubartsev, A. P. AMBER-ii: New combining rules and force field for perfluoroalkanes. *J. Phys. Chem. B* **2015**, *119*, 14563.
(286) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. *J. Comput. Chem.* **1998**, *19*, 535.
(287) Schuler, L. D.; van Gunsteren, W. F. On the choice of dihedral angle potential energy functions for n-alkanes. *Mol. Simul.* **2000**, *25*, 301.
(288) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205.
(289) Gross, K. C.; Hadad, C. M.; Seybold, P. G. Charge competition in halogenated hydrocarbons. *Int. J. Quantum Chem.* **2012**, *112*, 219.
(290) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(291) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
(292) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
(293) Manz, T. A.; Limas, N. G. Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology. *RSC Adv.* **2016**, *6*, 47771.
(294) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117.
(295) Hohm, U.; Zarkova, L. Extending the approach of the temperature-dependent potential to the small alkanes CH_4, C_2H_6, C_3H_8, n-C_4H_{10}, i-C_4H_{10}, n-C_5H_{12}, $C(CH_3)_4$, and chlorine, Cl_2. *Chem. Phys.* **2004**, *298*, 195.
(296) Zarkova, L.; Hohm, U.; Damyanova, M. Comparison of Lorentz-Berthelot and Tang-Toennies mixing rules using an isotropic temperature-dependent potential applied to the thermophysical properties of binary gas mixtures of CH_4, CF_4, SF_6, and $C(CH_3)_4$ with Ar, Kr, and Xe. *Int. J. Thermophys.* **2004**, *25*, 1775.
(297) Al-Matar, A. K.; Tobgy, A. H.; Suleiman, I. A. The phase diagram of the Lennard-Jones fluid using temperature dependent interaction parameters. *Mol. Simul.* **2008**, *34*, 289.
(298) Al-Matar, A. K.; Binous, H. Vapor-liquid phase equilibrium diagram for uranium hexafluoride (UF6) using simplified temperature dependent intermolecular potential parameters (TDIP). *J. Radioanal. Nucl. Chem.* **2016**, *310*, 139.
(299) Gong, Z.; Sun, H.; Eichinger, B. E. On the temperature transferability of force field parameters for dispersion interactions. *J. Chem. Theory. Comput.* **2018**, *14*, 3595.
(300) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. Electrical response in chemical potential equalization schemes. *J. Chem. Phys.* **1999**, *111*, 8569.
(301) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Assessment of atomic charge models for gas-phase computations on polypeptides. *J. Chem. Theory Comput.* **2012**, *8*, 661.
(302) Chen, J. H.; Martinez, T. J. QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chem. Phys. Lett.* **2007**, *438*, 315.
(303) Warren, G. L.; Davis, J. E.; Patel, S. Origin and control of superlinear polarizability scaling in chemical potential equalization methods. *J. Chem. Phys.* **2008**, *128*, No. 144110.
(304) Poier, P. P.; Jensen, F. Describing molecular polarizability by a bond capacity model. *J. Chem. Theory Comput.* **2019**, *15*, 3093.
(305) Yang, Z.-Z.; Wang, J.-J.; Zhao, D.-X. Valence state parameters of all transition metal atoms in metalloproteins. Development of ABEEM$\sigma\pi$ fluctuating charge force field. *J. Comput. Chem.* **2014**, *35*, 1690.
(306) Gilson, M. K.; Gilson, H. S. R.; Potter, M. J. Fast assignment of accurate partial atomic charges: An electronegativity equalization method that accounts for alternate resonance forms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1982.
(307) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253.
(308) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; McGraw-Hill: New York, 1981.
(309) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577.
(310) Hünenberger, P. H. Lattice-Sum Methods for Computing Electrostatic Interactions in Molecular Simulations. In *Simulation and Theory of Electrostatic Interactions in Solution: Computational Chemistry, Biophysics, and Aqueous Solution*; Hummer, G.; Pratt, L. R., Eds.; American Institute of Physics: New York, 1999; Vol. *492*, pp 17−83.
(311) Hünenberger, P. H.; van Gunsteren, W. F. Alternative schemes for the inclusion of a reaction-field correction into molecular dynamics simulations: Influence on the simulated energetic, structural, and dielectric properties of liquid water. *J. Chem. Phys.* **1998**, *108*, 6117.
(312) Kubincová, A.; Riniker, S.; Hünenberger, P. H. Reaction-field electrostatics in molecular dynamics simulations: Development of a conservative scheme compatible with an atomic cutoff. *Phys. Chem. Chem. Phys.* **2020**. DOI: 10.1039/D0CP03835K.
(313) Rappé, A. K.; Goddard, W. A. Charge equilibration for molecular-dynamics simulations. *J. Phys. Chem. B.* **1991**, *95*, 3358.
(314) Wells, B. A.; De Bruin-Dickason, C.; Chaffee, A. L. Charge equilibration based on atomic ionization in metal-organic frameworks. *J. Phys. Chem. C* **2015**, *119*, 456.
(315) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force field: Application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141.
(316) Rick, S. W.; Berne, B. J. Dynamical fluctuating charge force fields: The aqueous solvation of amides. *J. Am. Chem. Soc.* **1996**, *118*, 672.
(317) Rick, S. W. Simulations of ice and liquid water over a range of temperatures using the fluctuating charge model. *J. Chem. Phys.* **2001**, *114*, 2276.
(318) Banks, J. L.; Kaminski, G. A.; Zhou, R. H.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parameterizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *J. Chem. Phys.* **1999**, *110*, 741.
(319) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. Fluctuating charge, polarizable dipole, and combined models: Parameterization from ab initio quantum chemistry. *J. Phys. Chem. B* **1999**, *103*, 4730.
(320) Patel, S.; Brooks, C. L., III CHARMM fluctuating charge force field for proteins: I. Parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1.
(321) Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III CHARMM fluctuating charge force field for proteins. II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504.
(322) Patel, S.; Brooks, C. L., III Fluctuating charge force fields: recent developments and applications from small molecules to macromolecular biological systems. *Mol. Simul.* **2006**, *32*, 231.
(323) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102.
(324) Yaws, C. L. *The Yaws Handbook of Vapor Pressure. Antoine Coefficients*, 2nd ed.; Gulf Professional Publishing (Elsevier): Oxford, U.K., 2015.
(325) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Verlag der Fachvereine: Zürich, Switzerland, 1996.
(326) van Gunsteren, W. F. The GROMOS Software for Biomolecular Simulation. http://www.gromos.net (accessed May 05, 2011).
(327) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656.
(328) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719.
(329) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843.

(330) Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46*, 269.
(331) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.
(332) Barker, J. A.; Watts, R. O. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789.
(333) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451.
(334) van Gunsteren, W. F.; Daura, X.; Mark, A. E. GROMOS force field John. In *Encyclopedia of Computational Chemistry*; Schleyer, P., Ed.; John Wiley & Sons: Chichester, U.K., 1998; Vol. 2, pp 1211−1216.
(335) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; vanGunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596.
(336) Gasteiger, J.; Marsili, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, *19*, 3181.
(337) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity - A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219.
(338) Marsili, M.; Gasteiger, J. π charge distribution from molecular topology and π orbital electronegativity. *Croat. Chem. Acta* **1980**, *53*, 601.
(339) No, K. T.; Grant, J. A.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem. C.* **1990**, *94*, 4732.
(340) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem. D.* **1990**, *94*, 4740.
(341) Park, J. M.; No, K. T.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. III. Application to halogenated and aromatic molecules. *J. Comput. Chem.* **1993**, *14*, 1482.
(342) Park, J. M.; Kwon, O. Y.; No, K. T.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. IV. Application to hypervalent sulfur- and phosphorous-containing molecules. *J. Comput. Chem.* **1995**, *16*, 1011.
(343) Suk, J. E.; No, K. T. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. V. Application to silicon-containing organic molecules and zeolites. *Bull. Korean Chem. Soc.* **1995**, *16*, 915.
(344) Abraham, R. J.; Griffiths, L.; Loftus, P. Approaches to charge calculations in molecular mechanics. *J. Comput. Chem.* **1982**, *3*, 407.
(345) Nalewajski, R. F. Electrostatic effects in interactions between hard (soft) acids and bases. *J. Am. Chem. Soc.* **1984**, *106*, 944.
(346) Mortier, W. J.; Van Genechten, K.; Gasteiger, J. Electronegativity equalization: Application and parametrization. *J. Am. Chem. Soc.* **1985**, *107*, 829.
(347) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315.
(348) York, D. M.; Yang, W. A chemical potential equalization method for molecular simulations. *J. Chem. Phys.* **1996**, *104*, 159.
(349) Bret, C.; Field, M. J.; Hemmingsen, L. A chemical potential equalization model for treating polarization in molecular mechanical force fields. *Mol. Phys.* **2000**, *98*, 751.
(350) Chelli, R.; Procacci, P. A transferable polarizable electrostatic force field for molecular mechanics based on the chemical potential equalization principle. *J. Chem. Phys.* **2002**, *117*, 9175.
(351) Shimizu, K.; Chaimovich, H.; Farah, J. P. S.; Dias, L. G.; Bostick, D. L. Calculation of the dipole moment for polypeptides using the generalized Born-electronegativity equalization method: Results in vacuum and continuum-dielectric solvent. *J. Phys. Chem. B* **2004**, *108*, 4171.
(352) Chelli, R.; Procacci, P. Comment on "Calculation of the dipole moment for polypeptides using the generalized Born-electronegativity equalization method: Results in vacuum and continuum-dielectric solvent" [J. Phys. Chem. B 108, 4171 (2004)]. *J. Phys. Chem. B* **2004**, *108*, 16995.
(353) Verstraelen, T.; Sukhomlinov, S. V.; Van Speybroeck, V.; Waroquier, M.; Smirnov, K. S. Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models. *J. Phys. Chem. C* **2012**, *116*, 490.
(354) Yang, Z.-Z.; Wang, C.-S. Atom-bond electronegativity equalization method. 1. Calculation of the charge distribution in large molecules. *J. Phys. Chem. A* **1997**, *101*, 6315.
(355) Wang, C.-S.; Yang, Z.-Z. Atom-bond electronegativity equalization method. II. Lone-pair electron model. *J. Chem. Phys.* **1999**, *110*, 6189.
(356) Cong, Y.; Yang, Z.-Z. General atom-bond electronegativity equalization method and its application in prediction of charge distributions in polypeptide. *Chem. Phys. Lett.* **2000**, *316*, 324.
(357) Vařeková, R. S.; Koča, J. Software news and update optimized and parallelized implementation of the electronegativity equalization method and the atom-bond electronegativity equalization method. *J. Comput. Chem.* **2006**, *27*, 396.
(358) Vařeková, R. S.; Jiroušková, Z.; Vaněk, J.; Suchomel, Š.; Koča, J. Electronegativity equalization method: Parameterization and validation for large sets of organic, organohalogene and organometal molecule. *Int. J. Mol. Sci.* **2007**, *8*, 572.
(359) Oliferenko, A. A.; Palyulin, V. A.; Pisarev, S. A.; Neiman, A. V.; Zefirov, N. S. Novel point charge models: Reliable instruments for molecular electrostatics. *J. Phys. Org. Chem.* **2001**, *14*, 355.
(360) Oliferenko, A. A.; Pisarev, S. A.; Palyulin, V. A.; Zefirov, N. S. Atomic charges via electronegativity equalization: Generalizations and perspectives. *Adv. Quantum Chem.* **2006**, *51*, 139.
(361) Chen, J. H.; Martinez, T. J. Erratum to: QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics [Chem. Phys. Lett. 438 (2007) 315]. *Chem. Phys. Lett.* **2008**, *463*, 288.
(362) Nistor, R. A.; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **2006**, *125*, No. 094108.
(363) Knippenberg, M. T.; Mikulski, P. T.; Ryan, K. E.; Stuart, S. J.; Gao, G.; Harrison, J. A. Bond-order potentials with split-charge equilibration: Application to C-, H-, and O-containing systems. *J. Chem. Phys.* **2012**, *136*, No. 164701.
(364) Dapp, W. B.; Müser, M. H. Towards time-dependent, non-equilibrium charge-transfer force fields. *Eur. Phys. J. B* **2013**, *86*, No. 337.
(365) Smirnov, K. S. Assessment of split-charge equilibration model for development of polarizable force fields. *Modell. Simul. Mater. Sci. Eng.* **2015**, *23*, No. 074006.
(366) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An extended charge equilibration method. *J. Phys. Chem. Lett.* **2012**, *3*, 2506.
(367) Martin-Noble, G. C.; Reilley, D.; Rivas, L. M.; Smith, M. D.; Schrier, J. EQeq+C: An empirical bond-order-corrected extended charge equilibration method. *J. Chem. Theory Comput.* **2015**, *11*, 3364.
(368) Poier, P. P.; Jensen, F. Including implicit solvation in the bond capacity polarization model. *J. Chem. Phys.* **2019**, *151*, No. 114118.
(369) Poier, P. P.; Lagardère, L.; Piquemal, J. P.; Jensen, F. Molecular dynamics using nonvariational polarizable force fields: Theory, periodic boundary conditions implementation, and application to the bond capacity model. *J. Chem. Theory Comput.* **2019**, *15*, 6213.
(370) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138*, No. 074108.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555

(371) Sanderson, R. T. An interpretation of bond length and a classification of bonds. *Science* **1951**, *114*, 670.
(372) Sanderson, R. T. Polar Covalence I: Electronegativity Equalization, Partial Charge, and Bond Length. *Chemical Bonds and Bond Energy*; Loebl, E. M., Ed.; Academic Press: New York, 1976; Vol. *21*, pp 75−94.
(373) Parr, R. G. Density functional theory. *Ann. Rev. Phys. Chem.* **1983**, *34*, 631.
(374) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The electronegativity equalization method I: Parametrization and validation for atomic charge calculations. *J. Phys. Chem. A* **2002**, *106*, 7887.
(375) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. The electronegativity equalization method II: Applicability of different atomic charge schemes. *J. Phys. Chem. A* **2002**, *106*, 7895.
(376) Bultinck, P.; Vanholme, R.; Popelier, P. L. A.; De Proft, F.; Geerlings, P. High-speed calculation of AIM charges through the electronegativity equalization method. *J. Phys. Chem. A* **2004**, *108*, 10359.
(377) Berente, I.; Czinki, E.; Naray-Szabo, G. A combined electronegativity equalization and electrostatic potential fit method for the determination of atomic point charges. *J. Comput. Chem.* **2007**, *28*, 1936.
(378) Johnson, R. D., III. *NIST Standard Reference Database* Number 101, release 17b, September 2015; Eds.; NIST, 2015 (accessed 2015).
(379) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.
(380) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384.
(381) Kolb, A.; Dünweg, B. Optimized constant pressure stochastic dynamics. *J. Chem. Phys.* **1999**, *111*, 4453.
(382) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.
(383) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. Stochastic dynamics for molecules with constraints. Brownian dynamics of *n*-alkanes. *Mol. Phys.* **1981**, *44*, 69.
(384) Guàrdia, E.; Padró, J. A. Generalized Langevin dynamics simulation of interacting particles. *J. Chem. Phys.* **1985**, *83*, 1917.
(385) van Gunsteren, W. F.; Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1988**, *1*, 173.
(386) Yun-yu, S.; Lu, W.; van Gunsteren, W. F. On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastic dynamics simulation techniques. *Mol. Simul.* **1988**, *1*, 369.
(387) van Gunsteren, W. F. Molecular Dynamics and Stochastic Dynamics Simulation: A Primer. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Eds.; ESCOM Science Publishers, B.V.: Leiden, The Netherlands, 1993; Vol. *2*, pp 3−36.
(388) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327.
(389) Berendsen, H. J. C.; van Gunsteren, W. F.; Zwinderman, H. R. J.; Geurtsen, R. G. Simulations of proteins in water. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 269.
(390) Heinz, T. N.; Hünenberger, P. H. Combining the lattice-sum and reaction-field approaches for evaluating long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **2005**, *123*, No. 034107.
(391) Huggins, M. L. Bond Energies and Polarities. *J. Am. Chem. Soc.* **1953**, *75*, 4123.
(392) Ghosh, D. C.; Chakraborty, T. Gordy's electrostatic scale of electronegativity revisited. *J. Mol. Struct.: THEOCHEM* **2009**, *906*, 87.
(393) Sanderson, R. T. Electronegativity and bond energy. *J. Am. Chem. Soc.* **1983**, *105*, 2259.
(394) Allred, A. L. Electronegativity values from thermochemical data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215.

https://dx.doi.org/10.1021/acs.jctc.0c00683
J. Chem. Theory Comput. 2020, 16, 7525−7555