Automatic Parameterization of Force Fields for Liquids by Simplex Optimization

ROLAND FALLER, HEIKO SCHMITZ, OLIVER BIERMANN, FLORIAN MÜLLER-PLATHE

Max-Planck Institut für Polymerforschung, Ackermannweg 10, D-55128 Mainz, Germany

Received 22 December 1998; accepted 11 February 1999

ABSTRACT: In this study we demonstrate an automatic method of force field development for molecular simulations. Parameter tuning is taken as an optimization problem in many dimensions. The parameters are automatically adapted to reproduce known experimental data such as the density and the heat of vaporization. Our method is more systematic than guessing parameters and, at the same time, saves human labor in parameterization. It was applied successfully to several molecular liquids. As a test, force fields for 2-methylpentane, tetrahydrofurane, cyclohexene, and cyclohexane were developed. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 1009–1017, 1999

Keywords: force fields; molecular dynamics; parameter optimization; molecular liquids; simulation techniques

Introduction

In atomistic molecular dynamics (MD) simulations, one of the central problems is the choice of proper parameters for modeling the desired system. There are many approaches to this problem. *Ab initio* quantum chemistry would be an ideal tool for this purpose if it were able to handle interactions of large molecules in reasonable time. The standard solution, however, is quite pragmatic. One either chooses a force field that reproduces certain experimental data or one takes standard values for the different atoms. Hence, force field design is either a cumbersome trial-and-error procedure or relies heavily on the transferability of parameters.

There have been attempts to allow the computer to do this job; for instance, force field development by weak coupling.[1,2] However, that procedure relies on the requirement that one force field param-

Correspondence to: F. Müller-Plathe; e-mail: mplathe@mpip-mainz.mpg.de

CCC 0192-8651 / 99 / 101009-09

eter dominates the behavior of one property and that the relationship is monotonic. As, in more complex force fields, one property may be influenced significantly by several parameters, a more general multidimensional optimization algorithm is needed. In our approach, we consider the experimentally measured properties as multidimensional functions of the parameters. We then use the well-known simplex algorithm[3] to find the optimum parameter set.

Algorithm and Implementation

SIMPLEX ALGORITHM

The simplex method is a well-known algorithm for minimization in many dimensions.[3] It is not constrained by conditions like monotonicity, convexity, or differentiability of the function being optimized. It minimizes any single-valued function of an arbitrary number of variables. In addition, it is very robust in finding a local optimum. Its main drawback is the large number of necessary function evaluations; that is, in our case, MD simulation runs, which are quite time consuming. In what follows, we briefly summarize the simplex algorithm used in this work.

A simplex (a ''d-dimensional distorted tetrahedron'') is a set of $d + 1$ points in the d-dimensional parameter space. It is transformed geometrically depending upon the ''quality'' of the function values. There are three geometric transformations in the algorithm:

1. In a reflection, the point $\mathbf{x}_i$ with the highest function value is reflected through the hyperplane defined by the other points (see Fig. 1a):

$$\mathbf{x}'_i = \frac{2}{d}\sum_{j=1}^{d+1}\mathbf{x}_j - \left(\frac{2}{d} + 1\right)\mathbf{x}_i \qquad (1)$$

2. An expansion by the factor λ is a linear transformation of one point along the normal of the hyperplane defined by the others (Fig. 1b):

$$\mathbf{x}'_i = \frac{1-\lambda}{d}\sum_{j=1}^{d+1}\mathbf{x}_j - \left(\frac{1-\lambda}{d} + 1\right)\mathbf{x}_i \qquad (2)$$

   Thus, a reflection is just the special case $\lambda = -1$.

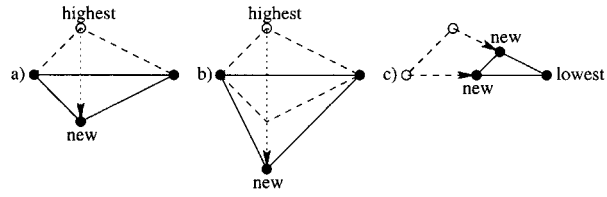

FIGURE 1. Transformations of the simplex used during the algorithm. (a) Reflection. (b) Expansion. (c) Contraction.

3. A (d-dimensional) contraction is a linear transformation of all but one point $\mathbf{x}_j$ toward the lowest point (Fig. 1c). Contractions by a factor of two are applied:

$$\mathbf{x}'_i = \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j), \qquad \forall i \neq j \qquad (3)$$

The algorithm runs iteratively. Each iteration starts with a reflection of the highest point. Depending on the function value at the new point, an expansion or a contraction is performed. If the new point is better than the best point an additional expansion with the factor $\lambda = 2$ (i.e., the distance to the hyperplane of the others is doubled) is applied to explore further into this ''promising'' direction. If the new point is very far away from the minimum (i.e., worse than the second worst point up to now) an expansion with $\lambda = 0.5$ is applied. If this resulting point is still very poor (in the aforementioned sense) a contraction around the best point is performed. Then, the next iteration, again starting with a reflection, follows.

TARGET FUNCTION EVALUATION

Because the algorithm only knows about scalar functions in $\mathscr{R}^d$, we must construct a single-valued function $f_{target}(p_1, \ldots, p_d)$ of our force field parameters $p_1, \ldots, p_d$. The function to be minimized should indicate the deviation of physical properties of the simulated model system from the real system as observed in experiments. Typically, one chooses a set of physical properties $\{P_i\}$, which have been well characterized experimentally and converge rapidly in simulations. A natural choice for f_{target} is the square root of the weighted sum of relative squared deviations:

$$f_{target}(\{p_n\}) = \left(\sum_i w_i\left(1 - \frac{P_i(\{p_n\})}{P_{i,\,target}}\right)^2\right)^{1/2} \qquad (4)$$

where $P_{i,\,target}$ is the experimental value of property P_i. The square root is chosen because it comes

steeper to the minimum. The weights w_i account for the fact that some properties may be easier to reproduce than others. Thus, the algorithm can be forced to focus on the difficult properties. Typically, the density ρ is reproduced more easily than the enthalpy of vaporization, ΔH_{vap}, which comprise the two properties we optimize our force fields against. They converge rapidly and experimental data are readily available for many fluids (see, e.g., references 4, 5).

If the number of parameters to be optimized is about two to four, the flexibility to fit the data is normally sufficient and the computational time is still manageable. If there are more target properties it may be necessary to increase the dimensionality of the optimization space, but at the cost of more computer time.

At the outset, a simplex of parameter sets must be constructed by the user. These data may be estimated from parameters for similar compounds or from standard force fields.[6–8] Furthermore, a starting configuration of the system is needed which should be close to the supposed real state. This means that geometry and density should be almost correct. The starting configuration is relaxed by some picoseconds with a guessed force field to obtain a proper liquid structure. The target function for the initial parameter sets is first evaluated before the simplex algorithm starts.

PARAMETERS TO OPTIMIZE

Because the dimensionality of parameter space is limited, we must decide which parameters of the force field we intend to optimize. This number is limited mainly by the available computing resources.

Typically, a Lennard–Jones potential is used to model the nonbonded interactions:

$$V_{\mathrm{LJ}} = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \qquad (5)$$

The density ρ depends quite strongly on the Lennard–Jones (LJ) radius, σ, whereas the enthalpy of vaporization, ΔH_{vap}, depends more strongly on ϵ. It is recommended to optimize nonbonded interaction parameters or charges and not the molecular geometry, because of simulational stability. The fact that the geometry is, for the most part, quite well known supports this choice. There are several experimental methods to determine geometries; for example, X-ray or neutron diffraction in the crystal, or electron or microwave diffraction in the gas phase. *Ab initio* quantum chemistry also gives molecular structures with useful accuracy. These geometries can, in most cases, be used for the liquid phase as well. Hence, we did not try to implement the algorithm on geometry optimization, although, in principle, this may be possible. Our simulations focus on the liquid phase, whose macroscopic properties depend only weakly on internal force field parameters. Therefore, the force field parameters for angles and dihedral angles may be adopted from similar force fields.

EQUILIBRATION

A MD run can produce reliable results only if the system has been equilibrated. Therefore, we need a scheme to test for equilibration that fulfills several requirements: it must reliably reject nonequilibrated configurations, otherwise all subsequent results are meaningless; it must work fully automatically inside the overall algorithm; and it must equilibrate as rapidly as possible to avoid wastage of resources.

If the force field parameters (i.e., the Hamiltonian) of a simulation change between iterations, like in our case, a configuration equilibrated with respect to the old parameters is no longer equilibrated with respect to the new ones. Hence, after each change of parameters (i.e., each step of the simplex algorithm) we must re-equilibrate with respect to the actual parameters. To do this, we take, as the starting configuration, the final configuration from a simulation with a parameter set, which is close to the new one. As ''distance'' in parameter space we define the sum of squared deviations:

$$|\{p^{(new)}\} - \{p^{(old)}\}|^2 := \sum_{i=1}^{n} (p_i^{(new)} - p_i^{(old)})^2 \qquad (6)$$

If, for some reason, the equilibration does not converge for that set, or some other problem occurs, a standard configuration is used.

Using the configuration selected in this way we begin a number of successive equilibration runs (typical length 50 ps with a 1-fs timestep). These runs are analyzed for equilibration until they are either accepted or a maximum number of runs (in our case 10) is exceeded. In the latter case, the parameters are considered not useful and the target function, f_{target}, is set to an arbitrarily high value to indicate failure.

How does the automatic determination of equilibration work? To our knowledge, there is no

strict criterion for equilibration. The standard procedure is to inspect visually the time development of a typical quantity (like the density for low-molecular-weight liquids). Then, one decides if it has ''settled'' to stochastic oscillations around a converged mean value. In our case, we use the following test: The time series of the density is cut into three to five intervals, for each of which the mean and the standard error are calculated. If all these averages agree within their errors the configuration is considered equilibrated. In comparison to the ''human-eye'' method, this method has proven to be rather strict. However, this is necessary because we cannot accept nonequilibrated configurations, which would mislead the simplex algorithm. The equilibration scheme has been shown to work well and led, on average, to an equilibrated configuration in about three to four runs. Naturally, the number of runs decreases during the optimization, because the changes in the parameters become less drastic. We also checked a second equilibration test in which the last third of the simulation was fitted by linear regression. If the slope was zero within its error, the configuration was considered equilibrated. The outcomes of the two tests differed only slightly.

Only very few parameter sets (less than 10%) had to be discarded due to nonequilibration. Even fewer led to instabilities in the simulation.

CONVERGENCE CRITERION

The simplex algorithm is completed if the target function falls below a given threshold, l, which is usually set to about 1% (i.e., $f_{target} < l \approx 0.01$). If this is achieved the parameters are considered satisfactory. It does not make sense to reproduce experimental data more closely because typical simulation errors limit the reliability anyway. In addition, the target values themselves carry some uncertainty.

If the desired accuracy, l, is not achieved, and the simplex ends up in a local minimum, the algorithm is aborted. Therefore, the highest value and lowest value of the target function in the actual simplex are compared. Hence, if:

$$\max(f_{target}) - \min(f_{target}) < \delta f \approx 0.001 \quad (7)$$

is achieved, further optimization makes no sense. In this case, either the number of parameters is too small to reproduce the desired number of properties (overdetermination) or the appropriate parameter values are far off the initial guess. We note that other convergence and abortion criteria are possible, for example, based on the size of the simplex. However, our criteria have proven to work well in practice.

IMPLEMENTATION

The parts of the algorithm were implemented in different programming languages. The backbone is a *tcsh* script, which calls all auxiliary programs and controls the overall flow of the procedure. It uses standard UNIX utilities like *awk*. The routine for producing a new topology from a set of parameters is a *Perl* script, whereas the programs for calculating the distance in parameter space and the determination of equilibration are implemented in C++. Several programs from the YASP simulation package[9] are used: the MD program itself as well as the utilities for calculating enthalpy of vaporization and density. Any program or utility may be easily exchanged without affecting the overall structure (e.g., when using another MD program or a different equilibration scheme).

The structure of the procedure for obtaining a function value from a given set of parameters is shown in the flow diagram in Figure 2.

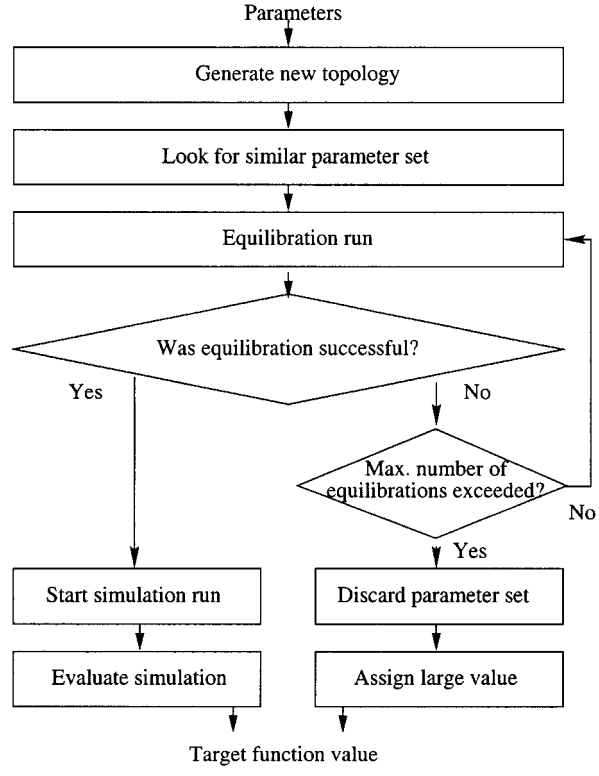

FIGURE 2. Flow diagram of the algorithm, one iteration.

Examples

The optimization procedure was tested with different model systems to assess its ability to produce proper force fields.

The all-atom nonbonded force field consists of a Lennard–Jones 12-6 potential and an electrostatic potential using reaction field and a finite cutoff (of 0.9 nm):

$$V^{(nonb)} = 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right) + \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left(\frac{1}{r} + \frac{\epsilon_{RF}-1}{2\epsilon_{RF}+1}\frac{r^2}{r_{cutoff}^3}\right) \quad (8)$$

This potential is applied to atoms belonging to different molecules, and internal nonbonded interactions are excluded in our test cases except for methylpentane. The Lennard–Jones parameters between unlike atoms are derived by the Lorenz–Berthelot mixing rules[10]:

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{\frac{1}{2}}, \qquad \sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad (9)$$

A bond angle potential:

$$V^{(angle)} = \frac{k^{(angle)}}{2}(\Theta - \Theta_0)^2, \qquad \Theta\text{: bond angle} \quad (10)$$

and, for some molecules, torsional potentials with threefold symmetry:

$$V^{(tors)} = \frac{k^{(tors)}}{2}(1 - \cos(3\tau)), \qquad \tau\text{: dihedral angle} \quad (11)$$

or a harmonic dihedral potential:

$$V^{(hd)} = \frac{k^{(hd)}}{2}(\tau - \tau_0)^2 \quad (12)$$

are applied to keep the correct molecular shape.

The bond lengths are constrained using the SHAKE algorithm.[11,12] Our systems are subject to cubic periodic boundary conditions. The simulations are run at ambient conditions (T = 298 K, p = 1013 hPa). The neighbor-list[10] is calculated up to 1.0 nm every 10 to 15 timesteps. We use the Berendsen algorithm for constant pressure and temperature.[13] The coupling times are 0.2 ps and 2 ps, respectively. The simulation runs last 50 ps at a timestep of 1 fs for each equilibration run and 100 ps at a timestep of 2 fs for the evaluation runs. The errors of the properties are obtained by binning analysis.[10]

METHYLPENTANE

As a first test, a system of 125 uncharged 2-methylpentane molecules was optimized with respect to density, ρ, and enthalpy of vaporization, ΔH_{vap}. All Lennard–Jones parameters are subject to optimization. However, all like atoms (C and H) are constrained to have the same LJ parameters. The internal part of the force field is taken from the AMBER force field[6]. This comprises the rule that the Lj ϵ (Equation 5) is scaled by a factor of 0.5 for 1–4 interactions.

We use the following parameter file to start the algorithm. The number ''4'' in the second line indicates the dimensionality of the parameter space. The following five lines are the guesses of the parameters, the initial simplex. The last column shows the results after evaluation of the target function. The Lennard–Jones energies (ϵ and ΔH_{vap}) are measured in kilojoules per mole, the radii (σ) in nanometers, and the density (ρ) in kilograms per cubic meter:

##	ϵ_C	σ_C	ϵ_H	σ_H	f_{target}	ΔH_{vap}	ρ
4							
0.291643	0.339215	0.154545	0.258859	0.030287	29.42	636.1	
0.290554	0.340351	0.151371	0.260571	0.052871	28.88	626.5	
0.290167	0.340656	0.151116	0.260825	0.057492	28.81	623.8	
0.290545	0.341183	0.149968	0.260762	0.043120	29.18	629.6	
0.290421	0.341161	0.150191	0.260763	0.051831	28.94	626.2	

These parameters produce properties which are already quite close to the target values. The simplex algorithm now does the fine tuning. First, the simplex is reflected away from parameter set 3. The new set is

##	ϵ_C	σ_C	ϵ_H	σ_H	f_{target}	ΔH_{vap}	ρ
0.291414	0.340299	0.151922	0.259653	0.045721	29.04	629.6	

After 11 optimization steps, which took about 2 weeks altogether on a DEC 433MHz processor, the optimization finally finished with the following values:

##	ϵ_C	σ_C	ϵ_H	σ_H	f_{target}	ΔH_{vap}	ρ
0.294477	0.336339	0.162144	0.254668	0.008629	30.15	652.7	

Figure 3 illustrates the progress of the optimization by means of a previous run. The circles show

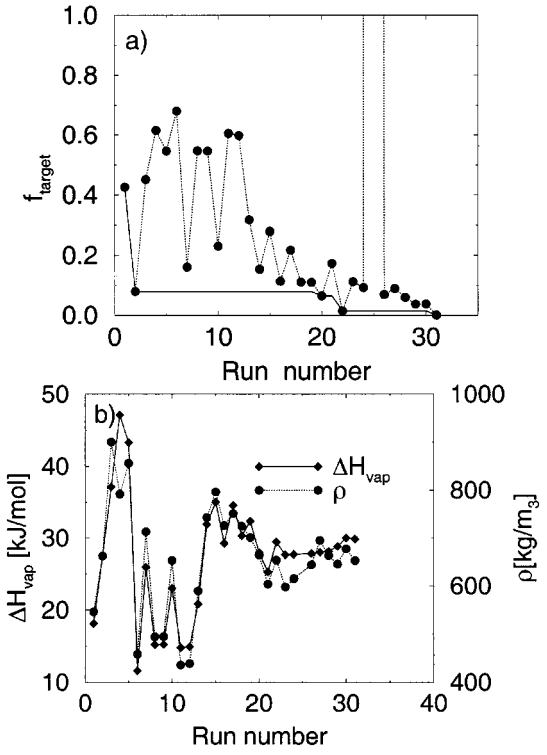

FIGURE 3. Convergence of the methylpentane run. (a) Target function (solid line: best value of f_{target}; circles / dotted line: actual value of f_{target}). (b) Properties: density and enthalpy of vaporization.

the results of function evaluations and the solid line shows the best current values of f_{target}. At the outset, the function values scatter quite strongly. During the run of the algorithm, this starts to decrease. Figure 3b shows how density and enthalpy of vaporization reach their target values. The only maintenance required involves restarting the algorithm after shutdown of the computer system. The whole algorithm has proven to be stable and works fully automatically. Equilibration failed only due to exceeding the limit of 10 runs. This is shown by the spike in Figure 3 (which goes up to 100,000). The final force field is shown in Table I. These values reproduced the experimental data in a satisfactory way (Table II).

TETRAHYDROFURANE

As a different test system we used tetrahydrofurane (THF). Here, we focused particularly on the optimization of the partial charges. The hydrogens did not carry any partial charges but oxygen and carbon did. The charges of carbons 2 and 5 and carbons 3 and 4 are the same for symmetry reasons. With the constraint of electroneutrality, there were two charge parameters to be optimized. We chose q_O and $q_{C2/C5}$, then we had $q_{C3/C4} = \frac{1}{2}(-q_O - 2q_{C2/C5})$. In addition, the oxygen parameters, ϵ_O and σ_O, were included in the optimization. The first guess for the partial charges was taken from a quantum-chemical Hartree–Fock calculation with a 6-311G** basis set using GAUSSIAN-94 (Mulliken charges with hydrogens summed into heavy atoms).[14] This also yielded the bond angle values. The bond lengths were taken from electron diffraction.[5] The simulated system contained 216 molecules. The electrostatic interactions were simulated with a reaction field correction ($\epsilon_{RF} = 7.5$) using the same cutoff $r_c = 0.9$ nm as for the Lennard–Jones potential. Here the following starting simplex was taken:

##	$-q_O$	$q_{C2/C5}$	ϵ_O	σ_O	f_{target}	ΔH_{vap}	ρ
4							
	0.581241	0.225443	0.516818	0.208594	0.084176	32.72	816.96
	0.658970	0.251733	0.325788	0.300391	0.137257	35.39	811.81
	0.480765	0.251793	0.635725	0.316797	0.186735	27.68	774.02
	0.684265	0.276431	0.729962	0.264345	0.232852	39.29	847.92
	0.582970	0.220928	0.535152	0.192715	0.088840	33.57	823.39

TABLE I.
Details of Methylpentane Force Field.

Bonded Parameters		Nonbonded Parameters	
Parameter	Value	Parameter	Value
|C—C|	0.1526 nm	m_C	12.01 amu
|C—H|	0.109 nm	m_H	1.00782 amu
$k^{(angle)}_{C-C-C}$	$167.47\frac{\text{kJ}}{\text{mol rad}^2}$	ϵ_C	0.294 kJ / mol
$k^{(angle)}_{C-C-H}$	$209.34\frac{\text{kJ}}{\text{mol rad}^2}$	ϵ_H	0.162 kJ / mol
$k^{(angle)}_{H-C-H}$	$146.54\frac{\text{kJ}}{\text{mol rad}^2}$	σ_C	0.336 nm
H — C — H	109.5°	σ_H	0.254 nm
C — C — C	109.5°		
C — C — H	109.5°		
$k^{(tors)}_{C-C-C-C}$	11.5 kJ / mol		
$k^{(tors)}_{methyl}$	11.5 kJ / mol		

TABLE II.
Experimental and Simulated Properties of 2-Methylpentane.

	Exp.	Sim.
ΔH_{vap} [kJ / mol]	29.89[5]	29.92 ± 0.03
ρ [kg / m^3]	653.0[5]	653.4 ± 0.5
D [10^{-5} cm^2 / s]		2.5 ± 0.2

The first optimization attempt ended up in a local minimum with $f_{target} \approx 0.07$ after 53 evaluations. The experimental liquid density could not be reproduced satisfactorily; it was systematically too low. Therefore, the best parameters so far were frozen and a new optimization was started where only the Lennard–Jones radii of all species were optimized. Finally, convergence ($f_{target} \leq 0.01$) was achieved. The resulting THF force field is detailed in Table III.

These parameters led to the physical properties shown in Table IV. Our force field has about the same accuracy as an earlier Monte Carlo simulation of a united-atom OPLS model for THF.[15]

CYCLIC HYDROCARBONS

Finally, the method was applied to obtain force fields for cyclohexene and cyclohexane with 125 molecules in the periodic box. The geometries were taken from electron diffraction data.[5] They are shown in Table V. In the cyclohexene force field, harmonic dihedral angles were used to keep the atoms around the double bond in the plane, because sp^2 hybridization prevented the double bond from rotating. In addition, standard torsional potentials with threefold symmetry were used. Cyclohexane was simulated without any dihedral angle potentials. For the angular force constants we used a standard value, because they are believed to be of minor importance for the desired properties. Furthermore, they may be compensated by the nonbonded parameters.

The optimized Lennard–Jones 12-6 parameters are shown in Table VI. The parameters included in the optimization procedure are denoted by ''opt.'' No charges were used. All parameters not optimized, as well as the initial simplices, were taken from similar force fields.

Except for the Lennard–Jones ϵ of the hydrogens, the resulting final parameters are very similar for the two molecules. This shows that force

TABLE III.
Optimized Force Field for Tetrahydrofurane.[a]

Nonbonded Parameters		Bonded Parameters	
Parameter	Value	Parameter	Value
ϵ_O	0.509 kJ / mol	|C—O|	0.1428 nm
ϵ_H	0.200 kJ / mol	|C—H|	0.1115 nm
ϵ_C	0.290 kJ / mol	|C—C|	0.1536 nm
σ_O	0.243 nm	$k^{(angle)}$	450.0 $\frac{\text{kJ}}{\text{mol rad}^2}$
σ_H	0.193 nm	C—O—C	111.2°
σ_C	0.306 nm	O—C—C	106.1°
q_O	$-0.577e$	C — C — C	101.4°
q_{C2}	$0.228e$	O—C—H	109.0°, 109.3°
q_{C3}	$0.061e$	C_3—C_2—H	111.0°, 113.2°
m_O	15.9949 u	H—C_2—H	108.2°
m_C	12.0 u	H—C_3—H	108.1°
m_H	1.00787 u	C_2—C_3—H	110.4° (2 ×)
			112.8° (2 ×)
		C_3—C_4—H	113.7° (2 ×)
			110.4° (2 ×)

[a] In the case of two angles, in one line one is applied to the first hydrogen, and the other to the second hydrogen—otherwise, the angles would not be consistent with each other.

TABLE IV.
Properties of Tetrahydrofurane.

	Experiment[5]	Simulation (This Work)	Simulation[15]
ρ	889.0 kg / m^3	886.0 ± 1.3 kg / m^3	882 ± 1 kg / m^3
ΔH_{vap}	31.99 kJ / mol	32.0 ± 0.1 kJ / mol	31.57 ± 0.08 kJ / mol

TABLE V.
Geometry of Cyclic Hydrocarbons and Their Intramolecular Potentials.

Property	C_6H_{10}	C_6H_{12}
$\lvert C_{sp^2}=C_{sp^2}\rvert$	0.1334 nm	
$\lvert C_{sp^2}—C_{sp^3}\rvert$	0.150 nm	
$\lvert C_3—C_4\rvert, \lvert C_5—C_6\rvert$	0.152 nm	
$\lvert C_4—C_5\rvert$	0.154 nm	
$\lvert C—C\rvert$		0.1526 nm
$\lvert C_{sp^2}—H\rvert$	0.108 nm	
$\lvert C_{sp^3}—H\rvert$	0.109 nm	
$k^{(angle)}_{C—C—C}$	450 kJ / mol rad^2	335 kJ / mol rad^2
$k^{(angle)}_{C=C—C}$	500 kJ / mol rad^2	
$k^{(angle)}_{C—C—H}$	500 kJ / mol rad^2	420 kJ / mol rad^2
$k^{(angle)}_{H—C—H}$	500 kJ / mol rad^2	290 kJ / mol rad^2
C=C—C	112.0°	
C—C—C	110.9°	109.5°
C_{sp^2}—C—C	123.45°	
C—C—H	109.5°	
C—C_{sp^2}—H	119.75°	
H—C—H	109.5°	
$k^{(hd)}_{C—C=C—C}$	250 kJ / mol rad^2	
$k^{(hd)}_{H—C=C—C}$	200 kJ / mol rad^2	
$k^{(tors)}_{C—C—C—C}$	10 kJ / mol	

TABLE VI.
Cyclohexene and Cyclohexane Nonbonded Parameters.

Parameter	Opt / Fix	C_6H_{10}	C_6H_{12}
ϵ_C	opt	0.296 kJ / mol	0.299 kJ / mol
ϵ_H	opt	0.265 kJ / mol	0.189 kJ / mol
σ_H	opt	0.252 nm	0.258 nm
σ_C	opt		0.328 nm
$\sigma_{C_{sp^2}}$	fix	0.321 nm	
$\sigma_{C_{sp^3}}$	fix	0.311 nm	
m_C	fix	12.01 amu	
m_H	fix	1.00787 amu	

field parameters are not a unique description of a certain atom type, but rather they are only a part of the overall molecular description. For the most part, however, the same atoms in similar environments may be described by similar parameters.

We compare our thermodynamic data with experiment in Table VII. A more detailed analysis of transport properties of these cyclic hydrocarbons will be published elsewhere.[16] The cyclohexane force field yields a slightly better comparison to experiment than in a recent study using a commercial force field,[8] whereas the present study of cyclohexene is the first to our knowledge.

Conclusions

We applied the simplex algorithm to the problem of force field optimization for MD simulations. Given a good initial guess for the force field parameters and the experimental data for some properties, our method tunes the parameters to optimum values. Once the routine is set up, very little human interference is required for maintenance. The algorithm proved to be robust and it successfully found local minima if set up properly. The resulting force fields are able to reproduce experimental data of low-molecular-weight liquids in a satisfactory manner.

In the examples, we typically optimized four force field parameters against two observables. Hence, the solutions were most likely not unique. This, however, is a feature of the problem of finding a force field given a small number of observables, not of the algorithmic solution presented. Density and enthalpy of vaporization are the two properties most commonly used to derive force fields, as they are experimentally available for many fluids and will converge quickly in a simulation. At present, our method must be used with a judicious choice of starting values of parameters to prevent it from optimizing an unphysical, nontransferable set of parameters. It shares this restriction with all other methods of finding force fields,

TABLE VII.
Properties of Cyclohexene and Cyclohexane.

	Cyclohexene		Cyclohexane		
	Exp.	Sim.	Exp.[5]	Sim. (This Work)	Sim.[8]
ρ [kg / m^3]	805.8[17]	806.0 $\pm$ 1.5	777.6	775.9 $\pm$ 0.8	774 $\pm$ 2
ΔH_{vap} [kJ / mol]	33.47[5]	33.3 $\pm$ 0.1	33.33	33.46 $\pm$ 0.05	33.41

including ''optimization by hand.'' Nevertheless, it is generally simple to come up with a reasonable first guess for the parameters. What is time consuming is the fine tuning, and here our method offers help.

A possible way out of the dilemma is to increase the base of experimental observables used in the target function. In a few selected cases we used liquid properties other than ρ and ΔH_{vap}, together with a different refinement scheme.[1,2] However, clearly there are few suitable fluid properties. Some properties are of similar character to what we already have. For example, the excess chemical potential, μ_{ex}, probes almost the same regions of the force field as ΔH_{vap} and, thus does not add much independent information. Dynamic properties often converge too slowly in simulations to be useful (shear viscosity, dielectric constant) or the experimental data are not of sufficient quality (tracer diffusion coefficient, molecular reorientation times). We, therefore, follow the strategy of optimizing toward ρ and ΔH_{vap} and subsequently checking the final force field against other liquid properties. For our models of cyclic hydrocarbons we have, for instance, calculated tracer and binary diffusion coefficients as well as molecular reorientation times for both pure liquids and binary mixtures, and the results agree well with experimental data where available.[16]

The automatic parameterization scheme presented has the minor disadvantage of probably requiring slightly more computer time than an optimization by hand. This is more than offset by the invaluable advantage of freeing researchers from parameter optimization. In reasonable computation time (a few weeks workstation time) one is able to cope with dimensionalities of parameter space of about four. This depends, however, on the actual simulations performed. On the other hand, the full potential of speeding up our algorithm has not yet been realized. We foresee the possibility of substantial improvement by using a less rigorous and more adaptive equilibration scheme by substituting the simplex algorithm with a faster converging optimizer (e.g., Fletcher) in the final stages of minimization. This remains an attractive starting point for future research.

References

1. Njo, S. L.; van Gunsteren, W. F.; Müller-Plathe, F. J Chem Phys 1995, 102, 6199–6207.
2. Berweger, C. D.; van Gunsteren, W. F.; Müller-Plathe, F. Chem Phys Lett 1995, 232, 429–436.
3. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed.; Cambridge University Press: New York, 1992.
4. Landolt, H. Börnstein Numerical Data and Functional Relationships in Science and Technology; Springer: Berlin, 1993.
5. Lide, D. R., ed. CRC Handbook of Chemistry and Physics, 76th Ed.; CRC Press: Boca Raton, FL, 1995.
6. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M.; Ferguson, D. M.; Spellmeyer, C. D.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179–5197.
7. 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 GROMOS Manual and User Guide; Vdf-Hochschul-Verlag: Zürich, 1996.
8. Sun, H. J Phys Chem B 1998, 102, 7338–7364.
9. Müller-Plathe, F. Comput Phys Commun 1993, 78, 77–94.
10. Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Clarendon Press: Oxford, 1987.
11. Ryckaert, J.-P.; Cicotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327–341.
12. Müller-Plathe, F.; Brown, D. Comput Phys Commun 1991, 64, 7–14.
13. Berendsen, H. J. C.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. J Chem Phys 1984, 81, 3684–3690.
14. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. GAUSSIAN-94; Gaussian: Pittsburgh, PA, 1995.
15. Briggs, J. M.; Matsui, T.; Jorgensen, W. L. J Comput Chem 1990, 11, 958–971.
16. Schmitz, H.; Faller, R.; Müller-Plathe, F. Submitted to J Phys Chem B.
17. Harris, K. R.; Dunlop, P. J. Ber Bunsenges Phys Chem 1994, 98, 560–562.