JCTC Journal of Chemical Theory and Computation

Article

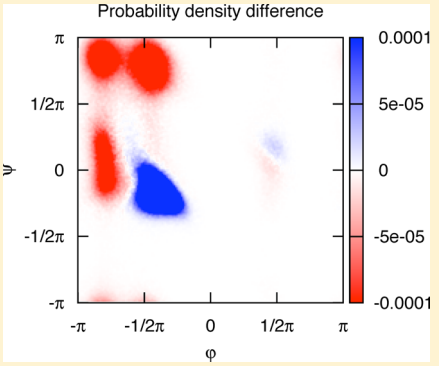
pubs.acs.org/JCTC

Automated Optimization of Potential Parameters

Michele Di Pierro[†] and Ron Elber*[,†,‡]

[†]Institute for Computational Engineering and Sciences and [‡]Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

ABSTRACT: An algorithm and software to refine parameters of empirical energy functions according to condensed phase experimental measurements are discussed. The algorithm is based on sensitivity analysis and local minimization of the differences between experiment and simulation as a function of potential parameters. It is illustrated for a toy problem of alanine dipeptide and is applied to folding of the peptide WAAAH. The helix fraction is highly sensitive to the potential parameters, while the slope of the melting curve is not. The sensitivity variations make it difficult to satisfy both observations simultaneously. We conjecture that there is no set of parameters that reproduces experimental melting curves of short peptides that are modeled with the usual functional form of a force field.



I. INTRODUCTION

One of the prime goals of molecular dynamics (MD) is to compute ensemble averages of measured quantities, say O, and to compare the results to experiments.[1] These comparisons are used to validate the results of computer simulations and to fine-tune the computational models. These averages are computed in MD by a time series

$$\langle O\rangle = \lim_{t\to\infty}(1/t)\int_0^t O(t')\,\mathrm{d}t' \tag{1}$$

The averages or the expectation values are compared to the experimental value O_{exp}. If the trajectory is sufficiently long and ergodic and is computed according to dynamics that produce the canonical distribution of coordinates (e.g., iso-kinetic ensemble[2]), then the time series can be replaced by a desired coordinate average

$$\langle O\rangle = \frac{\int \mathrm{d}x\cdot O(x)\,\exp[-\beta U(x)]}{\int dx\cdot\exp[-\beta U(x)]} \tag{2}$$

where U is the potential energy of the system and $\beta = 1/kT$. Predictions by MD of experimental observations have two sources of errors: (i) statistical and (ii) systematic (biased) errors. MD calculations require long simulations to converge the above averages. If the sampling is from the correct distribution the (statistical) errors decrease as $\sqrt{t}$, where t is the time. Longer simulations can further reduce the statistical errors, but systematic errors remain.

If the statistical sampling is adequate, the remaining errors are a result of inaccurate potential $U(x)$. The functional form may be imprecise. For example, most empirical potentials use point charge models to compute electrostatic energies while the problem at hand may require a better formulation of the charge density that includes polarization. Alternatively, the functional form of the potential is retained and only the parameters of the energy are modified.

Interestingly, widely used force fields such as CHARMm,[3] AMBER,[4] and OPLS[5] retain highly similar functional forms. Refinements of the energy function are mostly made by adjustments of the parameters of the energy function. Given the above experience, it makes sense to automate the process of parameter refinement, assuming a fixed functional form. This is what we have done and describe in the present manuscript.

Considerable development to test and improve empirical energy functions have already been made as described in the above manuscripts. Therefore, a valid question is whether further adjustments are needed and what are the target observables that require these adjustments?

In the present manuscript, we focus on conformational transitions and folding of small peptides. We illustrate that the current functional form is sufficiently flexible so that parameters can be found that reproduce the folding fraction at room temperature. In fact, the folding fraction is an extremely sensitive function of the torsion parameters, so it is not obvious that a particular set of parameters that we fit for one peptide will be transferrable to other folding problems.

We also demonstrate that peptide parameters that reproduce the melting curve of the peptides are much harder to identify. We conjecture, based on our numerical optimization, that no peptide parameters could be found that reproduce the correct melting curve of peptide folding. This is over a significant temperature range, given the current functional form of the peptide force field and the use of the TIP3P water model.[6] These observations illustrate the usefulness of the proposed approach that not only provide a suitable set of parameters for a

Received: April 17, 2013
Published: June 18, 2013

particular case but also examine the feasibility and sensitivity of the fitting process.

II. METHOD

II.1. Potential Energy. The basic functional form for MD potentials is a summation of bonding and nonbonding terms. The nonbonding terms are represented by pair potentials: the Lennard–Jones potential and the point charge Coulombic energy. The nonbonding terms are instead represented by two, three, and four body potentials (bonds, angles, and proper and improper torsions):

$$U = \sum_b U_b + \sum_\theta U_\theta + \sum_\phi U_\phi + \sum_I U_I + \sum_{\mathrm{LJ}} U_{\mathrm{LJ}} + \sum_{\mathrm{elec}} U_{\mathrm{elec}}$$

$$U_b = \frac{k_b}{2}(b - b_0)^2 \quad U_\theta = \frac{k_\theta}{2}(\theta - \theta_0)^2$$

$$U_\phi = \sum_n a_n \cos(n\phi + \delta_n)$$

$$U_I = \frac{k_I}{2}(\phi_I - \phi_{I0})^2 \quad U_{\mathrm{LJ}} = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$$

$$U_{\mathrm{elec}} = k_{\mathrm{el}}\frac{q_i q_j}{r_{ij}} \tag{3}$$

For the Lennard–Jones potential, a mixing rule for different chemical species must be defined; OPLS force field for a pair of atom types (i,j) defines $\sigma_{ij} = (\sigma_i\sigma_j)^{1/2}$ and $\varepsilon_{ij} = (\varepsilon_i\varepsilon_j)^{1/2}$ [another possible mixing rule used by other force fields is $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$]. In general, the potential energy depends on the set of parameters $\{\{k_b,b_0\},\{k_\theta,\theta_0\},\{k_I,\phi_{I0}\},\{a_n\},\{\sigma,\varepsilon\},\{q\}\}$; the complete set of parameters is denoted by $\{\pi\}$. We call a force field that uses the functional form of eq 3 the standard force field.

II.2. Optimization of the Target Function. The distance between the experimental result and the average that we obtain through our mathematical model is the error associated with modeling; ideally, it should be zero. This distance is our target function Θ, and we want to minimize it in the space of parameters. If the observable $O(x,p)$ is measured for a range of thermodynamic variables [e.g., temperature (T)], we define

$$\Theta(\pi) = \int_{T_1}^{T_2} [\langle O\rangle_{(\pi,T)} - O_{\mathrm{exp}}(T)]^2 \, \mathrm{d}T \tag{4}$$

where $O_{\mathrm{exp}}(T)$ is the experimental value and the experiment is performed in the temperature range from T_1 to T_2. The number of parameters that we wish to optimize can be large and an exhaustive grid search in π-space to optimize $\Theta(\pi)$ may be costly. Knowledge of the gradient of the target function $\nabla_\pi\Theta(\pi)$ allows us to use more efficient minimization algorithms. The gradient is computed in a single ensemble average,

$$\nabla_\pi\Theta(\pi) = \int_{T_1}^{T_2} 2[\langle O\rangle_{(\pi,T)} - O_{\mathrm{exp}}(T)]\nabla_\pi\langle O\rangle_{(\pi,T)} \, \mathrm{d}T,$$

$$\nabla_\pi\langle O\rangle_{(\pi,T)} = \langle\nabla_\pi O(x,p,\pi)\rangle_{(\pi,T)} - \beta[\langle\nabla_\pi H\cdot O(x,p,\pi)\rangle_{(\pi,T)} - \langle\nabla_\pi H\rangle_{(\pi,T)}\langle O(x,p,\pi)\rangle_{(\pi,T)}] \tag{5}$$

where we have also used $\nabla_\pi\langle O\rangle_{(\pi,T)}$, to denote the sensitivity of the macroscopic observable $\langle O\rangle_{(\pi,T)}$ with respect to the parameters π. The sensitivity measures show how much the observable depends on a parameter; if it is zero, the observable does not depend on that parameter and no improvement in the target function can be achieved by varying it.

It is most likely that the dependence of the observable on the parameters would not be direct but through the potential energy (e.g., the observable can be the potential energy itself or, for example, the heat capacity). In that case we write,

$$\langle\nabla_\pi O(x,p,\pi)\rangle_{(\pi,T)} \equiv \langle\nabla_U O(x,p,U)\nabla_\pi U\rangle_{(\pi,T)} \tag{6}$$

An approach to force field optimization is to manually modify one parameter at a time and check the effect on the quantity of interest in a simulation. Clearly, the use of a gradient promises a more efficient, accurate, and automated optimization procedure. This is at the cost of the development of appropriate software.

The use of sensitivity in the optimization process solves the two problems at once. First, chemical intuition is substituted by a quantitative measurement whose value is affected only by a tractable statistical error. Second, all the parameters are simultaneously optimized, allowing the capture of subtle cooperative effects that could not be detected previously. There is no practical restriction on the number of parameters that can be handled simultaneously. The disadvantage of the proposed algorithm is that the computed ensemble averages can be expensive, while the use of chemical intuition can cut many corners and determines the correct set of parameters more quickly.

The information contained in the gradient can be used to implement several optimization algorithms, like conjugate gradient and steepest descent[7] minimization or overdamped Langevin dynamics; the first two are deterministic local minimizers, while the latter is a stochastic algorithm designed to be a global minimizer. Global minimization requires (of course) a significantly larger number of function evaluations.

The force fields that are widely used today (e.g., refs 3–5,8) are useful for many modeling tasks and have been refined and tested over a span of many years. Therefore, a large change in their parameters is not desired when optimizing for a specific observable. The smaller the modifications we make the more likely the new potential will be transferable to other systems and other observables. For this reason, we choose to use a local optimization algorithm (i.e., steepest descent). Modifying the set of parameters along the direction of the gradient guarantees a maximum change in the value of the target function with a minimum displacement of the parameter values. Steepest descent minimization is the simplest optimization algorithm and operates following this recipe:

$$\Delta\pi_i = \pi_{i+1} - \pi_i = -\alpha_i\nabla_\pi\Theta(\pi_i) \tag{7}$$

where α is a scalar parameter. We estimate the parameter α using the following heuristic method. First, we calculate the expected change in the target function $\Theta(\pi)$ for a displacement $\Delta\pi$ in the linear regime:

$$\Delta\Theta(\pi_i) \simeq \nabla_\pi\Theta(\pi_i)\Delta\pi_i = -\alpha_i\nabla_\pi\Theta(\pi_i)\cdot\nabla_\pi\Theta(\pi_i) \tag{8}$$

Then, we pick the desired change Δ_i^* for the function $\Theta(\pi)$; finally, we choose α such that

$$\Delta_i^* = -\alpha_i\nabla_\pi\Theta(\pi_i)\cdot\nabla_\pi\Theta(\pi_i) \tag{9}$$

Deciding the optimal value of Δ_i^* is not obvious, and it varies from case to case; in our first illustration, shifting the conformational equilibrium of alanine dipeptide, we choose

$\Delta_i^* = -\Theta(\pi_i)$. This is the maximal change we can have (the minimum of the target function is at zero). This is a large step. However, the torsion potentials that make the largest contribution to the alanine dipeptide problem are linear in their parameters (note that the thermodynamic average is not linear). Therefore, the regression expression is approximately quadratic, which is appropriate for large Newton−Raphson-like steps. In this case, the minimization is equivalent to the Newton−Raphson algorithm to find roots of a real valued function

$$\alpha_i = \frac{\Theta(\pi_i)}{\nabla_\pi\Theta(\pi_i)\cdot\nabla_\pi\Theta(\pi_i)},$$
$$\pi_{i+1} = \pi_i - \frac{\Theta(\pi_i)\nabla_\pi\Theta(\pi_i)}{\nabla_\pi\Theta(\pi_i)\cdot\nabla_\pi\Theta(\pi_i)} \quad (10)$$

In our second application of optimizing the potential for a pentapeptide, the system deviates significantly from the quadratic description and we used progressively smaller Δ_i^* with the smallest value being just 1% of the $\Theta(\pi_i)$.

We call the optimization method outlined above POP (parameter optimization). The POP algorithm consists of iterating the steps in Table 1:

Table 1. The POP Algorithm

1	compute $\Theta(\pi_i)$ and $\nabla_\pi\Theta(\pi_i)$ by MD simulations. If $\Theta(\pi_i)$ is lower than a threshold ε, stop. Otherwise continue to 2.
2	Compute α_i.
3	Compute the new parameters, π_{i+1}.
4	Compute $\Theta(\pi_{i+1})$.
	If $\Theta(\pi_{i+1}) < \Theta(\pi_i)$, accept the new parameters. Go to 1.
	If $\Theta(\pi_{i+1}) \geq \Theta(\pi_i)$, reject the new parameters, reduce α_i, and go to 3.

II.3. Alanine Dipeptide Setup. We consider the "toy problem" of solvated alanine dipeptide that was simulated as follows. The program used for molecular dynamics simulations was MOIL.[9] The initial force field parameters were the united atom version of OPLS[5b] (the covalent parameters are from AMBER[10]). The water model was TIP3P.[6] The blocked alanine dipeptide was solvated in a periodic box of 19.6^3 Å^3 with 248 water molecules. The simulations were conducted at a constant temperature of 300 K (NVT ensemble), using velocity scaling. The equations of motion were integrated with a RESPA algorithm,[11] employing two time steps (1 fs and 4 fs[9a]). Water molecules were matrix SHAKED[12] with allowed relative errors in bond lengths of 10^{-12}. SHAKE[13] was introduced on the bonds involving hydrogen atoms with tolerance 10^{-12} Å. The time step was 1 fs. The Lennard−Jones interactions were truncated as at 8.5 Å using a buffer cutoff of 1 Å . Electrostatic interactions were computed with Particle Mesh Ewald[14] that was used with an error tolerance of 10^{-9} in reciprocal space and the grid was 20 × 20 × 20. For each of the iterations of parameter optimization, 100 ns of a MD simulation was run to compute the sensitivity. A structure was saved each 1 ps, thus providing a total of 100000 structures that were processed by the program POP to calculate the observable and its gradient and optimize the potential parameters, as described in Table 1.

II.4. WH5 Setup. *II.4.1. WH5 Helical Fraction.* WH5 is a shortcut for the blocked peptide Ac-WA_3H^+-NH_2, (the histidine is in its protonated form), which was studied experimentally[15] and computationally[15b,16] as a model for the formation of a single helical turn and minimal secondary structure. We seek parametrization of the potential for this peptide that reproduces a critical experimental curve for this system, the temperature-dependent helical fraction.[16] The simulation setup was the same as for alanine dipeptide, unless explicitly stated otherwise. The program MOIL in its GPU version[9a] was used in the simulation. The all-atom version of the OPLS force field[5a] (OPLSAA) started the optimization procedure. The molecule was solvated with 851 TIP3P water molecules in a periodic box size of 30.6 Å^3, and the simulations were conducted in the NVT ensemble using velocity scaling. The Particle Mesh Ewald grid was 32 × 32 × 32.

Simulations were run at temperatures of 280, 300, and 320 K for 360 ns each optimization cycle. Each picosecond of simulation, a structure was recorded and saved. The program POP processed the collected structures. It computes the helical fraction and its derivative with respect to the parameters and modifies the parameters. The process was iterated 4 times, following the algorithm of Table 1, always using the same three temperatures and the same time of simulation (360 ns).

II.4.2. WH5 Melting Curve. The helical fraction is an integrated measure over a range of temperatures. It does not capture the shape of the melting curve or the melting temperature. To study the last we also consider as a target function the derivative of the helical content as a function of temperature. The simulation parameters were the same as in

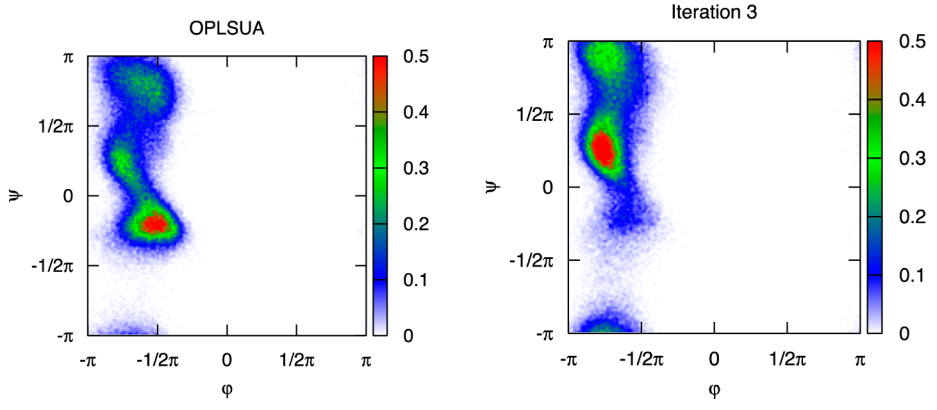


Figure 1. Evolution of the probability density of the dihedral angles (φ and ψ) for alanine dipeptide. The graph on the left shows results obtained with the force field OPLS with united atoms; a propensity to be in the helix configuration is indicated by the high-density region in quadrant III. The optimization process reduces the amount of structures in the helical configuration (quadrant III) to the prescribed value (right).

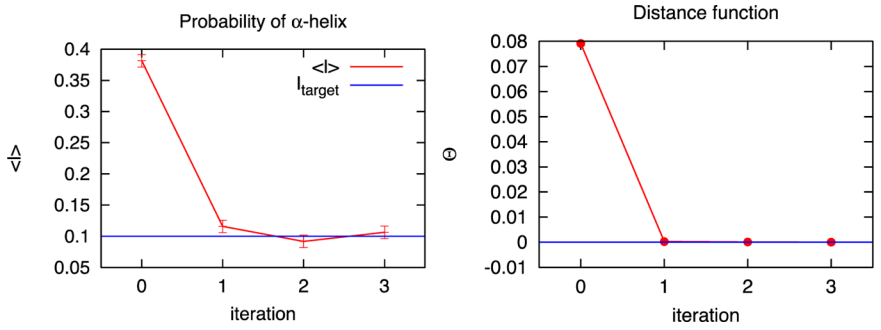


Figure 2. Evolution of probability of an α-helical configuration as a function of the iteration number (left). The target value is achieved in 3 iterations. The distance function, $\Theta(\pi)$, as a function of the iteration number (right).

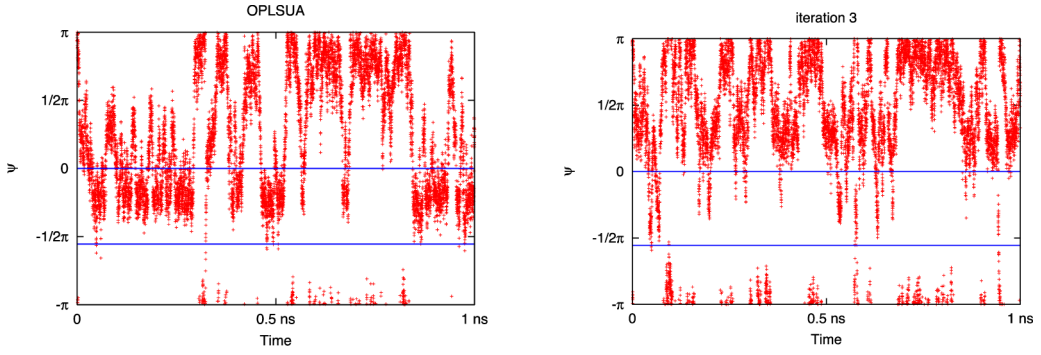


Figure 3. A comparison between a trajectory obtained with the force field OPLSUA (top) and a trajectory obtained with the last iteration of the modified force field (bottom). Both the trajectories are 1 ns long. The original force field shows frequent transition in and out of the helical state whose boundaries (as we defined them) are indicated in blue. The molecule spends ~38% of the time in the helical configuration. The modified force field in the bottom figure shows fewer transitions to the helical state where the molecule spends ~10% of the time.

II.4.1, except that twelve temperatures equally distributed between 280 and 390K were considered and replica exchange simulations were used.[17] Temperature swaps were attempted each 10 steps. The average running time between two temperature swaps was ~2.8 ps. This time, the simulations were run for 400 ns for each step of parameter optimization.

III. RESULTS

III.1. Alanine Dipeptide. Solvated alanine dipeptide has been a model system in computational molecular biophysics for a long time. Here, we examine it as a toy model for which the answer can easily be guessed. We obtain the optimal potential using our automated POP procedure. The conformational state of alanine dipeptide can be characterized by two soft degrees of freedom, the dihedral angles (φ, ψ).

Figure 1 shows the probability density of configurations as a function of (φ, ψ) calculated with the united atom OPLS:[5b] all the populated states are located in the region with negative φ (quadrants II and III). The region with highest probability density is in quadrant III, and it corresponds to helical states.

We define our microscopic observable $O(x,p)$ to be the following indicator function:

$$I(\psi) = \begin{matrix} 1 & -100^\circ \le \psi \le 0^\circ \\ 0 & \text{elsewhere} \end{matrix} \qquad (11)$$

The ensemble average of the function $I(\psi)$ is the probability of finding the molecule in the defined region. At 300 K, approximately 38% of the structures analyzed have a ψ angle that falls in the range from -100° to 0°, therefore

$$\langle I \rangle_{(\text{OPLSUA},300\text{K})} = 0.38 \qquad (12)$$

As a proof of principle, we use POP to modify the energy parameters to make the helix region unfavorable. This change does not imply (of course) that the preferred conformation of alanine dipeptide in nature is of an extended chain. Nevertheless, it is a useful exercise to illustrate the POP algorithm. We seek to modify the force field in such a way that exactly 10% of the structures lie in the helical region, therefore for the function $I(\psi)$, we set up a target value $I_{\text{TARGET}} = 0.1$. We will use a single temperature, so the distance function that we use is

$$\Theta(\pi) = [\langle I \rangle_{(\pi,300K)} - I_{\text{TARGET}}]^2 \qquad (13)$$

We minimized $\Theta(\pi)$ as a function of the parameters of the proper torsions and Lennard–Jones interactions. These energy terms are typical targets for optimization. Charges are frequently determined by quantum mechanic calculations, and we have them fixed in the present manuscript. The total number of parameters that we consider is 72.

One iteration reduces the target function significantly; after three iterations, the $\Theta(\pi)$ distance function is 2×10^{-5}, where $\langle I \rangle_{(\text{POP},300\text{K})} = 0.106 \pm 0.005$ and the target was considered achieved (Figure 2).

Figure 3 shows the difference in the dynamics of a trajectory obtained with united atom OPLS, and one obtained with the last version of the modified potential. When simulated with the original force field, the molecule has frequent transitions between the helical state and the extended chain state, spending a significant amount of time in the helical state; when simulated with the modified potential, the molecule undergoes fewer transitions into the helical state and the amount of time spent in that state is reduced to the one prescribed as a target.

dx.doi.org/10.1021/ct400313n | J. Chem. Theory Comput. 2013, 9, 3311−3320

To explore variations in the target function, we repeated the POP optimization using a different target; this time, instead of reducing the probability of the helical state, we increase the probability of the extended chain state. We defined the extended chain state by a range of the ψ angle (between 90° and 180°) and set I_{TARGET} = 0.9. The results are shown in Figure 4, the target was essentially achieved with just one optimization step.

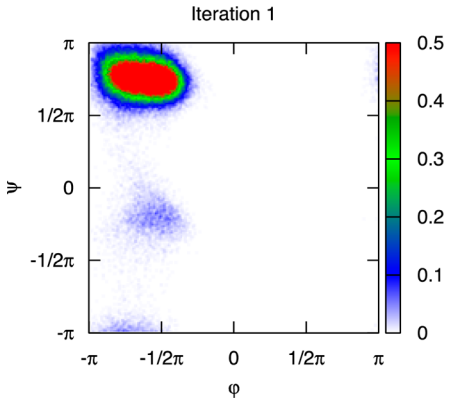


Figure 4. The probability density of states for alanine dipeptide. In this second test, we increased the target probability of the extended chain to 0.9.

The probability of helix as well as the probability of β sheet is largely determined by the torsion potential; the sensitivity with respect to the Lennard−Jones parameters is very small and little to no change was made to those parameters.

In MOIL, all possible torsions are generated based on the bond structure that is computed from the system sequence and prior database of amino acid connectivity. In a second step, these torsions are assigned values according to the force field at hand. Torsions that do not have contributions to the energy are removed from the list to save computational resources. However, during the optimization, it is important to also consider torsions that have no contributions to the current force field but may have one in a refined force field. The POP code was therefore modified to take into account torsion terms with zero coefficients and to compute their parameter derivatives.

Alanine dipeptide has 16 torsions subdivided into 13 torsion types (a torsion type is defined by all four types of atoms in sequence, either in the forward or backward directions). Of the 13 torsion types, 5 torsions are initially set as having no contribution to the energy. Perhaps not surprisingly, most of the changes in the force field following POP are in those torsions. The POP procedure returns more nonzero torsion coefficients. The stiff torsions (rotations of the amide planes) remain stiff as they should, and the free torsions remain relatively soft, even if the coefficients are no longer zero (Table 2 and Figure 5). It is striking that the small, distributed changes of the force field have dramatic effects on the general behavior of the molecule, as demonstrated by the density plots.

III.2. WH5. We applied the POP method also to the pentapeptide Ac-WAAAH-NH_2 (WH5) with the aim of reproducing the experimental helical fraction as a function of the temperature. WH5 is forming an α-helical turn efficiently and rapidly.[16] WH5 consists of only 5 amino acids (i.e., a minimum number of amino acids to form a complete turn of an α-helix). Its secondary structure has been determined by nuclear magnetic resonance (NMR) and circular dichroism (CD), and the kinetics of the helix ↔ coil transition has been studied by T-jump experiments.[15a] The cited reference reports the far UV CD spectra of WH5 and the derived helical fraction as a function of temperature. The small size of WH5 and its fast kinetics make this molecule the ideal candidate to undergo optimization by POP. The required sampling of conformational space should not be a problem for modern computer hardware and software.

The experimentally determined helical fraction for WH5 is reproduced from the experimental data of ref 16 and is shown in Figure 6 in the temperature range 280−360. In reference 16, the CD spectra was empirically matched against known fingerprints of a helix spectrum. The match provided the plot of helical content in Figure 6. The curve is similar in its overall shape to the melting curve measured for the peptide $(AAQAA)_3$ by NMR.[18] The experimental helical fraction decays from 56% to almost zero. We focus on three temperatures 280, 300, and 320 K and conduct the required simulations, as described in II.4.1

Table 2. Coefficients in the Expansion of a Torsion Energy Term $\alpha U_\alpha(\phi) = \sum_n a_{n,\alpha}[1 + \cos(n\phi - \delta_{n,\alpha})]$[a]

torsion type	starting parameters			POP		
	a_1	a_2	a_3	a_1	a_2	a_3
43	0.0000	2.5000	0.0000	−0.0001	2.4998	−0.0005
47	0.0000	2.5000	0.0000	0.0000	2.4999	0.0002
45	0.0000	2.5000	0.0000	0.0000	2.4999	0.0004
51	0.0000	2.5000	0.0000	0.0000	2.4999	−0.0003
17	0.0000	0.0000	0.0000	0.0586	−0.1096	0.0305
76	0.0000	0.0000	0.0000	−0.0682	0.0873	0.0434
30	0.0000	0.0000	0.0000	−0.0604	−0.1110	−0.0289
79	0.0000	0.0000	0.0000	0.0704	0.0883	−0.0408
64	0.0000	0.0000	0.1000	0.2167	−0.0354	0.0375
140	0.0000	0.0000	0.0000	−0.2157	0.2981	0.0548
60	0.0000	2.5000	0.0000	0.0000	2.4999	−0.0002
56	0.0000	2.5000	0.0000	0.0000	2.5001	−0.0002
55	0.0000	2.5000	0.0000	0.0000	2.4999	0.0005

[a]Only the coefficients a_n are optimized. δ_n is fixed at π. Sample of torsion types, their atom types, and their corresponding indices are shown in Figure 5.

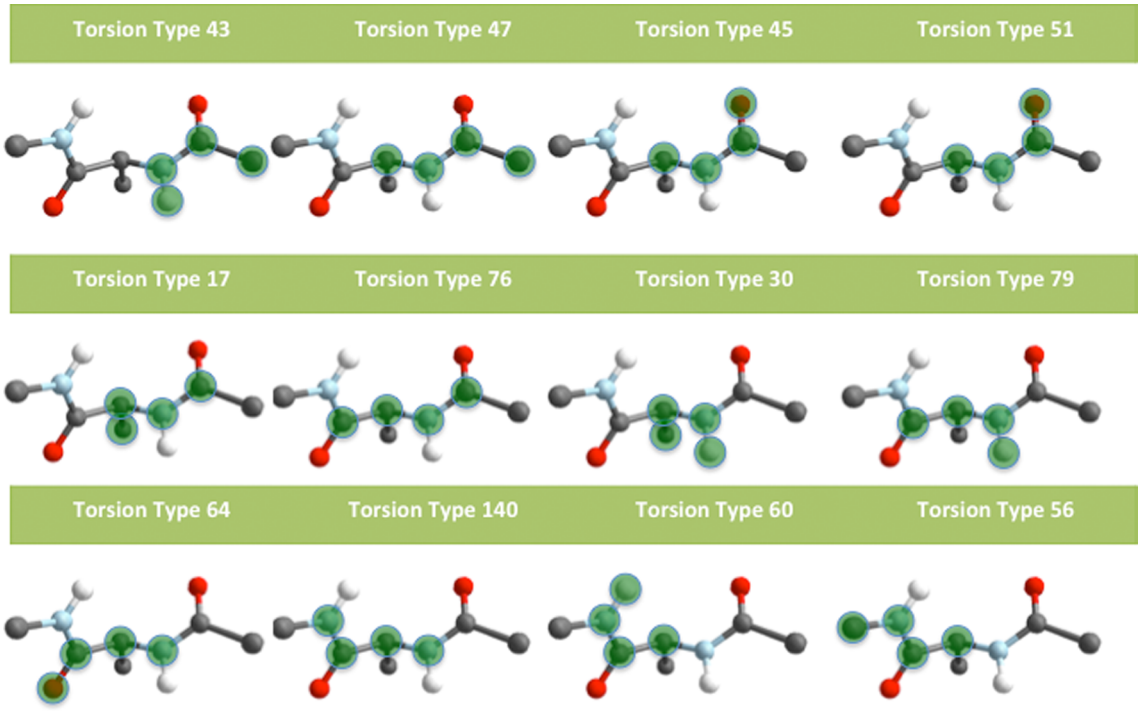


Figure 5. Sampled torsions. In green circles, we show the four atoms that are active in a particular torsion. Sometimes torsion parameters are modeled only according to the central two atoms. Here, we take the broader view and consider torsions with even one of the four atoms different as distinct.

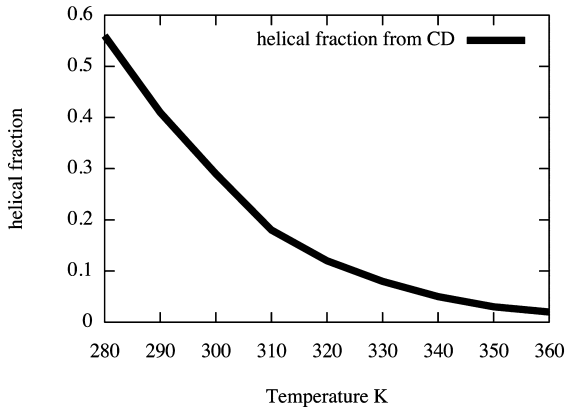


Figure 6. Experimentally measured fraction of helix conformation. See text for more details.

In general, the average value of the (φ, ψ) angles for the α-helix is peaked near the point $(-62°, -41°)$; we considered an amino acid to be in helical conformation if its backbone dihedral angles fall in a circle of radius 20° from the ideal value. Other definitions are possible. In the Lifson–Roig model, emphasis is made on the hydrogen bond, and therefore, a helix is defined if three sequential amino acids are in the helical state. It is not clear if this is the appropriate definition for CD spectra. We therefore try them both. Below, we present the result only for the single-amino-acid definition. The use of three (φ, ψ) pairs does not improve the results.

We define our microscopic observable $O(x,p)$ to be the helical fraction:

$$H(x) = \frac{N_{\text{helix}}}{N_{\text{total}}} \tag{14}$$

where N_{helix} is the number of amino acids in the helical conformation and N_{total} is the total number of amino acids (5 for WH5).

As shown in Figure 7, in this temperature range, the simulation with the starting potential gave helical fractions that are at a significant variance with the experiment. First, the helical fraction calculated from the simulation is less than the experimental value over the complete temperature range. Second, $\langle H \rangle_{(\text{OPLSAA},T)}$ does not exhibit the sharp dependence on temperature seen in the experimentally determined curve. The computational curve is almost constant in the range considered.

We defined the target function to be

$$\Theta(\pi) = \sum_i [\langle H \rangle_{(\pi, T_i)} - H_{\text{experiment}}(T_i)]^2 \tag{15}$$

The index i is running over the three temperatures, 280, 300, and 320 K.

POP optimization was performed on the van der Waals and torsions parameters to minimize $\Theta(\pi)$; water parameters and charges were left untouched.

After the first iteration, the helical fraction was significantly improved reproducing the correct value at ∼300 K (Figure 7). Three more POP iterations were performed with no substantial improvement; correct temperature dependence (or the overall shape of the melting curve) of the in silico helical fraction could not be achieved. The optimization process acted mostly on the parameters of the 207 torsions present in the molecule, thus confirming the intuitive idea that the helical fraction mostly depends on torsions; not one of the Lennard–Jones parameters was modified more the 0.1%.

Remarkably, the changes in torsions parameters were small too. The molecule has 207 torsions subdivided in 97 torsion types. Of the 97 torsion types, just 7 had parameters modified

dx.doi.org/10.1021/ct400313n | J. Chem. Theory Comput. 2013, 9, 3311−3320

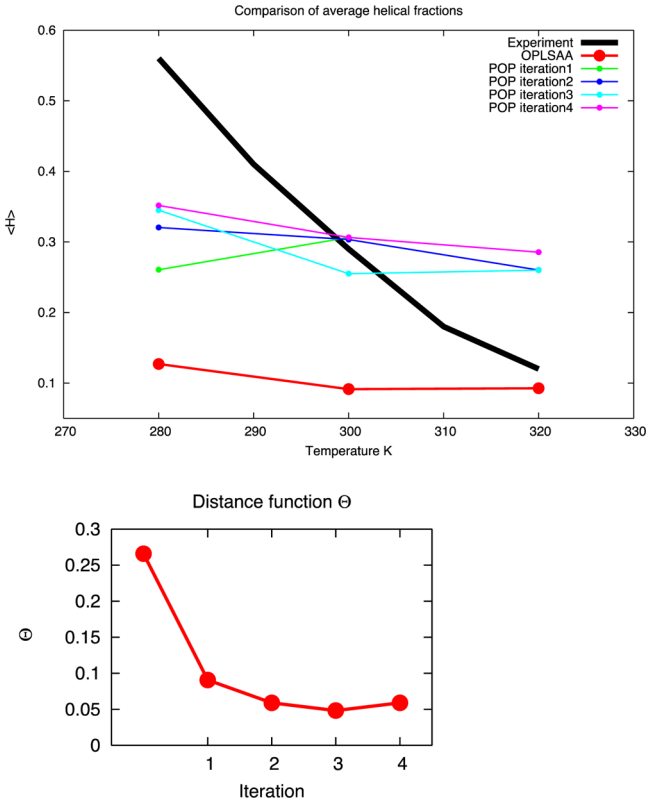


Figure 7. A comparison between the experimental folded fraction (black) and the results obtained from simulations with the OPLSAA force field (red) and optimized POP force fields (top). Distance function $\Theta(\pi)$ as a function of the iteration (bottom).

in the range of 5−10%, 9 torsions had parameters modified in the range of 2−5%, while the rest had changes of less than 2%. These small, distributed changes in the parameter's set had a large effect on the helical fraction, increasing it by more than a factor 2, Figure 8.

Figure 9 shows the probability density of all five amino acids of WH5, in different positions of the Ramachandran map. The figure shows a comparison of probability densities for the starting and optimized force fields. The calculation presented is at temperature 300 K. The result of the optimization is (as designed) an increased population of helical conformations. The differences between the two probability densities are also reported.

A second attempt to optimize the shape of the melting curve was made by modifying the target function. We focused not on the helical fraction but on its derivative with respect to the temperature. We defined the target function to be

$$\Theta(\pi) = \sum_{1}^{2} (D_{(i,\pi)} - D_{(i,\text{experiment})})^2$$

$$D_{(i,\pi)} = \frac{\langle H\rangle_{(\pi.T_{i+1})} - \langle H\rangle_{(\pi.T_i)}}{T_{i+1} - T_i}$$

$$D_{(i,\text{experiment})} = \frac{H_{\text{experiment}}(T_{i+1}) - H_{\text{experiment}}(T_i)}{T_{i+1} - T_i} \quad (16)$$

where T_1, T_2, and T_3 are, as before, the temperatures 280, 300, and 320 K.

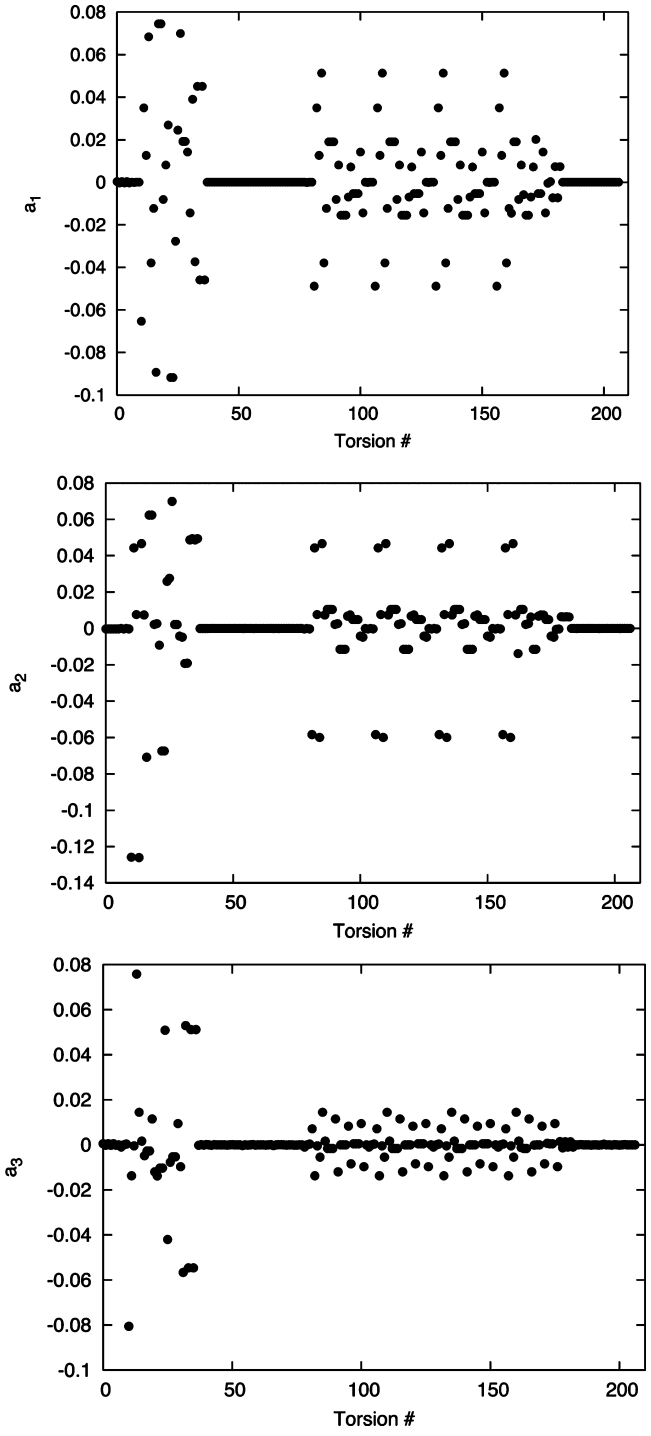


Figure 8. The difference in the coefficients of the torsion potential (a_1, a_2, and a_3) of the OPLS-AA and POP potentials. The differences are small, but they cause significant changes in helical content. Note that all torsions are displayed, which means that torsion types that repeat are shown more than once. The zero differences are mostly for side-chain torsions.

We are therefore optimizing the numerical derivative of the helical fraction with respect to the temperature calculated at the two midtemperatures 290 and 310 K.

The convergence of the sensitivity was verified using independent simulations. Two molecular dynamics trajectories generate configurations to compute the average and the derivatives independently for each of the trajectories. Figure

dx.doi.org/10.1021/ct400313n | J. Chem. Theory Comput. 2013, 9, 3311−3320

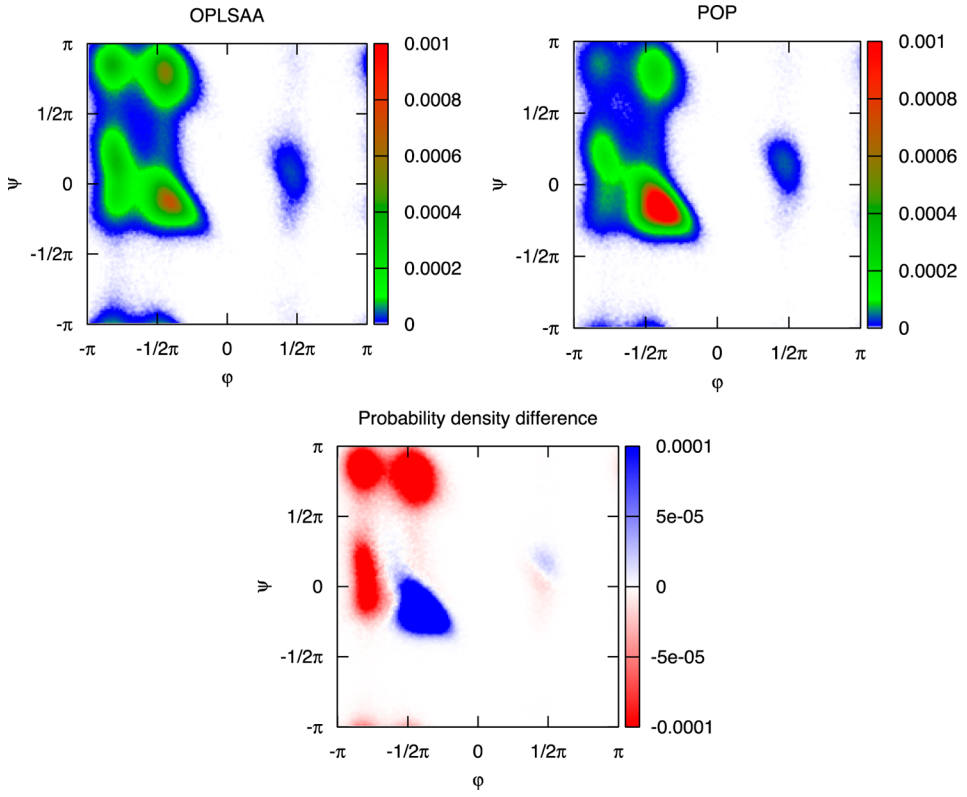


Figure 9. The three graphs show the probability density of finding any of the amino acids of WH5 in a given configuration defined by its dihedral backbone angles. The top figures show the probability density calculated at 300 K from the potential OPLSAA and from the optimized potential POP (iteration 4). The lower figure shows the probability density difference between the two force fields. The probability density is normalized with respect to integration over the dihedrals φ and ψ.

10 illustrates the high correlations between the vectors of derivatives, suggesting that the calculations of the derivatives indeed converge.

The final value of the target function is $\Theta = 2.6 \times 10^{-4}$. No further improvement in the target function was obtained, and the slope was not improved with additional cycles. We concluded that within the error bars of the POP method, it is not possible to adjust the shape of the melting curve.

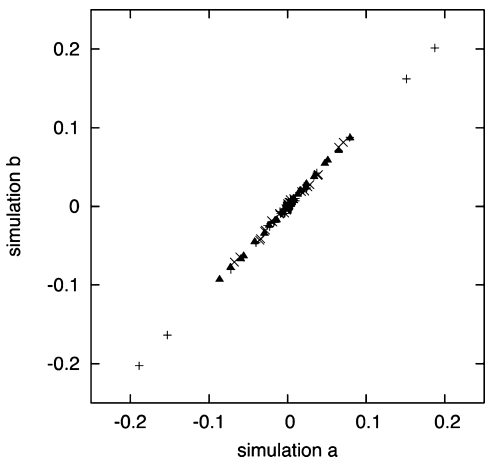


Figure 10. The gradient of the target function computed by ensemble averages of two simulations. We plot the correlation between the two gradients in which every element of the gradient vector is projected onto the axes to obtain the gradient value at each simulation. The high quality linear correlation suggests convergence.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

In this manuscript, we illustrate a combination of sensitivity calculations and direct local optimization of potential parameters for condensed phase simulations. This procedure, called POP, is a potentially useful approach for fine-tuning an energy function for applications in which statistical mechanic averages are compared with experimental measurements. The advantage of the proposed approach is that it can handle a large number of parameters in a single automated calculation. This is not to say that human decisions are completely factored out. Especially for cases in which multiple experimental observables are of interest, we expect to have weight coefficients for a sum of measured observables $\Theta = \sum_i w_i(\langle O_i \rangle - O_{\mathrm{exp},i})^2$, where the weights, w_i, are determined by the user. Experimental uncertainties as well as the importance of a particular observable for future projects impact the choice of the weights.

Another direction that we did not consider in the present investigation is of stochastic sampling, searching for multiple solutions of parameter sets, and potentially a global minimum. In the current investigation, we focused on one observable and a minimal change to the parameters. However, one may apply the same protocol to a large number of observables (weighted appropriately) and a large number of parameters to be modified in a substantial way. The feasibility of conducting converged statistical mechanic simulations on moderately complex molecules makes the present approach attractive.

Finally, we comment on the results for the peptide WH5 and their implications. We found that the helical content was a highly sensitive function of the backbone torsion parameters. In a way, this is good news since the optimization of existing

potentials to obtain the correct helical content is not difficult, provided that the direction of parameter changes is chosen correctly. Only minute adjustments in the potential parameters are required in order to bring the system to the correct conformations in equilibrium. Since the changes in the parameters are small, it is likely that other observables will not be affected. A concern is, however, of transferability. If indeed such small adjustment in parameters cause such large shifts in helical content, it is likely that transferability to other peptide systems (and proteins) would be hard to achieve, since different peptides may require other small adjustments.

Another interesting observation in the simulation of WH5 is our inability to reproduce the melting curve of the peptide. These widely used experimental curves to determine stability energies and entropies of protein folding are insensitive functions of the parameters at hand. In fact, we gave up on the optimization, since we did not seem to make progress with the procedure we use. We stress that our failure is relevant also to single temperature properties, since quantities like entropy are related to the derivatives of the free energy and, hence, are incorrect.

The failure of our efforts to produce a potential that correctly predicts the melting curve discourages us from testing it on other systems. This negative result is supported by other attempts in the field to understand the unusual stability of short helices and to adjust force fields to reproduce this experimental observation. Arguments were made in the past that interactions with charged side chains add to the strong co-operativity of the helix.[19] The current optimization does not include all types of side chains and, hence, our exploration is limited. Nevertheless, some coupling of the backbone conformation is observed for the side-chain torsion of tryptophan, supporting general interaction of side chains to the backbone.[21]

A number of extensions to the standard force field were proposed in the past. First, the water model (which we held fixed) may be inappropriate for a range of temperatures. Water models as TIP4P[22] and TIP5P[23] as well as a polarizable water model[24] are of significant promise since they better describe water properties and water anomalies. An interesting study was published by Best and Mittal[25] that illustrated the complexity of the task. The use of TIP4P/2005[26] showed moderate improvement after the peptide potential was reparameterized. Hence, the peptide parameters are strongly coupled to the water model, which in most force fields are fitted against TIP3P or a similar model. Second, adjustments to the backbone torsion potentials were made. The CMAP approach has shown promise and the progress is encouraging.[21,27] In the CMAP approach, a coupling between the (ϕ, ψ) dihedral angles is introduced, taking into account experimentally determined protein structures. This term is added to the usual Fourier expansion of the torsion energies that considers a torsion at a time. Another extension was proposed by Sakae and Okamoto,[28] in which the torsion energy is made dependent on the amino acid type. The last approach, however, was not illustrated to improve peptide-melting curves. Third, polarizable force fields[29] may also play a future role, and their evaluation is desired.

Finally, we make a cautionary remark. While our results support the designs beyond the standard force field, they are a conjecture and not a proof. Our sampling of parameter space is incomplete. The optimization is based on a local minimization in which a nearby optimum of the target function is searched with a minimal change in the parameter values. The parameter space is rather smooth (especially compared to the coordinate space), since the energy dependence on the parameters that we optimize is either linear (torsion) or quadratic (Lennard–Jones). For simple (quadratic) target functions, the local minimization is exact. However, the Boltzmann factor makes the target function more complex, and if exploration of multiple minima is required we will miss it.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ron@ices.utexas.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by NIH grant GM59796 and Welch grant F-1783 to R.E.

REFERENCES

(1) Frenkel, D.; Berend, S. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, 1996.

(2) Abrams, J. B., Tuckrman, M. E., Martyna, G. J. In *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer: Berin, 2006; Vol. *1*, pp 139–192.

(3) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.

(4) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The AMBER biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(5) (a) Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105* (28), 6474–6487. (b) Jorgensen, W. L.; Tiradorives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666.

(6) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(7) Nocedal, J.; Wright, S. J. *Numerical Optimization*; Springer: New York, 1999.

(8) (a) Chen, A. A.; Pappu, R. V. Parameters of monovalent ions in the AMBER-99 forcefield: Assessment of inaccuracies and proposed improvements. *J. Phys. Chem. B* **2007**, *111* (41), 11884–11887. (b) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712–725.

(9) (a) Ruymgaart, A. P.; Cardenas, A. E.; Elber, R. MOIL-opt: Energy-conserving molecular dynamics on a GPU/CPU system. *J. Chem. Theory Comput.* **2011**, *7* (10), 3072–3082. (b) Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Li, H. Y.; Verkhivker, G.; Keasar, C.; Zhang, J.; Ulitsky, A. Moil a program for simulations of macrmolecules. *Comput. Phys. Commun.* **1995**, *91* (1–3), 159–189.

(10) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784.

(11) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular-dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990−2001.
(12) (a) Weinbach, Y.; Elber, R. Revisiting and parallelizing SHAKE. *J. Comput. Phys.* **2005**, *209* (1), 193−206. (b) Ruymgaart, A. P.; Elber, R. Revisiting molecular dynamics on a CPU/GPU system: Water kernel and SHAKE parallelization. *J. Chem. Theory Comput.* **2012**, *8* (11), 4624−4636.
(13) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of cartesian equations of motion of a system with constraints: Molecular dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(14) 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* (19), 8577−8593.
(15) (a) Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H. Speed limit of protein folding evidenced in secondary structure dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (40), 16622−16627. (b) Hegefeld, W. A.; Chen, S. E.; DeLeon, K. Y.; Kuczera, K.; Jas, G. S. Helix formation in a pentapeptide experiment and force-field dependent dynamics. *J. Phys. Chem. A* **2010**, *114* (47), 12391−12402. (c) Mohammed, O. F.; Jas, G. S.; Lin, M. M.; Zewail, A. H. Primary peptide folding dynamics observed with ultrafast temperature jump. *Angew. Chem., Int. Ed.* **2009**, *48* (31), 5628−5632.
(16) Jas, G. S.; Hegefeld, W. A.; Majek, P.; Kuczera, K.; Elber, R. Experiments and comprehensive simulations of the formation of a helical turn. *J. Phys. Chem. B* **2012**, *116* (23), 6598−6610.
(17) (a) Earl, D. J.; Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7* (23), 3910−3916. (b) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151. (c) Kirmizialtin, S.; Elber, R. Computational exploration of mobile ion distributions around RNA duplex. *J. Phys. Chem. B* **2010**, *114* (24), 8207−8220.
(18) Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of helicity within the model peptide acetyl$(AAQAA)_3$ amide. *J. Am. Chem. Soc.* **1994**, *116* (18), 8288−8293.
(19) (a) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (24), 13075−13079. (b) Garcia, A. E.; Sanbonmatsu, K. Y. alpha-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (5), 2782−2787.
(20) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004−9015.
(21) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257−3273.
(22) Jorgensen, W. L.; Madura, J. D. Temperature and size dependence for monte-carlo simulations of tip4p water. *Mol. Phys.* **1985**, *56* (6), 1381−1392.
(23) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112* (20), 8910−8922.
(24) Ren, P. Y.; Ponder, J. W. Temperature and pressure dependence of AMOEBA water model. *J. Phys. Chem. B* **2004**, *108* (35), 13427−13437.
(25) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114* (46), 14916−14923.
(26) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505.
(27) (a) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400−1415. (b) Best, R. B.; Mittal, J.; Feig, M.; MacKerell, A. D. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. *Biophys. J.* **2012**, *103* (5), 1045−1051.
(28) Sakae, Y.; Okamoto, Y. Amino-acid-dependent main-chain torsion-energy terms for protein systems. *J. Chem. Phys.* **2013**, *138* (6).
(29) (a) Wu, J. C.; Chattree, G.; Ren, P. Y. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3); (b) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D. Development of CHARMM polarizable force field for nucleic acid bases based on the classical drude oscillator model. *J. Phys. Chem. B* **2011**, *115* (3), 580−596.