Journal of
**COMPUTATIONAL
CHEMISTRY**

Development and Use of an Atomistic CHARMM-Based Forcefield for Peptoid Simulation

Dina T. Mirijanian, Ranjan V. Mannige, Ronald N. Zuckermann, and Stephen Whitelam*

Peptoids are positional isomers of peptides: peptoid sidechains are attached to backbone nitrogens rather than α-carbons. Peptoids constitute a class of sequence-specific polymers resistant to biological degradation and potentially as diverse, structurally and functionally, as proteins. While molecular simulation of proteins is commonplace, relatively few tools are available for peptoid simulation. Here, we present a first-generation atomistic forcefield for peptoids. Our forcefield is based on the peptide forcefield CHARMM22, with key parameters tuned to match both experimental data and quantum mechanical calculations for two model peptoids (dimethylacetamide and a sarcosine dipeptoid). We used this forcefield to demonstrate that solvation of a dipeptoid substantially modifies the conformations it can access. We also simulated a crystal structure of a peptoid homotrimer, H-(N-2-phenylethyl glycine)$_3$-OH, and we show that experimentally observed structural and dynamical features of the crystal are accurately described by our forcefield. The forcefield presented here provides a starting point for future development of peptoid-specific simulation methods within CHARMM. © 2013 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23478

Introduction

Peptoids are structurally identical to peptides in all respects but one: peptoid sidechains (*R*-groups) are attached to backbone nitrogens instead of α-carbon atoms (Fig. 1). This difference confers upon peptoids structural properties very different to those of their peptoid counterparts. For example, peptoids lack the strong intra-backbone hydrogen bonding ability of proteins, and so do not form structures stabilized by hydrogen bonds.[1] Instead, peptoids achieve a similar range of structural variety through other mechanisms, including steric-, hydrophobic-, and electrostatic interactions. Examples include the formation of secondary structures such as helices[2,3] and sheets,[4] tertiary structures such as helical bundles,[5] and quaternary structures such as ordered bilayer nanosheets.[6–8] Peptoid and peptide chemical properties also differ: peptoids are resistant to protease degradation, for instance, but retain the ability to bind to target proteins.[9,10] This combination makes peptoids potentially useful as therapeutics[11] and diagnostics.[12] The peptoid world can furnish molecules useful for medicine,[13–16] catalysis,[17] and nanotechnology[5,6,18] that complement the peptide world's repertoire.

However, despite substantial progress in experimental peptoid chemistry,[2–6,9,11] our theoretical picture of peptoid behavior is incomplete. While quantum mechanical treatments are useful for exploring structures available to small peptoids – such as dimers – in vacuum,[19–21] understanding the solution- and condensed-phase structures, dynamics and self-assembly of peptoids calls out for simulation methods that can access larger lengthscales and timescales. Among these methods is classical atomistic molecular dynamics, in which atoms are treated as point masses undergoing Newtonian dynamics.[22–27] Atomistic simulation of peptoids has been done using peptide forcefields.[28,29] However, given the significant differences in structural and physical properties between peptoids and peptides, accurate simulation of peptoids requires a forcefield tailored to describe their particular atomistic behaviors.

Here, we introduce an atomistic forcefield for peptoids. This forcefield is rendered in CHARMM, a molecular dynamics simulation framework originally designed for proteins,[30] and subsequently extended to treat nucleotides,[31] carbohydrates,[32] and drugs.[33] Our parameterization scheme (overviewed in Fig. 2) follows closely the scheme used to parameterize peptides in CHARMM22.[30,34] Starting with the peptide forcefield, we used target data from both quantum mechanical calculations and experiments to tune the parameters that we expect to differ most between peptide and peptoid (see Fig. 1).

Using this forcefield, we carried out all-atom simulations of a peptoid in vacuum and in explicit water, demonstrating that solvating a sarcosine dipeptoid changes its accessible backbone conformations. We also simulated a crystal structure of a peptoid homotrimer (bearing *N*-2-phenylethyl sidechains),[29] a molecule previously shown to be not well described by peptide-based forcefields.[29] The structure and dynamics of the crystal seen in simulations agree qualitatively with properties observed experimentally.[29] We anticipate that the first-generation forcefield presented here will enable a growing community of researchers to do accurate atomistic simulations

D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, S. Whitelam
Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720
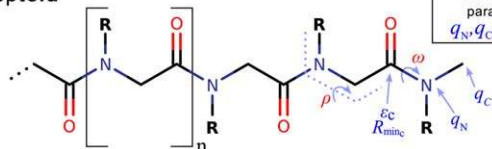E-mail: swhitelam@lbl.gov

ChemistryViews
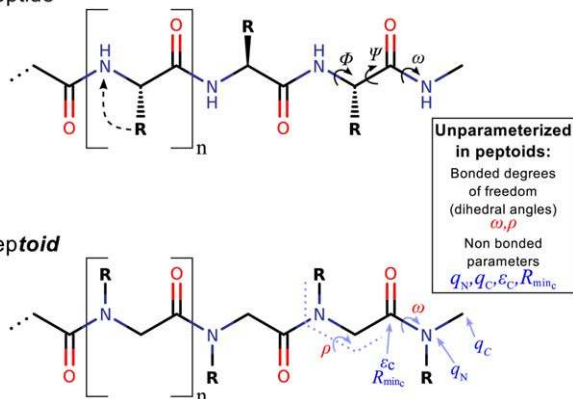
Figure 1. Peptoids are positional isomers of peptides whose sidechains (R) are attached to the backbone N instead of the α-carbon (C). To describe peptoids within the CHARMM forcefield, we tuned the parameters shown in blue and tuned (or introduced) parameters associated with the degrees of freedom colored red (see Figs. 2 and 3).

of peptoids, and will provide a starting point for the development of future peptoid forcefields.

Parameterization Philosophy

Parameters that differ between peptides and peptoids

The relocation of the sidechain from the α-carbon (Fig. 1a) to the amide nitrogen (Fig. 1b) results in key differences between peptides and peptoids. For example, the reduction of the

double-bonded character of the tertiary amide bond[21] renders peptoid backbone dihedral angles ω more flexible than their peptide counterparts. Further, the absence of an amide polar hydrogen atom prevents the formation of strong backbone-backbone hydrogen bonds. Our parameterization scheme, therefore, focused on tuning the CHARMM forcefield parameters that describe the amide core. These parameters, shown in Figure 1, are the partial charge of the amide nitrogen atom q_N; the partial charge of its associated sidechain carbon q_C; the Lennard-Jones "well depth" ε_C and radius R_{min_C} parameters for the carbonyl carbon atom; and the parameters associated with the amide dihedral angle ω. Also, while the dihedral angles ϕ and Ψ are not directly affected by the transformation from peptide to peptoid, the presence of a sidechain on the amide nitrogen atom affects the behavior of the backbone generally. We therefore introduced a set of parameters associated with the interaction between a sidechain's β carbon and the backbone (called ρ in Fig. 1b). Figure 1b lists the peptide forcefield elements whose associated parameters we tuned; Figure 3b lists the values of these parameters in the peptide CHARMM22 forcefield,[30] and in its peptoid counterpart.

CHARMM forcefield terms

The complete CHARMM forcefield functional form is reproduced in eq. (4). Here, we consider only the forcefield energy terms associated with the parameters we tuned. The first terms of interest are the bonded dihedral angle terms describing the behavior of ω and ρ. These terms are of the form
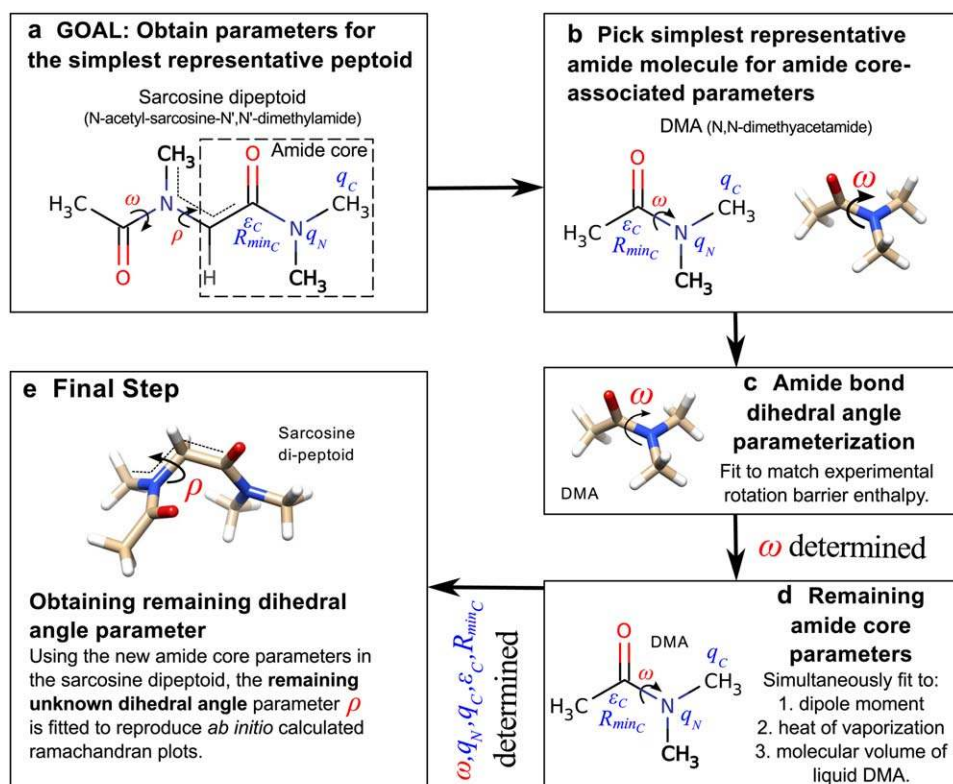

Figure 2. Overview of peptoid forcefield parameterization scheme.

a

b

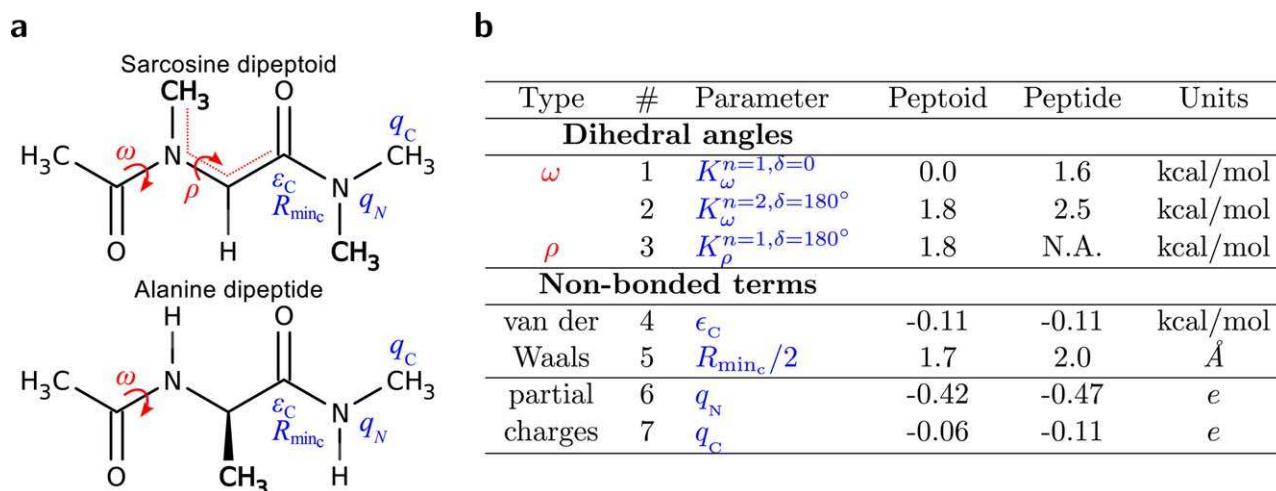| Type	#	Parameter	Peptoid	Peptide	Units
Dihedral angles					
ω	1	$K_\omega^{n=1,\delta=0}$	0.0	1.6	kcal/mol
	2	$K_\omega^{n=2,\delta=180°}$	1.8	2.5	kcal/mol
ρ	3	$K_\rho^{n=1,\delta=180°}$	1.8	N.A.	kcal/mol
Non-bonded terms					
van der	4	ϵ_C	-0.11	-0.11	kcal/mol
Waals	5	$R_{min_c}/2$	1.7	2.0	Å
partial	6	q_N	-0.42	-0.47	e
charges	7	q_C	-0.06	-0.11	e

Figure 3. Tuned peptoid forcefield parameters and their peptide analogs. a) Backbone peptoid model (top) and its peptide analog (bottom),[30] showing degrees of freedom and parameters. b) Values of the associated parameters in our peptoid forcefield and in the peptide forcefield. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$U_D(\xi) = K_\xi(1 + \cos(n\xi - \delta)). \tag{1}$$

Here ξ ($=\omega$ or ρ) is a dihedral angle; K_ξ is the force constant associated with that angle (the energy barrier between minima is $2K_\xi$; n defines the number of minima accessible to ξ; and δ (given n) determines the angles at which minima are found. The Lennard-Jones potential for the carbonyl carbon atom has the form

$$U_{LJ}^{ij} = \epsilon_{ij}\left[2\left(\frac{R_{min}^{ij}}{r_{ij}}\right)^6 - \left(\frac{R_{min}^{ij}}{r_{ij}}\right)^{12}\right], \tag{2}$$

where r_{ij} is the distance between atoms i and j; R_{min}^{ij} is the distance between i and j at the potential minimum; and $\epsilon_{ij} < 0$ is the binding energy at this minimum. Finally, the Coulomb potential describing charge-charge interactions is

$$U_C^{ij} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}, \tag{3}$$

where q_i is the partial atomic charge, ϵ_0 is the permittivity of free space, and ϵ_1 is the relative permittivity. As in the CHARMM22 peptide forcefield, ϵ_1 was set to unity during parameterization.

Model molecules used for parameterization

Consistent with the scheme used to parameterize the peptide backbone,[30] we tuned peptoid backbone parameters using the model molecule N-acetyl-sarcosine-N′,N′-dimethylamide (Fig. 2a). This molecule possesses all backbone dihedral features of longer peptoids, but is small enough to be easily simulated. Although the molecule is not what a synthetic chemist would call a peptoid dimer, we refer to it as "sarcosine dipeptoid," in imitation of its closest peptide analog, alanine dipeptide. Although the latter is too small to reproduce the full range of peptide behaviors, it is a good model for obtaining peptide parameters.[35] By analogy, we expect the same to be true of sarcosine dipeptoid.

This molecule has seven degrees of freedom whose associated parameters need to be tuned simultaneously (Fig. 3), which

for the present study is an intractable search space. To narrow this space, we chose in addition to work with the smallest molecule that possesses the amide core environment of a large polypeptoid: dimethylacetamide or DMA (Fig. 2b). This is a tertiary amide capped by methyl groups on both sides (on the amide N and carbonyl C atoms). The methyl caps serve, in an approximate manner, to simulate the interior of a peptoid, which allows DMA to represent the tertiary amide core of the peptoid backbone more faithfully than do simpler amides such as formamide, acetamide, and N-methylacetamide (NMA). In using DMA to model the amide core of peptoids, we were motivated in part by the use of NMA to model the amide core of peptides.[36]

In the next section, we describe the computational details involved with our parameterization schematic and simulations. In the Results section, we describe how we obtained the amide core parameters by fitting simulated DMA properties to experimental target data (Figs. 2c and 2d). We also describe how we set the parameters associated with the dihedral angle parameter ρ, by comparing sarcosine dipeptoid simulations with quantum mechanical calculations (Fig. 2e). We adhered closely to the general approach used in developing the CHARMM peptide forcefield, requiring after every trial parameter change that all target data were compared against forcefield simulations.

Computational Details

Complete functional form of the forcefield and potential energy scans

The energy function used in CHARMM,[30] given a set of atom positions ($\vec{R}$), is

$$U(\vec{R}) = + \sum_{bonds} K_b(b - b_0)^2 + \sum_{UB}(S - S_0)^2 + \sum_{angle} K_\theta(\theta - \theta_0)^2$$
$$+ \sum_{dihedrals} K_\xi(1 + \cos(n\xi - \delta)) + \sum_{impropers} K_{imp}(\omega - \omega_0)^2 \tag{4}$$
$$+ \sum_{nonbond} \epsilon_{ij}\left[2\left(\frac{R_{min}^{ij}}{r_{ij}}\right)^6 - \left(\frac{R_{min}^{ij}}{r_{ij}}\right)^{12}\right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}},$$

where $K_b, K_{UB}, K_\theta, K_\xi$, and K_{imp} are the force constants associated with bonds, Urey-Bradley angles, dihedral angles, and improper dihedral angles, respectively; b and S are the Urey-Bradley and 1,3-distances; and θ, ξ, and ω are bond-, dihedral-, and improper torsion angles, respectively. δ describes the values in dihedral angle space at which energy is minimum, and subscripts "0" indicate reference values. In nonbonded terms, R_{min}^{ij} is the minimum energy separation, with associated energy $\varepsilon_{ij} < 0$. q_i is the partial atomic charge of atom i, ε_0 is the permittivity of free space, ε_1 is the relative permittivity, and r_{ij} is the distance between atoms i and j.

We performed a one-dimensional (1D) potential energy scan along the ω dihedral angle of DMA and along the ρ dihedral angle of the sarcosine dipeptoid. We also performed a 2D potential energy scan along the dihedral angles of ϕ and Ψ of the sarcosine dipeptoid. In all cases, a constrained structure was minimized with a steepest decent algorithm for 50 steps followed by an adopted basis Newton Raphson method minimization of 1000 steps.[37] Once minimization was complete, we removed the constraint and calculated the energy of the structure.

Molecular dynamics simulations

We performed molecular dynamics simulations using the CHARMM simulation package.[37] We used the SHAKE algorithm[22] to constrain bonds associated with hydrogen atoms. For simulations done using periodic boundary conditions, we used the Particle Mesh Ewald summation method[38] to evaluate long-ranged electrostatic interactions, with a real space cutoff of 12 Å, a sixth-order cubic spline and a kappa value of 0.34. The Lennard-Jones interactions were cut off at 10 Å and had applied to them a tapering-switching function over the range 10–12 Å. It has been shown that TIP3P water with dispersion cutoffs of 10 and 12 Å that have been tapered over 1 Å differ in density from water with no Lennard-Jones cutoff by 0.4 and 0.3%, respectively.[39] Because we do not expect differences of this order to significantly affect the trends we see in our data, no long range correction was applied to the Lennard-Jones interactions.

Liquid-state DMA simulations were carried out in the Isothermal-isobaric (NPT) ensemble ensemble (at 298 K and 1 atm) using a cubic periodic box containing 216 DMA molecules, initially 32 Å in length. A Langevin barostat and Nosé-Hoover thermostat were applied, and a leapfrog algorithm with a time step of 1 fs was used to integrate equations of motion. The system was equilibrated for 1 ns, followed by another 1 ns for averaging to determine the molecular volume. The average simulation box length so determined was then used in the 1 ns Canonical (NVT) ensemble ensemble simulation to determine the average dipole moment for a single DMA molecule and the internal energy per DMA molecule in the liquid state. A single DMA molecule was simulated in vacuum using a Langevin integrator with a friction coefficient of $5\,ps^{-1}$. The gas-phase energy of this molecule was obtained from a 4 ns simulation, which followed a 1 ns equilibration period. The time step used was 1 fs. The average energies calculated from the liquid-phase simulation and the gas-phase simulation were used to determine the heat of vaporization of DMA.

The model molecule sarcosine dipeptoid was solvated in TIP3P water,[40] modified for the CHARMM forcefield.[36] The software suite Visual Molecular Dynamics was used for initial placement of water molecules.[41] Simulation boxes contained 829 atoms for trans state calculations, and 931 atoms for cis state calculations. Systems were equilibrated in the NPT ensemble (at 298 K and 1 atm) using the Langevin barostat and Nose-Hoover thermostat for 1 ns. The average simulation box lengths so determined were used for umbrella sampling simulations in the NVT ensemble, to calculate 2D free energy profiles[22] (using 2D-WHAM[42] to unbias histograms). CHARMM was used to analyze simulation trajectories. The 1D cis/trans free energy profile in Figure 7c was calculated using umbrella sampling with simulation bin sizes of 10 degrees. A time step of 2 fs was used for all free energy sampling simulations. We compared 1D free energy profiles from simulations using a timestep of 1 and 2 fs, and found that sampling with both time steps had converged (in the Supporting Information Fig. S1). For 2D free energy calculations, the ϕ and Ψ angles for each bin sample were constrained with a harmonic potential of force constant 100 kcal/mole. The angle ω was constrained for the cis and trans conformations using the dihedral potential shown in eq. (4), with a single well and a force constant of 40 kcal/mole. In the final 2D histograms, ϕ ranged from $-165.9°$ to $171.7°$ with 50 bins; ψ ranged from $5.7°$ to $177.4°$ with 20. We imposed a convergence tolerance of 10^{-4}.

Crystal structures were equilibrated starting from crystal coordinates taken from experiment. A single unit cell was replicated three times in each direction, resulting in 27 repeat units and 216 molecules. Systems were equilibrated using initial simulations of 5 ns; subsequent sampling simulations of 6 ns were done in the NPT ensemble at 300 K and 1 atm.

The blue histogram in panel c of Figure 9 was obtained by simulating a single trimer (extracted from the crystal structure) in implicit hexane (a generalized Born solvation model was used with simple switching[43] and a dielectric constant of 2). The χ and ξ values for the central sidechain were recorded for 2 ns following equilibration.

Ab initio calculations

Initial coordinates for sarcosine dipeptoid vacuum energy calculations were taken from energy-minimized structures given in Ref. [19]. We used the Qchem quantum chemistry package to optimize local geometries following angle rotations with second-order Möller–Plesset perturbation theory (MP2) and a 6-31+G* basis, within a frozen core approximation.[44] The geometry optimizations were considered converged when an energy tolerance of 10^{-6} Hartree and gradient tolerance of 3×10^{-4} Hartree/Å were reached. We used these geometries within the Molpro quantum chemistry package[45] to calculate single-point energies. These energies were calculated using

density-fitted second-order Möller–Plesset perturbation theory (DF-MP2)[46] with an aug-cc-pVQZ basis set. Using the double slash notation, the single-point energy calculations can be represented as DF-MP2/aug-cc-pVQZ//MP2/6-31+G*.

Results and Discussion

Parameterization steps

This section describes the steps taken to arrive at a suitable parameter set for peptoids; these steps are shown schematically in Figure 2.

Amide bond dihedral angle parameter. Given that the flexibility of the amide dihedral angle ω is an important feature of peptoids, we started by tuning forcefield parameters associated with this angle, using the model molecule DMA. Experimental nuclear magnetic resonance (NMR) data of solvated DMA[47] show the rotational energy barrier for its ω dihedral angle to be 17 kcal/mol. This was the target for our parameterization.

CHARMM22 has two terms associated with the ω dihedral angle, each of the form eq. (1). The first is a single well potential term ($n=1$, $\delta=0°$) describing an energy bias of magnitude $2K_\omega^{n=1, \delta=0}$ in favor of the trans ω conformation ($\omega=180°$); the second is a double well potential ($n=2$, $\delta=180°$) that ensures minima at cis ($\omega=0°$) and trans positions, separated by an energy barrier of $2K_\omega^{n=2, \delta=180°}$. Because both cis and trans conformations are observed in DMA with roughly equal probability,[21] we set the trans bias parameter $K_\omega^{n=1, \delta=0}$ to zero; it is 1.6 kcal/mol in the peptide forcefield. We note that while the "bare" backbone properties do not impose a preference for cis or trans, a bias of this nature can emerge through nonbonded interactions (e.g., sidechain steric interactions), as we will demonstrate. We found that a force constant $K_\omega^{n=2, \delta=180°} = 1.6$ kcal/mol (peptide forcefield value 2.5 kcal/mol) gave rise to a rotational energy barrier between cis and trans conformations of 16.8 kcal/mol, acceptably close to the 17 kcal/mol activation energy and 16 kcal/mol enthalpy change seen in experiment.

Amide core nonbonded terms. With the new force constants for the amide dihedral angle ω in hand, we tuned the amide core nonbonded parameters q_N, $R_{minc}/2$, and ε_C, using as target data the experimental values of DMA's heat of vaporization and liquid molecular volume,[48,49] and its calculated liquid-phase dipole moment.[49] In the CHARMM22 peptide forcefield, optimization of the van der Waals parameters for the NMA amide group was limited to adjustment of the carbonyl carbon radius and well depth; the parameters for all other atoms were taken from previous studies on nucleic acids and alkanes.[36] Guided by this approach, we adjusted for DMA only the Lennard-Jones parameters of the carbonyl carbon atom. The DMA molecule's tertiary amide nitrogen atom partial charge was also reparameterized; for its Lennard-Jones parameters, we took the values appropriate for nitrogens of primary and secondary amides (these choices follow those made in parameterizing tertiary amides in the polarizable intermolecular potential function (PIPF) forcefield for DMA)[49]. Given these

choices, Figure 4 describes how we arrived at the parameter values $q_N=-0.42\,e$, $R_{minc}/2=1.7Å$, and $\varepsilon_C=0.11$ kcal/mol. We assigned the methyl carbon's partial charge $q_C=-0.06\,e$ by requiring that the molecule be neutral. We point out that an alternate model for DMA now exists in the CHARMM General Force Field (CGENFF).[50] The average gas-phase dipole moment of a single DMA molecule from simulation using the parameters determined here, 4.40 D, and that found with using the CGENFF parameters, 4.55 D, are very similar. Both are within 20% of the experimental value of 3.7 D.[51] We also compared the dipole moment of a minimized structure of DMA to its experimental gas-phase value (see Supporting Information).

Sarcosine dipeptoid backbone rotation refinement. We transferred these tuned parameters to our second peptoid backbone model, sarcosine dipeptoid. We expect no major structural differences between peptoids and proteins with respect to the bonds involved in ϕ and Ψ dihedral angles (Fig. 3), and so we chose to keep the CHARMM22 parameters associated with these angles (we dispensed with the grid-based, peptide-specific CHARMM $\phi-\psi$ CMAP backbone cross terms). This leaves us with the dihedral angle ρ (see Fig. 3), which influences the interaction of the first carbon of the sidechain (R-group) with the backbone. We determined the force constant K_ρ by comparing vacuum energy calculations using the forcefield with the ϕ-dependent energy profile of sarcosine dipeptoid calculated using quantum mechanical methods (using the aug-cc-pVQZ basis set within the DF-MP2 level of theory). This profile is shown as a solid black line in Figure 5a; symbols show the results of forcefield energy calculations using four values of K_ρ. These calculations reveal three local energy minima, corresponding to the structures (depicted in Fig. 5b) called[20] $C_{7\beta}$, α, and $C_{7\beta}^{\dagger}$ (the "†" indicates that the first and third structures are related by symmetry).[1] From this set of calculations, we chose $K_\rho=1.8$ kcal/mol, thereby completing our parameterization scheme.

Simulation of the Peptoid Forcefield

In this section, we describe our use of the peptoid forcefield to (1) generate a sarcosine dipeptoid vacuum energy landscape and free energy landscape; (2) generate a solvated sarcosine dipeptoid free energy landscape; and (3) to simulate the crystal structure[29] of a tripeptoid that was found to challenge structure prediction methods that used peptide forcefields.[29]

---

[1]The α_D conformation was found to be locally stable by previous quantum mechanical calculations that used a smaller basis set (6-31+G*) within the MP2 level of theory[20] (black dotted line in Fig. 5a). We note that the larger basis set used here shows no such minimum, confirming that even qualitative features of quantum calculations of biomolecules are sensitive to fine details of their implementation.[52] We find the α_D conformation to be a local minimum only in terms of molecular free energy, and to be stabilized by solvation.
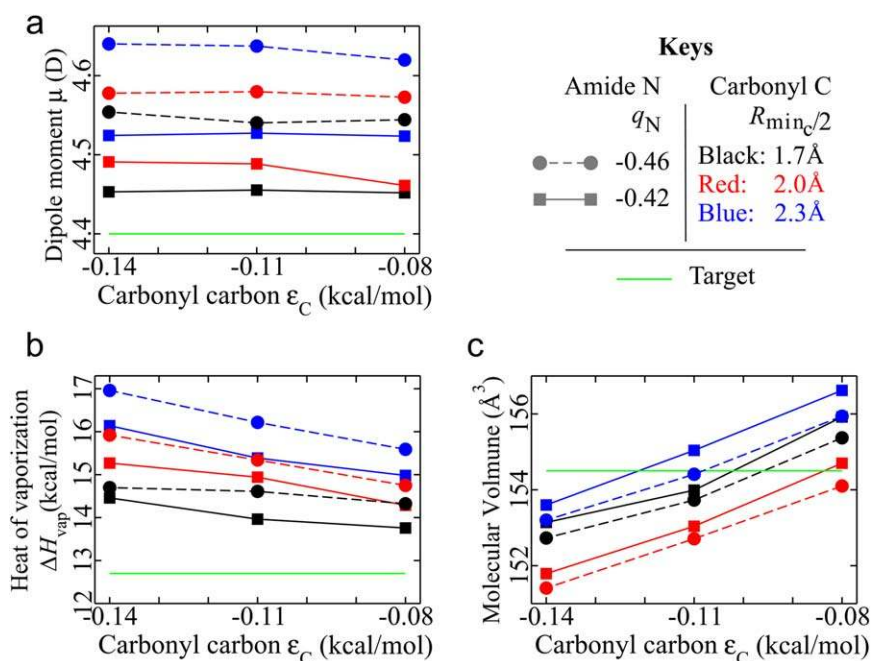
Figure 4. Determination of the key nonbonding amide core parameters using DMA data. Forcefield simulation data of DMA's a) liquid-phase dipole moment, b) heat of vaporization, and c) liquid molecular volume, shown on respective vertical axes, are plotted against various choices of carbonyl carbon well-depths ε_C (horizontal axes). On each panel, we show the effect of varying amide nitrogen partial charges q_N (solid lined squares vs. dashed circles) and carbonyl carbon equilibrium distances R_{min_c} (black, red, and blue). The target values (green lines: $\mu=4.4D$, $\Delta H_{vap}=12.7$ kcal/mol and molecular volume 154.5 Å^3/mol) were obtained from polarizable forcefield simulations of liquid DMA ($\mu^{[49]}$ and from experiment) ($\Delta H^{[49]}$ and molecular volume)$^{[48]}$. The largely horizontal lines in (a) and (b) show that the dipole moment and heat of vaporization do not depend strongly upon ε_C, while the pair of values $q_N=-0.42$ and $R_{min_c}=1.7$Å (indicated by the solid black line with squares as symbols) are closest to the target values in both panels (a) and (b). The parallel sloped lines that all intersect the target value in panel (c) indicate that the target liquid volume is met with a number of triplet parameter sets. However, given that the targets (a) and (b) are best met by $q_N=-0.42$ and $R_{min_c}=1.7$, we chose that combination (solid black line with square symbols), which sets $\varepsilon_C=-0.11$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Energy landscape of sarcosine dipeptoid in vacuum

We used quantum mechanical methods within a DF-MP2/aug-cc-pVQZ//MP2/6-31+G* level of theory to calculate a Ramachandran-like $\psi-\phi$ energy landscape for sarcosine dipeptoid in vacuum. Taking as an initial configuration the molecule in energy-minimized trans $C_{7\beta}$ conformation ($\phi=20°$, $\psi=-77°$,

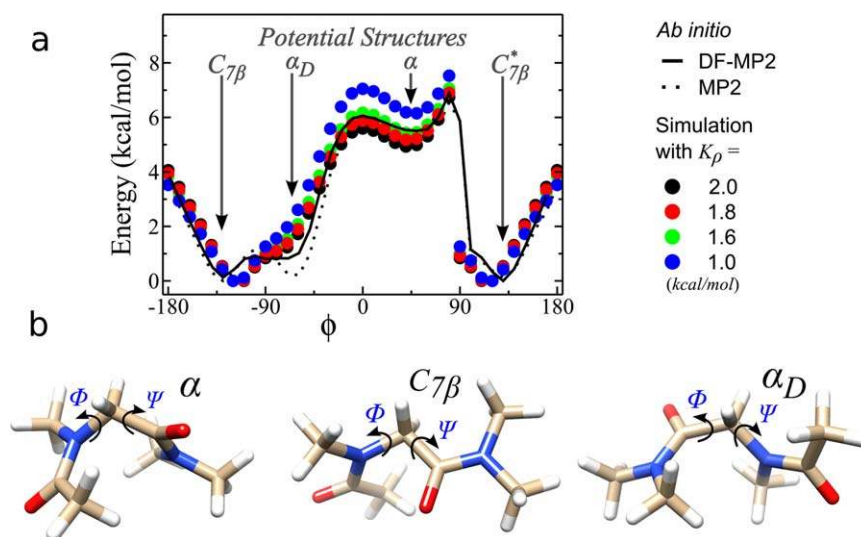**Figure 5.** Obtaining the dihedral angle force constant K_ρ from sarcosine dipeptoid's φ-profile. a) φ-dependent energy profiles (φ-profiles) were obtained from simulations using various values of the ρ dihedral angle force contant K_ρ (shown as series of colored circles). From these profiles, the one obtained from simulations with $K_\rho=1.8$ kcal/mol best matches key features of the target φ-profile calculated from quantum mechanical methods (solid black line is for DF-MP2/aug-cc-pVQZ//MP2/6-31+G* single-point energy calculation; the dashed black line for MP2/6-31+G*//MP2/6-31+G* single-point calculation is shown for reference). Examples of the backbone configurations that describe the three structures $C_{7\beta}$, α_D, and α are shown in b). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
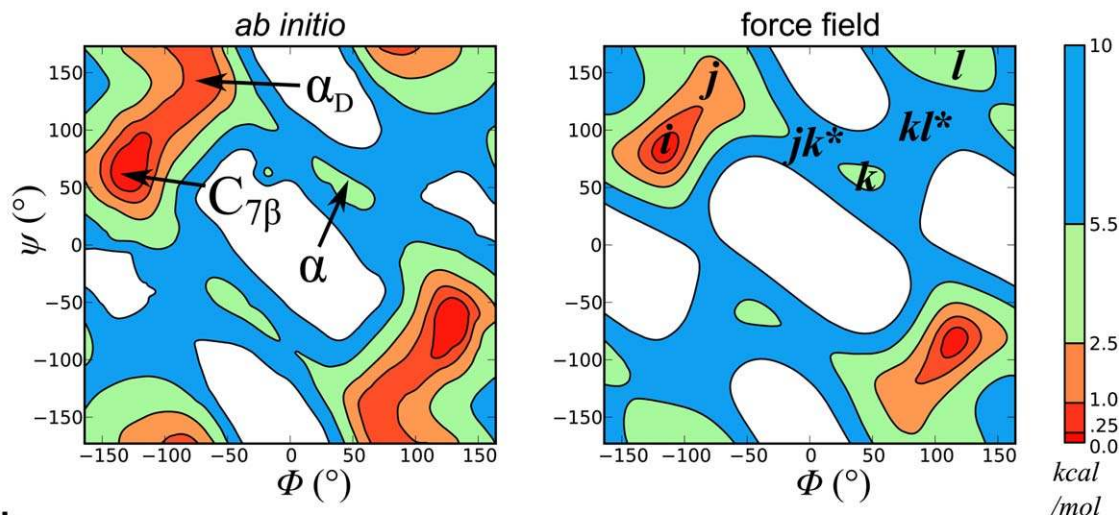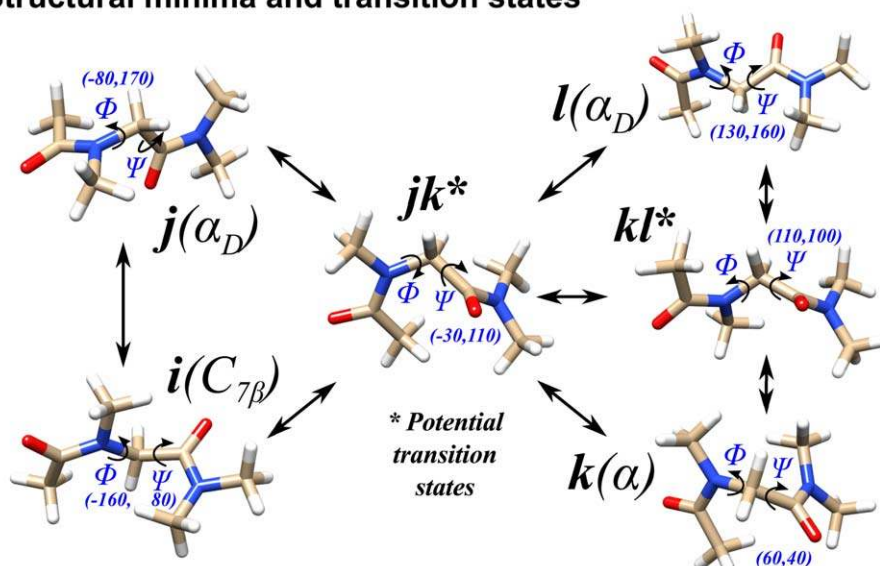
Figure 6. Energy landscapes for a trans sarcosine dipeptoid in vacuum from quantum mechanics and simulations of the forcefield demonstrate the accuracy of the latter. a) Sarcosine dipeptoid energy landscapes obtained from quantum mechanics (left) and forcefield energy terms (right) display similar minima at dihedral angle values describing the $C_{7\beta}$ and α structures,[21] and a low-energy plateau where the α_D conformation would lie. b) Configurations of the backbone describing the structures $C_{7\beta}$, α, and α_D, along with possible transition states in which the ω dihedral angle lies in the cis conformation. Areas in the plots shown in white stand for energies larger than 10 kcal/mol. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and $\omega = 180°$), we calculated energies of the ϕ and Ψ backbone dihedral angles in steps of $20°$ (Fig. 6a). The corresponding energy diagram calculated using our peptoid forcefield (Fig. 6b) is similar to Figure 6a in several key respects, particularly in respect of the placement and depth of the minima describing the $C_{7\beta}$ and α conformations, and the low-lying plateau where the α_D conformation would lie. The favorable comparison of quantum mechanical and forcefield energy landscapes is a first indication that this forcefield describes sarcosine dipeptoid with a reasonable level of accuracy.

Effect of solvation on sarcosine dipeptoid conformations

To determine the influence of water on peptoid conformations, we calculated free energy surfaces of the ϕ and Ψ dihedral

angles for the trans and cis states of sarcosine dipeptoid in vacuum and in water (Fig. 7). The vacuum trans free energy surface is different from the corresponding energy plot (Fig. 6a) in one important respect: the α_D conformation is a local free energy minimum. The vacuum cis free energy profile displays a similar minimum, and shows the α conformation to be globally stable (it is strongly disfavored in the trans state). Solvation changes the molecule's accessible conformations. The trans α configuration becomes slightly more stable. For both cis and trans modes, the $C_{7\beta}$ conformations are disfavored, and constitute free energy plateaux rather than minima. And the cis α_D conformation is changed in location, and made slightly less stable.

We can compare the relative likelihood of cis and trans modes inferred from our simulations with those deduced from
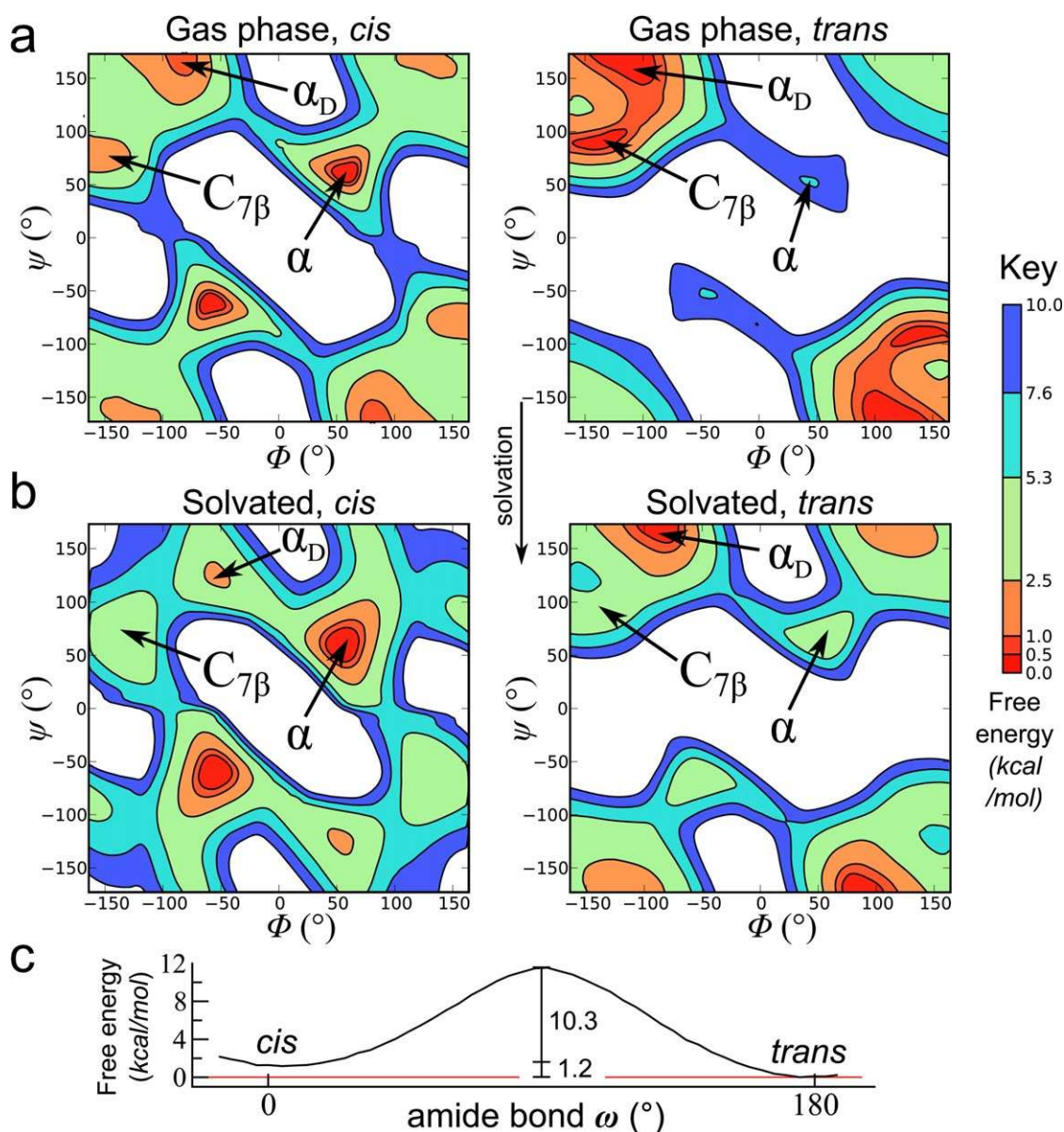
Figure 7. Free energy landscapes obtained from forcefield simulations of sarcosine dipeptoid in a) vacuum and b) water show the considerable effect of solvation on relative stabilities of various backbone configurations. a) Ramachandran-like free energy plots of the gas-phase sarcosine dipeptoid in cis and trans modes. These are different: the $C_{7\beta}$ conformation is most stable in the *trans* mode, while the α conformation is most stable in the *cis* mode. b) Solvation changes the molecule's accessible conformations. The trans α configuration becomes slightly more stable; in both cis and trans modes, the $C_{7\beta}$ conformations are disfavored, and constitute free energy plateaux rather than minima; and the cis α_D conformation is changed in location, and made slightly less stable. Areas shown in white denote free energies larger than 10 kcal/mol. c) A 1D free energy ω– profile of solvated sarcosine dipeptoid shows the trans mode to lie about 1.2 kcal/mol lower than cis. The two states are separated by a barrier of about 10 kcal/mol. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

NMR spectroscopy of peptoid monomers. NMR experiments show that the nine-residue [N-(S)-(1-phenyethyl)glycine]$_9$ peptoid populates its cis and trans modes equally,[53] and show the cis composition of benzyl- and butyl sidechain peptoid monomers to be 37.7 and 28.6%, respectively.[54] Umbrella sampling simulations along the ω dihedral angle coordinate show the cis mode of sarcosine dipeptoid to be higher in free energy than the trans mode by 1.2 kcal/mol (Fig. 7c), from which we infer that a molecule in solution at 298 K should adopt the cis form with a probability of about 12%. This value fits the general experimental trend of reduced cis probability

with decreasing sidechain size, providing additional confidence in the ability of this forcefield to describe peptoids.

Tripeptoid crystal simulation results

To test our forcefield in the condensed phase, we simulated the crystal structure[29] of an N-2-phenylethyl tripeptoid (asymmetric unit shown in Fig. 8a) in known to challenge structure prediction methods that used peptide-based forcefields.[29] We carried out simulations of 27 repeat units of the crystal (one such unit, composed of eight tripeptoids, is shown in the
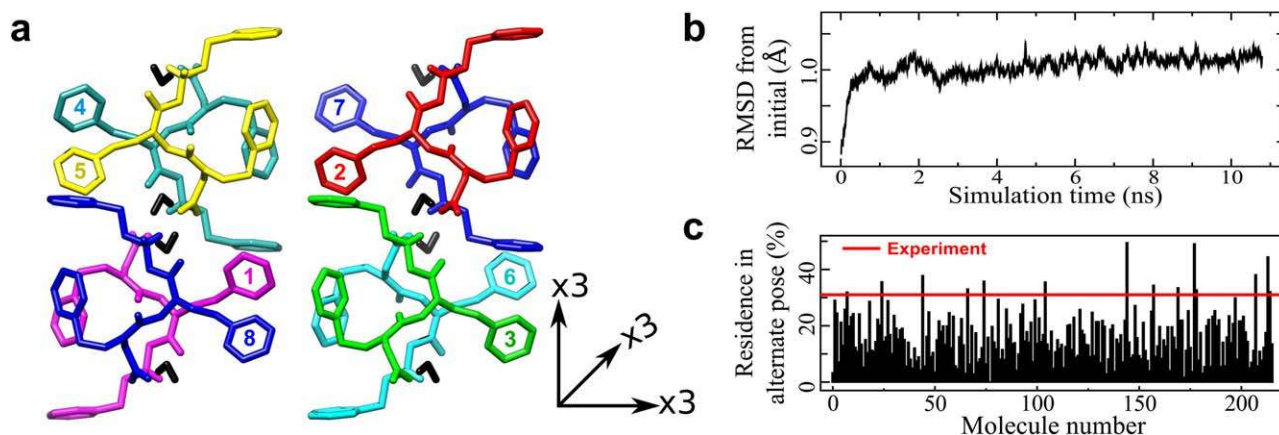
Figure 8. Simulation of crystalline *N*-2-phenylethyl tripeptoid. a) We simulated 27 repeat units of an experimentally obtained[29] crystal structure of *N*-phenylethyl tripeptoid (snapshot depicts one repeat unit). b) At 300 K, structural equilibration occurs within about 5 ns. The crystal space group is preserved. c) Each molecule's medial sidechain takes one of two conformations, called "dominant" and "alternate" (see text). The alternate pose is seen in experiment to occur with a likelihood of about 30% (red line), reasonably close to the likelihood with which individual molecules in our simulations achieve this pose (black bars). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

figure). As shown in Figure 8b, simulations at 300 K begun from the experimental structure equilibrate within about 5 ns. Equilibrium configurations preserve the symmetry of the crystal (space group Pbca), demonstrating that the forcefield reproduces key aspects of peptoid-peptoid interactions.

Further, our simulations show crystal structure dynamics whose time-averaged consequences are consistent with experimental results. The central residue's sidechain of each peptoid in the crystal structure sits either in a "dominant" or an "alternate" pose. For molecules labeled 1, 3, 5, and 7 in each repeat unit of the crystal (see Fig. 8a), the dominant pose corresponds to a middle sidechain angle χ of $90°$, and the alternate pose corresponds to $\chi = -90°$. For molecules 2, 4, 6, and 8 (and their counterparts throughout the crystal), the dominant and alternate poses are instead $-90°$ and $+90°$, respectively. X-ray measurements of the real crystal structure at 100 K indicate that molecules are found on average in their alternate pose with a likelihood of about 30%.[29] In simulations, we observed each of the 216 molecules to flip between dominant and alternate poses. They are resident in each pose, on average, with a likelihood similar to that seen in experiment (Fig. 8c). We verified statistically that no dynamic pairwise molecule-molecule correlations exist. We plan to investigate, for a future report, higher-order dynamic correlations, to determine if middle sidechain flippings occur in a concerted fashion. Sidechain rotation has been postulated to be a potential means of information transfer in proteins.[55] Simulations done at 100 K were too sluggish to show flipping events, and so we resorted to comparing observations made at 100 K with simulations done at 300 K (the temperature at which crystals were grown in experiment).

There also exist substantial intramolecule correlations between rotameric degrees of freedom in the crystal structure. Figure 9a depicts a tripeptoid within the crystal structure in both its dominant and alternate (teal) pose. When the sidechain dihedral angle χ flips from its dominant to its alternate pose, the benzene ring responds by varying its position (via the angle ξ subtly (panel b; we calculated a Pearson correla-

tion coefficient of 0.75 between χ and ξ, and a *p*-value that is indistinguishable from 0). Both angles show bimodal distributions (panel c). Further, the distribution of angles ξ (panel c, red histogram) is markedly different from the corresponding distribution obtained from simulations of a single tripeptoid in implicit hexane solvent (panel c, blue histogram). This difference highlights the importance of simulating molecules in a crystal structure, rather than in solution, to interpret experiments done on a crystal structure.

Conclusions

Peptoid-specific simulation tools are needed to enable researchers to close the gap between current synthetic and structure prediction capabilities, to realize the full potential of these molecules in medicine,[13–16] catalysis,[17] and nanotechnology.[5,6,18] We have introduced a first-generation CHARMM-based atomistic forcefield for peptoids. This forcefield reproduces a range of quantum mechanical and experimental data for small model peptoids with reasonable accuracy. We have used the forcefield to show that the solvation of small model peptoids influences their available conformations, and have used it to simulate a recently obtained tripeptoid crystal structure. The simulation using the forcefield preserves the crystal structure of the tripeptoid, and resolves subtle sidechain dynamics whose time-averaged consequences are consistent with experimental results.

A central concern in the development of a forcefield is of "transferability"[24] or generality: can forcefield parameters obtained from model molecules be used to simulate other molecules within that class? Although no guarantees can be made in this regard, we have used a parameterization scheme that follows as closely as possible the one used to represent the peptide backbone.[30] The successful simulation of thousands of proteins over the last four decades shows the peptide scheme to be modular to a considerable degree. We hope, therefore, that peptoid backbone parameters identified here are similarly modular, leaving only the parameterization of
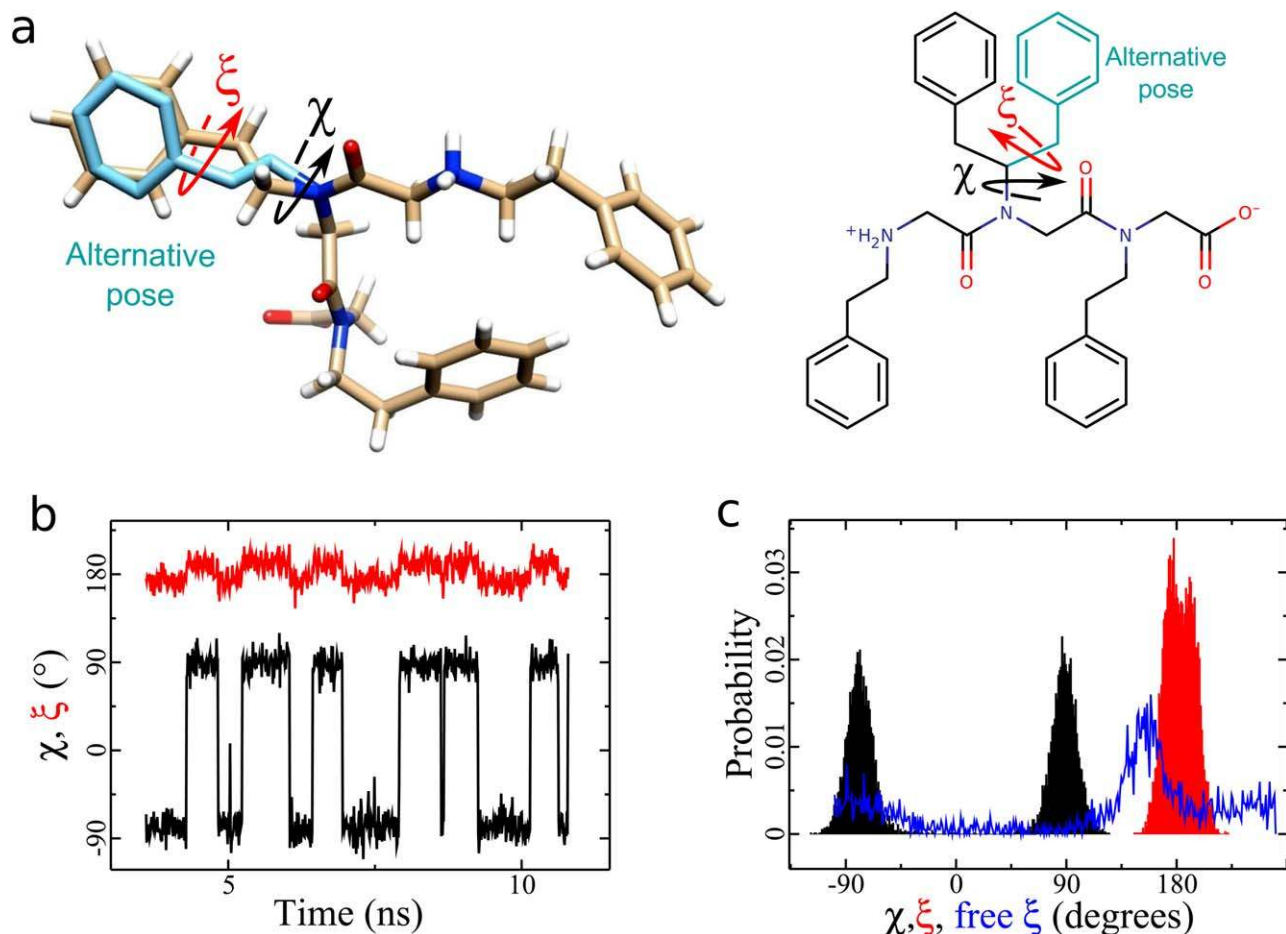
Figure 9. Resolving the flipping of individual sidechains within the crystal and in solution. a) Sketch of a single tripeptide in both dominant and alternate (teal) poses. b) A time series for a single peptoid in our equilibrated crystal (at 300 K) shows correlated flipping of the angles labeled χ and ξ. c) The resulting angular distributions (averaged over all molecules in the simulation box) are bimodal. The probability distribution of ξ for a free peptoid in implicit hexane (blue line) is markedly different from the distribution seen in the crystal structure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

sidechains to be done to simulate a given peptoid molecule. The first-generation forcefield presented here is a resource for the growing community of peptoid researchers, and should be seen as a starting point for the development of future peptoid forcefields. We propose that the present forcefield be called MFTOID (*em-eff-toyd*), after "Molecular Foundry" and "peptoid."

Acknowledgments

Keywords: peptoid · forcefield · parameterization · CHARMM · simulation

[1] A. Rosales, B. McCulloch, R. Zuckermann, R. Segalman, *Macromolecules* **2012**, *45*(15), pp. 6027–6035.

[2] K. Kirshenbaum, A. E. Barron, R. A. Goldsmith, P. Armand, E. K. Bradley, K. T. V. Truong, K. A. Dill, F. E. Cohen, R. N. Zuckermann, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4303.

[3] P. Armand, K. Kirshenbaum, R. A. Goldsmith, S. Farr-Jones, A. E. Barron, K. T. V. Truong, K. A. Dill, D. F. Mierke, F. E. Cohen, R. N. Zuckermann, E. K. Bradley, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4309.

[4] J. A. Crapster, J. R. Stringer, I. A. Guzei, H. E. Blackwell, *Biopolymers* **2011**, *96*, 604.

[5] B. -C. Lee, T. K. Chu, K. A. Dill, R. N. Zuckermann, *J. Am. Chem. Soc.* **2008**, *130*, 8847.

[6] K. T. Nam, S. A. Shelby, P. H. Choi, A. B. Marciel, R. Chen, L. Tan, T. K. Chu, R. A. Mesch, B. -C. Lee, M. D. Connolly, C. Kisielowski, R. N. Zuckermann, *Nat. Mater.* **2010**, *9*, 454.

[7] B. Sanii, R. Kudirka, A. Cho, N. Venkateswaran, G. Olivier, A. Olson, H. Tran, R. Harada, L. Tan, R. Zuckermann, *J. Am. Chem. Soc.* **2011**, *133*(51), pp. 20808–20815.

[8] R. Kudirka, H. Tran, B. Sanii, K. Nam, P. Choi, N. Venkateswaran, R. Chen, S. Whitelam, R. Zuckermann, *Pept. Sci.* **2011**, *96*, 586.

[9] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9367.

[10] R. Zuckermann, *Pept. Sci.* **2011**, *96*, 545.

[11] R. N. Zuckermann, T. Kodadek, *Curr. Opin. Mol. Ther.* **2009a**, *11*, 299.

[12] M. Reddy, R. Wilson, J. Wilson, S. Connell, A. Gocke, L. Hynan, D. German, *T. Kodadek, Cell* **2011**, *144*, 132.

[13] T. Schröder, A. Quintilla, J. Setzler, E. Birtalan, W. Wenzel, S. Bräse, *WSEAS Trans. Biol. Biomed.* **2007**, *4*, 145.

[14] N. P. Chongsiriwatana, J. A. Patch, A. M. Czyzewski, M. T. Dohm, A. Ivankin, D. Gidalevitz, R. N. Zuckermann, A. E. Barron, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2794. Available at: http://dx.doi.org/10.1073/pnas.0708254105.

[15] R. N. Zuckermann, T. Kodadek, *Curr. Opin. Mol. Ther.* **2009b**, *11*, 299.

[16] X. Chen, J. Wu, Y. Luo, X. Liang, C. Supnet, M. W. Kim, G. P. Lotz, G. Yang, P. J. Muchowski, T. Kodadek, I. Bezprozvanny, *Chem. Biol.* **2011a**, *18*, 1113.

[17] G. Maayan, M. D. Ward, K. Kirshenbaum, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13679. Available at: http://dx.doi.org/10.1073/pnas.0903 187106.

[18] C. -L. Chen, J. Qi, R. N. Zuckermann, J. J. DeYoreo, *J. Am. Chem. Soc.* **2011b**, *133*, 5214.

[19] K. Moehle, H. -J. Höfmann, *Biopolymers* **1996**, *38*, 781.

[20] K. Möhle, H. -J. Höfmann, *J. Mol. Model.* **1996**, *2*, 307.

[21] G. L. Butterfoss, P. D. Renfrew, B. Kuhlman, K. Kirshenbaum, R. Bonneau, *J. Am. Chem. Soc.* **2009**, *131*, 16798.

[22] M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids; Oxford University Press: Oxford, **1989**.

[23] J. M. Haile, Molecular Dynamics Simulation: Elementary Methods; Wiley-Interscience, Chichester, West Sussex, **1997**.

[24] A. Leach, Molecular Modeling: Principles and Applications, 2nd ed.; Prentice Hall, Upper Saddle River, New Jersey, **2001**.

[25] D. Frenkel, B. Smit, Understanding Molecular Simulation: From Algorithms to Applications, 2nd ed.; Academic Press, Waltham, Massachusetts, **2001**.

[26] A. Hinchliffe, Molecular Modelling for Beginners, 2nd ed.; Wiley, Chichester, West Sussex, **2002**.

[27] C. J. Cramer, Essentials of Computational Chemistry: Theories and Models, 2nd ed.; Wiley, Chichester, West Sussex, **2004**.

[28] S. H. Park, I. Szleifer, *J. Phys. Chem. B* **2011**, *115*, 10967. Available at: http://dx.doi.org/10.1021/jp2025957.

[29] G. Butterfoss, B. Yoo, J. Jaworski, I. Chorny, K. Dill, R. Zuckermann, R. Bonneau, K. Kirshenbaum, V. Voelz, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14320.

[30] A. D. Mackerell, Jr., M. Feig, C. L. Brooks, III, *J. Comput. Chem.* **2004**, *25*, 1400.

[31] A. MacKerell, Jr., N. Banavali, N. Foloppe, *Biopolymers* **2000**, *56*, 257. Available at: http://dx.doi.org/3.0.CO;2-W.

[32] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, 2nd, A. D. Mackerell, Jr., *J. Chem. Theory Comput.* **2011**, *7*, 3162. Available at: http://dx.doi.org/10.1021/ct200328p.

[33] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., *J. Comput. Chem.* **2010**, *31*, 671. Available at: http://dx.doi.org/10.1002/jcc.21367.

[34] R. Best, X. Zhu, X. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, Jr., *J. Chem. Theory Comput.* **2012**, *8*, 3257.

[35] M. Feig, *J. Chem. Theory Comput.* **2008**, *4*, 1555.

[36] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. III Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera J, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.

[37] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karpulus, *J. Comput. Chem.* **1983**, *4*, 187.

[38] T. Darden, D. York, L. Pederson, *J. Chem. Phys.* **1993**, *98*, 10089.

[39] M. Shirts, D. Mobley, J. Chodera, V. Pande, *J. Phys. Chem. B* **2007**, *111*, 13052.

[40] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

[41] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.

[42] A. Grossfield, WHAM: The Weighted Histogram Analysis Method, Version 2.0.6, Available at: http://membrane.urmc.rochester.edu/content/wham.

[43] W. Im, M. S. Lee, C. L. Brooks, III, *J. Comput. Chem.* **2003**, *24*, 1691.

[44] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Jr Distasio, R. C. Lochan, T. Wang, G. J. O. Beran GJO, N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. an Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, Y. M. Rhee, J. Ritchie, E. Rosta, Sherrill C. D., A. C. Simmonett, J. E. Subotnik, H. L. III Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. III Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.

[45] H. -J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Shao, Y. Rauhut, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Jr Distasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. an Voorhis, SH Chien, A Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A .D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. III Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. III Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schutz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Koppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, K. Pfluger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, A. Wolf, Molpro, Version 2010.1, A Package of Ab Initio Programs, **2010**. Available at: http://www.molpro.net.

[46] H. -J. Werner, F. R. Manby, P. J. Knowles, *J. Chem. Phys.* **2003**, *118*, 8149.

[47] F. P. Gasparro, N. H. Kolodny, *J. Chem. Educ.* **1977**, *54*, 258.

[48] W. Zielenkiewicz, J. Poznański, *J. Mol. Liq.* **1999**, *81*, 37.

[49] W. Xie, P. Jingzhi, A. MacKerrel, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1878.

[50] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Jr. Mackerell, *J. Comput. Chem.* **2009**, *31*, 671.

[51] W. Xie, J. Pu, A. D. Mackerell, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1878.

[52] L. F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **2007**, *442*, 42.

[53] K. Huang, C. Wu, T. Sanborn, J. Patch, K. Kirshenbaum, R. Zuckermann, A. Barron, I. Radhakrishnan, *J. Am. Chem. Soc.* **2006**, *128*, 1733.

[54] Q. Sui, D. Borchardt, D. L. Rabenstein, *J. Am. Chem. Soc.* **2007**, *129*, 12042. Available at: http://dx.doi.org/10.1021/ja0740925.

[55] K. DuBay, J. Bothma, P. Geissler, *PLoS Comput. Biol.* **2011**, *7*, e1002168.