

A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics

Cite as: J. Chem. Phys. **153**, 064101 (2020); <https://doi.org/10.1063/5.0019054>

Submitted: 19 June 2020 . Accepted: 19 July 2020 . Published Online: 10 August 2020

Anthony T. Bogetti , Hannah E. Piston , Jeremy M. G. Leung , Chino C. Cabaltea , Darian T. Yang , Alex J. DeGrave , Karl T. Debiec , David S. Cerutti , David A. Case , W. Seth Horne , and Lillian T. Chong 

COLLECTIONS

Paper published as part of the special topic on [Classical Molecular Dynamics \(MD\) Simulations: Codes, Algorithms, Force fields, and Applications](#)

Note: This paper is part of the JCP Special Topic on Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force Fields, and Applications.



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations](#)

The Journal of Chemical Physics **153**, 054123 (2020); <https://doi.org/10.1063/5.0013849>

[Toward empirical force fields that match experimental observables](#)

The Journal of Chemical Physics **152**, 230902 (2020); <https://doi.org/10.1063/5.0011346>

[Scalable molecular dynamics on CPU and GPU architectures with NAMD](#)

The Journal of Chemical Physics **153**, 044130 (2020); <https://doi.org/10.1063/5.0014475>

Lock-in Amplifiers
up to 600 MHz



A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics

Cite as: J. Chem. Phys. 153, 064101 (2020); doi: 10.1063/5.0019054

Submitted: 19 June 2020 • Accepted: 19 July 2020 •

Published Online: 10 August 2020



View Online



Export Citation



CrossMark

Anthony T. Bogetti,¹ Hannah E. Piston,¹ Jeremy M. C. Leung,¹ Chino C. Cabalteja,^{1,2}
Darlan T. Yang,² Alex J. DeGrave,³ Karl T. Debiec,^{1,2} David S. Cerutti,⁴ David A. Case,⁴
W. Seth Horne,^{1,a)} and Lillian T. Chong^{1,a)}

AFFILIATIONS

¹Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

²Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, Pennsylvania 15260, USA

³School of Computer Science and Engineering, University of Washington, Seattle, Washington 98115, USA

⁴Department of Chemistry and Chemical Biology, Rutgers University, New Brunswick, New Jersey 008854, USA

Note: This paper is part of the JCP Special Topic on Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force Fields, and Applications.

^{a)}Authors to whom correspondence should be addressed: horne@pitt.edu and ltchong@pitt.edu

ABSTRACT

We present a new force field, AMBER ff15ipq-m, for simulations of protein mimetics in applications from therapeutics to biomaterials. This force field is an expansion of the AMBER ff15ipq force field that was developed for canonical proteins and enables the modeling of four classes of artificial backbone units that are commonly used alongside natural α residues in blended or “heterogeneous” backbones: chirality-reversed D - α -residues, the C_{α} -methylated α -residue Aib, homologated β -residues (β^3) bearing proteinogenic side chains, and two cyclic β residues (β^{cy} ; APC and ACPC). The ff15ipq-m force field includes 472 unique atomic charges and 148 unique torsion terms. Consistent with the AMBER IPolQ lineage of force fields, the charges were derived using the Implicitly Polarized Charge (IPolQ) scheme in the presence of explicit solvent. To our knowledge, no general force field reported to date models the combination of artificial building blocks examined here. In addition, we have derived Karplus coefficients for the calculation of backbone amide J -coupling constants for β^3 Ala and ACPC β residues. The AMBER ff15ipq-m force field reproduces experimentally observed J -coupling constants in simple tetrapeptides and maintains the expected conformational propensities in reported structures of proteins/peptides containing the artificial building blocks of interest—all on the μ s timescale. These encouraging results demonstrate the power and robustness of the IPolQ lineage of force fields in modeling the structure and dynamics of natural proteins as well as mimetics with protein-inspired artificial backbones in atomic detail.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0019054>

I. INTRODUCTION

The extraordinary characteristics of proteins originate in the diversity of amino acid sequences found in nature specifying a wide array of precise folded shapes. While proteins in nature are constructed primarily from a palette of 20 α -amino acids, chemists have long sought to demonstrate what is possible by expanding beyond this set, with respect to both side chain identity and backbone composition. Many oligomers with artificial amide-based backbones are

capable of discrete folding behavior,¹ and these entities, termed foldamers,² have found widespread applications.³ After the inception of the field, work on folding in artificial backbones focused primarily on isolated secondary structures (e.g., helices) and, later, their assemblies (e.g., helix bundles).^{4,5} More recently, attention has turned to developing artificial backbones able to recreate more sophisticated tertiary folding patterns that are typical of proteins.⁶ Advancing from secondary structure mimicry (peptidomimetics) to tertiary structure mimicry (proteomimetics) is important,

as a tertiary structure is essential to function for most natural proteins. As the complexity of folds targeted for mimicry by artificial backbones has increased, so has the challenge of design and elucidating folding behavior in the resulting scaffolds. Computational tools have proven invaluable in shedding light on such issues in the context of natural proteins; however, the application of such methods to artificial backbones has been limited.

Some prior efforts have adapted computational tools to treat various non-canonical building blocks for structure prediction, enabling the design of oligo-oxopiperazines, oligo-peptoids, β -peptides, hydrogen bond surrogate helices, and oligosaccharides.⁷ Other work has sought to broaden the scope of backbone compositions amenable to the atomistic MD simulation in the form of CHARMM force fields for peptoids⁸ and for *N*- and α -methylated amino acids.⁹ Among experimental strategies for creation of artificial backbone proteomimetics with predictable and complex folding patterns, a particularly useful approach that has emerged is the blending of many artificial building block types alongside natural α residues in a single chain.¹⁰ When such “heterogeneous backbone” oligomers display a side chain sequence derived from a natural protein, the resulting agents can manifest a range of interesting folds and functions.^{11–18} Biophysical analysis of folding behavior in some of these tertiary structure mimetics has revealed interesting thermodynamic effects accompanying backbone alteration.^{14,19–21} MD simulation would be an invaluable tool to shed light on issues related to structure and dynamics in these and related systems; however, this requires a force field that is able to treat a range of artificial residue classes in a general manner that is internally consistent among all the residues involved, both natural and artificial.

Here, we report the development and validation of a general force field, AMBER ff15ipq-m, for the simulation of heterogeneous backbone protein mimetics. The force field is an expansion of the AMBER ff15ipq force field for canonical proteins,²² which features implicitly polarized atomic charges in the presence of explicit solvent, i.e., the best that fixed-charge force fields can offer toward modeling condensed-phase electrostatics in the absence of explicit polarization (e.g., Drude oscillators²³ and inducible multipoles²⁴). In ff15ipq-m, the original force field is expanded to include four new classes of artificial backbone units commonly used in protein mimetics (see Fig. 1): *D*- α -, C_{α} -methylated α - or aminoisobutyric acid (Aib), β^3 , and the cyclic β - (β^{cyc}) residues, *trans*-aminopyrrolidine

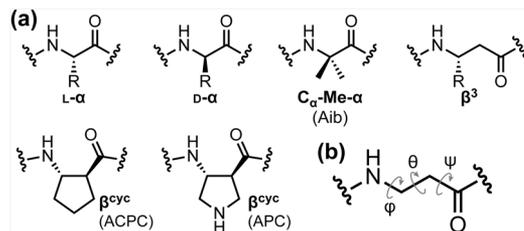


FIG. 1. (a) Chemical structures of an *L*- α residue alongside four classes of artificial backbone units that can be modeled using the AMBER ff15ipq-m force field. In the case of β^3 and *D*- α residues, the identity of the R group matches that of the corresponding *L*- α residue (i.e., R=CH₃ for Ala, *D*-Ala, and β^3 Ala). (b) Definition of backbone torsion angles ϕ , θ , and ψ for the β -residue classes.

carboxylic acid (APC) and *trans*-2-aminocyclopentane-1-carboxylic acid (ACPC). Given that the *D*- α -residues required no new parameterization, the expansion includes parameters for a total of 25 backbone units, consisting of 22 β^3 residues (side chains corresponding to the 20 canonical amino acids including, three protonation states for His), Aib, APC, and ACPC. As shown in Fig. 1, β residues exhibit greater flexibility relative to α -residue counterparts due to the additional carbon in the backbone, which introduces a third backbone torsion θ in addition to the usual ϕ and ψ backbone torsions. With the aid of the GPU-accelerated AMBER molecular dynamics (MD) engine,^{25–28} we have extensively validated the force field by simulating both peptides and proteins that include each backbone modification type—all on the μ s timescale, yielding 30 μ s of aggregate simulation time. Collectively, these results provide a useful new computational model for enabling the simulation of protein-inspired artificial backbones at the atomic level and describe a general workflow for expanding the force field further to encompass additional monomer classes in future work.

II. THEORY

A. The AMBER ff15ipq force field for canonical proteins

The AMBER ff15ipq force field is the latest in the lineage of Implicitly Polarized Charge (ipq) protein force fields that are based almost entirely on quantum mechanical data.^{22,29} A hallmark of ipq force fields is that the atomic charges are implicitly polarized, meaning that the charges derived are intended to reproduce the mean field electron density of a molecule in explicit solvent.³⁰ For ff15ipq, the explicit SPC/E_b (extended simple point charge) water solvent model³¹ was chosen to take advantage of (i) the model's accurate parameterization of rotational protein diffusion in solution and (ii) the relatively low cost of a three-point model compared to other four-point alternatives.

The ff15ipq force field exhibits a number of features which, when taken together, distinguish it from other contemporary force fields.²² First, the force field yields reasonable propensities for salt-bridge formation, as it was originally developed to address the over-stabilization of salt bridges, a common limitation of many force fields.^{22,32} In particular, newly derived atomic charges at the MP2/cc-pVTZ level of theory and atomic radii for polar hydrogens reproduce experimental probabilities of salt-bridge formation. Second, ff15ipq is far from a limited adjustment of its predecessor, ff14ipq,²⁹ but rather a complete rederivation that yields the expected balance of secondary structures for both the globular and non-globular (“disordered”) peptides/proteins. The force field not only includes new atomic charges and radii but also a greatly expanded torsion parameter set and new angle parameters. Finally, as a benefit of its parameterization with SPC/E_b explicit water, ff15ipq yields reasonable NMR observables, including *J*-coupling constants for the Ala₅ peptide that are comparable to those of force fields that were specifically parameterized to reproduce such experimental values (e.g., ff03w³³ and ff14SB³⁴). Moreover, a recent modification of ff15ipq methylene side chain torsions has yielded accurate NMR relaxation involving methyl groups.³⁵ The ff15ipq force field has been implemented in dynamics engines that take advantage of GPU acceleration, e.g., AMBER²⁸ and OpenMM.³⁶

B. The AMBER ff15ipq-m force field for protein mimetics

New parameters for each of the 25 artificial residues in the AMBER ff15ipq-m force field were derived using a workflow that was as consistent as possible with that of the parent ff15ipq protein force field.²² The same fitting classes and convergence criteria were used for deriving atomic charges and torsion parameters. The fitting classes consisted of the following: neutral, negatively charged, positively charged, glycine, and proline. The neutral class was divided into four subclasses based on the type of the first carbon of the side chain (the C_γ atom in β^3 residues): alanine, two-branch residues whose C_γ atoms are attached to two heavy atoms, three-branch residues whose C_γ atoms are attached to three heavy atoms, and aromatic residues. Atomic charges for each residue were derived in the same context, i.e., within a blocked dipeptide containing acetyl (Ace) and N-methyl (NMe) caps, in the presence of explicit SPC/ E_b water molecules. One minor difference from the original ff15ipq workflow is in the generation of dipeptide conformations for the derivation of both new charges and torsion parameters. In particular, the diversity of the conformations was improved by progressively restraining the backbone torsions at evenly spaced intervals and energy-minimizing the conformation rather than using high-temperature simulations.

A total of 12 new atom types were introduced for the ff15ipq-m force field (see Fig. S1). For the β^3 residues, new atom types were assigned to the additional backbone methylene carbon between the carbonyl carbon and the carbon bearing the side chain. For the C_α -methylated α -residue Aib, a new atom type was introduced for the C_α atom. For the β^{cyc} residue APC, a new atom type was assigned to the ring nitrogen of its side chain since this form of sp^3 -hybridized nitrogen is not present in the side chains of other residues. For the other β^{cyc} residue, ACPC, it was not necessary to introduce any new atom types for its side chain since the ring carbons are sufficiently similar to those present in proline.

III. METHODS

A. Derivation of IPolQ atomic charges

For each artificial residue, IPolQ atomic charges were derived by (i) generating a set of conformations for a blocked dipeptide containing the artificial residue with Ace and NMe caps; (ii) calculating the electrostatic potential for each conformation at the MP2/cc-pVTZ level of theory in both vacuum and explicit solvent; (iii) using the average electrostatic potential for all conformations to fit two sets of atomic point charges, one in vacuum and one in solvent; (iv) averaging the vacuum-phase and solvent-phase point charges to obtain an “implicitly polarized” set of charges; and (v) repeating steps (i) through (iv) until a desired convergence of the calculated charges was achieved.

To generate a set of conformations for each blocked dipeptide (20 for Aib; 30 for all other residues), ϕ , ψ , and θ backbone torsions were progressively restrained in vacuum between -180° and 180° using harmonic restraints with a force constant of 32 kcal/mol and subjected to energy minimization. Each conformation was then solvated in a truncated octahedral box of SPC/ E_b water molecules with a 12 Å buffer and subjected to energy minimization while

applying harmonic restraints to the solute (the dipeptide). To equilibrate the solvent, the solvated system was subjected to 100 ps of dynamics at constant temperature (25°C) and pressure (1 atm) while applying harmonic restraints to the solute. A further 500 ps equilibration of the solvent was carried out during which the coordinates of surrounding solvent molecules were collected to generate a collection of point charges representing the solvent reaction field potential. This collection consists of an inner cloud of point charges based on the coordinates of the solvent molecules within 5 Å of the solute and three outer shells of point charges fitted to reproduce longer-range contributions to the solvent reaction field potentials.

Next, electrostatic potentials were calculated for each peptide conformation in two environments: a vacuum-phase environment and a solvent-phase environment, which includes the solvent reaction field potential as modeled by the collection of point charges representing the surrounding explicit solvent molecules. The calculations were carried out at the MP2/cc-pVTZ level of theory using the ORCA 4.1.2 software package.^{37,38}

The resulting electrostatic potentials in the vacuum and solvent phases were used to fit two corresponding sets of nuclear-centered atomic partial charges. For each artificial residue, the fitting procedure utilized the MP2/cc-pVTZ electron densities for all 30 dipeptide conformations and was carried out using the FitQ fitting procedure in the mdgx program.²⁵ Consistent with the parent ff15ipq force field, negatively charged, positively charged, and neutral residues were fit as separate groups. To account for differences in molecular charge densities due to the addition of a backbone ring, β^3 Pro and ACPC were fit separately from the other neutral residues, and APC was fit separately from the other positively charged residues. In addition, the C_α -methylated α -residue Aib was fit separately. Atomic charges of the Ace and NMe capping groups for each blocked dipeptide were fixed to their net neutral values in the ff15ipq force field.

In the final step of the IPolQ method of charge derivation, the vacuum-phase and solvent-phase charges were averaged to obtain an “implicitly polarized” set of point charges. The resulting set of charges were used as a starting point for further iterations of the above procedure, whereby another 30 conformations were generated using charges from the previous iteration, and an improved set of implicitly polarized atomic point charges were generated. This procedure was repeated until the charges remained within a 10% deviation of those from the previous iteration.

B. Generation of the torsion fitting dataset and fitting of torsion terms

For each artificial residue, backbone torsion parameters were derived by (i) obtaining initial trial parameters from an existing force field for each angle and torsion involving a new atom type; (ii) generating a set of 1000 conformations for the artificial residue in the context of a blocked dipeptide with Ace and NMe caps in vacuum; (iii) calculating the QM energy of each conformation at the MP2/cc-pVTZ level of theory alongside the MM energy using the current force field's torsion parameters; (iv) adjusting the backbone torsion parameters of the residue in order to minimize the error between the QM and MM energies; and (v) repeating steps (i)–(iv) until the error between the QM and MM energies does not change.

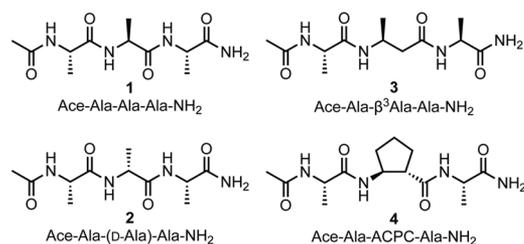


FIG. 2. Structures of peptides 1–4.

Initial angle and backbone torsion parameters were preferably taken from the parent ff15ipq force field if a direct analog existed. In the event that an angle or torsion term involving a new atom type did not have an exact analog in the ff15ipq force field, starting parameters were taken from the most similar parameters in the GAFF2 force field.³⁹ Angle terms were not fit in our force field and in most cases retained their fitted values from either ff15ipq or GAFF2.

An array of conformations of each blocked dipeptide were generated in vacuum by progressively restraining the ϕ , ψ , and θ angles at 36 evenly spaced intervals between -180° and 180° with a force constant of 32 kcal/mol and then energy minimized in the absence of restraints. In this way, the trial parameters allowed the dipeptide to extensively explore its conformational space. All conformations were inspected using ϕ/ψ , ϕ/θ , and θ/ψ plots to ensure that the conformations did not sample regions of configurational space that were too unfavorable and that the conformations did not all fall into only a few minima. Across all iterations of torsion fitting >125 000 total conformations were generated.

Next, the QM energy of each conformation was calculated at the MP2/cc-pVTZ level of theory to match the original ff15ipq protocol and subsequently fit to a function that describes the molecules MM energy as a function of torsions around the three backbone angles. The calculated QM energies were screened to exclude conformations whose energies were >10% different from the average in order to ensure a better fit to MM energies. The QM energies of all conformations were then used in a fitting procedure to generate torsion terms that gave MM energies as close to the QM energies as possible. The addition of new atom types for the residues, as previously discussed, has led to the torsion classes and subclasses by decoupling

the torsion angles. Backbone torsion ϕ , θ , and ψ classes were separated for the following groupings of residues: neutrals, negatively charged, positively charged, glycine, proline, ACPC, APC, and Aib. The ϕ' and θ' torsion angles of the neutral residues were further separated into groups to account for differences in the first carbon in the side chain (see Table S3). The new torsion terms were then used to begin a new iteration of conformation generation. The conformation generation and fitting procedure were repeated until the root-mean-square error (RMSE) between the QM and MM energies deviated by <1% from the previous iteration.

C. Preparation of validation systems for simulation

Starting coordinates for simulations of peptides 1–4 (see Fig. 2) were built from scratch using Avogadro.^{40,41} Starting coordinates for protein systems were obtained from x-ray crystal structures previously deposited in the RCSB Protein Data Bank (PDB; accession codes in Table I).^{19,20,42–45} All systems were solvated in truncated octahedral boxes of explicit SPC/E_b water molecules with a clearance of at least 16 Å for the peptides and 12 Å for the proteins. Heavy-atom coordinates for ubiquitin, GB1, and their variants were taken from the corresponding experimental structures (see Table I). Any missing atoms were built in using the SCAP program,^{46–48} and net system charges were neutralized by adding Na⁺ or Cl[−] ions, which were treated with Joung/Cheatham ion parameters intended for the SPC/E water model.⁴⁹ Hydrogens were added according to ionizable states present at pH 5.6 for GB1 and its two variants; pH 2 for peptides 1–4; and pH 7 for ubiquitin and its D-Gln35 mutant.

D. Unrestrained simulations of validation peptides and proteins

All simulations were carried out using the GPU-accelerated pmemd module of the AMBER 18 software package.^{25–28} After preparing the system as described above and subjecting the system to energy minimization, the solvent was equilibrated in two stages, with harmonic restraints implemented on the proteins. In the first stage, a 20 ps equilibration was carried out at 25 °C at constant volume. In the second stage, a 1 ns equilibration was carried out at a constant temperature of 25 °C and a constant pressure of 1 atm. Finally, unrestrained 1 ns simulations were carried out at 25 °C and 1 atm for the peptides and proteins in Table I before data were

TABLE I. Peptide and protein systems used for force field validation.

System	Sequence/PDB code	No. residues	No. simulations	μ s per simulation
Peptide 1	Ace-Ala-Ala-Ala-NH ₂	3	10	0.5
Peptide 2	Ace-Ala-(D-Ala)-Ala-NH ₂	3	10	0.5
Peptide 3	Ace-Ala-β ³ Ala-Ala-NH ₂	3	10	0.5
Peptide 4	Ace-Ala-ACPC-Ala-NH ₂	3	10	0.5
Wild-type ubiquitin	1UBQ	76	1	2
Ubiquitin variant	1YJ1 with D-Gln35	76	1	2
Wild-type GB1	1PGB	57	1	2
GB1 variant A	5HI1 (first chain) with Aib24, β ³ Lys28, β ³ Lys31, and Aib35	57	1	2
GB1 variant B	4OZB (first chain) with ACPC24, β ³ Lys28, β ³ Lys31, and ACPC35	57	1	2

collected for 0.5 μ s and 2 μ s, respectively. Consistent with NMR experiments performed with peptides 1–4 (see Sec. III F), each peptide was capped with Ace and NH₂. For each of the peptides, five 500 ns simulations were started from an extended conformation and five 500 ns simulations were started from an α -helical conformation, enforcing the total number of atoms in each simulation to be the same and yielding an aggregate simulation time of 5 μ s per peptide. Both sets of five simulations result in similar distributions of φ backbone torsion angles (see Fig. S2), demonstrating convergence of the simulations. Temperatures were maintained using the Langevin thermostat with a frictional constant of 1 ps⁻¹, while pressure was maintained using the Monte Carlo barostat⁵⁰ with 100 fs between attempts to adjust the system volume. van der Waals and short-range electrostatic interactions were truncated at 10 Å, while long-range electrostatic interactions were calculated using the particle mesh Ewald method.⁵¹ A 2 fs time step was enabled by constraining bonds to hydrogen to their equilibrium values using the SHAKE algorithm.⁵² Coordinates were saved every 100 ps for analysis.

E. Umbrella sampling of tetrapeptides

To generate Ramachandran plots of backbone conformations favored for the various artificial residues (φ , θ , and ψ torsions), umbrella sampling was carried out for a series of tetrapeptides Ace-Ala-Xaa-Ala-NMe (Xaa = each of 25 new monomers in the force field). Next, the weighted histogram analysis method (WHAM)⁵³ was applied to the umbrella sampling simulations to yield unbiased free energy profiles as a function of the torsion angles of interest for the central residue. To avoid the ring puckers of the APC and ACPC β^{yc} residues becoming kinetically trapped in minor conformations, the θ torsion angles of these residues were set to 70° in the starting models of the corresponding tetrapeptides. Prior to umbrella sampling, each tetrapeptide was solvated and equilibrated as described above for unrestrained simulations with the only difference being that the duration of the second stage of the equilibration was 100 ps instead of 1 ns. For each tetrapeptide, umbrella sampling involved 1296 windows for each pair of torsions, and each window corresponded to the application of a harmonic penalty function with a force constant of 8 kcal/mol rad² to restrain the pair of torsions to angles at 10° intervals. For C _{α} -methylated tetrapeptides, the φ and ψ torsions of the Aib residue were restrained; for β^3 residues and β^{yc} residues (APC and ACPC), the φ , θ , and ψ torsions were restrained. Each window was subjected to a 0.2 ns incrementally restrained equilibration prior to a 1 ns restrained simulation at a constant temperature (25 °C) and pressure (1 atm). Coordinates of the tetrapeptide were saved every 0.05 ps. For the collective set of 25 tetrapeptides, umbrella sampling involved an aggregate simulation time of 208 μ s. All umbrella sampling simulations were carried out using the GPU-accelerated pmemd module of the AMBER 18 software package.

F. Calculation of NMR J -coupling constants

Backbone amide J -coupling constants for peptides 1–4 were calculated from our simulations using the following Karplus relation:⁵⁴

$$J(\varphi) = A \cos^2(\varphi + \Delta) + B \cos(\varphi + \Delta) + C.$$

$^3J_{\text{HN-H}\alpha}$ values were calculated for α residues and $^3J_{\text{HN-H}\beta}$ for β^3 residues to probe the φ backbone torsion angles and compare to NMR measurements (see Sec. III A).^{55–57} We calibrated separate sets of Karplus coefficients for β^3 Ala, ACPC, and Ala based on quantum mechanical calculations using the H–N–C–H torsion that is directly measured and setting Δ to 0°. For each residue, these calculations were based on a set of 39 dipeptide conformations containing the residue of interest. Consistent with the material used in the NMR experiments (see Sec. III A), the dipeptides were blocked with Ace and NH₂ caps (Ace-Xaa-NH₂). Conformations were generated by first restraining the φ backbone torsion to angles at 30° intervals between –180° and 180° and then at 3° intervals between –180° and –90° while subjecting the dipeptide to energy minimization with the AMBER ff15ipq-m force field. Quantum mechanical (DFT) calculations were carried out in Gaussian 09revD1⁵⁸ using the OLYP functional⁵⁹ and the pcJ-1 (triple- ζ) basis set optimized for spin–spin coupling calculations.⁶⁰ Test calculations showed that going to the larger pcJ-2 basis set changed the results by <0.2 Hz, as did changing to the hybrid functions such as PBE0 or B3LYP. The Karplus coefficients were determined by a least-squares fit to the J -coupling values computed using DFT (see Fig. S3). As noted below (see Table II and Fig. S4), results for the Ala residue in this new parameterization (which we call DFT-3) are very close to earlier fits (DFT-1 and DFT-2)⁶¹ despite differences in the density functional, basis set, and method of calculation. We have too few examples here to draw any conclusions from these small differences, especially given the experimental uncertainties in couplings of ~0.4 Hz.

G. Backbone conformations observed for Aib and β residues in the PDB

For each residue of interest, a ligand structure search was performed to determine if the monomer was present in a PDB deposited structure and, if so, the corresponding three-letter identifier. All monomers, except the β^3 analogs of His and Ser, were present in at least one structure. The resulting list of identifiers (AIB, B3A, BCX, B3D, B3E, 3FB, BAL, BIL, B3K, B3L, B3M, B3X, EOE, B3Q, HMR, B3T, 1VR, HT7, B3Y, XCP, and XPC) was used to construct a query for structures containing any of these monomers as part of a polymer chain (i.e., not a free ligand). The dataset was filtered for structures that were determined by x ray and refined to a resolution of ≤ 2.5 Å. The resulting dataset consisted of 107 structures of backbone modified proteins. Backbone φ , θ , and ψ torsion angles were then calculated for each artificial residue, omitting the φ and ψ torsion angles for C-terminal and N-terminal residues, respectively.

H. Peptide synthesis and purification

Peptides 1–4 (see Table I and Fig. 2) were synthesized via the microwave-assisted Fmoc solid-phase peptide synthesis (CEM MARS microwave reactor) on Sieber amide resin (0.05 mmol scale). Protected amino acids were activated *in situ* by combining *N*- α -Fmoc-protected amino acid (4 equiv), HCTU (4 equiv), and DIEA (6 equiv) in *N*-methyl-2-pyrrolidone (final concentration 0.1M amino acid). Couplings were performed at 70 °C for 4 min. Deprotections of the Fmoc group were performed at 80 °C for 2 min using 20% v/v 4-methylpiperidine in DMF. Each peptide was capped at the *N*-terminus with an acetyl group. DMF/DIEA/acetic anhydride (8:2:1 by volume; 2 ml) was added to the resin and allowed to stir at

TABLE II. Backbone amide J -coupling constants for peptides 1–4 in Hz from the simulation and experiment. J -coupling constants reported are $^3J_{\text{HN-H}\alpha}$ for α residues and $^3J_{\text{HN-H}\beta}$ for β^3 residues. Results from the simulation are shown for three sets of DFT-based Karplus coefficients: two previously published sets, DFT-1 and DFT-2,⁶¹ and the DFT-3 set derived in this study. Parentheses indicate results that are less accurate since the DFT-1 and DFT-2 coefficients were solely derived for canonical residues.

Peptide	Residue	Simulation ^a			Experiment ^b
		DFT-1	DFT-2	DFT-3	
1	Ala ¹	6.1 ± 0.1	6.7 ± 0.1	6.4 ± 0.1	5.9 ± 0.4
	Ala ²	6.2 ± 0.2	6.7 ± 0.1	6.5 ± 0.2	6.2 ± 0.4
	Ala ³	6.5 ± 0.1	7.0 ± 0.1	6.8 ± 0.1	6.4 ± 0.4
2	Ala ¹	6.3 ± 0.3	6.8 ± 0.3	6.8 ± 0.4	5.9 ± 0.4
	D-Ala ²	6.1 ± 0.1	6.7 ± 0.1	6.4 ± 0.1	6.4 ± 0.4
	Ala ³	6.8 ± 0.2	7.4 ± 0.3	7.1 ± 0.2	6.8 ± 0.4
3	Ala ¹	6.4 ± 0.2	6.4 ± 0.2	6.7 ± 0.3	6.3 ± 0.4
	β^3 Ala ²	(7.5 ± 0.1)	(8.0 ± 0.1)	8.6 ± 0.1	9.2 ± 0.4
	Ala ³	6.5 ± 0.1	7.0 ± 0.1	6.8 ± 0.1	6.5 ± 0.4
4	Ala ¹	6.4 ± 0.2	7.0 ± 0.2	6.7 ± 0.2	6.1 ± 0.4
	ACPC ²	(6.9 ± 0.1)	(7.5 ± 0.1)	8.5 ± 0.1	8.4 ± 0.4
	Ala ³	6.6 ± 0.1	7.2 ± 0.1	6.9 ± 0.1	6.4 ± 0.4

^aReported uncertainties are twice the standard error of the mean from ten simulations.

^bReported uncertainties are estimated as half the spectral resolution of the experiment.

room temperature for 20 min. Post synthesis, the resin was successively washed with DMF, DCM, and methanol and allowed to dry in a vacuum desiccator for 30 min. Cleavage of the peptide from the resin was performed by treatment of the resin with 1% v/v TFA in DCM for 20 min–30 min with gentle stirring. The cleavage flowthrough was collected in a scintillation vial, and excess DCM evaporated under nitrogen. For peptide 4, the resin was washed an additional five times with cleavage solution after incubation to maximize the cleavage yield. Cleaved peptides were purified by preparative HPLC on a Phenomenex Luna Prep C18 column (10 μ M particle size and 100 Å pore size). The identity of each peptide was confirmed by electrospray-ionization mass spectrometry (ESI-MS; see Table S1), and the purity of each peptide was determined by analytical HPLC (see Fig. S5) and NMR (see Figs. S7–S10). Purified peptides were lyophilized prior to NMR analysis.

1. NMR data acquisition and analysis

Lyophilized peptides were weighed and dissolved in 9:1 H₂O:D₂O with 0.2 mM DSS and the pH adjusted to 2. NMR experiments were performed on a Bruker Avance 700 MHz spectrometer at 300 K. Spectra acquired for each peptide consisted of 1D ¹H (128 scans), 2D NOESY (200 ms mixing time and 16 scans), and 2D COSY (8 scans). All spectra were processed using TOPSPIN, employing DSS as an internal standard for chemical shift and concentration. Final concentrations were 10 mM, 9 mM, 17 mM, and 20 mM for peptides 1–4, respectively. J -coupling values were measured on TOPSPIN through deconvolution and peak fitting of amide H_N doublets (see Fig. S6) using a mixture of Lorentzian and Gaussian curves; reported experimental uncertainties (0.4 Hz) were estimated as half the spectral resolution. Variable-concentration 1D ¹H experiments for peptide 1 in the range of 1 mM–10 mM showed

that spectra were concentration-independent in this range, demonstrating that the peptide is monomeric under the conditions of the measurement (see Fig. S11).

IV. RESULTS AND DISCUSSION

A. Accuracy of torsion and angle parameters

To assess the accuracy of the newly derived torsion and angle parameters for each of the 25 artificial amino acid residues, we monitored the RMSE of the AMBER ff15ipq-m molecular mechanical (MM) potential energy surface from the target quantum mechanical (QM) potential energy surface. As shown in Fig. 3, the RMSE values range from 1.3 kcal/mol to 2.3 kcal/mol. These values are slightly higher, by 0.1 kcal/mol–1 kcal/mol, relative to those of their canonical counterparts in the parent AMBER ff15ipq force field²² due, in part, to the increased backbone flexibility that results from the additional methylene group in the β^3 residues and a greater likelihood for forming backbone hydrogen bonds.

B. Conformational preferences of individual artificial residues

To characterize the backbone conformational preferences of each artificial residue modeled by the ff15ipq-m force field, we carried out umbrella sampling of Ace-Ala-Xaa-Ala-NMe tetrapeptides with the artificial residue at the central position, progressively restraining the backbone torsion angles. Given that the simultaneous variation of all three backbone torsion angles in the β^3 and β^{yc} residues is computationally prohibitive within the same umbrella sampling simulation, only two of the backbone torsion

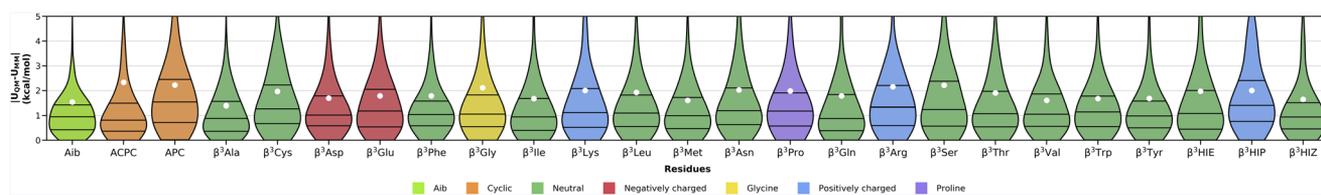


FIG. 3. Violin plots showing distributions of residuals of relative molecular mechanical (MM) energies, with respect to their quantum mechanical (QM) energies, for 25 Ace-Xaa-NMe dipeptides where Xaa is one of the newly parameterized residues in the force field. Distributions are colored according to the seven backbone classes indicated in the legend. Residue names for $\beta^3\text{HIE}$, $\beta^3\text{HIZ}$, and $\beta^3\text{HIP}$ are specified for neutral $\beta^3\text{His}$ protonated at the ϵ -nitrogen, neutral $\beta^3\text{His}$ protonated at the ζ -nitrogen, and positively charged $\beta^3\text{His}$ protonated at both nitrogens, respectively. The 25th, 50th, and 75th percentiles are represented by horizontal lines, while root-mean-square values are indicated by white dots. Each dataset is colored based on its corresponding backbone torsion class.

angles were restrained at a time, yielding a total of three free energy surfaces (i.e., φ/ψ , φ/θ , and ψ/θ Ramachandran plots) after the application of the weighted histogram analysis method (WHAM; see Sec. III). To validate the calculated Ramachandran plots from

the simulations, we compared predicted torsional preferences to experimentally observed conformations from 107 x-ray structures of proteins containing one or more of the newly parameterized residues from the PDB (see Data S1).

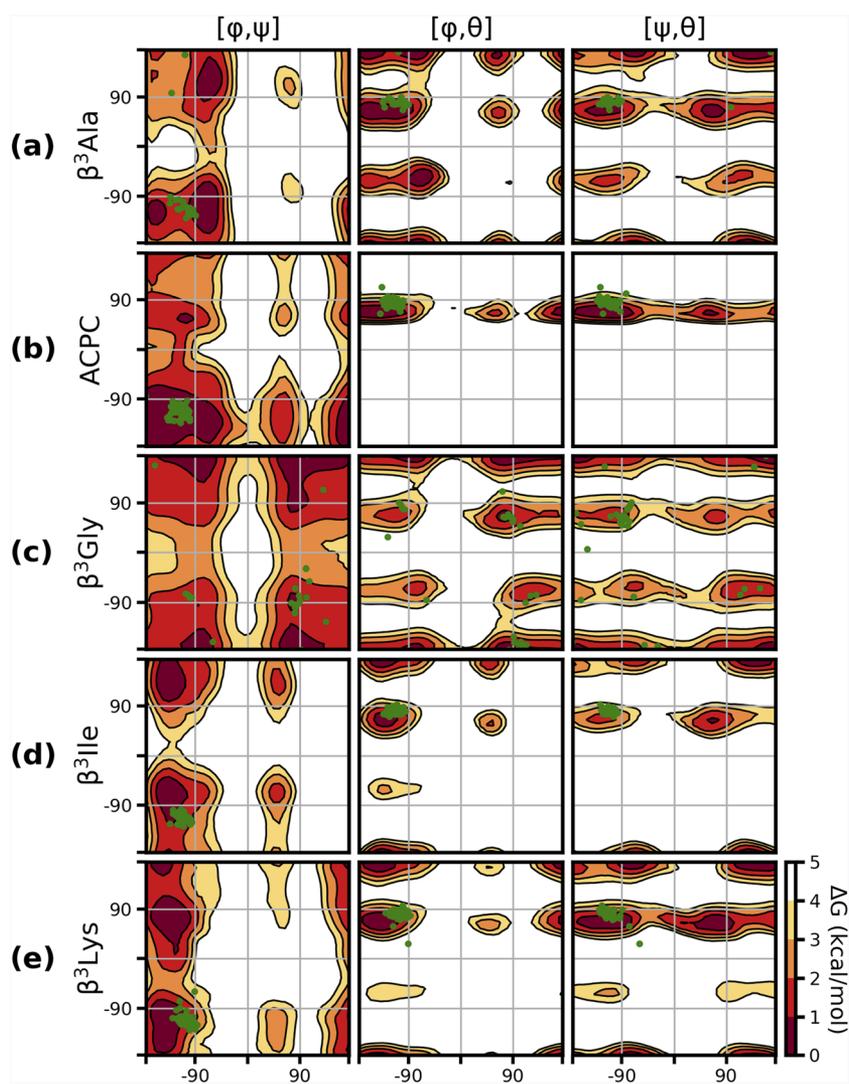


FIG. 4. Ramachandran plots of backbone torsion angles adopted by five artificial residues: (a) $\beta^3\text{Ala}$, (b) ACPC, (c) $\beta^3\text{Gly}$, (d) $\beta^3\text{Ile}$, and (e) $\beta^3\text{Lys}$. The free energy surfaces as $[x, y]$ functions of $[\varphi, \psi]$, $[\varphi, \theta]$, and $[\psi, \theta]$ are based on conformational preferences of each artificial residue at the central positions of an Ace-Ala-Xaa-Ala-NMe tetrapeptide from umbrella sampling and subsequent application of WHAM. Plots were constructed using fine histogram bins ($2^\circ \times 2^\circ$), and no smoothing was applied to the contour lines. The green filled circles indicate backbone torsions observed for the corresponding residue in experimental structures from the PDB.

Overall, the backbone torsional preferences observed for each artificial residue in our tetrapeptide simulations are in qualitative agreement with the behavior of the corresponding residue seen in the PDB dataset (see Fig. 4 and Fig. S12). Of note, the experimental dataset is dominated by structures where the artificial residue is in an α -helical context as α -helix mimicry has historically been a major focus in development of protein mimetics. As a result, the corresponding regions in the β -residue Ramachandran plots (at around $\phi/\psi -110^\circ/-110^\circ$, $\psi/\theta -110^\circ/90^\circ$, and $\psi/\theta -110^\circ/90^\circ$) are heavily populated by experimental data points.

Consistent with the principles that informed earlier application to control β -residue conformational preferences in homogeneous β -peptide backbones⁶² and heterogeneous backbones displaying biological side chain sequences,⁶³ the Ramachandran plot for the β^{Gly} residue ACPC is highly restricted and favors a right-handed helical conformation. Ramachandran plots for β^3 residues bearing proteinogenic side chains show more accessible minima, many outside the helical fold seen in experimental structures. As expected, β^{Gly} exhibits the greatest diversity of conformations and falls within the same free energy well that is populated by the PDB dataset. Conformations of representative γ -branched ($\beta^3\text{Ile}$) and charged ($\beta^3\text{Lys}$) residues also fall within the free energy wells populated by the PDB dataset. Taken together, the above findings validate the conformational preferences of the newly parameterized artificial residues modeled by ff15ipq-m. Furthermore, given that the

PDB dataset consists of structures for entire proteins containing the artificial residues, the qualitative agreement of our simulation-based Ramachandran plots with those from the PDB dataset indicates that the conformational preferences observed by simulations in the context of tetrapeptides are predictive of those in the context of proteins.

C. J -coupling constants of short peptides

To complement our qualitative comparisons for the individual artificial residues, we quantitatively assessed the accuracy of backbone conformational predictions by ff15ipq-m for selected artificial residues by calculating backbone amide J -coupling constants for tetrapeptides 1–4 (see Fig. 2 and Table I) and comparing to the observations in NMR experiments on the same compounds. J -coupling constants corresponding to the ϕ torsion angle ($^3J_{\text{HN-H}\alpha}$ for α residues and $^3J_{\text{HN-H}\beta}$ for β residues) were calculated using the Karplus relation and three sets of DFT-based Karplus coefficients: DFT-1, DFT-2, and DFT-3 (see Table S2). The DFT-1 and DFT-2 coefficients were previously derived for canonical residues,⁶¹ whereas the DFT-3 coefficients were derived in this study for the $\beta^3\text{Ala}$ and ACPC residues and re-calibrated for the Ala residue.

As shown in Table II, when using the newly derived DFT-3 coefficients, the ff15ipq-m force field yields J -coupling constants that

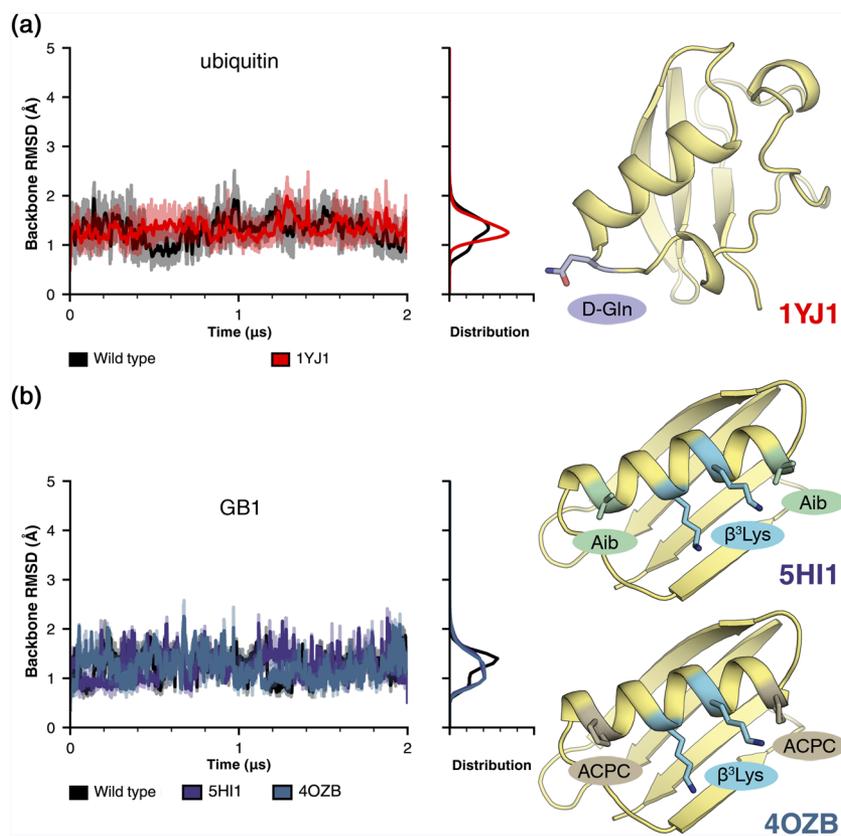


FIG. 5. Stability of proteins containing artificial residues as monitored by backbone RMSD between simulations relative to experimental structures for (a) wild-type ubiquitin (PDB code 1UBQ)⁴³ and a variant of ubiquitin with D-Gln35 in place of L-Gln35 (PDB code 1YJ1)⁴⁴ and (b) wild-type GB1 (PDB code 1PGB);⁴⁵ GB1 variant with Aib24, $\beta^3\text{Lys}28$, $\beta^3\text{Lys}31$, and Aib35 (PDB code 5H11);²⁰ and GB1 variant with ACPC24, $\beta^3\text{Lys}28$, $\beta^3\text{Lys}31$, and ACPC35 (PDB code 4OZB).¹⁹ Locations of the artificial residues in the corresponding crystal structures are indicated by stick representations in the ribbon diagrams, which were generated using PyMOL.⁶⁴ The ubiquitin variant (PDB code 1YJ1) exhibited an average backbone RMSD of $1.3 \pm 0.2 \text{ \AA}$ (mean \pm one standard deviation) compared to $1.3 \text{ \AA} \pm 0.3 \text{ \AA}$ for wild-type ubiquitin. The two GB1 variants with PDB codes 5H11 and 4OZB exhibited average backbone RMSDs of $1.2 \text{ \AA} \pm 0.3 \text{ \AA}$ and $1.2 \text{ \AA} \pm 0.3 \text{ \AA}$, respectively, compared to $1.3 \text{ \AA} \pm 0.3 \text{ \AA}$ for wild-type GB1.

are generally within error of experiment. The only exception is that the J -coupling constant for β^3 Ala is just outside of the error for the experimental value, indicating a conformational ensemble with a slightly less extended character in the simulation compared to the experiment. The β^3 Ala and ACPC β^{cyc} residues appear to have significantly different J -coupling constant distributions than those of the Ala and D-Ala residues (see Fig. S4), likely, due to the preference toward an extended conformation and the increased flexibility of the peptide backbone. Relative to the DFT-1 and DFT-2 coefficients, the DFT-3 coefficients yield improved agreement with experimental J -coupling values for the β^3 Ala and ACPC residues, while preserving the expected trends in the ensemble-averaged values and distributions (see Fig. S4) of J -coupling constants for the Ala, D-Ala, β^3 Ala, and ACPC residues.

D. Stability of proteins

To assess the kinetic stability of protein mimetics with the ff15ipq-m force field, we carried out μs -timescale simulations for a set of three backbone modified proteins that have high-resolution crystal structures available and together contain the various types of artificial backbone units that have been parameterized for the force field (i.e., D- α -, Aib, β^3 , and β^{cyc} residues). The set of protein mimetics consists of (i) a ubiquitin variant that contains D-Gln35 in place of L-Gln35; (ii) a GB1 variant that contains Aib24, β^3 Lys28, β^3 Lys31, and Aib35; and (iii) a GB1 variant that contains ACPC24, β^3 Lys28, β^3 Lys31, and ACPC35. For comparison, simulations were also performed for the corresponding wild-type proteins. All simulations were carried out at 25 °C. For further simulation details, see Table I and Sec. III.

As shown in Fig. 5, the overall structures of both the wild-type and variant proteins remained stable over their entire simulations. Only small deviations from the corresponding crystal structures were observed, and these were similar in magnitude to deviations observed in simulations of the corresponding wild-type proteins. The average backbone root-mean-square deviation (RMSD) from the corresponding crystal structure is <2.3 Å for all systems examined. Notably, the secondary structures of both wild-type ubiquitin and GB1 are maintained throughout the simulations in their corresponding variants (see Fig. S13).

V. CONCLUSIONS

In summary, we have reported here the development and validation of a general force field, AMBER ff15ipq-m, for the molecular modeling of protein mimetics with backbones bearing combinations of canonical α residues and four classes of artificial backbone units: D- α -, Aib, β^3 , and β^{cyc} residues (APC and ACPC). The ff15ipq-m force field is an expansion of the AMBER ff15ipq force field for natural proteins and includes 472 unique atomic charges and 148 unique backbone torsion terms. As was the case for creating the ff15ipq force field, the derivation of new IPolQ parameters was greatly facilitated by the mdgx program of AMBERTools via a “sweeping optimization” approach.

The AMBER ff15ipq-m force field features all the following for the modeling of the newly parameterized artificial backbone units: (i) a charge set that accounts directly for water-induced polarization, as consistent with the AMBER IPolQ lineage of force field;

(ii) conformational preferences of individual residues that are consistent with those found in crystal structures of protein mimetics; (iii) strong agreement with NMR J -coupling constants for short peptides using newly derived Karplus coefficients; and (iv) maintenance of globular protein folds with altered backbones on the μs timescale.

The new force field reported here and the path to its development have a number of potential implications with respect to understanding folding behavior in protein-inspired artificial backbones. The ff15ipq-m force field can be used to conduct atomistic simulations on an array of experimentally characterized systems, shedding light on fundamental issues related to the impact of altered backbone composition on folding and function. Furthermore, simulations on simple *de novo* systems may provide insights into regions of backbone configurational space that are accessible to heterogeneous backbones but not previously observed experimentally due to the historical focus on mimicry of α -helical structures. Finally, the present results demonstrate the applicability of the IPolQ methodology to backbones beyond a peptide. The workflow described should prove readily applicable to additional residue classes, raising the possibility of future expansion of ff15ipq-m to encompass a broad array of macromolecules in chemical space that are adjacent to but distinct from proteins.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for new atom types for artificial backbone units; probability distributions of torsion angles and J -coupling constants for central residues of peptides 1–4; DFT-3 Karplus coefficients and least-squares fits for extracting these coefficients; HPLC traces, ESI-MS data, and NMR spectra for peptides 1–4; Ramachandran plots for all 25 artificial amino acid residues in the context of tetrapeptides; secondary structure content of ubiquitin and GB1 variants in simulations on the μs timescale; backbone torsion classes of the AMBER ff15ipq-m force field; PDB codes of the 107 crystal structures used for qualitative validation of computed Ramachandran plots; and ff15ipq-m force field files for use with the AMBER software package.

AUTHORS' CONTRIBUTIONS

A.T.B., H.E.P., and J.M.G.L. contributed equally to this work.

ACKNOWLEDGMENTS

This work was supported by the University of Pittsburgh Arts and Sciences Graduate Fellowship to A.T.B., the NSF (Grant No. CHE-1807301) to L.T.C. and W.S.H., the NIH (Grant No. 1R01GM115805-05) to L.T.C., and the NIH (Grant No. U54AI50470) private clusters in the laboratory of D.A.C. Computational resources were provided by the University of Pittsburgh's Center for Research Computing and private clusters in the CASE Laboratory.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#). All AMBER

ff15ipq-m force field parameters will be included as an update to the AmberTools20 distribution and are provided in the [supplementary material](#).

REFERENCES

- ¹D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, and J. S. Moore, *Chem. Rev.* **101**, 3893 (2001).
- ²S. H. Gellman, *Acc. Chem. Res.* **31**, 173 (1998).
- ³I. M. Mándity and F. Fülöp, *Expert Opin. Drug Discov.* **10**, 1163 (2015).
- ⁴V. Azzarito, K. Long, N. S. Murphy, and A. J. Wilson, *Nat. Chem.* **5**, 161 (2013).
- ⁵M. Pelay-Gimeno, A. Glas, O. Koch, and T. N. Grossmann, *Angew. Chem., Int. Ed. Engl.* **54**, 8896 (2015).
- ⁶W. S. Horne and T. N. Grossmann, *Nat. Chem.* **12**, 331 (2020).
- ⁷K. Drew, P. D. Renfrew, T. W. Craven, G. L. Butterfoss, F.-C. Chou, S. Lyskov, B. N. Bullock, A. Watkins, J. W. Labonte, M. Pacella, K. P. Kilambi, A. Leaver-Fay, B. Kuhlman, J. J. Gray, P. Bradley, K. Kirshenbaum, P. S. Arora, R. Das, and R. Bonneau, *PLoS One* **8**, e67051 (2013).
- ⁸D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, and S. Whitelam, *J. Comput. Chem.* **35**, 360 (2014).
- ⁹K. Vanommeslaeghe and A. D. MacKerell, *Biophys. J.* **108**, 156a (2015).
- ¹⁰K. L. George and W. S. Horne, *Acc. Chem. Res.* **51**, 1220 (2018).
- ¹¹H. S. Haase, K. J. Peterson-Kaufman, S. K. Lan Levensgood, J. W. Checco, W. L. Murphy, and S. H. Gellman, *J. Am. Chem. Soc.* **134**, 7652 (2012).
- ¹²Z. E. Reinert, G. A. Lengyel, and W. S. Horne, *J. Am. Chem. Soc.* **135**, 12528 (2013).
- ¹³J. W. Checco, D. F. Kreidler, N. C. Thomas, D. G. Belair, N. J. Rettko, W. L. Murphy, K. T. Forest, and S. H. Gellman, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4552 (2015).
- ¹⁴K. L. George and W. S. Horne, *J. Am. Chem. Soc.* **139**, 7931 (2017).
- ¹⁵S. K. Mong, F. V. Cochran, H. Yu, Z. Graziano, Y.-S. Lin, J. R. Cochran, and B. L. Pentelute, *Biochemistry* **56**, 5720 (2017).
- ¹⁶A. G. Kreutzer and J. S. Nowick, *Acc. Chem. Res.* **51**, 706 (2018).
- ¹⁷C. C. Cabaltea, D. S. Mihalko, and W. S. Horne, *ChemBioChem* **20**, 103 (2019).
- ¹⁸H. M. Werner, S. K. Estabrooks, G. M. Preston, J. L. Brodsky, and W. S. Horne, *ChemBioChem* **20**, 2346 (2019).
- ¹⁹Z. E. Reinert and W. S. Horne, *Chem. Sci.* **5**, 3325 (2014).
- ²⁰N. A. Tavenor, Z. E. Reinert, G. A. Lengyel, B. D. Griffith, and W. S. Horne, *Chem. Commun.* **52**, 3789 (2016).
- ²¹S. R. Rao and W. S. Horne, *Pept. Sci.* **112**, e24177 (2020).
- ²²K. T. Debiec, D. S. Cerutti, L. R. Baker, A. M. Gronenborn, D. A. Case, and L. T. Chong, *J. Chem. Theory Comput.* **12**, 3926 (2016).
- ²³P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. Mackerell, Jr., *J. Chem. Theory Comput.* **9**, 5430 (2013).
- ²⁴Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, *J. Chem. Theory Comput.* **9**, 4046 (2013).
- ²⁵D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, and P. A. Kollman, AMBER 18, University of California, San Francisco, 2018.
- ²⁶A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, *J. Chem. Theory Comput.* **8**, 1542 (2012).
- ²⁷R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, *J. Chem. Theory Comput.* **9**, 3878 (2013).
- ²⁸S. Le Grand, A. W. Götz, and R. C. Walker, *Comput. Phys. Commun.* **184**, 374 (2013).
- ²⁹D. S. Cerutti, W. C. Swope, J. E. Rice, and D. A. Case, *J. Chem. Theory Comput.* **10**, 4515 (2014).
- ³⁰D. S. Cerutti, J. E. Rice, W. C. Swope, and D. A. Case, *J. Phys. Chem. B* **117**, 2328 (2013).
- ³¹K. Takemura and A. Kitao, *J. Phys. Chem. B* **116**, 6279 (2012).
- ³²K. T. Debiec, A. M. Gronenborn, and L. T. Chong, *J. Phys. Chem. B* **118**, 6561 (2014).
- ³³R. B. Best and J. Mittal, *J. Phys. Chem. B* **114**, 14916 (2010).
- ³⁴J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, *J. Chem. Theory Comput.* **11**, 3696 (2015).
- ³⁵F. Hoffmann, F. A. A. Mulder, and L. V. Schäfer, *J. Chem. Phys.* **152**, 084102 (2020).
- ³⁶P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, *PLoS Comput. Biol.* **13**, e1005659 (2017).
- ³⁷F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 73 (2012).
- ³⁸F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1327 (2018).
- ³⁹J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- ⁴⁰S. Ali, M. Banck, R. Braithwaite, J. Bunt, D. Curtis, N. Fox, M. Hanwell, G. Hutchison, B. Jacob, D. Lonie, J. Mantha, T. Margraf, C. Niehaus, S. Ochsenreither, K. Tokarev, and T. Vandermeersch, *Avogadro: An Open-Source Molecular Builder and Visualization Tool* (Avogadro Chemistry, 2016).
- ⁴¹M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, *J. Cheminf.* **4**, 17 (2012).
- ⁴²H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.* **28**, 235 (2000).
- ⁴³S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, *J. Mol. Biol.* **194**, 531 (1987).
- ⁴⁴D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, and S. B. Kent, *Angewandte Chem. Int.* **44**, 3852 (2005).
- ⁴⁵T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland, *Biochemistry* **33**, 4721 (1994).
- ⁴⁶Z. Xiang and B. Honig, Scap (n.d.).
- ⁴⁷Z. Xiang and B. Honig, *J. Mol. Biol.* **311**, 421 (2001).
- ⁴⁸M. P. Jacobson, R. A. Friesner, Z. Xiang, and B. Honig, *J. Mol. Biol.* **320**, 597 (2002).
- ⁴⁹I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B* **112**, 9020 (2008).
- ⁵⁰M. P. Allen and D. J. Tildesley, *Oxford Scholarship Online* (Oxford University Press, 2017).
- ⁵¹U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- ⁵²J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- ⁵³A. Grossfield, WHAM: The Weighted Histogram Analysis Method (n.d.).
- ⁵⁴M. Karplus, *J. Chem. Phys.* **30**, 11 (1959).
- ⁵⁵D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.* **9**, 3084 (2013).
- ⁵⁶J. J. Chou, D. A. Case, and A. Bax, *J. Am. Chem. Soc.* **125**, 8959 (2003).
- ⁵⁷C. Pérez, F. Löhner, H. Rüterjans, and J. M. Schmidt, *J. Am. Chem. Soc.* **123**, 7081 (2001).
- ⁵⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 09, Gaussian, Inc., Wallingford, CT, 2016.

⁵⁹W.-M. Hoe, A. J. Cohen, and N. C. Handy, *Chem. Phys. Lett.* **341**, 319 (2001).

⁶⁰F. Jensen, *Theor. Chem. Acc.* **126**, 371 (2010).

⁶¹D. A. Case, C. Scheurer, and R. Brüschweiler, *J. Am. Chem. Soc.* **122**, 10390 (2000).

⁶²D. H. Appella, L. A. Christianson, D. A. Klein, D. R. Powell, X. Huang, J. J. Barchi, Jr., and S. H. Gellman, *Nature* **387**, 381 (1997).

⁶³W. S. Horne, J. L. Price, and S. H. Gellman, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9151 (2008).

⁶⁴The PyMOL Molecular Graphics System, Schrodinger, LLC, 2010.