

Development and Validation of Fluorinated Amino Acid Parameters for use with the AMBER ff15ipq Protein Force Field

Darian T. Yang^{1, 2, 3}, Angela M. Gronenborn^{2*}, Lillian T. Chong^{3*}.

¹ Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, Pennsylvania 15260, United States

² Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, United States

³ Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

*Corresponding authors: ltchong@pitt.edu and amg100@pitt.edu

Abstract

We developed force field parameters for fluorinated aromatic amino acids enabling molecular dynamics (MD) simulations of fluorinated proteins. These parameters are tailored to the AMBER ff15ipq protein force field and enable the modeling of 4, 5, 6, and 7F-tryptophan, 3F- and 3,5F-tyrosine, and 4F- or 4-CF₃-phenylalanine. The parameters include 181 unique atomic charges derived using the Implicitly Polarized Charge (IPolQ) scheme in the presence of SPC/E_b explicit water molecules and 9 unique bond, angle, or torsion terms. Our simulations of benchmark peptides and proteins maintain expected conformational propensities on the μ s-timescale. In addition, we have developed an open-source Python program to calculate fluorine relaxation rates from MD simulations. The extracted relaxation rates from protein simulations are in good agreement with experimental values determined by ¹⁹F NMR. Collectively, our results illustrate the power and robustness of the IPolQ lineage of force fields for modeling structure and dynamics of fluorine containing proteins at the atomic level.

I. Introduction

While nuclear magnetic resonance (NMR) spectroscopy using the ^{19}F nucleus is emerging as a powerful tool for measuring various structural and dynamical properties of fluorine-labeled biomolecules¹, the availability of force fields for modeling the corresponding conformational dynamics at the atomic level has been limited. To address this unmet need, we have expanded upon the standard amino acids available in the AMBER ff15ipq protein force field² and present a comprehensive set of parameters for the most commonly used fluorinated aromatic amino acids in ^{19}F NMR experiments. Coupled with GPU-accelerated molecular dynamics (MD) simulations^{3,4}, we can now begin to effectively integrate the results from ^{19}F NMR experiments with atomistic MD simulations.

The fluorine nucleus has several features that render it a particularly useful NMR probe. Specifically, ^{19}F is a 100% naturally abundant isotope with a high gyromagnetic ratio, which makes it almost as sensitive as ^1H . Importantly, ^{19}F chemical shifts span a very large range and are exquisitely responsive to the local chemical and electronic environment around the atom⁵. Fluorine is absent from virtually all naturally occurring biomolecules. Therefore, studies of fluorinated biopolymers can be carried out in any routine buffer system or environment without interference from background signals⁶. In general, only one or a handful of fluorine atoms are introduced into a biopolymer, overcoming the common handicap of spectral overlap in proton spectra that necessitated uniform labelling with ^{13}C and ^{15}N , in conjunction with 3- and 4D heteronuclear spectroscopy, for resonance assignments in bioNMR. These desirable properties have propelled ^{19}F NMR based studies of biomolecular systems forward, such as investigations of thermodynamics⁷ and kinetics⁸, protein structure and dynamics^{9–11}, or protein-protein and protein-ligand interactions^{12–17}. Efforts are also ongoing towards ^{19}F NMR method development, including relaxation optimization approaches^{18,19} to extend applicability to larger systems and to exploit ^{19}F paramagnetic relaxation enhancement to determine long distances up to 35 Å²⁰. ^{19}F NMR is particularly useful in the pharmaceutical arena, guiding fragment-based screening²¹ and drug discovery/design²². Finally, a more recent and exciting direction is the movement of ^{19}F NMR into a physiological environment, measuring spectra of fluorinated proteins directly in *Escherichia coli*²³, oocyte²⁴, and mammalian cells²⁵.

Naturally, ^{19}F NMR of proteins is not a panacea and has shortcomings and limitations. The 1.47 Å van der Waals radius of the fluorine atom lies between those of hydrogen (1.2 Å) and oxygen (1.52 Å). Therefore, fluorine is frequently substituted for hydrogens, hydroxyl groups, or carbonyl oxygens. While such substitutions are only weakly perturbing, with little to no effect on a protein's structure and biological activity^{7,10,11,26}, this is not necessarily the case and needs to be verified for each system under study²⁷. Compared to more traditional NMR probes (^1H , ^{15}N , ^{13}C), introducing a highly electronegative fluorine atom changes electronic surroundings and thereby local interactions, especially if the substituted H or OH was involved in hydrogen bonding. In addition, although resonance overlap is avoided, having only a few probes available, compared to thousands with traditional ^1H , ^{15}N , ^{13}C labeling, limits the information content and several samples may need to be prepared where single fluorinated amino acids are judiciously placed to probe different parts of the biomolecule. MD simulations can help bridge this gap in information content at the all-atom level.

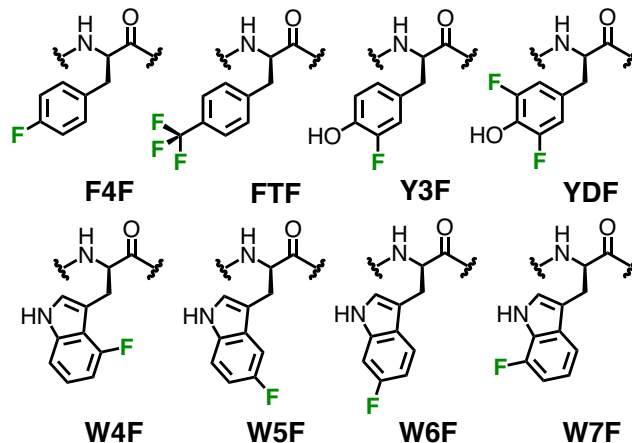


Figure 1. Eight commonly used fluorinated aromatic amino acids with their respective 3-letter identifier codes. The fluorine atom is highlighted in green.

Here, we have developed new force field parameters for eight commonly used fluorinated aromatic amino acids, facilitating more accurate atomic-level simulations of ^{19}F NMR observables. As indicated in **Figure 1**, our study focuses on 4, 5, 6, and 7 fluoro-tryptophan (W4F, W5F, W6F, W7F); 3, and 3, 5 fluoro-tyrosine (Y3F, YDF); as well as 4 fluoro- and 4 trifluoromethyl-phenylalanine (F4F, FTF). These fluorinated derivatives of tryptophan, tyrosine, and phenylalanine were selected because they are readily available and can be easily incorporated into proteins for ^{19}F NMR^{28–31}. Our parameters are intended for use with the AMBER ff15ipq force field and were derived using a general workflow that has been used previously for deriving new classes of non-canonical residues³². Consistent with the ff15ipq force field, we derived implicitly polarized atomic charges in the presence of explicit solvent. The implicit polarization of atomic charges has been demonstrated to be essential for modelling condensed-phase electrostatics in the absence of explicit polarization by fixed-charge force fields^{2,33}, such as Drude oscillators³⁴ and inducible multipoles³⁵. Furthermore, application of the ff15ipq force field with the intended SPC/E_b water model yields more accurate rotational diffusion times of proteins, enabling direct calculation of NMR observables^{2,36}. Using the GPU-accelerated AMBER MD engine^{37,38}, we have extensively validated our force field parameters by simulating both peptides and proteins that include each of the eight fluorinated amino acids, yielding over 47 μs of aggregate simulation time. Our parameters maintain expected conformational propensities of both fluorinated peptide- and protein-based systems on the μs -timescale, and our relaxation rates from simulation agree with those from ^{19}F NMR experiments.

Until now, parameters for the full set of fluorinated aromatic amino acids listed above for the AMBER ff15ipq protein force field have not been available, although recently, parameters for over 400 non-standard amino acids were added to the CHARMM36 protein and CHARMM general force fields³⁹, including some for several fluorinated amino acids. The CHARMM parameters were derived by isolating only the sidechain atoms of the amino acid of interest and optimizing select water interactions based on quantum mechanical calculations at either the MP2/6-31G(d) or the HF/6-31G(d) level of theory. However, CHARMM parameters are not available for 7-fluoro-tryptophan and 4-fluoro-phenylalanine. In addition, parameters for apolar,

non-aromatic, fluorinated amino acids are available for use with the AMBER ff14SB force field, which selectively optimized fluorine specific Lennard-Jones parameters and derived atomic charges at the RHF/6-31G* level of theory^{40,41}. Also, for use with the ff14SB force field, a set of unnatural phenylalanine and tyrosine derivatives was parameterized which included 3, 5 fluoro-tyrosine⁴². Atomic charges for these parameters were based on electrostatic potential calculations of a dipeptide molecule at the HF/6-31G* level of theory. Other efforts to develop parameters for simulating fluorinated proteins have been primarily ad hoc and single-use cases⁴³⁻⁴⁵.

II. The AMBER ff15ipq force field

Parameters for the eight different fluorinated amino acids were developed for use with the AMBER ff15ipq protein force field, the latest version of the Implicitly Polarized Charge (ipq) force field lineage.^{2,46} These parameters are also compatible with the ff15ipq-m force field for protein mimetics³². Ipq force fields feature implicitly polarized atomic charges, with each charge optimized to reproduce the mean field electron density of the molecule in explicit solvent⁴⁷. Notably, the ff15ipq force field is parameterized using the three-point, explicit SPC/E_b (extended simple point charge) water model⁴⁸, which reproduces the experimental rotational diffusion of proteins and can therefore yield reasonable dynamical observables such as those measured by NMR without the computational burden associated with four-point water model alternatives.

The original motivation behind the ff15ipq force field was to yield more accurate propensities for salt bridge formation, which is a common limitation of most contemporary fixed-charge force fields³³. The final ff15ipq implementation was a complete rederivation of its predecessor⁴⁶, including a greatly expanded torsion and angle parameter set, atomic charges derived at the MP2/cc-pVTZ level in explicit solvent, and adjustments to atomic radii for polar hydrogens. The ff15ipq force field reproduces the experimental probabilities of salt-bridge formation, while maintaining the secondary structure of stably folded and disordered systems on the μ s-timescale, and faithfully predicts J-coupling constants for a penta-alanine peptide as well as NMR relaxation rates for protein systems².

Additional developments involved a modification of methyl side chain rotational barriers that help to accurately predict methyl relaxation rates⁴⁹, as well as an expanded force field (ff15ipq-m) for modeling four classes of artificial backbone units, including D- and C α -methylated amino acids, β -amino acids and two cyclic β residues³². These developments are implemented in both the AMBER^{37,38,50} and open-source OpenMM⁵¹ GPU-accelerated biomolecular simulation software packages, and are accessible to other such packages through the ParmEd program⁵² for format conversion.

Our new parameters for the eight fluorinated aromatic amino acids were derived using a workflow designed to be consistent with the parent ff15ipq derivation process. As done for the ff15ipq-m force field, one minor difference pertained in the generation of dipeptide conformations for both the charge and bonded parameter derivation: to increase and safeguard conformational diversity, we progressively restrained the backbone torsions at evenly spaced intervals before energy-minimizing each restrained conformation³². In the parent ff15ipq force field, conformations were generated using high-temperature MD simulations².

III. Methods

A. Derivation of IPolQ atomic charges

For each fluorinated residue, IPolQ atomic charges were derived using a four-step iterative procedure until convergence was reached (**Figure 2**):

- 1. Generate a set of conformations.** For a blocked dipeptide (a single amino acid with capping groups) containing the fluorinated residue of interest with acetyl (Ace) and N-methylamide (Nme) N- and C-terminal capping groups, respectively, 20 conformations were generated by progressively restraining backbone Φ/Ψ torsion angles from -180° to 180° using a force constant of 32 kcal/mol. Only for the first iteration, the initial set of atomic charges were derived using the AM1-BCC charge method⁵³. Each restrained conformation was then subjected to energy minimization and solvated using a truncated octahedral box of SPC/E_B water molecules with at least a 12 Å clearance between the solute and the edge of the box. After another round of energy minimization, each solvated system underwent a two-stage equilibration which included a 10 kcal/mol Å² positional restraint on the fluorinated amino acid. In the first stage, 20 ps of dynamics were carried out at constant temperature (25°C) and volume. In the second stage, 100 ps of dynamics were carried out at constant temperature (25°C) and pressure (1 atm). A final 500 ps simulation was performed at constant temperature (25°C) and volume during which the solute remained fixed, and the solvent coordinates were used to generate a distribution of point charges that represent the solvent reaction field potential. This distribution consisted 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 that reproduce contributions to the solvent reaction field potential from the periodic system beyond 5 Å.
- 2. Calculate the electrostatic potential of each conformation in vacuum and explicit solvent.** Two sets of QM electrostatic potential calculations were carried out for each conformation of the fluorinated amino acid at the MP2/cc-pVTZ⁵⁴⁻⁵⁷ level of theory using the ORCA 4.2.0⁵⁸ software package. The first set of QM calculations were in vacuum, and the other set included the solvent reaction field potential to represent the surrounding explicit solvent molecules.
- 3. Fit the average electrostatic potential over all conformations to atomic charges, both in vacuum and explicit solvent.** All eight aromatic fluorinated amino acids were fit together using the fitq module of the mdgx program⁵⁰, with atomic charges of the Ace and Nme capping groups fixed to net neutral values during the process.
- 4. Average the vacuum-phase and solvent-phase point charges to obtain implicitly polarized atomic charges.** This averaged charge set was used as the starting point for the next iteration of charge generation and fitting as described above.

The above procedure was repeated four times to reach convergence of the partial atomic charge values within 10% of the previous iteration's values.

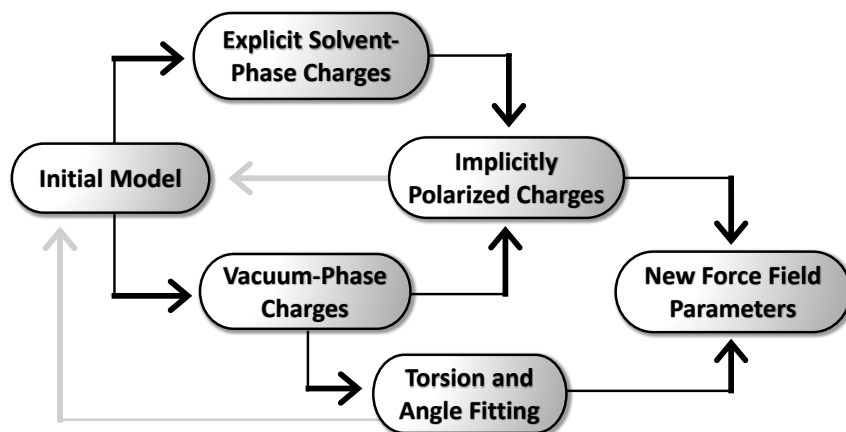


Figure 2. Workflow for developing ff15ipq force field parameters. Starting from a model consisting of multiple coordinate and topology files, the IPolQ charge derivation protocol is used to fit atomic charges from electrostatic potential calculations in both vacuum and explicit solvent phases. These charges are averaged, implicitly modeling solvent polarization effects. The vacuum-phase charges are used to fit bonded parameter terms to the vacuum-phase QM targets and used iteratively with the implicitly polarized charges, as indicated by the gray arrows, to generate a new initial model until the parameters are self-consistent.

B. Generation and fitting of the bonded parameter dataset

For each fluorinated amino acid, bonded parameters were derived using an iterative, multi-step procedure: (i) 1000 conformations of the blocked dipeptide were generated in vacuum, taking trial torsion and angle parameters from the parent ff15ipq force field, if available, or otherwise from the general amber force field ²⁵⁹. Atomic charges were obtained from the vacuum phase set of converged point charges as described above. Each conformation was generated by progressively restraining backbone Φ and Ψ torsion angles of the fluorinated dipeptide between -180° and 180° using a force constant of 32 kcal/mol. After energy minimization Φ and Ψ distributions of the conformation set were plotted to ensure extensive sampling of the configurational space. (ii) For each conformation, the quantum mechanical (QM) single point energy was calculated at the MP2/cc-pVTZ level of theory, along with the molecular mechanical (MM) energy. Linear least-squares fitting for the entire set of bond, angle, and torsion parameters was run using mdgx⁶⁰. (iii) Force field bonded parameters such as torsional barrier heights, angle equilibria, angle stiffness, bond length equilibria, and bond stiffness were adjusted to minimize the error between the QM and MM energies. Using the updated parameter set, a second round of conformer generation and QM energy calculations was carried out in the absence of restraints to prevent becoming trapped in local minima. This new set of conformations excluded redundant conformations, as defined by those with MM energies that differ by <0.01 kcal/mol. Another round of fitting was then performed using both sets of conformations and their respective energies to obtain the final parameter set for each iteration. Steps (i-iii) were repeated until the root-mean-square error (RMSE) between the QM and MM energies was less than 1% from the previous iteration (**Figure 2**).

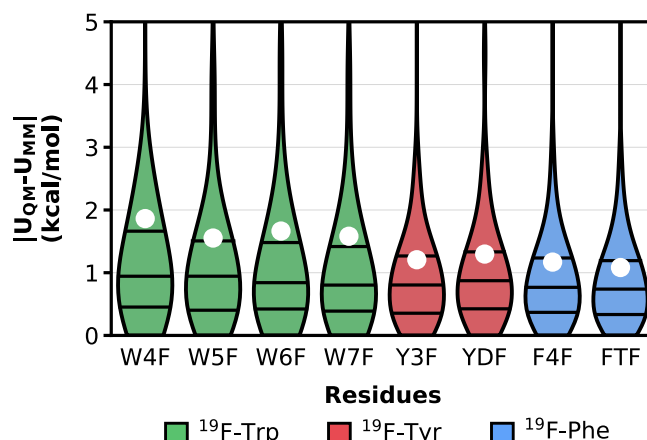


Figure 3. Distributions of residuals between quantum mechanical (U_{QM}) energies and molecular mechanical (U_{MM}) energies for the eight fluorinated Ace-Xaa-Nme dipeptides where Xaa is the fluorinated residue of interest. The 25th, 50th, and 75th percentiles are represented by horizontal lines and root-mean-square-error values are represented by white filled circles. Each dataset is colored according to the original residue type indicated in the legend.

The accuracy of the molecular mechanical energies (U_{MM}) produced using the optimized parameters for each fluorinated amino acid was assessed by comparing them to the respective QM energies (U_{QM}) for all generated conformations (**Figure 3**). During each iteration of the above fitting procedure, optimization of the bonded parameters was monitored by the RMSE, which reached final values of 1.08 - 1.86 kcal/mol for all eight residues. These errors are slightly higher, on average, than those of the non-fluorinated tryptophan, tyrosine, and phenylalanine counterparts from the canonical ff15ipq force field², which may be due to fewer parameters in the fitting procedure and/or the presence of electronegative fluorine atom(s).

C. Preparation of validation systems

Models of fluorinated and canonical tetrapeptides were built using Avogadro⁶¹ and tleap⁶⁰. To generate the 4, 5, 6, and 7F-Trp modified protein systems, the atomic coordinates of cyclophilin A (CypA; Protein Data Bank⁶² (PDB) ID: 3K0N)⁶³ were modified using Chimera⁶⁴, with fluorine atoms substituted individually for hydrogens in Trp121 at the respectively numbered positions on the indole ring. Each system was solvated in a truncated octahedral box of explicit SPC/E_b⁴⁸ water molecules with a 10 Å clearance between the solute and the edge of the box for the peptides and a 12 Å clearance for the proteins. All systems with unpaired charges were neutralized by adding Na⁺ or Cl⁻ ions, treated with Joung and Cheatham ion parameters⁶⁵. Protonation states for ionizable residues were adjusted to represent the major species present at pH 6.5 to match the experimental NMR conditions⁶⁶.

D. Umbrella sampling of fluorinated and canonical peptides

To validate the conformational preferences of each fluorinated residue, umbrella sampling simulations were carried out for a tetrapeptide (a tripeptide with capping groups (Ace-Ala-Xaa-Ala-NMe) containing a fluorinated amino acid at the central residue Xaa). Ramachandran plots were generated for each tetrapeptide by calculating the potential of mean force as a function of the Φ and Ψ torsion angles of the fluorinated residue. Prior to umbrella sampling, each tetrapeptide was solvated and equilibrated as described above for unrestrained simulations, differing only in the duration of the final equilibration stage, which was 100 ps instead of 1 ns. Each window was then subjected to a 200 ps incrementally restrained equilibration prior to a 2 ns restrained simulation at constant temperature (25 °C) and pressure (1 atm). The Φ and Ψ torsions of the central residue were restrained using a harmonic penalty function with a force constant of 8 kcal/mol rad² for each window with 10° intervals about each torsion angle. This restraint scheme resulted in a series of 1296 windows for each set of two torsions, cumulating into 22.8 μ s of aggregate simulation time for the eight fluorinated residue classes and 10,368 windows. From each set of 1296 windows, the unbiased potential of mean force was reconstructed using the weighted histogram analysis method (WHAM)^{67–69}.

E. MD Simulations of fluorinated and wild-type CypA

Simulations of the 4, 5, 6, 7F-Trp fluorinated and the wild-type CypA proteins were carried out using the GPU-accelerated pmemd module of the AMBER 18 software package^{37,38,50}, ff15ipq force field², and our new fluorinated amino-acid parameters. Each system was initially subjected to energy minimization followed by a three-stage equilibration. In the first stage, a 20 ps simulation was carried out at constant volume and temperature (25 °C) in the presence of solute heavy atom positional restraints using a harmonic potential with a force constant of 1 kcal/mol Å². In the second stage, a 1 ns simulation was carried out at constant temperature (25 °C) and pressure (1 atm) using the same harmonic positional restraints. Finally, an unrestrained 1 ns simulation was carried out before performing a 1 μ s production simulation with both constant temperature (25 °C) and pressure (1 atm). Five production simulations were run for each CypA protein, yielding 25 μ s of aggregate simulation time.

Temperatures were maintained using a Langevin thermostat with a frictional constant of 1 ps⁻¹, while pressure was maintained using a Monte Carlo barostat with 100 fs between system volume changes. 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 method⁷⁰. To enable a 2 fs time step, all CH and NH bonds were constrained to their equilibrium values using the SHAKE algorithm⁷¹. Coordinates were saved every ps for analysis in CPPTRAJ⁷².

F. Backbone conformations of fluorinated amino acids in the PDB

A “ligand expo” search⁷³ was carried out to determine whether each fluorinated amino acid of interest was present in PDB deposited structures. The corresponding three-letter identifiers in the PDB were as follows: 4FW, FTR, FT6, F7W, PFF, 55I, YOF, F2Y; which are equivalent to the following identifiers in the current study: W4F, W5F, W6F, W7F, F4F, FTF, Y3F, and YDF (**Figure 1**), respectively. Each identifier was associated with at least one

structure and was used to construct a query for structures that included the fluorinated residue as a part of a polymer chain, avoiding fluorinated ligand molecules. For structures determined by X-ray crystallography, only those with resolution $\leq 2.5 \text{ \AA}$ were included, resulting in a total of 56 structures with at least one fluorinated residue each. The backbone Φ and Ψ torsion angles were then calculated using a custom Python script, omitting the torsion angles belonging to fluorinated C-terminal and N-terminal residues.

G. Calculation of ^{19}F NMR relaxation rates

^{19}F longitudinal (R_1) and transverse relaxation rates (R_2) were calculated from MD simulations for each of the four ^{19}F -Trp CypA protein systems using equations 1-4^{5,74}. Both types of relaxation rates are affected by dipole-dipole (DD) interactions and chemical shift anisotropy (CSA). The influence of DD interactions is described by equations 1 and 2:

$$R_{1_{DD}} = \frac{1}{10} \cdot \frac{\gamma_F^2 \gamma_H^2 \hbar^2}{r_{FH}^6} \cdot \left(\frac{3}{1 + \omega_F^2 \tau_c^2} + \frac{1}{1 + (\omega_F - \omega_H)^2 \tau_c^2} + \frac{6}{1 + (\omega_F + \omega_H)^2 \tau_c^2} \right) \quad (1)$$

$$R_{2_{DD}} = \frac{1}{20} \cdot \frac{\gamma_F^2 \gamma_H^2 \hbar^2}{r_{FH}^6} \cdot \left(4 + \frac{3}{1 + \omega_F^2 \tau_c^2} + \frac{6}{1 + \omega_H^2 \tau_c^2} + \frac{1}{1 + (\omega_F - \omega_H)^2 \tau_c^2} + \frac{6}{1 + (\omega_F + \omega_H)^2 \tau_c^2} \right) \quad (2)$$

where γ is the gyromagnetic ratio of fluorine or hydrogen, \hbar is the reduced Plank's constant, ω is the Larmor frequency of fluorine or hydrogen, and τ_c is the rotational correlation time.

For each frame in the MD simulations, the distances between each hydrogen and the fluorine atoms within a 3 \AA radius around the fluorine were calculated (r_{FH}) and used in equations 1 and 2 to calculate the relaxation rate contribution to fluorine from each nearby hydrogen. These contributions were summed to account for the influence of all surrounding hydrogen dipoles.

The influence of CSA effects is described by equations 3 and 4:

$$R_{1_{CSA}} = \frac{3}{10} \cdot \omega_F^2 \delta_\sigma^2 \tau_c \left(1 + \frac{\eta^2}{3} \right) \frac{1}{1 + \omega_F^2 \tau_c^2} \quad (3)$$

$$R_{2_{CSA}} = \frac{1}{20} \cdot \omega_F^2 \delta_\sigma^2 \tau_c \left(1 + \frac{\eta^2}{3} \right) \left(4 + \frac{3}{1 + \omega_F^2 \tau_c^2} \right) \quad (4)$$

where δ_σ is the reduced anisotropy and η is the asymmetry parameter, as described in Haeberlen⁷⁵ convention by the following equations:

$$\delta_{iso} = \frac{1}{3} (\delta_{xx} + \delta_{yy} + \delta_{zz}) \quad (5)$$

$$|\delta_{zz} - \delta_{iso}| \geq |\delta_{xx} - \delta_{iso}| \geq |\delta_{yy} - \delta_{iso}| \quad (6)$$

$$\delta_\sigma = \delta_{zz} - \delta_{iso} \quad (7)$$

$$\eta = \frac{\delta_{yy} - \delta_{xx}}{\delta_\sigma} \quad (8)$$

The final relaxation rate values were then calculated as the sum of the individual DD and CSA based components, where $R_1 = R_{1_{DD}} + R_{1_{CSA}}$ and $R_2 = R_{2_{DD}} + R_{2_{CSA}}$.

These equations assume the following approximations: (i) the protein tumbles isotropically with the rotational correlation time remaining constant and the fluorinated side chain motion is governed by the same overall protein rotational correlation; (ii) fluorine-hydrogen distances close the F atom remain constant and can be described by a single value; (iii) cross-correlation

interactions⁷⁶ are not accounted for, even though a 10-25% cross-term contribution to the total R_2 relaxation rate may be possible⁶⁶.

For all CypA variants a single rotational correlation time (τ_c) of 8.2 ns⁷⁷ was used, along with the ¹⁹F CSA values from solid-state magic angle spinning (MAS) NMR experiments for 4, 5, 6, or 7-fluoro-tryptophans⁷⁸ (**Table S2**) to calculate δ_σ and η . To sample a representative ensemble, only the last 800 ns of each trajectory, from each independent 1 μ s production simulation, was used. All fluorine relaxation calculations were carried out using a custom-made Python program, which is open-source and freely available online (<https://github.com/darianyang/fluorelax>).

IV. Results and Discussion

Our parameters for the eight fluorinated aromatic amino acids were derived using the IPolQ workflow (**Figure 2**). Briefly, a set of conformations for a capped peptide with the fluorinated residue of interest was generated, and electrostatic potentials of each conformation were calculated quantum mechanically, both in vacuum and in the presence of explicit solvent. The resulting set of vacuum and solvent phase atomic charges were then optimized to reproduce the respective electrostatic potential calculations, before being averaged to obtain an implicitly polarized charge set. Along with van der Waals interactions, these atomic charges estimate the non-bonded contributions of the force field. With the vacuum optimized charges, the bonded terms of the force field, such as bond angles and dihedrals, were fit to minimize the error between the molecular mechanical and the quantum mechanical energies of each conformation. Both the atomic partial charges and the bonded parameter derivation steps were repeated until they were self-consistent. To validate our force field parameters, we initially carried out peptide simulations to explore the changes in the conformational free energy landscape of each fluorinated residue. We then performed protein-based simulations of 4, 5, 6, and 7F-Trp substituted Cyclophilin A (CypA) and compared our simulation results with those of the native CypA protein. Finally, we calculated NMR relaxation rates from our MD simulation data of CypA and compared these rates to the respective, experimentally determined ¹⁹F relaxation rates from NMR.

A. Conformational preferences of individual fluorinated residues

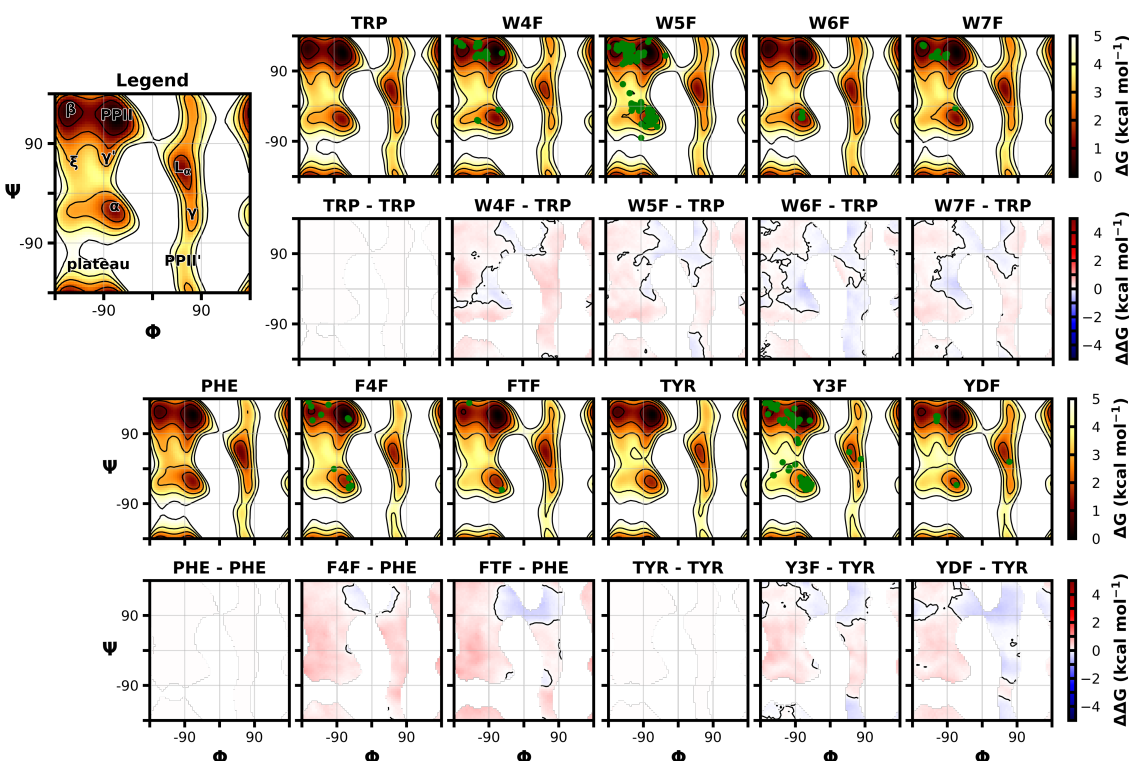


Figure 4. Free energy profiles for the central residue of tetrapeptides as a function of backbone Φ and Ψ torsions about the central residue, as well as the difference in free energy relative to the canonical non-fluorinated residue containing tetrapeptide. Each system is an Ace-Ala-Xaa-Ala-NMe peptide which underwent umbrella sampling simulations with subsequent application of WHAM. The green filled circles represent the backbone torsion angles observed for crystal structures that contain the same fluorinated residue as the tetrapeptide in their protein structure.

Backbone conformational preferences for the tetrapeptide systems (Ace-Ala-Xaa-Ala-Nme) were assessed using umbrella sampling simulations in which the central residue (Xaa) backbone dihedrals were progressively restrained (**Figure 4**). In umbrella sampling, a chosen reaction coordinate is initially divided into a series of windows or sections and a harmonic restraint is applied. In our case, this coordinate was the backbone Φ and Ψ dihedrals of the central residue. We choose 10° intervals from -180° to 180° using a force constant of 8 kcal/mol rad^2 to ensure that the reaction coordinate remained within the center of the window during MD simulations (see methods section E). A series of histograms along the reaction coordinate was generated, and because these distributions were overlapping, our umbrella sampling parameters were determined to be exhaustive and allowed for the corresponding unbiased free energy landscape to be recovered using WHAM^{67–69}. Umbrella sampling was carried out for the fluorinated amino acid containing peptides as well as their non-fluorinated counterparts (Trp, Tyr, Phe), as depicted in the relative free energy difference plots (**Figure 4**). In addition, we compared our simulation results to the experimentally observed conformations for each fluorinated variant, extracted from 56 protein structures deposited in the PDB.

Our results show that most free energy barriers between different secondary structures were either unperturbed, or only slightly increased for the fluorinated residues relative to their non-fluorinated counterparts, indicating less favorable sampling of that region. An exception is the area between the poly-proline II ($\Phi \approx -70^\circ$, $\Psi \approx 140^\circ$) and left-handed alpha helical ($\Phi \approx 60^\circ$, $\Psi \approx 40^\circ$) regions, which were more favorably sampled in all cases for fluorinated residues. For tyrosine and phenylalanine variants, we saw this favorable sampling amplified with more fluorine atoms present in the rings, such as with tyrosine possessing one or two fluorine atoms and phenylalanine with a single fluorine or the trimethyl fluorine group. For tryptophan, our resulting energy landscapes produced a similar trend, except that the alpha helical ($\Phi \approx -70^\circ$, $\Psi \approx -20^\circ$) and γ' ($\Phi \approx -80^\circ$, $\Psi \approx 60^\circ$) regions were more favorably sampled, with fluorination at the six-position of the indole ring exhibiting the largest difference. All dihedral angles of fluorinated residues remained consistent with those experimentally observed in structures deposited in the PDB. The backbone Φ and Ψ torsion angle energies for fluorine containing peptides are very similar to non-fluorinated peptides, with $\Delta\Delta G$ values ranging from -0.63 to 0.64 kcal/mol for the tryptophan variants, -0.55 to 0.65 kcal/mol for the tyrosine variants, and -0.53 to 0.88 kcal/mol for the phenylalanine variants (**Table S2**). The average $\Delta\Delta G$ of our non-fluorinated versus fluorinated peptides are all close to zero, further supporting our conclusion that only minimal perturbations are induced upon fluorine substitution. The only exception is for 4-fluoro-phenylalanine, for which an average $\Delta\Delta G$ value of 0.273 ± 0.186 kcal/mol was observed. Overall, these findings are consistent with previous studies where fluorinated tryptophan^{11,79}, tyrosine⁸⁰, and phenylalanine⁷ substitutions introduced minimal to no differences in the global protein structure or the local dihedral angles around the fluorinated residue.

B. Simulations of fluorinated Cyclophilin A

To evaluate our force field parameters in the context of a protein, we carried out multiple μ s-timescale simulations for each of four variants of cyclophilin A (CypA), a 18.3 kDa peptidyl-prolyl isomerase that is a known host-factor for HIV-1 infection⁸¹. The variants are singly fluorinated at four different indole ring positions of Trp 121, which is close to the active site of the protein. These fluorinated variants of CypA have been studied previously⁶⁶ and serve as useful benchmark proteins for our simulation studies.

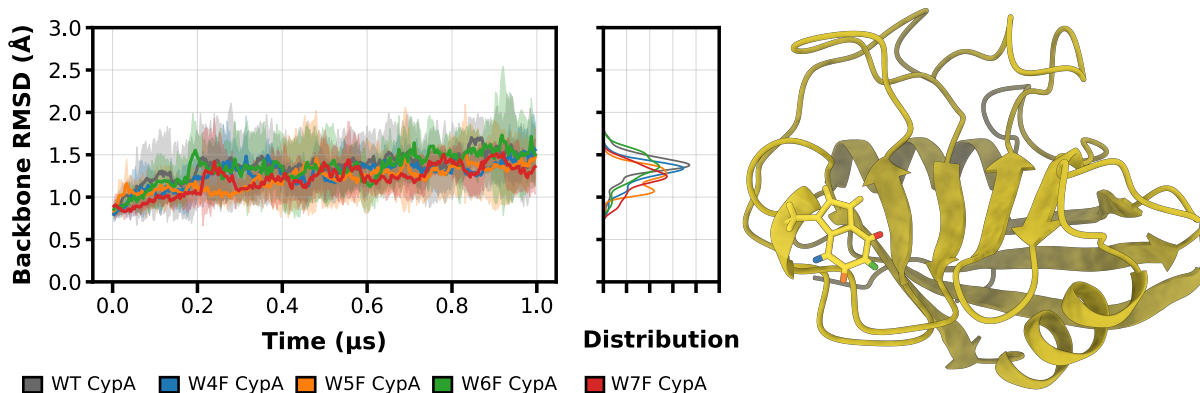


Figure 5. Average backbone RMSD values from the experimental X-ray structure (PDB: 3K0N)⁶³ coordinates for native and fluorinated CypA proteins as a function of simulation time. Data shown is based on five independent 1 μ s replicates, with one standard deviation of each time point shown in transparent coloration. The distribution of RMSD values is depicted in the middle panel and the protein structure being simulated is depicted on the right with individual fluorine positions on the TRP121 indole ring colored according to the legend.

Our data show that both wild-type and the fluorinated CypA variants all remained stable over the course of multiple μ s-timescale simulations (**Figure 5**), and only small deviations are noted. For the wild-type CypA simulations, the average backbone RMSD value is 1.37 ± 0.37 Å (average \pm one standard deviation), while the fluorinated Trp121 CypA variants exhibited average backbone RMSD values of 1.28 ± 0.23 Å (4F-Trp CypA), 1.22 ± 0.27 Å (5F-Trp CypA), 1.35 ± 0.31 Å (6F-Trp CypA), and 1.21 ± 0.20 Å, (7F-Trp CypA). RMSD values for each individual 1 μ s simulation as a function of simulation time are shown in **Figure S2**.

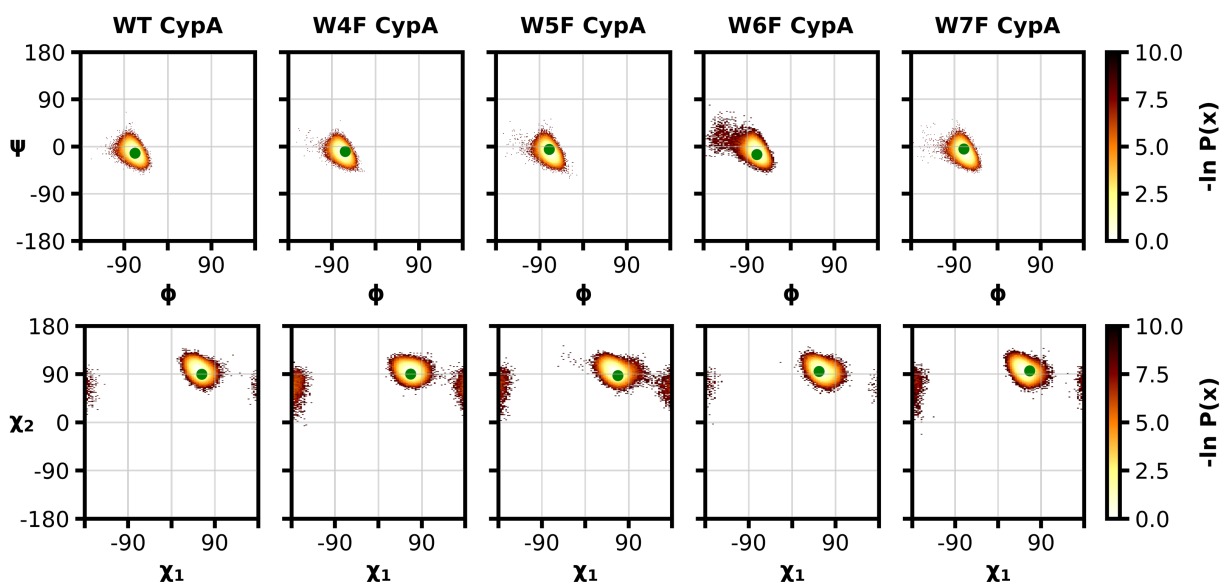


Figure 6. Ramachandran (Φ and Ψ) and Janin (χ_1 and χ_2) plots for an aggregate simulation time of 5 μ s for fluorinated or native CypA proteins. The green filled circles indicate the original dihedral angle in either the native CypA crystal structure⁶³, or the fluorinated CypA crystal structures at the respective 4, 5, 6, or 7 positions of the indole ring (unpublished data).

In addition to backbone RMSD values, we also assessed the dihedral angle propensities of our aggregate simulation data by generating Ramachandran (Φ and Ψ) and Janin (χ_1 and χ_2) plots. All dihedral probabilities for the fluorinated CypA proteins remained within the same distribution as that of wild-type CypA (**Figure 6**). Only for 6F-Trp, a slightly larger range of conformational sampling around the alpha helical region in the Ramachandran plot and a slightly restricted conformational sampling distribution in the Janin plots was noted. These findings are similar to those with our peptide systems, where the 6F-Trp substitution also

indicated a change in sampling of the free energy landscape near the alpha helical region, compared to the other fluorinated tryptophan moieties (**Figure 4**).

When evaluating the backbone RMSD and secondary structure predictions on a per-residue basis (**Figure S4**) excellent stability was maintained throughout multiple simulation replicates. However, a sharp increase in RMSD and corresponding change in secondary structure was noted between residues 146-153 and 102-107 during a fraction of both the wild-type and fluorinated CypA simulations. We isolated the characteristics of these increases in RMSD and found that they were largely driven by compensatory peptide-plane flips, which occur when changes in $|\Psi_i| + |\Phi_{i+1}|$ are large, while changes in $|\Psi_i + \Phi_{i+1}|$ are comparatively small⁸². We accessed the per-residue level torsion angles of each trajectory (**Figure S5**) and found that residues 146-153 are a part of a loop which undergoes a hinge motion that's compensated by a peptide-plane flip of residues 149 and 150 (top panel: **Figure S6**). Furthermore, residues 102-107 are a part of a beta turn that becomes twisted throughout our simulations and is compensated by peptide-plane flips of residues 103/104 and 107/108 (bottom and middle panel: **Figure S6**).

C. ¹⁹F Relaxation of CypA

In addition to structural properties, we also evaluated whether longitudinal and transverse fluorine NMR relaxation rates would be accessible from our aggregate CypA simulation data. NMR relaxation rates function as useful probes to study dynamics and depend on the properties of a nucleus, modulated by its local environment. Fluorine relaxation is governed by both dipole-dipole interactions with neighboring proton spins as well as chemical shift anisotropy (CSA)⁵. Fluorine CSA in solution is averaged due to molecular tumbling, although CSA associated motions can cause oscillations of the local magnetic field that contribute to relaxation. Thus, fluorine relaxation is complex and not well understood.

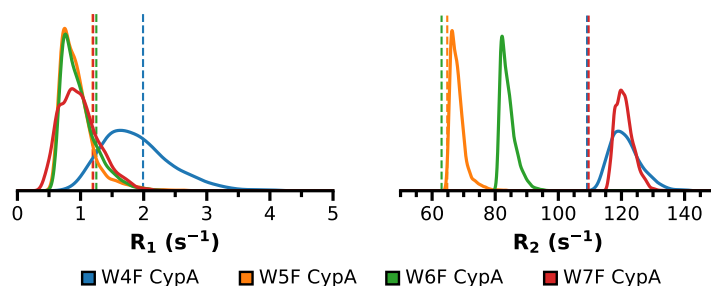


Figure 7. Fluorine longitudinal (R_1) and transverse (R_2) NMR relaxation rate distributions extracted from the final 80% of the 5 μ s of aggregate MD simulation data for each fluorinated CypA variant. Experimental relaxation rates⁶⁶ are shown as dotted vertical lines.

In a first approximation, we calculated longitudinal and transverse relaxation rates from our aggregate CypA simulation data using an overall rotational correlation time τ_c of 8.2 ns⁷⁷ (methods section G). We compared these calculated rates to the measured experimental rates⁶⁶ (**Figure 7**). Overall, our calculated relaxation rates from aggregate MD simulation data are in good agreement with the experimental values. The average calculated longitudinal rates (R_1)

are systematically somewhat smaller than the experimental values, although within a reasonable error (**Table S4**). 4F-Trp CypA exhibited a somewhat larger R_1 value ($\sim 2 \text{ s}^{-1}$) than all others, which are similar and grouped around 1 s^{-1} . The calculated transverse relaxation rates (R_2) are not as close to the experimental ones, but still followed the same trend. They can be grouped into two sets comprising 5F- and 6F-Trp CypA with R_2 values around $60\text{-}80 \text{ s}^{-1}$, and 4F- and 7F-Trp CypA with R_2 values around $110\text{-}120 \text{ s}^{-1}$. Overall, the calculated relaxation values follow the same trends and agree well with the experimental ^{19}F NMR relaxation rates.

At this juncture it should be pointed out that our methodology for calculating relaxation rates from MD simulation data includes several key assumptions. In particular, our calculations assume that (i) the tryptophan sidechain essentially moves like the overall protein, i.e. it does not exhibit internal motions in addition to the overall molecular tumbling of the protein (ii) only hydrogens within a radius of 3 \AA around the fluorine atom contribute to the relaxation rate, and (iii) any contribution from dipole-dipole and CSA cross-correlated relaxation⁷⁶ is unaccounted for, since they may only contribute 10-25% to the total R_2 relaxation rate⁶⁶. Furthermore, the experimental relaxation rates are extracted by using single exponential fitting of the signal intensity decays, although multi-exponential fitting may more accurately capture cross-correlation induced relaxation^{66,74,83}. Despite these differences between simulation and experiment, as well as their respective assumptions, the simulation-based calculated rates agree surprisingly well with the experimental relaxation rates. In the future, potentially more accurate methods⁸⁴ for fluorine relaxation rate calculations from MD simulations can be tested, which will help to dissect which components predominantly affect relaxation decay rates.

V. Conclusions

We have reported the development and validation of force field parameters for a set of eight fluorinated amino acids that are commonly used for ^{19}F NMR, for use with the AMBER ff15ipq protein force field. Our parameters include 181 implicitly polarized atomic charges and 9 unique bonded terms for 4, 5, 6, and 7 fluoro-tryptophan; 3, and 3, 5 fluoro-tyrosine; as well as 4 fluoro- and 4 trifluoromethyl-phenylalanine. We validated that our new parameters maintain the expected conformational propensities of the fluorinated amino acids consistent with both the respective canonical residues and previously characterized experimental X-ray structure-derived propensities extracted from structures deposited in the PDB. Fluorinated amino acid containing proteins, such as CypA, maintain the overall globular protein fold over multiple μs -timescale simulations and extracted ^{19}F NMR relaxation rates are in good agreement with the corresponding experimental rates.

Overall, our results demonstrate the robustness of the “sweeping optimization” approach using the mdgx program of AMBERTools20 distribution⁶⁰ and the power of the IPolQ lineage of AMBER force fields for modeling fluorinated proteins. Our workflow is readily applicable to other residue classes³² and can be easily expanded to include other fluorinated amino acids, if so desired. On the practical side, our force field parameters have numerous potential implications, particularly for use with complementary ^{19}F NMR studies and when considering structural ensembles. Since ^{19}F NMR data can be used to guide MD simulations and vice versa, our parameters will aid to derive an integrated, all-atom view of any fluorinated protein. Our parameters extend the macromolecular chemical space available to the AMBER ff15ipq force

field and bridge the gap between computation and experiment for the collaborative study of fluorinated proteins at the atomic level.

Similar to the IPolQ methodology, the Force Balance approach and the corresponding AMBER-FB15 force field⁸⁵ performs sweeping optimization of hundreds of parameters simultaneously, and includes nonlinear optimization methods which can incorporate alternative datasets, such as *in vitro* experiments, directly into the parameter optimization process. The FB15 force field has also recently been expanded to include parameters for phosphorylated amino acids⁸⁶, but at this time, does not have parameters for halogenated residues such as the fluorinated amino acids presented in this work. In contrast to the IPolQ workflow, which is fully physics-based, the burgeoning development of machine learning based force fields⁸⁷ are promising, but are heavily dependent on the training dataset being used. Thus, while physics-based approaches are more generalizable, deep learning methods can be more specialized for specific atoms or dependent on dataset properties. The goal of such machine learning based approaches is to narrow the gap between the efficiency of classical models and the accuracy of *ab initio* methods, by training a machine learning model to predict molecular potential energies, usually using quantum mechanical datasets. Examples of these approaches include the ANAKIN-ME^{88,89} neural network potential and the OrbNet⁹⁰ method, which, once fully trained, can accurately predict the energetic properties of organic molecules within chemical accuracy of *ab initio* density functional theory calculations, but at a fraction of the computational cost. In the future, machine learning based methods may be readily paired with the IPolQ method to further enhance the efficiency of the parameter fitting and optimization process.

Supporting Information

Supplemental figures include atom types for fluorinated amino acids and further analysis of CypA simulations: individual RMSD plots, carbonyl to backbone distances, secondary structure, per-residue RMSD, and per-residue backbone dihedral. Supplemental tables include individual force field bonded parameters, relative free energy differences between tetrapeptides, CSA tensors used for relaxation calculations, and comparison of average relaxation rates to experimental values.

Acknowledgments

This work was supported by the NIH Pittsburgh Aids Research Training (PART) program grant T32AI065380 to D.T.Y., NIH grant P50AI150481 and NSF grant CHE-1708773 to A.M.G., and NIH grant R01GM115805 and NSF grant CHE-1807301 to L.T.C. Computational resources were provided by the University of Pittsburgh's Center for Research Computing and through NSF XSEDE grant TG-BIO210161 to D.T.Y.

References

1. Gronenborn, A. M. Small, but powerful and attractive: 19F in biomolecular NMR. *Structure* **0**, (2021).
2. Debiec, K. T. *et al.* Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.*

- 12**, 3926–3947 (2016).
3. Friedrichs, M. S. *et al.* Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J. Comput. Chem.* **30**, 864 (2009).
 4. Stone, J. E., Hardy, D. J., Ufimtsev, I. S. & Schulten, K. GPU-accelerated molecular modeling coming of age. *J. Mol. Graph. Model.* **29**, 116–125 (2010).
 5. Gerig, J. T. Fluorine NMR. 1–35 <https://www.biophysics.org/Portals/0/BPSAssets/Articles/gerig.pdf> (2001).
 6. Sharaf, N. G. & Gronenborn, A. M. ¹⁹F-Modified Proteins and ¹⁹F-Containing Ligands as Tools in Solution NMR Studies of Protein Interactions. *Methods Enzymol.* **565**, 67–95 (2015).
 7. Welte, H., Zhou, T., Mihajlenko, X., Mayans, O. & Kovermann, M. What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein. *Sci. Reports 2020 101* **10**, 1–12 (2020).
 8. Stadmiller, S. S., Aguilar, J. S., Waudby, C. A. & Pielak, G. J. Rapid Quantification of Protein-Ligand Binding via ¹⁹F NMR Lineshape Analysis. *Biophys. J.* **118**, 2537–2548 (2020).
 9. Sharaf, N. G., Ishima, R. & Gronenborn, A. M. Conformational plasticity of the NNRTI-binding pocket in HIV-1 reverse transcriptase: A fluorine nuclear magnetic resonance study. *Biochemistry* **55**, 3864–3873 (2016).
 10. Danielson, M. A. & Falke, J. J. USE OF ¹⁹F NMR TO PROBE PROTEIN STRUCTURE AND CONFORMATIONAL CHANGES. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163 (1996).
 11. Campos-Olivas, R., Aziz, R., Helms, G. L., Evans, J. N. S. & Gronenborn, A. M. Placement of ¹⁹F into the center of GB1: effects on structure and stability. *FEBS Lett.* **517**, 55–60 (2002).
 12. Peng, J. W. Cross-correlated ¹⁹F relaxation measurements for the study of fluorinated ligand-receptor interactions. *J. Magn. Reson.* **153**, 32–47 (2001).
 13. Marsh, E. N. G. & Suzuki, Y. Using ¹⁹F NMR to Probe Biological Interactions of Proteins and Peptides. *ACS Chem. Biol.* **9**, 1242–1250 (2014).
 14. Lu, M. *et al.* ¹⁹F Dynamic Nuclear Polarization at Fast Magic Angle Spinning for NMR of HIV-1 Capsid Protein Assemblies. *J. Am. Chem. Soc.* **141**, 5681–5691 (2019).
 15. Campomizzi, C. S., Ghanatios, G. E. & Estrada, D. F. ¹⁹F-NMR reveals substrate specificity of CYP121A1 in Mycobacterium tuberculosis. *J. Biol. Chem.* **297**, 101287 (2021).
 16. Didenko, T., Liu, J. J., Horst, R., Stevens, R. C. & Wüthrich, K. Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Curr. Opin. Struct. Biol.* **23**, 740–747 (2013).
 17. Loewen, M. C. *et al.* Solution ¹⁹F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4888–4892 (2001).
 18. Boeszoermenyi, A. *et al.* Aromatic ¹⁹F-¹³C TROSY: a background-free approach to probe biomolecular structure, function, and dynamics. *Nat. Methods* **16**, 333–340 (2019).
 19. Becette, O. B. *et al.* Solution NMR readily reveals distinct structural folds and interactions in doubly ¹³C- And ¹⁹F-labeled RNAs. *Sci. Adv.* **6**, (2020).
 20. Matei, E. & Gronenborn, A. M. ¹⁹F Paramagnetic Relaxation Enhancement: A Valuable Tool for Distance Measurements in Proteins. *Angew. Chemie* **128**, 158–162 (2016).
 21. Buchholz, C. R. & Pomerantz, W. C. K. ¹⁹F NMR viewed through two different lenses: Ligand-observed and protein-observed ¹⁹F NMR applications for fragment-based drug discovery. *RSC Chem. Biol.* **2**, 1312–1330 (2021).
 22. Dalvit, C. & Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **62**, 2218–2244 (2019).

23. Li, C. *et al.* Protein 19F NMR in Escherichia coli. *J. Am. Chem. Soc.* **132**, 321–327 (2010).
24. Ye, Y., Liu, X., Xu, G., Liu, M. & Li, C. Direct Observation of Ca²⁺-Induced Calmodulin Conformational Transitions in Intact *Xenopus laevis* Oocytes by 19 F NMR Spectroscopy. *Angew. Chemie* **127**, 5418–5420 (2015).
25. Veronesi, M. *et al.* Fluorine nuclear magnetic resonance-based assay in living mammalian cells. *Anal. Biochem.* **495**, 52–59 (2016).
26. Gerig, J. T. Fluorine NMR of proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 293–370 (1994).
27. Sun, X., Dyson, H. J. & Wright, P. E. Fluorotryptophan Incorporation Modulates the Structure and Stability of Transthyretin in a Site-Specific Manner. *Biochemistry* **56**, 5570–5581 (2017).
28. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of 19F-Tryptophan for protein NMR spectroscopy. *Chem. Commun.* **48**, 10681–10683 (2012).
29. Arntson, K. E. & Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J. Med. Chem.* **59**, 5158–5171 (2016).
30. Shu, Q. & Frieden, C. Urea-Dependent Unfolding of Murine Adenosine Deaminase: Sequential Destabilization As Measured by 19F NMR. *Biochemistry* **43**, 1432–1439 (2004).
31. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The Preparation of 19F-Labeled Proteins for NMR Studies. *Methods Enzymol.* **380**, 400–415 (2004).
32. Bogetti, A. T. *et al.* A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics. *J. Chem. Phys.* **153**, 64101 (2020).
33. Debiec, K. T., Gronenborn, A. M. & Chong, L. T. Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. *J. Phys. Chem. B* **118**, 6561–6569 (2014).
34. Lin, F. Y. *et al.* Further Optimization and Validation of the Classical Drude Polarizable Protein Force Field. *J. Chem. Theory Comput.* **16**, 3221–3239 (2020).
35. Shi, Y. *et al.* Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **9**, 4046–4063 (2013).
36. Koes, D. R. & Vries, J. K. Evaluating Amber force fields using computed NMR chemical shifts. *Proteins* **85**, 1944 (2017).
37. Götz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
38. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
39. Croitoru, A. *et al.* Additive CHARMM36 Force Field for Nonstandard Amino Acids. *J. Chem. Theory Comput.* **17**, 3554–3570 (2021).
40. Robalo, J. R. & Vila Verde, A. Unexpected trends in the hydrophobicity of fluorinated amino acids reflect competing changes in polarity and conformation. *Phys. Chem. Chem. Phys.* **21**, 2029–2038 (2019).
41. Robalo, J. R., Huhmann, S., Kocsch, B. & Vila Verde, A. The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids. *Chem* **3**, 881–897 (2017).
42. Wang, X. & Li, W. Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives. *Front. Mol. Biosci.* **7**, 350 (2020).
43. Tobola, F. *et al.* Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS Chem. Biol.* **13**, 2211–2219 (2018).
44. Panek, J. J., Ward, T. R., Jezierska, A. & Novic, M. Effects of tryptophan residue fluorination on streptavidin stability and biotin-streptavidin interactions via molecular

- dynamics simulations. *J. Mol. Model.* **15**, 257–266 (2009).
45. Rashid, S., Lee, B. L., Wajda, B. & Spyropoulos, L. Side-Chain Dynamics of the Trifluoroacetone Cysteine Derivative Characterized by ¹⁹F NMR Relaxation and Molecular Dynamics Simulations. *J. Phys. Chem. B* **123**, 3665–3671 (2019).
 46. Cerutti, D. S., Swope, W. C., Rice, J. E. & Case, D. A. Ff14ipq: A self-consistent force field for condensed-phase simulations of proteins. *J. Chem. Theory Comput.* **10**, 4515–4534 (2014).
 47. Cerutti, D. S., Rice, J. E., Swope, W. C. & Case, D. A. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **117**, 2328–2338 (2013).
 48. Takemura, K. & Kitao, A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J. Phys. Chem. B* **116**, 6279–6287 (2012).
 49. Hoffmann, F., Mulder, F. A. A. & Schäfer, L. V. Predicting NMR relaxation of proteins from molecular dynamics simulations with accurate methyl rotation barriers. *J. Chem. Phys.* **152**, 084102 (2020).
 50. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
 51. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **13**, e1005659 (2017).
 52. Shirts, M. R. *et al.* Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aided. Mol. Des.* **31**, 147–161 (2017).
 53. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).
 54. Dunning, J., Peterson, K. A. & Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **114**, 9244–9253 (2001).
 55. Woon, D. E. & Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **98**, 1358–1371 (1993).
 56. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
 57. Kendall, R. A., Dunning, T. H. & Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96**, 6796–6806 (1992).
 58. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).
 59. He, X., Man, V. H., Yang, W., Lee, T. S. & Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **153**, 114502 (2020).
 60. D. 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. Mermelste, and P. A. & Kollman. AMBER 18. (2018).
 61. Hanwell, M. D. *et al.* Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17 (2012).
 62. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
 63. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
 64. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

65. Joungh, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
66. Lu, M., Ishima, R., Polenova, T. & Gronenborn, A. M. 19F NMR relaxation studies of fluorosubstituted tryptophans. *J. Biomol. NMR* **73**, 401–409 (2019).
67. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multi dimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **16**, 1339–1 (1994).
68. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
69. Grossfield, A. WHAM: the weighted histogram analysis method.
70. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 234505 (1995).
71. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **23**, 321–341 (1977).
72. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
73. Feng, Z. *et al.* Ligand Depot: A data warehouse for ligands bound to macromolecules. *Bioinformatics* **20**, 2153–2155 (2004).
74. Goldman, M. Interference effects in the relaxation of a pair of unlike spin- 1 2 nuclei. *J. Magn. Reson.* **60**, 437–452 (1984).
75. Haeberlen, U. *High Resolution Nmr in Solids Selective Averaging. High Resolution Nmr in Solids Selective Averaging* (Elsevier, 1976). doi:10.1016/b978-0-12-025561-0.x5001-1.
76. Grace, R. C. R. & Kumar, A. Observation of Cross Correlations in a Weakly Coupled 19F-1H Four-Spin System. *J. Magn. Reson. Ser. A* **115**, 87–93 (1995).
77. Ottiger, M., Zerbe, O., Güntert, P. & Wüthrich, K. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* **272**, 64–81 (1997).
78. Lu, M. *et al.* 19F Magic Angle Spinning NMR Spectroscopy and Density Functional Theory Calculations of Fluorosubstituted Tryptophans: Integrating Experiment and Theory for Accurate Determination of Chemical Shift Tensors. *J. Phys. Chem. B* **122**, 6148–6155 (2018).
79. Lau, E. Y. & Gerig, J. T. Effects of fluorine substitution on the structure and dynamics of complexes of dihydrofolate reductase (*Escherichia coli*). *Biophys. J.* **73**, 1579–1592 (1997).
80. Ayala, I. *et al.* Hydrogen bonding in human manganese superoxide dismutase containing 3-fluorotyrosine. *Biophys. J.* **89**, 4171–4179 (2005).
81. Colgan, J., En, H., Yuan, H., Franke, E. K. & Luban, J. Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization. *J. Virol.* **70**, 4299 (1996).
82. Hayward, S. Peptide-plane flipping in proteins. *Protein Sci.* **10**, 2219 (2001).
83. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & Torchia, D. A. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson.* **97**, 359–375 (1992).
84. Hoffmann, F., Mulder, F. A. A. & Schäfer, L. V. Accurate Methyl Group Dynamics in Protein Simulations with AMBER Force Fields. *J. Phys. Chem. B* **122**, 5038–5048 (2018).
85. Wang, L. P. *et al.* Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **121**, 4023–4039 (2017).
86. Stoppelman, J. P., Ng, T. T., Nerenberg, P. S. & Wang, L. P. Development and Validation

- of AMBER-FB15-Compatible Force Field Parameters for Phosphorylated Amino Acids. *J. Phys. Chem. B* **125**, 11927–11942 (2021).
87. Unke, O. T. *et al.* Machine Learning Force Fields. *Chem. Rev.* **121**, 10142–10186 (2021).
 88. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
 89. Devereux, C. *et al.* Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
 90. Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R. & Miller, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153**, 124111 (2020).

Supplementary Figures

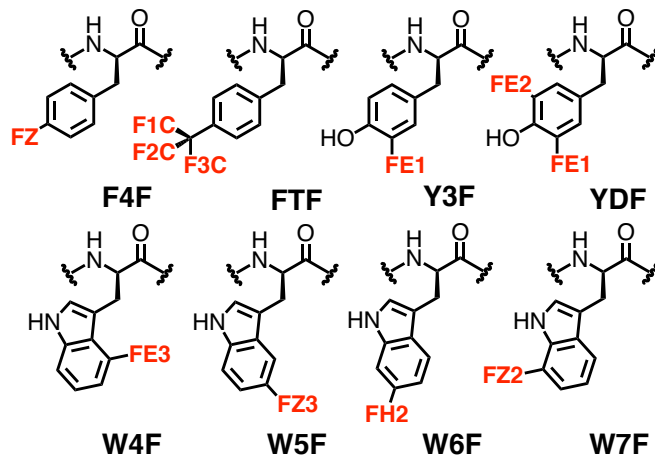


Figure S1. New fluorinated residue atom types presented in this work and their respective 3-letter identifier codes. The substituted fluorine atom-type is shown in red.

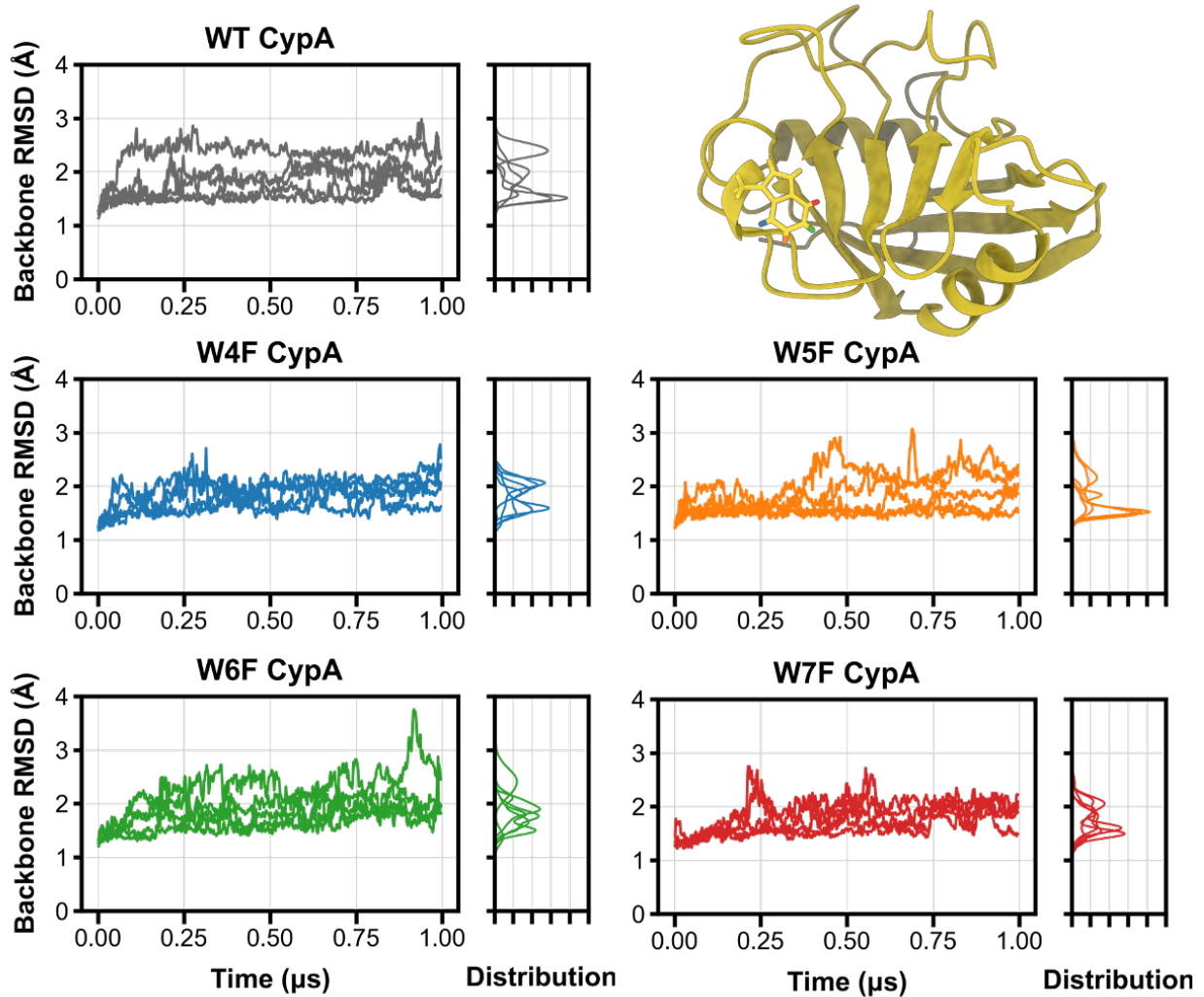


Figure S2. Backbone RMSD values with respect to the experimental X-ray structure (PDB: 3K0N)⁶³ coordinates for wild-type and fluorinated CypA proteins extracted from each of the five independent 1 μs simulation replicates. The distribution of RMSD values is shown for each protein. The protein X-ray structure is shown in ribbon representation in the upper-right with individual fluorine positions on the Trp121 indole ring color coded like the trajectory plots.

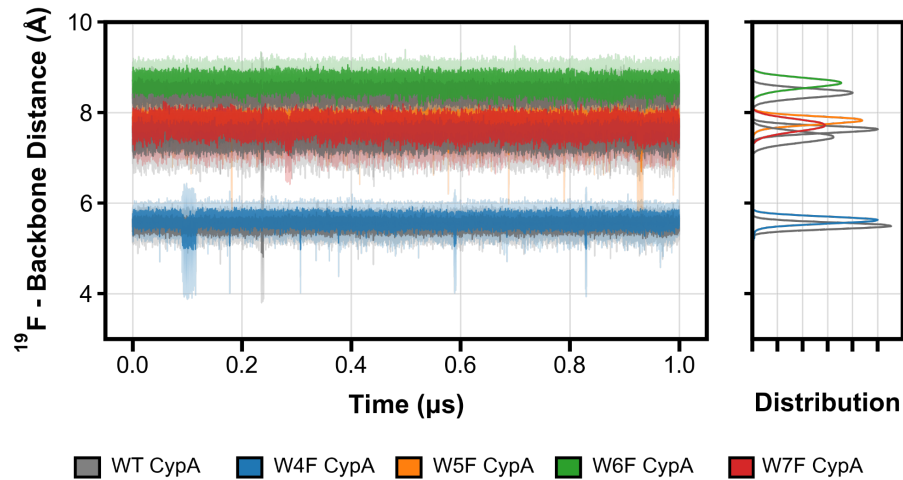


Figure S3. Average distance between the center of mass of the fluorine atom and the backbone carbonyl \pm the standard deviation over the simulation time course for the fluorinated CypA variants. The distance for the hydrogen atom and the backbone carbonyl for wild-type CypA is shown in grey. This data represents all five independent replicate simulations of 1 μs each for each system.

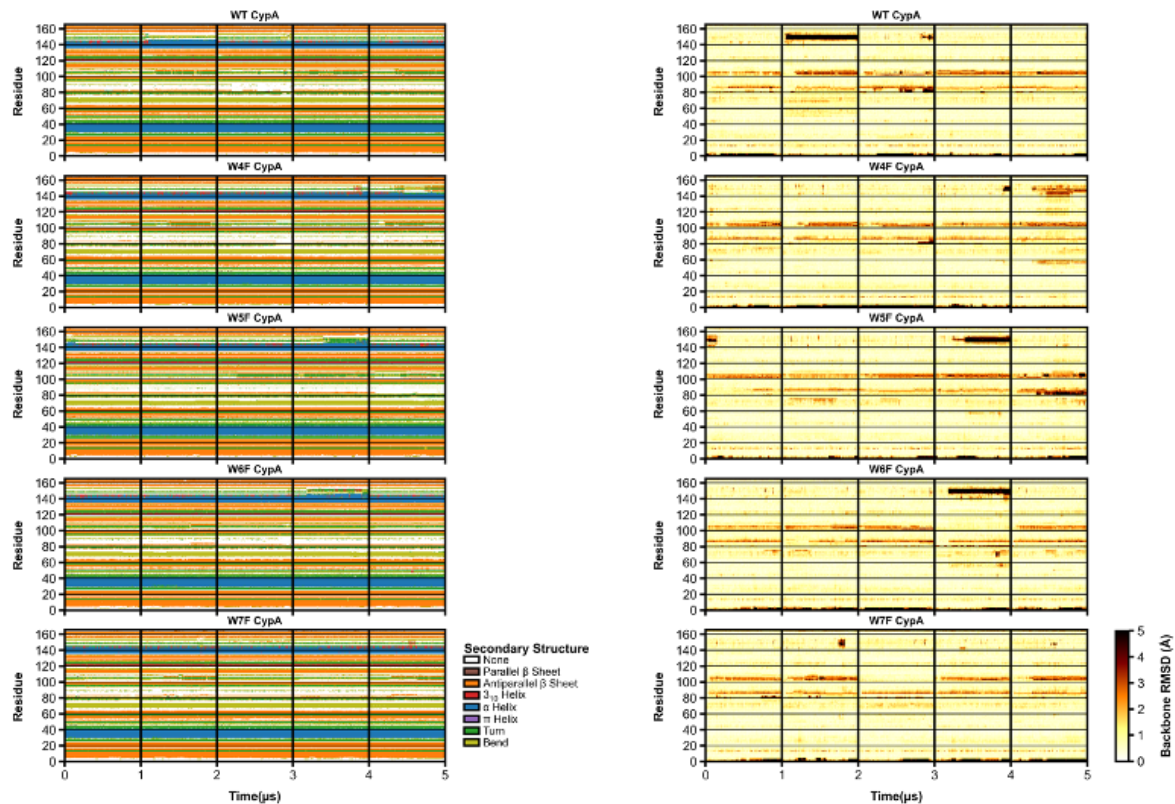


Figure S4. Secondary structure and per residue backbone RMSD values with respect to the crystal structure for both native CypA and the fluorinated variants. Each 1 μs simulation was

replicated five times (separated by vertical lines) for a discontinuous simulation time of 5 μs for each system.

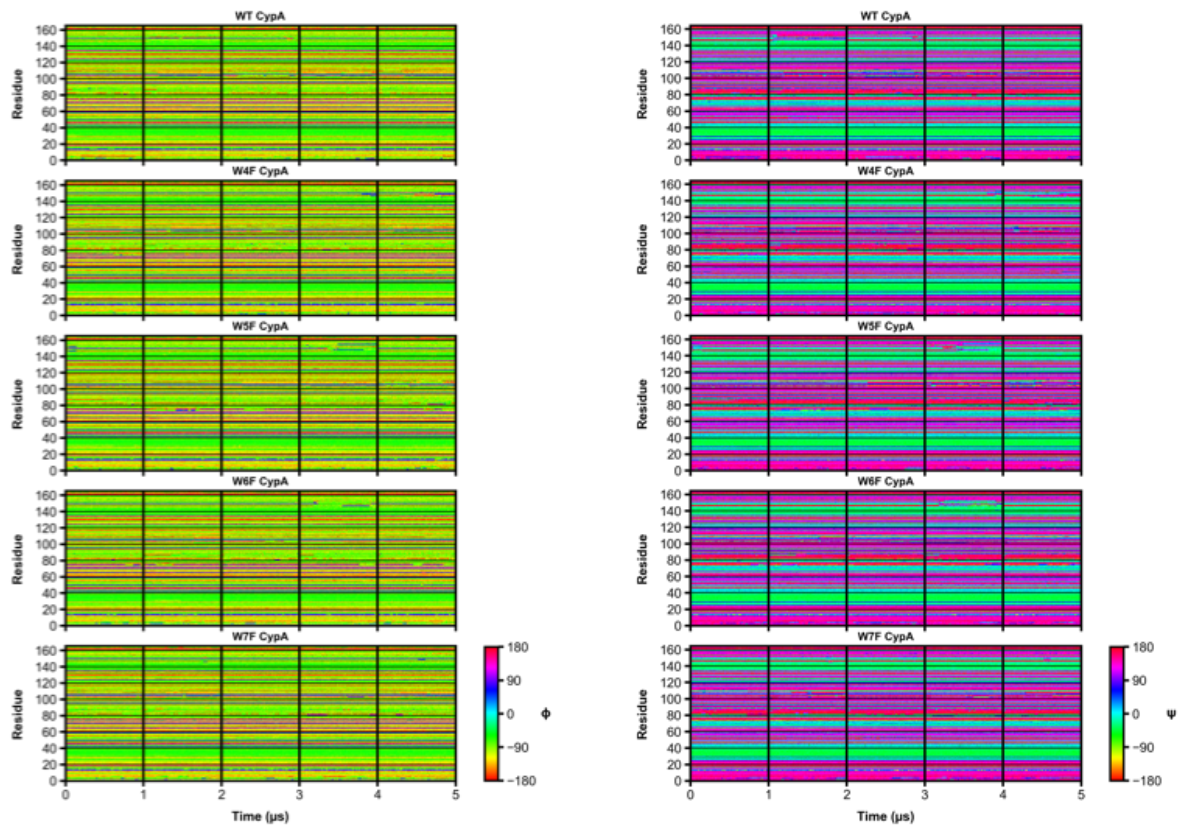


Figure S5. The per-residue Φ and Ψ dihedral angle variations for wild-type CypA and the fluorinated variants. Each 1 μs simulation was replicated five times (separated by vertical lines) for a discontinuous simulation time of 5 μs for each system.

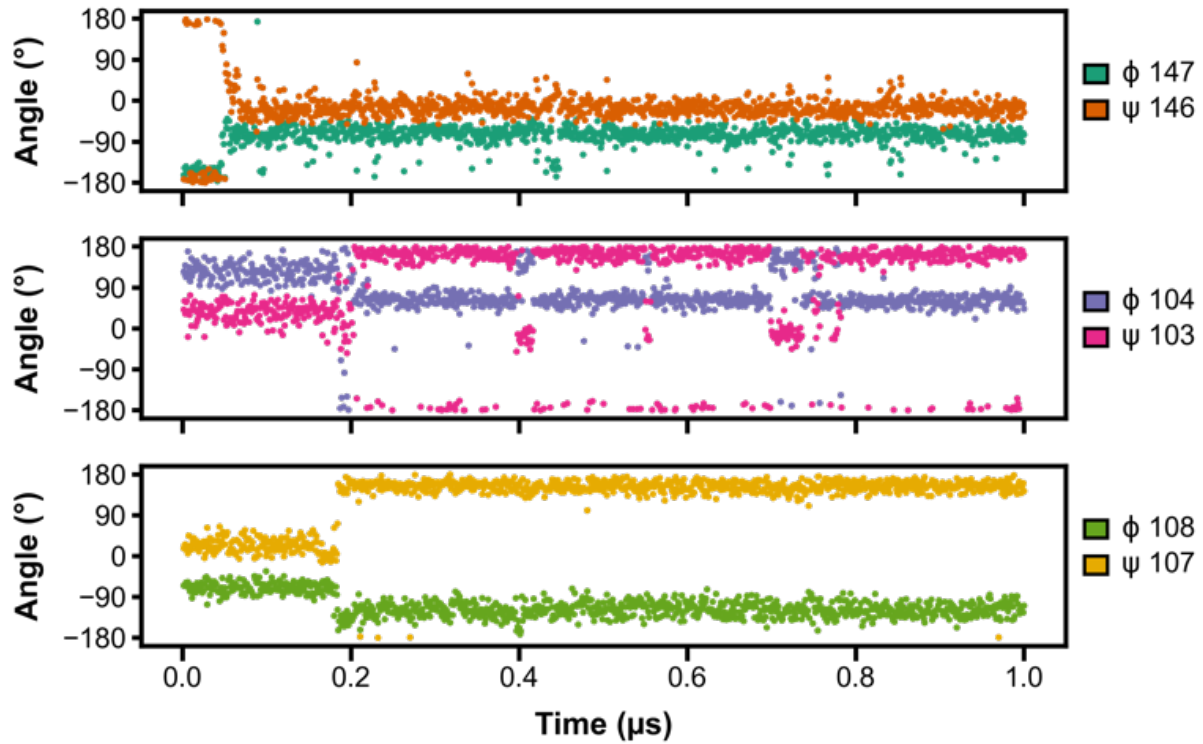


Figure S6. The Φ and Ψ dihedral angles of key neighboring residues around the amino acid exhibiting the jumps in backbone RMSD in **S4**, taken from a representative 1 μ s simulation of the non-fluorinated, wild-type CypA protein. Changes in adjacent dihedral angles are in opposite direction to the original jump, indicative of peptide-plane flips.

Supplementary Tables

Table S1. Bond, angle, and torsion terms for the current set of fluorinated amino acid parameters.

Bond	Force Constant (kcal/mol/Å ²)	Bond Length (Å)	
3C-CA	317.4204	1.5100	
3C-F	367.0233	1.3800	
Angle	Force Constant (kcal/mol/rad ²)	Angle (°)	
CB-CA-F	109.3865	133.44	
CN-CA-F	53.6158	118.76	
C-CA-F	55.5980	120.15	
3C-CA-CA	70.9193	119.54	
CA-3C-F	96.9340	111.90	
F-3C-F	98.1332	107.08	
Dihedral	Barrier Height (kcal/mol)	Phase Angle (°)	Periodicity
F-3C-CA-CA	3.83580	0.0	2.0

Table S2. Average \pm one standard deviation, minimum, and maximum $\Delta\Delta G$ values of the relative free energy difference of the tetrapeptide with a fluorine substituted residue, compared to the same peptide with a non-fluorinated amino acid.

Peptide	Avg $\Delta\Delta G$ (kcal/mol)	Min $\Delta\Delta G$ (kcal/mol)	Max $\Delta\Delta G$ (kcal/mol)
W4F	0.158 \pm 0.162	-0.312	0.637
W5F	0.089 \pm 0.137	-0.378	0.530
W6F	-0.007 \pm 0.139	-0.626	0.547
W7F	0.062 \pm 0.137	-0.484	0.474
Y3F	0.088 \pm 0.196	-0.457	0.651
YDF	0.030 \pm 0.228	-0.554	0.551
F4F	0.273 \pm 0.186	-0.367	0.881
FTF	0.203 \pm 0.268	-0.534	0.859

Table S3. Experimental NMR principal components of the CSA tensor, reduced anisotropy, and asymmetry parameters for fluoro-substituted tryptophan powders⁷⁸. These values were used to calculate the relaxation rates from MD simulations for each respective CypA variant.

	σ_{11} (ppm)	σ_{22} (ppm)	σ_{33} (ppm)	δ_{σ} (ppm)	η
4F-DL-Trp	11.2	-48.3	-112.8	62.8	0.9
5F-DL-Trp	4.8	-60.5	-86.1	52.1	0.5
6F-DL-Trp	12.9	-51.2	-91.6	56.2	0.7
7F-DL-Trp	4.6	-48.3	-123.3	67.6	0.8

Table S4. Comparison between the experimental ¹⁹F NMR relaxation rates and those extracted from MD simulations. Each value \pm one standard deviation is listed.

System	Experimental		Average $\pm \sigma$ from MD – All protons < 3Å	
	R ¹ (s ⁻¹)	R ² (s ⁻¹)	R ¹ (s ⁻¹)	R ² (s ⁻¹)
W4F <u>CypA</u>	1.99 \pm 0.05	109.1 \pm 4.9	1.89 \pm 0.55	121.3 \pm 4.94
W5F <u>CypA</u>	1.19 \pm 0.02	64.8 \pm 2.5	0.94 \pm 0.27	68.1 \pm 2.47
W6F <u>CypA</u>	1.25 \pm 0.01	63.0 \pm 3.9	0.97 \pm 0.27	84.1 \pm 2.47
W7F <u>CypA</u>	1.20 \pm 0.01	109.6 \pm 1.9	0.99 \pm 0.31	120.9 \pm 2.82