phosaa14SB and phosaa19SB: Updated Amber Force Field Parameters for Phosphorylated Amino Acids

Lauren E. Raguette[1,2], Abbigayle E. Cuomo[1], Kellon A.A. Belfon[1,2], Chuan Tian[1,2], Victoria Hazoglou[1], Gabriela Witek[3], Qin Wu[4], and Carlos Simmerling[*,1,2]*

1. Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States.

2. Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States.

3. Department of Pharmacological Sciences, Stony Brook University, Stony Brook, New York 11794, United States.

4. Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States.

ABSTRACT

Phosphorylated amino acids are involved in many cell regulatory networks; proteins containing these post-translational modifications are widely studied both experimentally and computationally. Simulations are used to investigate a wide range of structural and dynamic properties of biomolecules, such as ligand binding, enzyme-reaction mechanisms, and protein folding. However, the development of force field parameters for the simulation of proteins containing phosphorylated amino acids using the Amber program has not kept pace with the development of parameters for standard amino acids, and it is challenging to model these modified amino acids with accuracy comparable to proteins containing only standard amino acids. In particular, the popular ff14SB and ff19SB models do not contain parameters for phosphorylated amino acids. Here, the dihedral parameters for the side chains of the most common phosphorylated amino acids are trained against reference data from QM calculations adopting the ff14SB approach, followed by validation against experimental data. Library files and corresponding parameter files are provided, with versions that are compatible with both ff14SB and ff19SB.

INTRODUCTION

Phosphorylated proteins are the most common post-translational protein modification.[1] Coupled with dephosphorylation, these modifications are critical for regulating many cellular processes in biology, especially regarding intercellular communication and coordination of complex functions, as they activate (or deactivate) almost half of enzymes.[2] Most commonly occurring on serine, threonine, and tyrosine side chains, phosphorylation mediates signal transduction pathways for both eukaryotic and prokaryotic cells.[1] Phosphorylation forms a phosphoester linkage between the amino acid residue and the phosphate group at the site of the side chain hydroxyl group oxygen

for serine, threonine, and tyrosine.[3] Through a phosphoramidate bond, phosphorylation can also occur on histidine, which plays a role in a variety of processes, such as prokaryotic two-component signal transduction, bacterial carbohydrate transport systems, and as intermediates in metabolic pathways.[1, 4] Phosphorylation can either occur at the ND1 or NE2 site of the histidine, depending on other structural features of the molecule.[5] Understanding phosphorylated histidine is important, yet enigmatic, because the typical experimental procedures used to detect and preserve phosphoester amino acids can fail for phosphorylated histidine.[4] These limitations also can impose challenges for other phosphorylated amino acids.

Simulations can help fill gaps left by experiments, helping to better understand phosphorylation and dephosphorylation and their integral roles in activating or deactivating enzymes such as kinases and phosphatases. Many of these mechanisms and the associated ligand binding have the potential to be simulated with molecular dynamics (MD).[2-3] MD has been used in many studies involving phosphorylated amino acids, ranging from medicinal to agricultural research; these include governing molecular recognition,[6] altering G-protein signaling,[7] and causing conformational changes in protein loops and helices[8] with diverse applications, such as studying Alzheimer's disease[9] and breast cancer,[10-11] and increasing nutritional attributes of rice starch[12].

Amber[13] is a widely used package of molecular simulation programs, typically paired with Molecular Mechanics (MM) force fields for simulating biomolecules. Weiner et al.[14] created the first generation protein force field associated with Amber in 1984, followed by Cornell et al.'s[15] second generation force field a decade later (ff94). Cornell et al. created a new charge model using a 6-31G* basis set and restrained electrostatic potential (RESP) fitting.[16] Force fields continued to develop using this "fixed-charge" model, which assigns a single partial charge to each atom that includes implicit polarization. Using the same fixed-charge model helps to maintain compatibility

and transferability with other parameter sets, such as those that model nucleic acids or carbohydrates. Force fields continue to evolve as extensive testing and use reveals weaknesses. This is particularly true for dihedral parameters, which are employed to account for key orbital effects and other weaknesses in the model. Dihedral corrections alter the energy profile for rotation around a bond, and therefore need to be as accurate as possible and are frequently updated. For example, ff94 was succeeded by the ff99SB[17] force field developed by our group to improve protein backbone dihedral parameters. By training the protein backbone using tetrapeptide model systems rather than dipeptides, ff99SB remedied the over-stabilization of α-helices seen in previous force fields,[15] leading to widespread use.[17] In ff14SB,[18] we improved protein side chain behavior, as these dihedral values from ff99 (also used in ff99SB) were fit to small organic molecules rather than amino acids. ff14SB was trained against using quantum mechanics (QM) data using conformational grid scans of side chain rotamers for standard amino acids (with alternate protonation states) while improving transferability of side-chain parameters across alternate backbone conformations by including both alpha and beta peptide backbone conformations during training.[18] We later developed ff19SB,[21] which uses amino-acid specific 2D dihedral correction map profiles[22] (CMAPs) to improve sequence-specific behavior of protein simulations involving standard amino acids.

While some other force field lineages, such as CHARMM, have modern parameters[23] for phosphorylated amino acids, the most recent parameters for Amber[24] pre-date ff14SB (published by Homeyer et al.[24] in 2006 and referred to hereafter as the "HHLS" parameters). The HHLS parameters[24] were taken directly from the general Amber force field (gaff)[25] and the parm99 data set[15] when available, and the remainder were generated using QM calculations. HHLS includes

only generic dihedral parameters (where a bond between two specific atoms includes a wildcard on either end), whereas ff14SB-trained side-chain parameters use specific sets of four atom types[18].

Besides these libraries not employing the more recent optimization methods and training data, other complications arise from mixing these libraries with newer models for standard amino acids. In particular, ff14SB introduced new atom types, which improved accuracy by separating dihedral parameters at different locations where transferability was noted to be low. Especially relevant here, the protein backbone parameters were adjusted via a new CX atom type at the α-carbon. Older phosphorylated amino acid libraries lack the CX type and thus do not activate the use of the ff14SB backbone parameters; similar issues arise with ff19SB. This can lead to problems comparing simulations with a standard vs. a phosphorylated amino acid, since there are simultaneous changes in both the phosphorylation state and the φ/ψ parameters. Thus, new parameters that use training protocols and atom types consistent with ff14SB and ff19SB are needed.

PARAMETERIZATION GOALS AND STRATEGY

Specific dihedrals can increase the accuracy of the parameters, as generic dihedrals, such as those in HHLS, have limited periodicities and generally do not influence the relative energies of alternate rotamers. The overarching goal of this work is to extend the ff14SB treatment to the phosphorylated amino acids and provide updated, specific parameters, as well as library files that are compatible with ff14SB and ff19SB.

The dihedrals for the side chains of the most common phosphorylated amino acids (phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine**, Figure 1**) were parameterized. Both the singly protonated and unprotonated states were included for each amino

acid, as the equilibrium pK_a values of the phosphorylated amino acids are close to pH 7 **(Table S1)**.[26-28] Histidine can be phosphorylated at either ND1 or NE2, so parameters were trained for both isomers **(Table 1)**.[5]

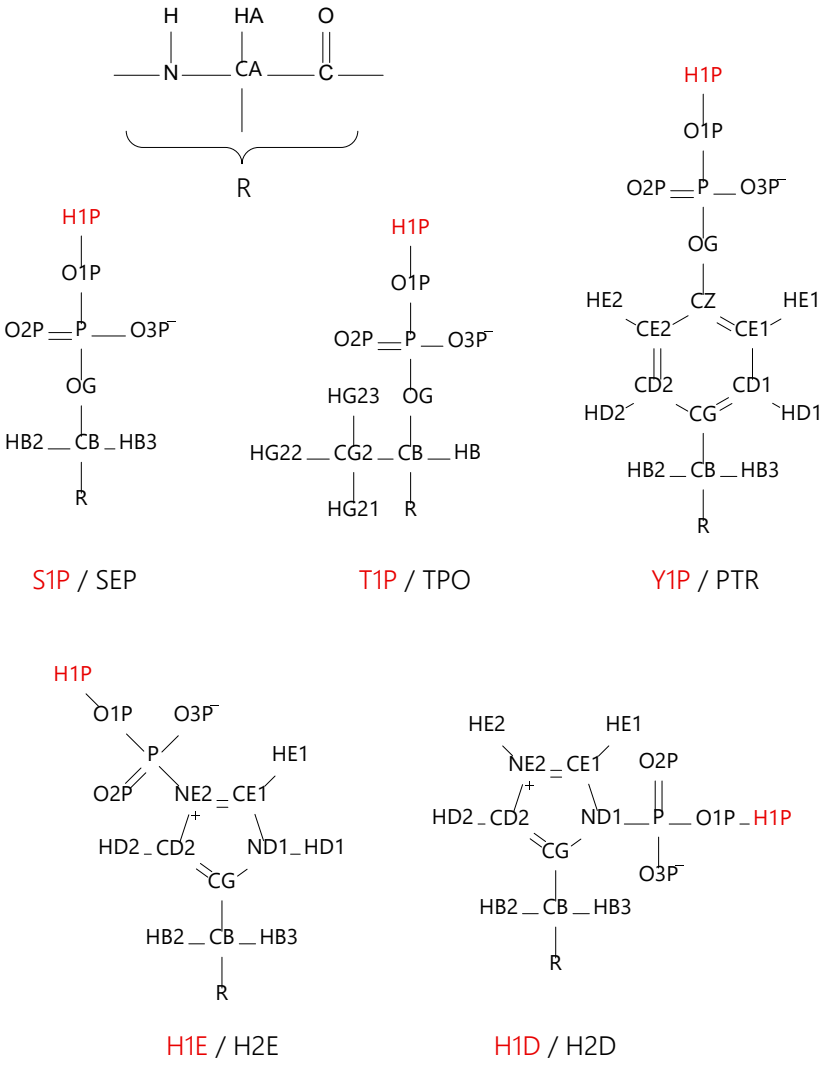

Figure 1. Phosphorylated amino acids for which parameters are reported here. Names for each atom and residue are indicated. The protonated phosphorylated amino acids have an additional hydrogen, shown in red. A corresponding figure with atom types is shown in **Figure S1**.

Table 1. Phosphorylated amino acids, charge states and residue names

	Charge of phosphate group	Total charge	Amber residue name
Phosphoserine	-1	-1	S1P
	-2	-2	SEP
Phosphothreonine	-1	-1	T1P
	-2	-2	TPO
Phosphotyrosine	-1	-1	Y1P
	-2	-2	PTR
Phosphohistidine (phosphorylated at ND1)	-1	0	H1D
	-2	-1	H2D
Phosphohistidine (phosphorylated at NE2)	-1	0	H1E
	-2	-1	H2E

An important goal of this work is to not only parameterize the dihedrals for common phosphorylated amino acids but also to maximize transferability. Transferability implies that force field parameters trained on one set of molecules will perform well on different (and often larger) molecules. Since the parameters between neighboring groups are not explicitly coupled, good transferability also implies that a set of parameters will work well in all situations where it is combined with neighboring groups with different chemistry and/or conformations than those that were present during training. Transferability can be improved by explicitly training parameters in the context of the diverse surroundings expected to be present during applications. In ff99SB, we

trained the protein backbone φ and ψ parameters in the context of a tetrapeptide that included neighboring amino acids rather than a single amino acid that lacks the context of a longer protein.[17] In ff14SB, we trained amino acid sidechain parameters in the context of several alternate backbone conformations to avoid having parameters that are accurate for the conformation used during training but not for another.[18] Where transferability is found to be low, new atom types can be added to reduce parameter sharing and allow for more specific parameterization, better reproducing the training data, and improved MM properties. In general, however, adding additional parameters should be avoided when possible due to the necessary increase in training data to avoid overfitting.

For the discussion below, it is important to distinguish between a rotatable bond and a dihedral. While a rotatable bond can be defined using two atoms that each have at least one additional atom to which they are bonded, a dihedral refers to a specific set of four sequentially bonded atoms (here, the "dihedral atom quartet") where the central two atoms in the quartet correspond to the bond being rotated. Since each atom in the central bond can have multiple additional bonded atoms, a single rotatable bond can have multiple dihedral atom quartets; thus, multiple dihedrals can contribute to the energy profile for rotation around that bond. Furthermore, each dihedral contribution typically uses a function with multiple terms and parameters (such as the truncated Fourier series used in many MM force fields where the parameters are the amplitudes and phases of a series of cosine terms with different periodicities).

The MM function calculates the energy for each dihedral, using the dihedral angle value and dihedral parameters for that dihedral atom quartet; the parameters are obtained from a library based on the atom types of the dihedral atom quartet (here, the "dihedral atom type quartet").

Importantly, these dihedral energies are only a correction to the intrinsic MM energy for this rotation, which may involve intramolecular nonbonded interactions between groups on either side of the bond or with other molecules such as solvent. Therefore, the dihedral energies are not the same as the bond rotational energy, and the dihedral amplitudes should not be regarded as rotational barrier heights. In some situations, the role of the dihedral terms is to shift or reduce the rotational barrier by being placed out of phase with the original barrier. Casually reducing dihedral amplitudes can lead to unexpected consequences such as raising barriers, shifting positions of minima, or changing the preferred rotamer.

Here, we retain the truncated Fourier series dihedral function employed in ff14SB for the sidechain rotamers of standard amino acids. As in ff14SB, each dihedral includes four cosine terms with integer periodicities of 1 to 4. Amplitudes and phases of each term were trained for each atom type quartet. Quartets including hydrogen bonded to carbon were not modified. To account for symmetry, the amplitudes of some periodicities were zeroed. For instance, the rotatable bonds connecting the standard side chain to the phosphorous in the deprotonated phosphate groups can use only three-fold dihedral corrections since there are duplicate sets of four atom types, each offset by 120°, and the energy profile must maintain that threefold symmetry.

Many of the atom type quartets present in phosphorylated amino acids are also present in standard amino acids. Any atom type quartets present in standard amino acids were not modified in this work and retain the same dihedral parameters as ff14SB; those not present in ff14SB were trained here **(Table 2)**. When multiple quartets were present for a given rotatable bond, the amplitudes of one of the atom type quartets were zeroed, and one was optimized. This avoids the instability that often arises in optimizing multiple quartets when many parameter combinations can give the same net energy for rotation around that bond. For example, the phosphohistidine

NA-P bond is the central rotatable bond in both the CC-NA-P-OP and CR-NA-P-OP dihedrals in H1D. Here, the amplitude of CR-NA-P-OP was zeroed, and CC-NA-P-OP was optimized. CC-NA-P-OP was chosen to be optimized instead of CR-NA-P-OP, as CR-NA-P-OP occurs in both protonation states of both histidines (phosphorylated at the ND1 and NE2 sites). This reduces the number of rotatable bonds the atom type quartet defines, improving specificity.

As in ff14SB, our training data include conformations with differences in multiple rotatable bonds to improve transferability; therefore, we simultaneously optimized all dihedral parameters in that amino acid since the energy differences between conformations cannot be assigned solely to a single rotatable bond.

Table 2. Information for the new dihedral parameters

Atom type quartets	Amino acids	Atom names	Periodicities	Fitting group
N-CX-CT-CG	Y1P	N-CA-CB-CG	4, 3, 2, 1	3
	PTR	N-CA-CB-CG		
N-CX-2C-OR	S1P	N-CA-CB-OG	4, 3, 2, 1	3
N-CX-3C-OR	T1P	N-CA-CB-OG	4, 3, 2, 1	3
CX-CT-CG-CA	Y1P	CA-CB-CG-CD1 CA-CB-CG-CD2	4, 3, 2, 1	3
	PTR	CA-CB-CG-CD1 CA-CB-CG-CD2		
OR-P-OQ-HO	S1P	OG-P-O1P-H1P	4, 3, 2, 1	3
	T1P	OG-P-O1P-H1P		
	Y1P	OG-P-O1P-H1P		

CA-C-OV-P	PTR	CE1-CZ-OH-P CE2-CZ-OH-P	4, 3, 2, 1	3
CX-2C-OR-P	S1P	CA-CB-OG-P	4, 3, 2, 1	3
CX-3C-OR-P	T1P	CA-CB-OG-P	4, 3, 2, 1	3
CA-C-OR-P	Y1P	CE1-CZ-OG-P CE2-CZ-OG-P	4, 3, 2, 1	3
C-OV-P-OT	PTR	CZ-OH-P-O1P CZ-OH-P-O2P CZ-OH-P-O3P	3	3
2C-OR-P-OP	S1P	CB-OG-P-O2P CB-OG-P-O3P	3	3
3C-OR-P-OP	T1P	CB-OG-P-O2P CB-OG-P-O3P	3	3
C-OR-P-OP	Y1P	CZ-OG-P-O2P CZ-OG-P-O3P	3	3
N-CX-2C-OZ	SEP	N-CA-CB-OG	4, 3, 2, 1	1
CX-2C-OZ-P	SEP	CA-CB-OG-P	4, 3, 2, 1	1
2C-OZ-P-OX	SEP	CB-OG-P-O1P CB-OG-P-O2P CB-OG-P-O3P	3	1
N-CX-3C-OZ	TPO	N-CA-CB-OG1	4, 3, 2, 1	2
CX-3C-OZ-P	TPO	CA-CB-OG1-P	4, 3, 2, 1	2
3C-OZ-P-OX	TPO	CB-OG1-P-O1P CB-OG1-P-O2P CB-OG1-P-O3P	3	2
CC-NA-P-OP	H1D	CG-ND1-P-O2P	3	4

		CG-ND1-P-O3P		
	H2D	CG-ND1-P-O1P		
		CG-ND1-P-O2P		
		CG-ND1-P-O3P		
CW-NA-P-OP	H1E	CD2-NE2-P-O2P	3	4
		CD2-NE2-P-O3P		
	H2E	CD2-NE2-P-O1P		
		CD2-NE2-P-O2P		
		CD2-NE2-P-O3P		
NA-P-OQ-HO	H1D	ND1-P-O1P-H1P	4, 3, 2, 1	4
	H1E	NE2-P-O1P-H1P		

Some atom type quartets are present in multiple amino acids, requiring simultaneous optimization of all amino acids that share these dihedral parameters. This improves transferability compared to training on a single amino acid. As with ff14SB, this is achieved here using "fitting groups," which include all amino acids that overlap in any of their atom type quartets; different fitting groups share no quartets and thus share no parameters. Based on the dihedral definitions, four fitting groups for the phosphorylated amino acids were required (**Table 2**).

Methods

Generation of diverse conformations for parameter optimization

To minimize backbone-dependence of side-chain parameters and maximize transferability, the ff14SB protocol was followed. For both alpha (φ = -60°, ψ = -45°) and beta (φ = -135°, ψ =

135°) backbone conformations, acetyl and *N*-methyl capped dipeptides of each phosphorylated amino acid were built using LEaP using their assigned atom types **(Figure 2**). Further details of protocols are available in the **Supporting Information**.

Diverse sets of side chain conformations were generated using two different methods, depending on the number of rotatable bonds to be optimized. Grid scans were used for side chains with less than three rotatable bonds to be optimized. The unprotonated phosphohistidines (H2D and H2E) only have one rotatable bond to be optimized, so grid scans were generated every 10° for that rotatable bond. For the singly protonated phosphohistidines (H1D and H1E), which have two rotatable bonds to be optimized, grid scans were generated in two dimensions at every 20° for each rotatable bond. Due to computational cost, 20° rather than 10° was used here to minimize the number of total structures. All structures from grid scans were saved for parameter optimization.

Phosphorylated serine, threonine, and tyrosine required optimization of more than two rotatable bonds. To minimize computational cost, we followed the ff14SB protocol of generating conformational diversity using high temperature MD simulations rather than grid scans on the amino acids with a larger number of rotatable bonds requiring optimization. Details of the high temperature simulations are provided in the **Supporting Information.**

The high-temperature simulations generate a large set of structures that is not uniform across dihedral space. A sparse grid approach was used to extract a smaller, more uniformly distributed subset of structures for training. For each phosphorylated amino acid (and each protonation state and backbone conformation), each side chain dihedral values in each structure were mapped onto a multidimensional grid spaced at 10° (where the number of dimensions is the number of rotatable bonds being optimized).

In ff14SB, the five lowest potential energy structures for each grid point were saved, and 500 of these structures were extracted randomly for parameter optimization.[18] However, when this protocol was followed for the phosphorylated amino acids, there were large gaps in the grid sampling for some dihedrals **(Figure S2)**, where no structures were chosen, although structures existed in these gaps before the random selection. Many grid points in these gaps had less than five structures, so they were underrepresented using the ff14SB approach. The protocol was thus altered to select only the single lowest potential energy structure for each grid point, and randomly choose 500 of these structures for optimization. This approach improved the sampling of the sparse areas **(Figure S3)**. Although there are still some gaps in the grids, forcing structures to adopt dihedral values in the gap regions resulted in large steric clashes (not shown).

To minimize distortions in the structures generated from the high temperature simulations, we performed MM relaxation. To reconcile the minor differences in the backbone conformations induced by the high temperature simulations, a single set of backbone dihedral values was enforced for each alpha and beta backbone set. Following the ff14SB[18] protocol, each structure was minimized for 100,000 cycles, with the backbone dihedrals restrained to the average value over all structures in that set. Restraints were also placed on every rotatable bond being optimized. Further details can be found in the **Supporting Information.**

QM Optimization

Next, the relaxed high-temperature structures and the structures generated from the 1-D and 2-D MM grid scans were optimized with QM. Orca 3.1[29] was used with default options except for the specification of TightOpt convergence criteria. Rotatable bonds being optimized were constrained to maintain the initial dihedral values. These constraints were placed on all the

backbone dihedrals and the side chain dihedrals being optimized. The QM geometries were optimized at HF/6-31G*[30], and single point energies were calculated with MP2/6-31+G**[31]. Selected structures for SEP and H2D were optimized using MP2/6-31+G**; differences from the HF/6-31G* were noted to range from 0.01 - 0.05 Å in bond lengths and 2 - 3 ° in bond angles, suggesting that the lower-level theory was sufficient for optimizing the larger set.

Filtering structures

During the development of ff14SB, we noted that structures from the high-temperature simulations could have unfavorable steric and/or very strong electrostatic interactions. MM may not model these interactions accurately. If these structures were used during parameter optimization, the dihedral parameters could end up correcting for the inaccurate nonbonded function, and these corrections would then apply to all structures even when the strong nonbonded interactions are not present. To avoid this transferability failure, ff14SB eliminated training structures with atoms that did not share a bond or angle yet adopted a distance less than the sum of their vdW radii divided by 1.3. Maier et al.[18] chose an additional scaling factor to remove structures where the unsigned Coulombic energy between a side chain atom and another atom exceeded 42 kcal/mol. Here, these cutoffs did not always leave sufficient structures for training. Instead, individual cutoffs for Coulomb and Lennard-Jones energies were chosen for each phosphorylated amino acid that underwent high temperature simulations, guided by the ff14SB protocol, while ensuring a reasonable number of structures remained **(Table S2)**.

MM Reoptimization

Maier et al. observed that the fundamental differences between QM and MM modeling of bonded and nonbonded interactions are likely worsened when MM energies are calculated from QM-optimized structures without MM reoptimization[18]. For example, optimal bond lengths differ between the models, and calculating MM energy for a QM-optimized structure could lead to large bond potential energies in MM that can contaminate the training data. Therefore, the QM-optimized structures were subjected to MM minimization, with the amplitudes set to zero for all dihedrals being optimized ("MM0" parameters). Minimization was performed in the gas phase for 1000 cycles or until the RMS gradient was less than 1.0×10^{-4} kcal mol^{-1} Å^{-1}. A nonbonded cutoff of 99.0 Å was used (no cutoff). All backbone dihedrals and one sidechain dihedral per rotatable bond were restrained to the values obtained from the QM optimization, with a force constant of 2.5×10^{5} kcal mol^{-1} rad^{-2}. MM energies for each conformation were obtained from the final step of the minimization.

Parameter Optimization

Parameter optimization followed the general protocol used for parameter optimization in ff14SB and used the relative energies between all pairs of conformations rather than absolute energies as a target.[18] This pairwise calculation is done to avoid using an arbitrary reference structure, whose choice can bias the resulting parameters.[17] Only pairs of the same amino acid with the same backbone conformation were included. A GPU-based genetic algorithm program, RAGTAG[32], was employed to optimize and report dihedral parameters that minimize the relative energy errors between MM and QM for all pairs of input conformations. Each chromosome is a full set of parameters (the amplitude and phase for each periodicity in the truncated Fourier series for each

atom type quartet). The relative energy error (REE) between a single pair of conformations *i* and *j* is defined as:

$$\mathrm{REE(i,j)} = (\mathrm{E_{QM,i}} - \mathrm{E_{QM,j}}) - (\mathrm{E_{MM,i}} - \mathrm{E_{MM,j}}) \qquad (1)$$

where E_{QM} and E_{MM} are the quantum and molecular mechanics energies of conformations i and j.

The average absolute error (AAE) is calculated as the error in pairwise conformational energies using MM vs. the energies for the same pair calculated using QM, averaged over conformation pairs. The AAE is defined as:

$$\mathrm{AAE} = \frac{2}{\mathrm{N(N-1)}} \sum_{\mathrm{i}} \sum_{\mathrm{j<i}} |\mathrm{REE(i,j)}| \qquad (2)$$

where N is the number of conformations.

A fitting group in RAGTAG may include multiple backbone conformations, amino acids, or protonation states, but the AAE includes only conformation pairs for which these properties match. The objective function, O, represents the error averaged over all backbone conformations and amino acids/protonation states, and is defined as:

$$\mathrm{O} = \frac{1}{\mathrm{n_{profiles}}} \sum_{\mathrm{r=1}}^{\text{amino acids}} \sum_{\mathrm{bb}=\alpha,\beta} \mathrm{AAE_{r,bb}} \qquad (3)$$

where $n_{profiles}$ is the number of AAE profiles in that fitting group.

RAGTAG is given the list of atom type quartets for which parameters are being optimized, and for each, the set of periodicities for which amplitudes and phases should be optimized. For each structure, RAGTAG is given the QM and MM energies and the values of each dihedral angle. The objective function is minimized using evolution, and the chromosome with the lowest score is selected as the final parameter set. An Amber frcmod file containing the optimized dihedral parameters is created upon completion, along with data that quantifies the accuracy of the optimization; these include the final scores, the relative energy error between each pair of conformations, and the overall error in the optimization for each dataset.

RAGTAG was run separately for each of the four fitting groups. For each, the population size was 2000 chromosomes, and each population evolved for 1000 generations with a probability of mutation of 0.002 and a crossover rate of 0.8.

Generating Test Set Structures Not Included in Training Data

An initial test of transferability is to ensure that the trained parameters reproduce the pairwise relative energies for amino acid structures not included in the training set. One hundred structures were generated for each backbone and protonation state. As with the structures in the training data, phosphorylated amino acids with one or two rotatable bonds to be optimized were generated using a QM grid scan, with dihedral values offset by 5° compared to the structures in the training data. For phosphorylated amino acids with more than two rotatable bonds to be optimized, an independent set of 100 structures was chosen randomly from the high temperature simulations using the same grid approach. All structures followed the same optimization and energy calculation protocols as the structures in the training data.

Generating libraries for use with ff19SB

Although parameterization generally followed the side chain parameterization protocol from ff14SB, these parameters are compatible with ff19SB as well since ff19SB adopted the side chain parameters directly from ff14SB. In ff19SB, standard amino acids have a CMAP term that replaces the ff14SB backbone.[21] Leucine was considered to be a better model than alanine, as all amino acids except alanine and glycine have a γ carbon. In ff19SB, histidine and tyrosine use the leucine CMAP; phosphorylated histidine and tyrosine were also assigned the leucine CMAP. In ff19SB,

serine and threonine each have their own CMAPs, which were used for phosphorylated serine and threonine.

Model Systems for Testing Parameter Accuracy

Several systems, including one for each phosphorylated amino acid, were selected to test the parameters in proteins **(Table 3)**. Analyses included comparing structural properties, such as the stability of secondary structures and hydrogen bonding that were observed in the experimental structure.

Table 3. Protein test systems

Protein	***PDB ID***	***Amber Residue Name***
Aspergillus oryzae cutinase	3QPD[33]	SEP
Aurora A Kinase	5DT3[34]	TPO
XLP Protein SAP	1D4W[35]	PTR
Mannose-specific phosphotransferase enzyme IIA component	1EUD[36]	H2E

Each system was simulated under three different conditions: once using the HHLS parameters[24] and ff14SB, once with phosaa14SB and ff14SB, and once with phosaa19SB and ff19SB. To make a fair comparison, some atom types in the HHLS parameters were modified in order to activate ff14SB backbone parameters. These modified library and parameter files are provided in the **Supporting Information.** Each simulation followed the equilibration and production protocols

provided in **Supporting Information.** Production runs were carried out in duplicate to estimate precision and reproducibility.

CPPTRAJ[37] was used to calculate dihedral angles, assign secondary structures, and evaluate hydrogen bonding and other key interactions in each simulation. A distance cutoff of 3.0Å was used for hydrogen bonds (acceptor to donor heavy atom), and a distance cutoff of 4.0Å between N-O atom pairs was used for salt bridges. VMD[38] was used for visual analysis and image generation.

RESULTS

We parameterized the side chain dihedrals for the most common phosphorylated amino acids (phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine), each at two different protonation states. After comparing the partial charges for atoms sharing the same atom type, a new atom type CG was created, representing Cγ in both the singly protonated and unprotonated phosphotyrosine, as described in **Supporting Information**.

Examining all rotatable bonds without parameters in ff14SB provided 48 dihedral atom quartets, but only 22 unique atom type quartets to be trained, since multiple dihedrals share atom type quartets (**Table 2**).

Parameter training

Training was carried out using acetyl and *N*-methyl amide capped dipeptides. Structures were built in both alpha and beta backbone conformations to minimize backbone dependence of side chain parameters in both methods. As with ff14SB, two different methods were used to generate diverse conformations of training structures. Grid scans were used to generate side chain variety

in phosphorylated amino acids with less than three rotatable bonds to be optimized. Otherwise, high temperature MD simulations generated conformational diversity. To relax distortions in these structures, we performed a relaxation with MM. Structures were filtered to remove those with unfavorable steric and/or very strong electrostatic interactions. Geometries of the structures were optimized with QM at HF/6-31G*, and single point energies were calculated with MP2/6-31+G**[31]. Additional details are provided in Methods.

RAGTAG,[32] a GPU-based genetic algorithm program based on the protocol employed for training ff14SB, was employed to optimize and report the dihedral parameters. The phosphorylated amino acids were split between four different fitting groups, ensuring that no dihedral parameters were shared between fitting groups (**Table 2**).

The average AAE (objective function) for the phosaa14SB parameters over all fitting groups is 0.98 kcal/mol (**Table 4**). This is comparable to the score of 0.98 kcal/mol obtained during the optimization of dihedral parameters for standard side chains in ff14SB, indicating that amino acid phosphorylation did not pose an additional challenge to the training protocol or MM functional form. A separate score for phosaa19SB was not calculated, as it uses the same side-chain parameters as phosaa14SB. This score reflects a significant improvement over the HHLS parameters, which produce an average AAE of 2.55 kcal/mol. However, it is important to note that the energies of these structures were used to train the new parameters but were not part of the HHLS training data.

Table 4. AAE calculated for individual amino acids, calculated using the new parameters and HHLS parameters[24] for structures in the training set and alternate conformations

		Training Set		***Test Set***	
Amber Residue Name	Backbone	phosaa14SB (kcal/mol)	HHLS Parameters (kcal/mol)	phosaa14SB (kcal/mol)	HHLS Parameters (kcal/mol)
S1P	Alpha	0.94	1.18	1.15	1.21
	Beta	1.41	2.54	1.91	2.51
SEP	Alpha	0.07	0.70	0.48	0.79
	Beta	0.19	2.82	0.65	2.66
T1P	Alpha	0.89	1.50	1.22	1.57
	Beta	1.31	2.05	1.82	1.94
TPO	Alpha	0.59	4.67	1.68	4.69
	Beta	0.46	3.56	1.51	3.47
Y1P	Alpha	1.97	1.61	1.94	1.69
	Beta	1.40	2.76	1.67	2.83
PTR	Alpha	1.85	3.37	2.13	3.48
	Beta	2.10	5.16	2.57	4.98
H1D	Alpha	1.49	1.92	1.66	1.99
	Beta	1.59	2.33	1.88	2.25
H2D	Alpha	0.70	3.50	1.41	3.51
	Beta	0.60	3.71	1.46	3.64
H1E	Alpha	0.96	1.20	1.10	1.28
	Beta	0.95	1.64	1.14	1.62
H2E	Alpha	0.11	2.34	0.46	2.38
	Beta	0.12	2.44	0.49	2.47
Average		0.98	2.55	1.42	2.55

We next compared objective functions (average AAE) across different fitting groups. Outliers can indicate whether a particular group contains amino acids that were difficult to model using shared parameters, suggesting that dividing them via new atom types could improve the fit. As expected, the fitting groups with only one protonation state of one phosphorylated amino acid result in the lowest scores since the parameters can be highly specific (Fitting Group 1: SEP with an AAE of 0.13 kcal/mol and Fitting Group 2: TPO with an AAE of 0.53 kcal/mol). These fitting groups also show the most improvement over the HHLS parameters. Fitting Group 4 is comprised of four different residues (H1D, H2D, H1E, and H2E), but phosaa14SB still provides a good average AAE of 0.82 kcal/mol. Fitting Group 3, comprised of S1P, T1P, Y1P, and PTR, has the highest average AAE of 1.48 kcal/mol. Furthermore, Y1P with an alpha backbone is also the only case where HHLS provides a lower AAE than phosaa14SB. These results suggest that Fitting Group 3 may share parameters too broadly. Although Groups 3 and 4 both include four residues, Group 4 includes only variations of histidine, while Group 3 includes dihedral parameters shared across phosphorylated versions of serine, threonine, and tyrosine. This suggests that the higher AAE for Group 3 might reflect poor transferability of parameters across the group. Looking at the AAE values for individual residues indicates reduced accuracy for singly protonated and unprotonated phosphotyrosine compared to the other amino acids (Y1P, PTR; **Table 4**). However, PTR shares no dihedral parameters with S1P or T1P, and PTR and Y1P share only parameters for $\chi 1$ and $\chi 2$, and these employ the newly-created CG atom type not shared with other amino acids. This minimal overlap in parameters in the fitting group suggests that the reduced accuracy for PTR and Y1P arises from sources other than dihedral parameter transferability, and may reflect other issues such as the lack of explicit aromatic ring polarization with repositioning of the phosphate.

In every case other than the alpha backbone of Y1P, the AAE score for phosaa14SB is lower than that of HHLS. Overall, the analysis suggests that force field accuracy improves with phosaa14SB for most amino acids, and that phosphotyrosine might be improved with a more complex treatment than explored here.

Evaluating structures not included in the training set

We evaluated the performance of parameters on amino acid conformations not included in the training data (**Table 4**). 100 additional structures for each residue in each backbone conformation were selected and optimized as described in *Methods*. The average AAE score for these alternate conformations using phosaa14SB is 1.42 kcal/mol, somewhat higher than the 0.98 kcal/mol for the training conformations. This continues to be an improvement over the average AAE score using the HHLS parameters, which is 2.55 kcal/mol**.** The only instance of the HHLS parameters resulting in a lower score than phosaa14SB is Y1P with an alpha backbone, consistent with the results from the training set.

Protein simulations

The ultimate test of the force field parameters is evaluating their performance in protein simulations. This evaluates the transferability of the parameters from isolated amino acids in the gas phase to the context of protein chains in water. Simulation accuracy was evaluated by analyzing structural stability, specifically focusing on maintaining the rotamers and local interactions observed in the experimental structure. The systems selected **(Table 3)** contain varying phosphorylated amino acids. Each system was built with the HHLS, phosaa14SB, and

phosaa19SB parameters for phosphorylated proteins. An equilibration protocol detailed in **Methods** was followed before duplicate 4 μs production MD.

Phosphoserine: Aspergillus oryzae cutinase

Cutin, a waxy substance found in plant cell walls, can be hydrolyzed by cutinases. *Aspergillus oryzae* cutinase is traditionally used in the fermentation and production of rice wine, soy sauce, and soybean paste.[39] Phosphorylation can occur at Ser101, as seen in PDB 3QPD[33]. In this structure, Ser101 is at the end of a 12-residue α-helix and points into the active site. The phosphate oxygens form hydrogen bonds with the NE2 atom of His169, and with the peptide nitrogen and side chain hydroxyl group of Ser23. We evaluated the ability of simulations using different parameters to maintain these interactions.

Following simulations, CPPTRAJ[37] was used to evaluate the presence of these hydrogen bonds in the structure, with the structure shown in **Figure S4** and simulation results in **Table S4**. Precision estimates reflect half the difference between the results of the duplicate runs. With phosaa14SB, phosaa19SB, and HHLS parameters, the hydrogen bonds between the phosphate oxygens and the NE2 of His169 and the hydroxyl group of Ser23 were all present during 100% of the trajectory (with exact data and uncertainties in **Table S4**). The interaction between the peptide nitrogen of Ser23 and the phosphate oxygens was observed 99.46% ± 0.05%, 96.7% ± 0.3%, and 96% ± 2% of the time with phosaa14SB, phosaa19SB, and HHLS parameters, respectively.

A more direct measure of the quality of the new dihedral parameters is evaluating their ability to maintain the rotamer present in the experimentally determined structure. For all three parameter sets, the dihedral angles for the parameterized atom type quartets, N-CX-2C-OZ (atom names N-CA-CB-OG), CX-2C-OZ-P (atom names CA-CB-OG-P), and 2C-OZ-P-OX (atom

names CB-OG-P-O1P, CB-OG-P-O2P, and CB-OG-P-O2P), remained stable at the crystallographic values throughout all of the simulations, as shown in **Figure S5**. With HHLS, the phosphate group rotates more frequently throughout the simulation (indicating a reduced bond rotation barrier), but the three peak values remain consistent with the crystal structure. For CX-2C-OZ-P, which corresponds to atom names CA-CB-OG-P, the range of angles is shifted modestly from the crystal structure value. However, the dihedral angles for the rotatable bonds on either side remain consistent with the crystal structure values. Together, these observations indicate that the HHLS and revised parameters all provide reasonably stable simulations of this system.

Phosphotyrosine: XLP Protein SAP

SAP is the product of the gene mutated in X-linked lymphoproliferative syndrome (XLP). With this mutation, the body cannot properly regulate the number of lymphocytes, which leads to the enlargement of the lymph nodes, liver, and spleen. SAP adopts a single SH2 domain that binds the cytoplasmic tail of the lymphocyte coreceptor SLAM, and SLAM can be phosphorylated at Tyr281. Structures have been solved for SAP bound to a peptide corresponding to the SLAM tail when phosphorylated (1D4W) or unphosphorylated (1D4T).[35] SAP binds a site encompassing Tyr281 in SLAM, irrespective of the phosphorylation state of this site in experiment.[35] At residues 279-281, the SLAM peptide backbone forms a short parallel β-sheet interaction with residues 51-53 in the βD strand of the SAP domain.[35]

In the crystal structure, the phosphate group of Ptr281 forms hydrogen bonds with the side chains of Arg32, Ser34, Ser36, and Cys42. The phosphate group also hydrogen bonds with the peptide nitrogen of Glu35. The aromatic ring of Ptr281 engages in a cation-pi stacking interaction with the guanidinium group of Arg55 **(Figure S6)**.

The side chain rotamer for Ptr281, as well as the presence of the β-sheet, cation-pi interaction, salt bridge, and hydrogen bonds in the simulations, were evaluated and compared between parameter sets. Most interactions between the Ptr281 side chain and the SAP protein are stable with all parameter sets **(Table S5)**. All maintained 100% population for the salt bridge between the phosphate group and Arg32 and the hydrogen bond with Ser34 in their duplicate runs. The Ser36 hydroxyl interacted with the phosphate group 93.5% ± 0.1%, 81% ± 1%, and 91% ± 1% of the time with phosaa14SB, phosaa19SB, and HHLS parameters, respectively. A hydrogen bond between the phosphate and the sulfur of Cys42 was stable in all simulations (95.5-99.7%).

In some cases, reduced stability of the phosphate interactions was noted. A hydrogen bond between the Glu35 backbone nitrogen and the phosphate group was present 99% of the time in the simulations using phosaa14SB (± 0.6%) and HHLS (± 1%) parameters, but only 55 ± 6% of the time with phosaa19SB. Both Glu35 and Ser36 are in a flexible loop and do not hydrogen bond with the phosphate oxygens simultaneously with phosaa19SB as often as phosaa14SB and HHLS. The cation-pi interaction with Arg55 was observed to be more stable in simulations with the new parameters. It was present in the simulations with phosaa14SB and phosaa19SB 98.5 ± 0.3% and 97 ± 1% of the time, respectively, but only 65 ± 6% of the simulation with the HHLS parameters. This is likely due to the deviation in the dihedral angle of atom names N-CA-CB-CG (**Figure S7**), which turns the Ptr ring away from Arg55, as discussed below.

The stability of the parallel β-sheet interaction with the peptide and the βD strand of the protein (**Figure S8)** was quantified by calculating secondary structure content with DSSP[40] in CPPTRAJ[37]. In all simulations, the short parallel β-sheet shows a similar population of about 75%, with consistent results between the three sets of parameters, as shown in **Figure S8.**

The dihedral angles for the parameterized atom type quartets, N-CX-CT-CG (atom names N-CA-CB-CG), CX-CT-CG-CA (atom names CA-CB-CG-CD1 and CA-CB-CG-CD2), and CA-C-OV-P (atom names CE1-CZ-OG-P and CE2-CZ-OG-P), were calculated throughout the simulations. In simulations using phosaa14SB and phosaa19SB, these rotamers remained stable and consistent with the values in the crystal structure, as shown in **Figure S7**. However, the dihedral angles using the HHLS parameters are less stable in two cases. Specifically for CA-C-OV-P (atom names CE1-CZ-OG-P and CE2-CZ-OG-P), the dihedral angles are unstable and do not align with the values of the crystal structure. The dihedral angles in the crystal structure are -75° and 105°. Using the HHLS parameters, the dihedral angles begin in alignment with the values from the crystal structure, but during both simulations become offset by about 90°. After this rotation, the dihedral angles average 15° and -165°. Despite this rotation, the phosphate oxygens maintain their formation of hydrogen bonds with adjacent residues.

The second deviation observed in simulations using the HHLS parameters is the dihedral angle of atom type quartet N-CX-CT-CG (atom names N-CA-CB-CG), which shifts 25 degrees from the crystal structure (**Figure S7**). This leads to a rotation of the ring away from the NH1 and NH2 nitrogens of Arg55, and significantly reduces the presence of the cation-pi interaction present in the other simulations (**Table S5**).

All the observations mentioned above suggest that the new parameters provide improved accuracy. With HHLS parameters, all three parameterized atom type quartets do not align with the values from the crystal structure as closely as phosaa14SB and phosaa19SB. The dihedral angles for atom type quartet CA-C-OV-P differ from the crystal structures by about 90° for most of the HHLS simulations, weakening the cation-pi interaction between the Ptr ring and Arg55. These differences are remedied with phosaa14SB and phosaa19SB. Comparing the duplicate runs with

each parameter set, the uncertainties in **Table S5** are typically ~ 1%, rising to ~ 6% only in cases where the interactions are less stable.

Phosphothreonine: Aurora A kinase

Aurora A kinase (AURKA) is a major regulator of mitosis, and its activation plays an important role in numerous cancers.[41] Autophosphorylation occurs at Thr164 in a flexible loop region (PDB 5DT3[34]). The Tpo164 phosphate forms a salt bridge with Arg57, ordering the activation loop, which begins activating the kinase.[34] Arg57 is located in the αC helix, along with Gln54, which also interacts with the Tpo164 phosphate. The phosphate forms an additional salt bridge with Arg132 from the catalytic loop, and a hydrogen bond with the peptide nitrogen of Arg163 (Figure S9). The simulation stability for these interactions provides a critical evaluation of parameter accuracy.

The dihedral angles for the parameterized atom type quartets, N-CX-3C-OZ (atom names N-CA-CB-OG), CX-3C-OZ-P (atom names CA-CB-OG-P), and 3C-OZ-P-OX (CB-OG-P-O1P, CB-OG-P-O2P, and CB-OG-P-O3P), are stable as shown in **Figure S10**. The dihedral angle for atom names N-CA-CB-OG is -25° in the crystal structure, and shifted by an average of 25-30° in the simulations. However, this shift is consistent throughout all of the simulations. In the crystal structure, the dihedral angle for atom names CA-CB-OG-P is 129°. With phosaa14SB and phosaa19SB, the average dihedral angle is 76° and 74°, respectively. While these dihedral angles do not align with the crystal structure, the experimental density in this region is weak. However, the key interactions of the phosphate are stable, as discussed below.

Using phosaa14SB, all 3 salt bridges were formed 80 to 100% of the time. Simulations with phosaa19SB adopted the salt bridges 80 to 90% of the time. Using HHLS parameters, the salt

bridges were present 97 to 98% of the time. Notably, the interaction between the peptide nitrogen of Arg163 and the phosphate group was observed 99-100% of the time in all simulations. (**Table S6).**

In summary, the stability of the dihedral values in simulation and the presence of the key interactions with the phosphate group support the accuracy of the new parameters, and deviations in specific dihedral values are likely reasonable given the uncertainty in the experimental model.

Phosphohistidine: Mannose-specific phosphotransferase enzyme IIA component

The mitochondrial enzyme succinyl-CoA synthetase (SCS) builds succinyl-CoA from succinate and coenzyme A using either ATP or GTP as the catalyst.[36] The GTP-specific isoform of SCS from pig heart can be phosphorylated (PDB 1EUD[36]) or dephosphorylated (PDB 1EUC[36]) at His259α.[42]

In the crystal structure, the phosphate group of H2e forms helix N-capping interactions with helices A and B **(Figure S11)**. In helix A, the phosphate oxygens act as hydrogen bond acceptors for the exposed peptide nitrogens of Ala273β and Gly274β. In helix B, the phosphate complements the traditional N-capping role of serine at Ser162α, acting as a hydrogen bond acceptor for the exposed amide N of Gly163α.

The dihedral angles for the parameterized atom type quartet, CW-NA-P-OP (atom names CD2-NE2-P-O1P, CD2-NE2-P-O2P, and CD2-NE2-P-O3P), remained stable and consistent with the values in the crystal structure, as shown in **Figure S12**. H2E dihedral angles for atom type quartets that were not parameterized (due to their presence in ff14SB) were also calculated (**Figure S12)** to verify that re-using the existing parameters is reasonable. These include N-CX-CT-CC

(atom names N-CA-CB-CG) and CX-CT-CC-NA (atom names CA-CB-CG-NA). With all force fields, these dihedral angles remain stable and consistent with crystal structure in every run.

In the simulations using all three parameter sets (phosaa14SB, phosaa19SB, and HHLS), a phosphate oxygen acted as a hydrogen bond acceptor for the peptide nitrogen of Ala273β 88-94% of the time **(Table S7)**. The phosphate group interacted with the peptide nitrogen of Gly274β 93-97% of the time in all simulations. Simulations with both phosaa14SB and HHLS parameters had the peptide nitrogen of Gly163α close enough to donate a hydrogen bond to a phosphate oxygen 96-99% of the time, while the simulations using phosaa19SB had these atoms close enough 83 ± 6% of the time. In the crystal structure, the Gly163 initiates helix B. However, this region is more flexible with phosaa19SB, resulting in reduced hydrogen bonding with the phosphate group.

The new parameters show remarkably similar results to the HHLS parameters here; this is expected, as H2E only has one atom type quartet that was updated with the new parameters (the three-fold symmetric phosphate rotation).

CONCLUSIONS

New sidechain dihedral parameters were developed for five phosphorylated amino acids, each in two protonation states. As with ff14SB, training used both alpha and beta peptide backbone conformations for training data. The parameters were optimized using QM data for a set of training structures, with multiple conformations for each combination of phosphorylated amino acid, protonation state, and backbone. A GPU-based genetic algorithm optimized the parameters, producing an average absolute error of 0.98 kcal/mol. This aligns with the parameterization of standard amino acids in ff14SB[18], which also had an average absolute error of 0.98 kcal/mol. Testing on structures outside the training set produced an average absolute error of 1.42 kcal/mol.

This demonstrates improvement over the HHLS parameters, which had an average absolute error of 2.55 kcal/mol for structures in and outside the training set.

Simulations of proteins containing phosphorylated amino acids using phosaa14SB and phosaa19SB were consistent with experimental structures, specifically in terms of hydrogen bonding, dihedral angle stability, and secondary structure. In the two instances where phosaa19SB did not replicate the observed hydrogen bonding in both the crystal structure and simulations with phosaa14SB, the standard amino acids that were the hydrogen bond donors were in flexible regions. However, the phosphate group and corresponding dihedral angles remained stable, and still formed hydrogen bonds with other nearby residues. In the Aurora A kinase system, phosaa14SB and phosaa19SB had an atom type quartet that did not align with the crystal structure. However, the density in the PDB structure was weak and, therefore, should not be overinterpreted. The dihedral angles remained stable, and the key interactions in the crystal structure and with the HHLS parameters were maintained.

PARAMETER AVAILABILITY

The optimized dihedral side chain parameters are provided in Amber frcmod files, and the library files are available for use with the Amber input preparation module LEaP. For use with ff14SB, users should use frcmod.phosaa14SB, leaprc.phosaa14SB, and phosaa14SB.lib. For use with ff19SB, users should use frcmod.phosaa19SB, leaprc.phosaa19SB, and phosaa19SB.lib.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge:

Additional information regarding training structure generation and selection, simulation methods, and additional figures (PDF)

Amber-format library, frcmod, and leaprc files for use with Amber and other simulation programs (text files), including updated HHLS library files suitable for use with ff14SB:

- frcmod.phosaa14SB
- leaprc.phosaa14SB
- phosaa14SB.lib
- frcmod.phosaa19SB
- leaprc.phosaa19SB
- phosaa19SB.lib
- frcmod.HHLS14
- leaprc.HHLS14
- HHLS14.lib

AUTHOR INFORMATION

Corresponding Author

* Email: carlos.simmerling@stonybrook.edu

Present Addresses

Abbigayle Cuomo is currently at Yale University, Victoria Hazoglou is currently at the University of Arizona, and Gabriela Witek is currently at the University of Pennsylvania.

Author Contributions

L.E.R. performed the parameter training, and L.E.R. and A.E.C. performed the parameter testing under the guidance of C.S. L.E.R., A.E.C., K.A.A.B., C.T., V.H., G.W., Q.W., and C.S. contributed to the project design, data analysis, and interpretation. All authors have given approval to the final version of the manuscript.

Funding Sources

The authors acknowledge funding from the National Science Foundation (grant 1665159 to C.S.), the URECA (Undergraduate Research and Creative Activities) Program at Stony Brook University, and the Dr. Kenneth M. Nicholas Undergraduate Fellowship through the Department of Chemistry at Stony Brook University. This research used resources from the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility at Brookhaven National Laboratory under Contract No. DE-SC0012704. The Laufer Center at Stony Brook University also supported this work.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

AAE, absolute average error; CMAP, correction map; ff14SB, force field 2014 Stony Brook; ff19SB, force field 2019 Stony Brook; ff94, force field 1994; ff99SB, force field 1999 Stony Brook; HF, Hartree-Fock; MD, Molecular Dynamics; MM, Molecular Mechanics; MP2, Møller-Plesset pertubation theory of the second order; PDB, Protein Data Bank; QM, Quantum Mechanics; REE, relative energy error; RESP, Restrained Electrostatic Potential

REFERENCES

1. Cieśla, J.; Fraczyk, T.; Rode, W., Phosphorylation of Basic Amino Acid Residues in Proteins: Important but Easily Missed. *Acta Biochim. Pol.* **2011,** *58* (2), 137-48.
2. Manning, G.; Plowman, G. D.; Hunter, T.; Sudarsanam, S., Evolution of Protein Kinase Signaling From Yeast to Man. *Trends Biochem. Sci.* **2002,** *27* (10), 514-520.
3. Klumpp, S.; Krieglstein, J., Phosphorylation and Dephosphorylation of Histidine Residues in Proteins. *Eur. J. Biochem.* **2002,** *269* (4), 1067-1071.
4. Fuhs, S. R.; Hunter, T., pHisphorylation: The Emergence of Histidine Phosphorylation as a Reversible Regulatory Modification. *Curr. Opin. Cell. Biol.* **2017,** *45*, 8-16.

5. Rajagopal, P.; Waygood, E. B.; Klevit, R. E., Structural Consequences of Histidine Phosphorylation: NMR Characterization of the Phosphohistidine Form of Histidine-Containing Protein from Bacillus subtilis and Escherichia coli. *Biochem.* **1994,** *33* (51), 15271-15282.
6. Kawade, R.; Kuroda, D.; Tsumoto, K., How the Protonation State of a Phosphorylated Amino Acid Governs Molecular Recognition: Insights From Classical Molecular Dynamics Simulations. *FEBS Lett.* **2020,** *594* (5), 903-912.
7. Bellucci, L.; Felline, A.; Fanelli, F., Dynamics and Structural Communication in the Ternary Complex of Fully Phosphorylated V2 Vasopressin Receptor, Vasopressin, and β-Arrestin 1. *Biochim. Biophys. Acta - Biomembr.* **2020,** *1862* (9), 183355.
8. Groban, E. S.; Narayanan, A.; Jacobson, M. P., Conformational Changes in Protein Loops and Helices Induced by Post-translational Phosphorylation. *PLoS Comput. Biol.* **2006,** *2* (4), e32.
9. Ahamad, S.; Kanipakam, H.; Kumar, V.; Gupta, D., A Molecular Journey to Check the Conformational Dynamics of Tau Tubulin Kinase 2 Mutations Associated With Alzheimer's Disease. *RSC Adv.* **2021,** *11* (3), 1320-1331.
10. You, W.; Huang, Y.-m. M.; Kizhake, S.; Natarajan, A.; Chang, C.-e. A., Characterization of Promiscuous Binding of Phosphor Ligands to Breast-Cancer-Gene 1 (BRCA1) C-Terminal (BRCT): Molecular Dynamics, Free Energy, Entropy and Inhibitor Design. *PLoS Comput. Biol.* **2016,** *12* (8), e1005057.
11. Stender, J. D.; Nwachukwu, J. C.; Kastrati, I.; Kim, Y.; Strid, T.; Yakir, M.; Srinivasan, S.; Nowak, J.; Izard, T.; Rangarajan, E. S.; Carlson, K. E.; Katzenellenbogen, J. A.; Yao, X. Q.; Grant, B. J.; Leong, H. S.; Lin, C. Y.; Frasor, J.; Nettles, K. W.; Glass, C. K., Structural and Molecular Mechanisms of Cytokine-Mediated Endocrine Resistance in Human Breast Cancer Cells. *Mol. Cell.* **2017,** *65* (6), 1122-1135.e5.
12. Chi, C.; Li, X.; Feng, T.; Zeng, X.; Chen, L.; Li, L., Improvement in Nutritional Attributes of Rice Starch with Dodecyl Gallate Complexation: A Molecular Dynamic Simulation and in Vitro Study. *Journal of Agricultural and Food Chemistry* **2018,** *66* (35), 9282-9290.
13. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.
14. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984,** *106* (3), 765-784.
15. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
16. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993,** *115* (21), 9620-9631.
17. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006,** *65* (3), 712-725.
18. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-713.

19. Janowski, P. A.; Liu, C.; Deckman, J.; Case, D. A., Molecular Dynamics Simulation of Triclinic Lysozyme in a Crystal Lattice. *Protein Sci.* **2016,** *25* (1), 87-102.
20. Huang, J.; MacKerell, A. D., Jr., CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013,** *34* (25), 2135-45.
21. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2019**.
22. Mackerell Jr., A. D.; Feig, M.; Brooks III, C. L., Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004,** *25* (11), 1400-1415.
23. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr., CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017,** *14* (1), 71-73.
24. Homeyer, N.; Horn, A. H.; Lanig, H.; Sticht, H., AMBER Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J. Mol. Model.* **2006,** *12* (3), 281-9.
25. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
26. Xie, Y.; Jiang, Y.; Ben-Amotz, D., Detection of Amino Acid and Peptide Phosphate Protonation Using Raman Spectroscopy. *Anal. Biochem.* **2005,** *343* (2), 223-230.
27. Robitaille, P.-M. L.; Robitaille, P. A.; Gordon Brown, G.; Brown, G. G., An Analysis of the pH-Dependent Chemical-Shift Behavior of Phosphorus-Containing Metabolites. *J. Magn. Reson.* **1991,** *92* (1), 73-84.
28. Gassner, M.; Stehlik, D.; Schrecker, O.; Hengstenberg, W.; Maurer, W.; Rüterjans, H., The Phosphoenolpyruvate-Dependent Phosphotransferase System of Staphylococcus aureus. *Eur. J. Biochem.* **1977,** *75* (1), 287-296.
29. Neese, F., The ORCA Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2012,** *2* (1), 73-78.
30. Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971,** *54* (2), 724-728.
31. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., A Direct MP2 Gradient Method. *Chem. Phys. Lett.* **1990,** *166* (3), 275-280.
32. Belfon, K. A. A.; Tian, C.; Maier, J.; Simmerling, C. RAGTAG: Rapid Amber Gpu Torsion pArameter Generation for Amber. (accessed June 4, 2024).
33. Lu, A.; Gosser, Y.; Montclare, J. K.; Liu, Z.; Kong, X., Structure of Aspergillus Oryzae Cutinase Expressed in Pichia Pastoris, Crystallized in the Presence of Paraoxon. 2012.
34. Janeček, M.; Rossmann, M.; Sharma, P.; Emery, A.; Huggins, D. J.; Stockwell, S. R.; Stokes, J. E.; Tan, Y. S.; Almeida, E. G.; Hardwick, B.; Narvaez, A. J.; Hyvönen, M.; Spring, D. R.; McKenzie, G. J.; Venkitaraman, A. R., Allosteric Modulation of AURKA Kinase Activity by a Small-Molecule Inhibitor of Its Protein-Protein Interaction With TPX2. *Sci. Rep.* **2016,** *6* (1), 28528.
35. Poy, F.; Yaffe, M. B.; Sayos, J.; Saxena, K.; Morra, M.; Sumegi, J.; Cantley, L. C.; Terhorst, C.; Eck, M. J., Crystal structures of the XLP protein SAP reveal a class of SH2

domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell.* **1999,** *4* (4), 555-61.
36. Fraser, M. E.; James, M. N. G.; Bridger, W. A.; Wolodko, W. T., Phosphorylated and Dephosphorylated Structures of Pig Heart, GTP-Specific Succinyl-CoA Synthetase. *J. Mol. Biol.* **2000,** *299* (5), 1325-1339.
37. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084-3095.
38. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.
39. Liu, Z.; Gosser, Y.; Baker, P. J.; Ravee, Y.; Lu, Z.; Alemu, G.; Li, H.; Butterfoss, G. L.; Kong, X.-P.; Gross, R.; Montclare, J. K., Structural and Functional Studies of Aspergillus oryzae Cutinase: Enhanced Thermostability and Hydrolytic Activity of Synthetic Ester and Polyester Degradation. *J. Am. Chem. Soc.* **2009,** *131* (43), 15711-15716.
40. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-637.
41. Du, R.; Huang, C.; Liu, K.; Li, X.; Dong, Z., Targeting AURKA in Cancer: Molecular Mechanisms and Opportunities for Cancer Therapy. *Mol. Cancer* **2021,** *20* (1), 15.
42. Ottaway, J. H.; McClellan, J. A.; Saunderson, C. L., Succinic Thiokinase and Metabolic Control. *Int. J. Biochem.* **1981,** *13* (4), 401-410.

For Table of Contents Only