



Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives

Xiaowen Wang^{1,2} and Wenjin Li^{1*}

¹ Institute for Advanced Study, Shenzhen University, Shenzhen, China, ² College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China

OPEN ACCESS

Edited by:

Yong Wang,
University of Copenhagen, Denmark

Reviewed by:

Yifei Qi,
East China Normal University, China
Andrea Spitaleri,
San Raffaele Hospital (IRCCS), Italy
Huiyong Sun,
China Pharmaceutical
University, China

*Correspondence:

Wenjin Li
liwenjin@szu.edu.cn

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 22 September 2020

Accepted: 23 October 2020

Published: 15 December 2020

Citation:

Wang X and Li W (2020) Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives.
Front. Mol. Biosci. 7:608931.
doi: 10.3389/fmolb.2020.608931

Theoretical analyses are valuable for the exploration of the effects of unnatural amino acids on enzyme functions; however, many necessary parameters for unnatural amino acids remain lacking. In this study, we developed and tested force field parameters compatible with Amber ff14SB for 18 phenylalanine and tyrosine derivatives. The charge parameters were derived from ab initio calculations using the RESP fitting approach and then adjusted to reproduce the benchmark relative energies (at the MP2/TZ level) of the α - and β -backbones for each unnatural amino acid dipeptide. The structures optimized under the proposed force field parameters for the 18 unnatural amino acid dipeptides in both the α - and β -backbone forms were in good agreement with their QM structures, as the average RMSD was as small as 0.1 Å. The force field parameters were then tested in their application to seven proteins containing unnatural amino acids. The RMSDs of the simulated configurations of these unnatural amino acids were approximately 1.0 Å compared with those of the crystal structures. The vital interactions between proteins and unnatural amino acids in five protein–ligand complexes were also predicted using MM/PBSA analysis, and they were largely consistent with experimental observations. This work will provide theoretical aid for drug design involving unnatural amino acids.

Keywords: unnatural amino acids, charge parameters, Amber ff14SB, relative energy, molecular dynamics, MM/PBSA

INTRODUCTION

As is well known, 20 natural amino acids are the main building blocks of proteins, the macromolecules that perform a broad spectrum of functions within organisms (Qin et al., 2015). Unnatural amino acids (UAAs) also called noncanonical amino acids are analogs or metabolic intermediates of the 20 natural amino acids with only minor structural differences—often just a chemical functional group—which is beneficial for analyzing their effects on enzyme functions (Zhao et al., 2020). Since UAAs are of high chemical diversity, possess strong site specificity, and introduce little disturbance to the protein structure, it is widely applied in protein engineering, virus vaccine development, and medical therapeutics (Minnihan et al., 2011; Si et al., 2016; Young and Schultz, 2018). For instant, biological catalysis and reaction mechanism of tyrosine in aminoacyl-tRNA synthetases (aaRS) were investigated through the incorporation of UAA fluorotyrosine, whose pKa was tuned by changing the number and the site of fluoro-substitution (Minnihan et al., 2011). Si and co-workers employed the UAA N^ε-2-azidoethoxy-carbonyl-L-lysine to produce replication-incompetent viral vaccines by introducing premature termination codon into the genome of influenza A virus, and these viral vaccines prevented further damage

inside conventional cells via immune response (Si et al., 2016). In addition, UAAs are utilized in the bio-orthogonal reactivity. For example, UAA-incorporated proteins (such as antibodies, growth factors, and cytokines) specifically interacted with diverse moieties to form bispecific antibodies, antibody-drug conjugates, and pegylated proteins, which provided effective treatments for various clinical testing (Young and Schultz, 2018).

The incorporation of UAAs into canonical proteins expanded significantly the genetic code library (Xiao et al., 2015). Natural UAAs occur commonly in plants, microorganisms, and animals, while those in human organisms must be chemically synthesized (Zou et al., 2018). Typically, an orthogonal amber suppressor aaRS/tRNA pair has been utilized to guide the incorporation of UAAs in response to a unique nonsense codon (Santoro et al., 2002; Liu and Schultz, 2010). Experimentally, many studies have reported the incorporation of UAAs into the designated sites of target proteins by means of popular residue-specific and site-directed mutagenesis approaches (Sakamoto et al., 2002; Fleissner et al., 2009; Xiao et al., 2015; Yuet et al., 2015). For example, Yuet et al. described a method for residue-specific labeling that enabled the use of the UAA *p*-azido-*l*-phenylalanine (AzF) to tag and analyze protein metabolism in specific cells based on the phenylalanyl-tRNA synthetase (Yuet et al., 2015). Schultz et al. utilized site-directed mutagenesis to mutate Val216 of TEM-1 β -lactamase into *p*-acrylamido-phenylalanine (AcrF), which enhanced the catalytic activity of the enzyme (Xiao et al., 2015). Although the two complementary methods involved in residue-specific and site-directed mutagenesis are widely used to incorporate UAAs into proteins, they often contend with certain technical difficulties.

To compensate for experimental obstacles, theoretical computational methods validated by experimental data offer a novel way to screen potential analogs for natural amino acids. A number of computational methods to study proteins containing UAAs have been successively reported by other groups in recent years (Renfrew et al., 2012; Petrov et al., 2013; Khoury et al., 2014a,b). For example, Renfrew et al. constructed a rotamer library containing 114 UAAs to study the interface of calpain and calpastatin, which was evaluated using a scoring function based on the Rosetta program (Leaverfay et al., 2011; Renfrew et al., 2012). New GROMOS54a7 force field parameters were developed by the Zagrovic group for processing post-translationally modified amino acids by means of molecular dynamics (MD) simulations executed by the GROMACS package (Petrov et al., 2013). In addition, a tool called “Forcefield_NCAA” created by the Floudas lab is now available for generating UAA parameters related to a library of 147 noncanonical amino acids compatible with the Amber ff03 parameters (Khoury et al., 2014a,b).

The aim of our work was to develop and test force field parameters for phenylalanine and tyrosine derivatives, most of which are not included in the reported literature. The structures of the involved UAAs in this study are displayed in **Figure 1**. The newly developed parameters were then applied to mutant proteins or protein–ligand interactions involving UAAs, as listed in **Supplementary Table 1**, by MD simulations and molecular mechanics–Poisson Boltzmann solvent accessible

surface area (MM/PBSA) calculations. Based on comparison with experimental data as the benchmark, the simulation results indicate that the new force field parameters can predict protein structures with incorporated UAAs well and generally describe the exact interplay that occurs in the binding pockets of proteins with UAAs as substrates.

SIMULATION STRATEGIES

We first constructed dipeptides of the α - and β -conformer of each UAA in the form of Ace-XXX-NMe using GaussView 6 (Dennington et al., 2016) (Step 1 in **Figure 2**). Here, XXX represents the analogs of phenylalanine and tyrosine shown in **Figure 1**. It is a popular way to employ α - and β -backbones of amino acids to fit parameters in current classical force fields, such as AMBER, CHARMM, and OPLS (Hornak et al., 2006; Best et al., 2012; Robertson et al., 2015), as these backbones dominate in the sterically allowed structural regions of the Ramachandran plot (Ramachandran et al., 1963). For the constructed dipeptides, structural optimization was performed at the B3LYP/6-31G* level, and single-point energy calculations were executed at the MP2/cc-pVTZ level using the Gaussian 16 program (Step 2 in **Figure 2**) (Frisch et al., 2016). Based on the optimized structures obtained at the B3LYP/6-31G* level, the electrostatic potential (ESP) charges at the HF/6-31G* level were further evaluated; this is a popular method to produce ESP charges because of the accurate reproduction of free energies of solvation and liquid enthalpies (Cornell et al., 1993; Wang et al., 2000). In Step 3, restrained electrostatic potential (RESP) charges were generated based on the ESP-fit charge model (Cornell et al., 1995). The general Amber force field (GAFF) is a useful molecular mechanics and is designed to be suitable for organic molecules, especially drug-like small molecules (Wang et al., 2004). In the following stage, we thus produced bonded and non-bonded parameters using GAFF based on the Antechamber tool (Case et al., 2020). The newly generated parameters can be transferred into the GMX format using the ACPYPE.py script for subsequent MD simulations in the GROMACS software package (Sousa da Silva and Vranken, 2012). In Step 5, the initial parameters of the structures of the α - and β -conformers of each UAA were tested. Accordingly, we optimized the charge parameters of the 18 analogs by estimating the relative energies of each UAA pair compared with the benchmark of quantum mechanics (QM) data at the MP2/cc-pVTZ level. In the final step, MD simulations and MM/PBSA calculations on the proteins or protein–ligand complexes involving UAAs were further performed to test the new parameters determined in this work. The complete workflow of parametrization is shown in **Figure 2**, and the detailed methodology for producing the parameters is described below.

QM Calculations

The 18 UAAs shown in **Figure 1** are analogs derived from the amino acids of phenylalanine (F) and tyrosine (Y). Based on the 18 UAAs, we constructed dipeptides of two backbone conformers for each UAA blocked with *N*-methyl and acetyl groups in the form of Ace-XXX-NMe in GaussView 6 (Dennington et al., 2016). The two backbone conformers were designed in the forms

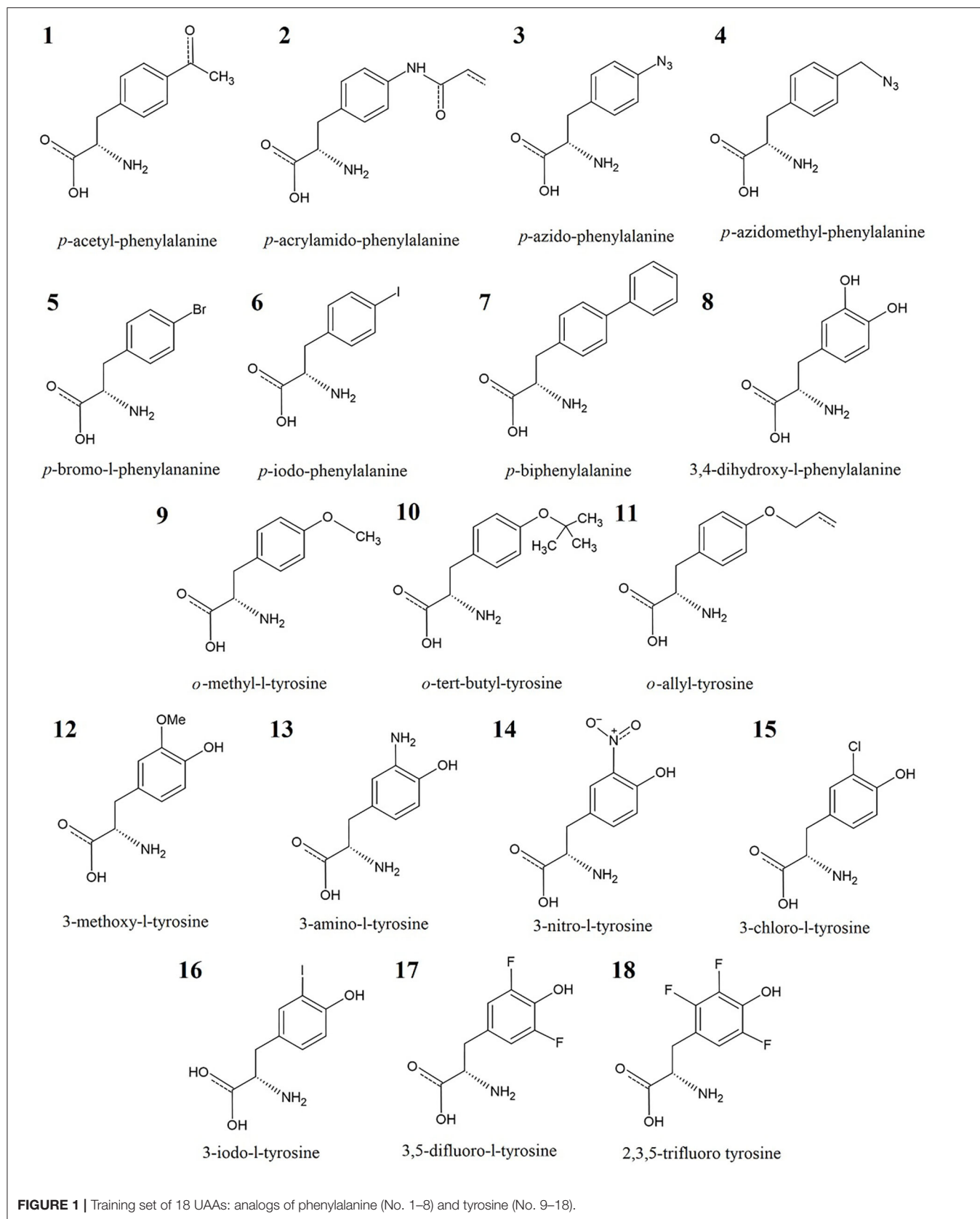
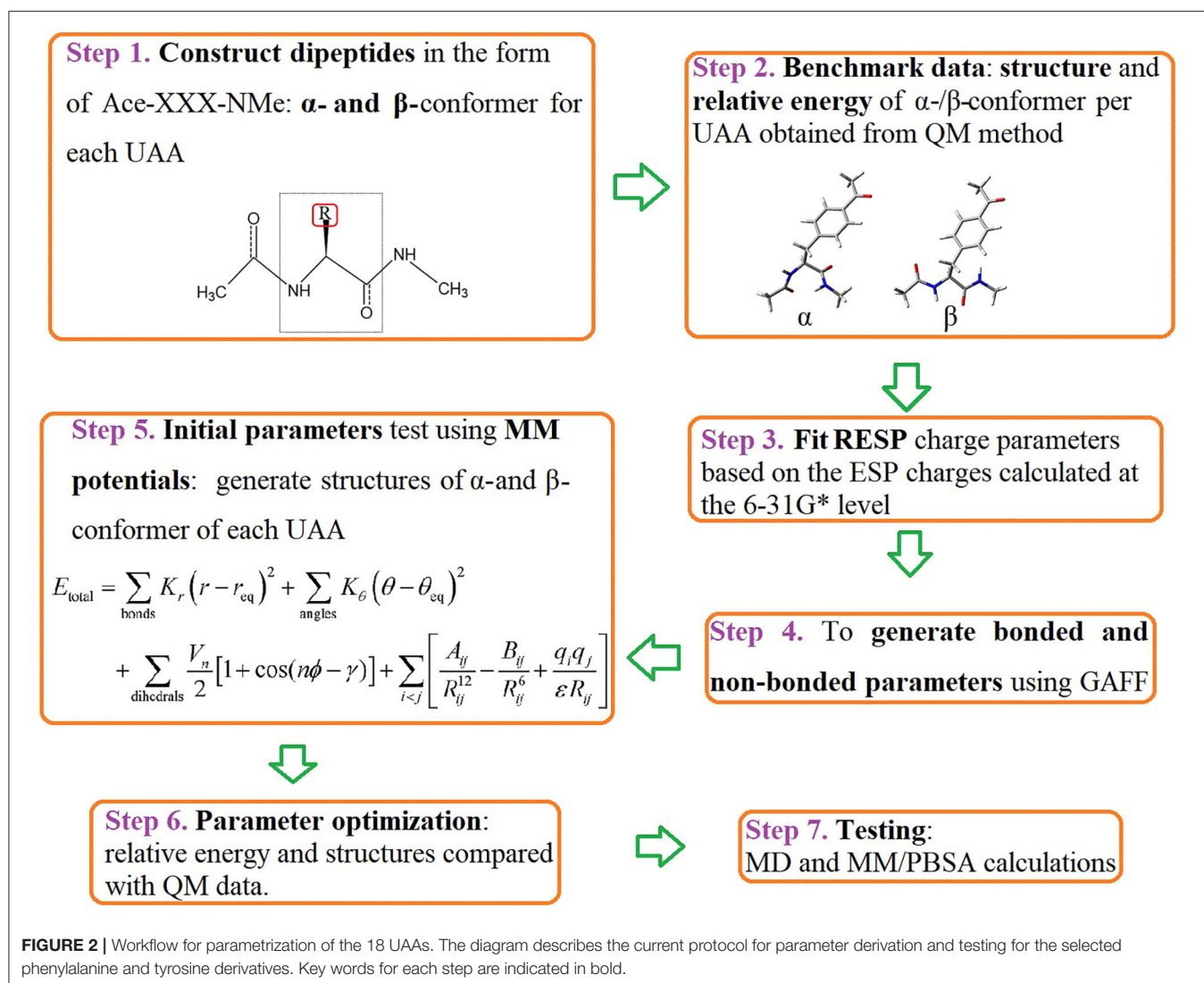


FIGURE 1 | Training set of 18 UAAs: analogs of phenylalanine (No. 1–8) and tyrosine (No. 9–18).



of an α -helix ($\phi = -60^\circ$, $\psi = -40^\circ$) and β -strand ($\phi = -180^\circ$, $\psi = 180^\circ$). Structural optimizations were performed at the B3LYP/6-31G* level (Mondal et al., 2007), followed by single-point energy calculations at the MP2/cc-pVTZ level (Harder et al., 2016). For comparison, an additional method for the structure and energy calculations was performed at the M06-2X level (Robertson et al., 2015). The pseudopotential for iodine-containing systems was assigned as the SDD basis set in this work (Yurieva et al., 2008). The missing van der Waals (vdW) radius for iodine atoms was chosen as the Pauling radius (2.15 Å) (Pauling, 1939). The QM calculations were performed using the Gaussian 16 program (Frisch et al., 2016).

Energy Model

The total pair potential energy used in this work is written as a sum of terms as follows:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{es}} + E_{\text{vdW}} \quad (1)$$

Our goal is to develop UAA charge parameters that are compatible with the Amber ff14SB parameter set for the 20 natural amino acids. The energy function (Equation 1) from the Amber force field is thus employed here (Maier et al., 2015). Generally, the vdW radius and epsilon parameters are derived from experimental data (Weiner et al., 1984). The charge parameters were adjusted using the protocol described below.

Parameter Optimization

The partial charges were fitted using RESP charges obtained at the HF/6-31G* level (Cornell et al., 1993, 1995; Wang et al., 2000). The initial bonding and vdW parameters were generated from GAFF using the Antechamber module in AmberTools20 (Case et al., 2020). The charge sets of the Ace and NMe groups are identical to the Amber ff14sb force fields. Together, we used bonded and non-bonded parameters to calculate the structures and relative energies of the α - and β -conformers of the 18 UAAs. By comparing the QM structures and relative energies, we adjusted the charge parameters of the UAA backbones and

side chains until good accordance was achieved with the QM data in terms of the root-mean-square (RMS) deviation, as shown in Equation 2.

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^N (RE_{QM}(i) - RE_{our}(i))^2}{N}} \quad (2)$$

where $RE_{QM}(i)$ and $RE_{our}(i)$ are the relative energies calculated by QM and the new parameters developed in our work for the i th training set, respectively, and N is the total number of training sets. We minimized the RMS values to obtain the charge parameters.

MD Simulations

MD simulations were performed using the 2019 version of the GROMACS program (Abraham et al., 2015). We chose the Amber ff14SB force field for proteins composed of natural amino acids (Maier et al., 2015). For the UAA components, the new parameters developed in this work were used. We placed the initial systems in the center of a cubic box 10 Å from the box edge. The box was then filled with a water solvent using the TIP3P water model (Jorgensen et al., 1983). The water molecules were randomly replaced by Na^+ and Cl^- ions to a 0.1 M concentration. For each model, energy minimization with a maximum of 5,000 steps was carried out without any restraints. After optimization, two short 200 ps MD simulations in the NVT and NPT ensembles were successively performed with the heavy-atom position restraint at a force constant of 500 kcal/(mol·Å²). The position restraints were gradually released via four steps of 100 ps NPT simulations with force constants of 250, 100, 50, and 10 kcal/(mol·Å²) for the heavy atoms. Finally, 20 ns production MD simulations were performed in the NPT ensemble. The time step was set to 2 fs, and the temperature and pressure were kept constant at 300 K and 1 bar, respectively. In the production runs, the velocity-rescaling thermostat was applied for temperature coupling (Berendsen et al., 1984; Bussi et al., 2007), while the Parrinello–Rahman approach was applied for constant pressure control (Parrinello and Rahman, 1981; Nosé and Klein, 1983). The SHAKE algorithm was used to constrain covalent bonds involving hydrogen atoms (Andersen, 1983; Miyamoto and Kollman, 1992). The particle mesh Ewald method was applied to the calculation of long-range electrostatic interactions (Darden et al., 1993). The cutoff values for vdW and electrostatic forces were set to 12 Å, and the simulation structures were saved every 100 ps to obtain the trajectories for analysis.

MM/PBSA Estimation

In general, the binding free energy for protein–ligand interactions can be expressed as

$$\Delta G_{\text{bind}} = \Delta E_{\text{vdW}} + \Delta E_{\text{ele}} + \Delta G_{\text{solv}} - T\Delta S \quad (3)$$

where ΔE_{vdW} and ΔE_{ele} are the non-bonded terms of the system total energy (ΔE_{MM}) due to vdW and electrostatic interactions, respectively. The bonded terms of ΔE_{MM} were assumed to be zero in the single-trajectory setup used in this procedure because

of its simplicity and accuracy similar to those of a multi-trajectory setup (Genheden and Ryde, 2015; Wang et al., 2018). ΔG_{solv} is the solvation free energy required to move the solute from a vacuum (dielectric constant of 1) into the solvent (dielectric constant of 80). It can be further decomposed into polar ($\Delta G_{\text{pb/solv}}$) and nonpolar ($\Delta G_{\text{np/solv}}$) contributions to solvation. T and ΔS are the absolute temperature and entropy, respectively. However, the entropy term was ignored in this study because of the significant time consumption, uncertainty of the contributions to the total free energy, and small improvement by comparison with the experimental results (Yang et al., 2011; Kumari et al., 2014).

Furthermore, the binding free energy decomposition of each residue was analyzed to understand the key residue impact at the activation region of the protein–inhibitor interaction. Hence, the free energy of each residue ($\Delta G_{\text{res}}^{\text{bind}}$) can be divided into three terms:

$$\Delta G_{\text{res}}^{\text{bind}} = \Delta E_{\text{res}}^{\text{MM}} + \Delta G_{\text{res}}^{\text{pb/solv}} + \Delta G_{\text{res}}^{\text{np/solv}} \quad (4)$$

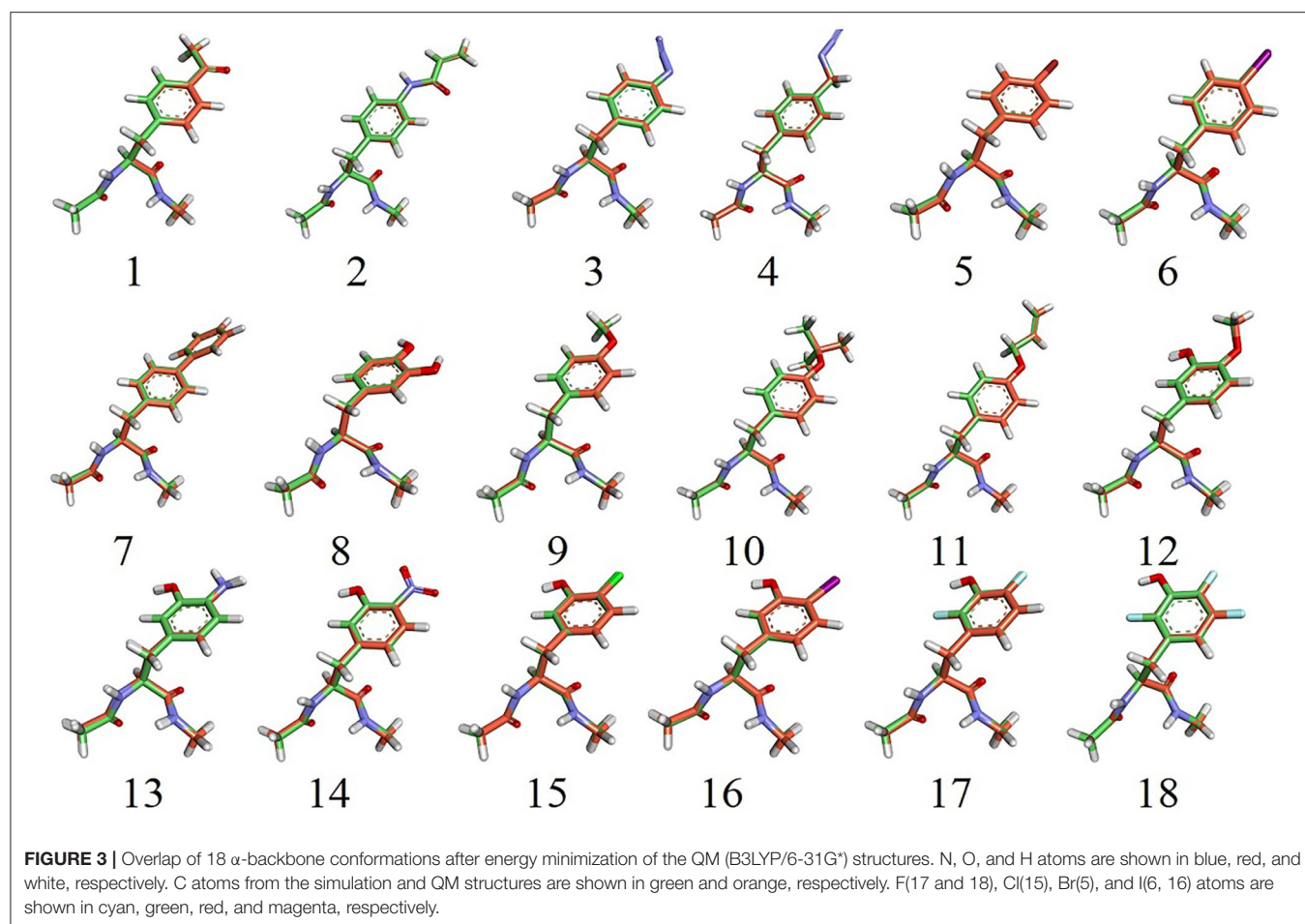
where $\Delta E_{\text{res}}^{\text{MM}}$ is the sum of the electrostatic and vdW interactions per residue in a vacuum, and $\Delta G_{\text{res}}^{\text{pb/solv}}$ and $\Delta G_{\text{res}}^{\text{np/solv}}$ are the polar and nonpolar parts of the per-residue solvation free energy, respectively.

In this work, the successive 20 ns trajectories produced were used to perform MM/PBSA calculations on the free energies using the `g_mmpbsa` tool (Kumari et al., 2014). Here, the system coordinates were saved for every 1 ns used for MM/PBSA analysis such that 20 snapshots for each trajectory were considered to calculate the binding free energies of the protein–inhibitor interactions. The Poisson–Boltzmann (PB) equation was applied to calculate $\Delta G_{\text{pb/solv}}$ (Honig and Nicholls, 1995). The temperature and grid spacing were set to 300 K and 0.5 Å, respectively, and the concentration of charged ions was 0.1 M with radii of 0.95 and 1.81 Å for Na^+ and Cl^- , respectively. The solvent accessible surface area (SASA) model was employed to estimate the nonpolar contributions ($\Delta G_{\text{np/solv}}$) from the function $\gamma\text{SASA} + b$ (Sitkoff et al., 1994). The radius value for SASA was 1.4 Å, and the constants γ and b were set to default values of 0.00542 kcal/(mol·Å²) and 0.92 kcal/mol, respectively.

RESULTS AND DISCUSSION

Initial Parameters Applied to α -/ β -Conformer Optimization

After the initial parameters (hereafter referred to as cycle-1 parameters) involved in the bonded and non-bonded terms were generated, we performed structural optimizations for the α - and β -conformers of each UAA. For comparison with the B3LYP/6-31G* structures, we depict the optimized structures of the 18 UAA dipeptides in the α -state from the initial parameters in **Figure 3**; the minimized structures for the β -state are shown in **Supplementary Figure 1**. As shown, the two backbone conformations in the α - and β -states of the training set are in good agreement with the QM structures. The initial parameters also performed well for the side-chain structures. Additionally, the determined heavy-atom and all-atom RMS displacements (RMSDs) for the 18 training



sets from **Table 1** are nearly $<0.1 \text{ \AA}$ (refer also to the RMSD distributions in **Supplementary Figure 2**). Among the systems, system 13 has the greatest RMSDs of 0.083–0.116 \AA . Meanwhile, **Supplementary Figure 2** shows that the all-atom RMSDs are comparable to the heavy-atom RMSDs but fluctuate to slightly higher values. Overall, the initial parameters yield good results for the 18 training sets, especially the bonded connections, but further improvements to the energies are necessary.

Testing of Optimized Parameters

Displayed in **Table 2** are the relative energies for the 18 training sets. We selected the relative energies evaluated at the MP2/cc-pVTZ//B3LYP/6-31G* level of theory as a benchmark (Mondal et al., 2007; Harder et al., 2016). For comparison, one density functional theory (DFT) method with a small basis set at the M06-2X/6-311++G**//M06-2X/6-31+G* level was used in this work (Robertson et al., 2015). For the parameter optimization process, four cycles were performed. First, we fixed the charges of the Ace and NMe groups in the 18 UAA dipeptides to remain the same as the corresponding Amber ff14sb force field parameter sets and made minor adjustments to the backbone RESP charges. As shown in **Table 2**, the relative energies from the cycle-1 parameters show a correlation of 0.8212 compared

with the MP2 energies, with a larger RMS deviation of 4.86 kcal/mol. In the next two cycles, we chose to treat the backbones and side chains as α -helical RESP charges and averaged RESP charges in the α - and β -states, respectively. In the third cycle adjustment, the RMS decreased to 2.33 kcal/mol with a 0.8072 correlation. At this point, we noted that the relative energies of most systems were comparable to the benchmark data except for those of systems 4, 9–12, 17, and 18. Therefore, the parameters for these systems were further optimized. In the final procedure, we chose the β -conformational charges from the first cycle as the determined parameters for systems 4 and 18. For systems 9–12 and 17, different proportions between the α - and β -conformational RESP charges were ultimately treated (see the footnotes in **Table 2**). For the remaining systems, we employed the averaged charges in the α - and β -states. Eventually, we observed a strong correlation between our work and the QM data, with $R^2 = 0.9407$ (**Supplementary Figure 3**). Therefore, the parameters from the fourth cycle were employed in subsequent calculations. Although the partial charges were obtained by fitting to the RESP of independent conformations for each UAA, the partial charges of the atoms in their common structures are quite close to each other (see **Supplementary Material**). Note that these UAAs are phenylalanine and tyrosine derivatives and share a common structure. The observation of such small differences

TABLE 1 | Initial parameter test for the 18 training sets evaluated by heavy-atom and all-atom RMSDs.

Training set	Heavy-atom RMSD (Å)		All-atom RMSD (Å)	
	α -Backbone	β -Backbone	α -Backbone	β -Backbone
1	0.041	0.052	0.045	0.059
2	0.072	0.055	0.083	0.060
3	0.054	0.052	0.051	0.053
4	0.046	0.039	0.043	0.046
5	0.068	0.052	0.067	0.058
6	0.080	0.042	0.079	0.048
7	0.084	0.041	0.082	0.053
8	0.084	0.045	0.082	0.057
9	0.046	0.046	0.048	0.053
10	0.044	0.042	0.048	0.048
11	0.054	0.045	0.060	0.051
12	0.064	0.048	0.068	0.059
13	0.096	0.083	0.116	0.094
14	0.056	0.047	0.058	0.055
15	0.073	0.055	0.074	0.063
16	0.072	0.058	0.075	0.065
17	0.067	0.040	0.069	0.046
18	0.046	0.039	0.047	0.045

indicates that the obtained RESP parameters for these UAAs are reliable and the charge parameters are well converged.

In addition, we noted that the β -backbone conformation of each UAA is more stable than the α -backbone as predicted by all employed methods. Here, our work shows a more favorable RMS deviation of 0.33 kcal/mol compared with M06-2X with an RMS deviation of 1.08 kcal/mol. Additionally, existing charge parameters from a reference were also tested on reported systems 7, 9, 11, and 13 (Khoury et al., 2014b). The relative energies of these four systems are 5.74, 8.52, 8.05, and 4.40 kcal/mol obtained from the reported parameters, which are comparable to those from the MP2 data of 6.76, 6.86, 6.79, and 8.26 kcal/mol, respectively, but produce absolute errors of approximately 1.5 kcal/mol or higher (Supplementary Table 2). Compared with the MP2 energies, as also shown in Supplementary Table 2, the RMS deviations obtained from the reference and our work were 2.25 and 0.38 kcal/mol, respectively, for these four systems. Therefore, the energetic performance of the new parameters determined in our work results in more satisfactory predictions. In addition, the structural optimizations from the cycle-4 parameters were again tested on the 18 dipeptides in the α - and β -states. Comparisons of heavy-atom RMS distributions between cycles 1 and 4 are provided in Supplementary Figure 4, which clearly shows that the new parameters produce smaller heavy-atom RMS deviations than the initial parameters. Overall, the new parameters show a good performance in terms of structural optimization and relative energy calculations for the 18 UAA models based on comparison with QM results, indicating that the new parameters determined in this work are appropriate for performing further tests via MD simulations.

TABLE 2 | Relative energies (kcal/mol) for the α - and β -conformers of the 18 UAA dipeptides obtained from QM calculations and our work.

Training set	MP2 ^a	M06-2X ^b	Our work			
			Cycle 1	Cycle 2	Cycle 3	Cycle 4 ^c
1	6.50	7.39	11.07	2.61	6.72	6.72
2	6.93	7.72	10.78	3.34	7.12	7.12
3	6.35	7.13	11.14	1.96	6.70	6.70
4	6.05	7.30	6.63	0.01	3.52	6.63
5	6.45	7.28	10.11	2.01	6.31	6.31
6	6.53	7.38	10.56	2.44	6.55	6.55
7	6.76	7.58	10.36	2.77	6.76	6.76 (5.74)
8	6.77	7.54	10.06	3.18	7.06	7.06
9	6.86	7.62	12.78	5.56	9.57	7.00 (8.52)
10	6.89	9.74	12.40	5.54	9.06	6.77
11	6.79	7.54	12.32	4.33	8.22	7.10 (8.05)
12	8.02	8.90	15.12	6.34	11.08	7.93
13	8.26	8.99	13.78	4.05	8.94	8.94 (4.40)
14	7.12	8.08	14.38	2.44	7.70	7.70
15	7.44	8.37	12.76	1.80	7.58	7.58
16	7.53	8.42	12.88	2.69	7.86	7.86
17	5.07	5.54	6.70	-6.75	0.47	4.96
18	4.80	5.34	4.45	-7.52	-1.53	4.45
^d RMS MP2	–	1.08	4.86	5.68	2.33	0.33
^e R ² (MP2)		0.7831	0.8212	0.6922	0.8072	0.9407

^aMP2/cc-pVTZ//B3LYP/6-31G*.

^bM06-2X/6-311++G**//M06-2X/6-31+G*.

^cValues in parentheses were obtained using charge parameters taken from the literature (Khoury et al., 2014b). The different proportions of charge parameters in the final cycle are $\beta/6 + \alpha^*/5/6$ for system 9, $\beta/5 + \alpha^*/4/5$ for system 10, $\beta/3 + \alpha^*/2/3$ for system 11, $\beta/6 + \alpha^*/5/6$ for system 12, $\beta^*/7/8 + \alpha/8$ for system 17, β for systems 4 and 18, and $\beta/2 + \alpha/2$ for the other systems.

^dAll units for RMS deviations are kcal/mol.

^eCorrelation between MP2 and the other methods involved in M06-2X and our work. The relative energy (RE) is defined as $E_\alpha - E_\beta$.

Testing

MD Simulations of Proteins Containing UAAs

Seven isolated protein systems containing UAAs were selected to identify the new parameters as the testing set. At present, crystals composed of noncanonical amino acids have rarely been recorded in the PDB. We attempted to search for the protein structures covering UAAs related to phenylalanine and tyrosine, which are T4 lysozyme (PDB ID: 3HWL) (Fleissner et al., 2009), CaM-peptide (PDB ID: 6HCS) (Creon et al., 2018), modified threonyl-tRNA synthetase (PDB ID: 4S0I) (Pearson et al., 2015), sphingosine-1-phosphate lyase (PDB ID: 3MBB) (Bourquin et al., 2010), birch pollen allergen Bet v 1.0101 (PDB ID: 4B9R) (Ackaert et al., 2014), ketosteroid isomerase (PDB ID: 5D82) (Wu et al., 2015), and acetyltransferase (PDB ID: 2Z10) (Sakamoto et al., 2009). Each protein mainly contains one UAA, dominated by secondary structures of α -helices, β -sheets, and γ -turns made of natural amino acids. Among them, the UAAs ACF131, AZF108, NIY150, CHY16, and IOY111 incorporated in the T4 lysozyme, CaM-peptide, Bet v 1.0101, ketosteroid isomerase,

and acetyltransferase, respectively, are mainly located at the α -helices. BFA11 and NIY5, 66, and 83 of the threonyl-tRNA synthetase and Bet v 1.0101 are distributed in the β -sheet regions, and AMY249 of sphingosine-1-phosphate lyase is located at the γ -turn. The remaining five systems involving UAAs recorded in the PDB are in regard to the protein–ligand interactions (**Supplementary Table 1**) and will be discussed in the next section, “MM/PBSA analysis of protein–UAA interactions.”

Here, we show each UAA fragment from the final MD structure compared with the crystal structure in **Figure 4**, and **Table 3** displays the averaged heavy-atom RMSD values of the UAAs and corresponding proteins in isolated systems. As shown in **Figure 4**, the backbone and side chains of the UAAs in the isolated proteins are generally well overlapped with the experimental structures, although there are slight structural derivations in the fragments of the ACF131 backbone and NIY66 side chain. **Table 3** also shows that the averaged RMSDs for ACF131 and NIY66 are the largest at 1.31 ± 0.11 and 0.95 ± 0.20 Å, respectively, corresponding to moderate RMSDs of 1.94 ± 0.24 and 2.43 ± 0.29 Å for their whole proteins. Simultaneously, we inspected the UAA motions by referring to one equilibrium structure during MD simulations, as listed in **Table 3** (column 3). Almost all the RMSDs of the UAAs are under 0.5 Å, with only NIY5 showing a larger RMSD of 0.66 ± 0.30 Å. In addition, we plotted the RMSD distributions for each UAA in the isolated protein systems compared with the crystal structures over time in **Supplementary Figure 5**. As shown, each trajectory reaches a balance after 20 ns MD simulations. However, the RMSD of NIY5 decreased by <0.5 Å between 5 and 10 ns. After 10 ns, all the UAAs reached equilibrium with RMSDs under 1.5 Å.

Additionally, the backbone conformations of seven UAAs in their isolated proteins obtained from the new parameters during the MD simulations were further investigated (**Figure 5**). As shown in **Figure 5E**, only NIY5 of birch pollen allergen Bet v 1.0101 (PDB ID: 4B9R) inclines toward the more stretched β -sheet backbone conformation during the MD simulation (black symbols). Compared to the crystal structures, the calculated backbone torsions of the remaining UAAs are generally well

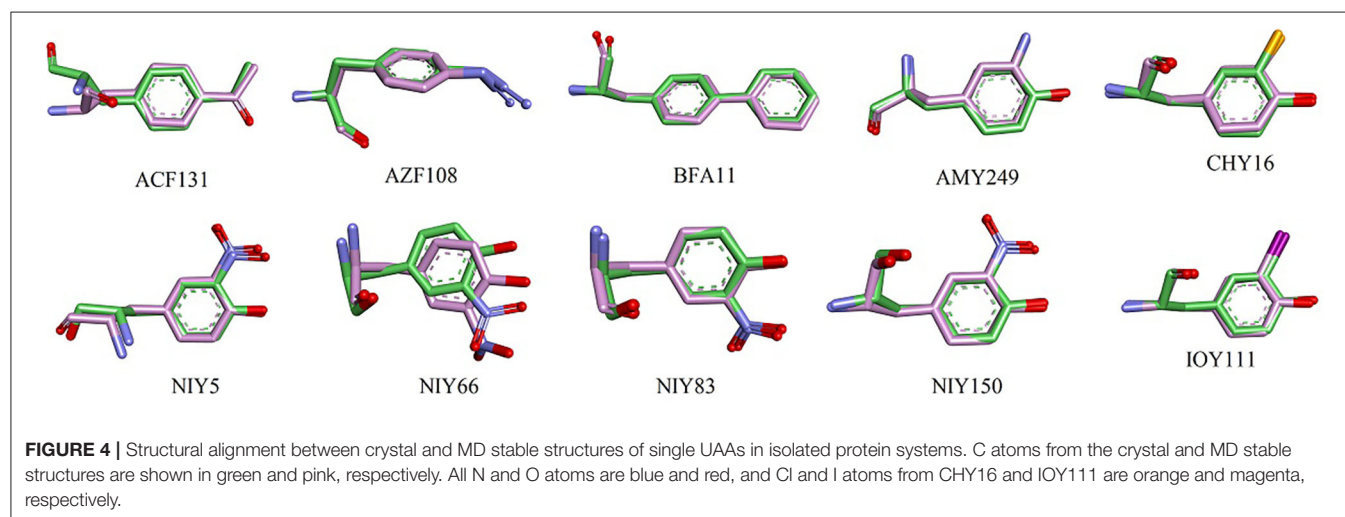
consistent. The backbones of ACF131, AZF108, AMY249, NIY150, CHY16, and IOY111 are in the form of α -helices, while those of BFA and NIY5, 66, and 83 are formed by β -strands.

MM/PBSA Analysis of Protein–UAA Interactions

Aside from the isolated proteins containing UAAs found in the PDB search, the UAAs were resolved as a ligand role in protein–UAA interactions (Turner et al., 2006; Moor et al., 2011; Takimoto et al., 2011; Li et al., 2013). To evaluate the quality of the new parameters determined in this work, five systems of protein–UAA interactions were studied by MM/PBSA analysis. The complexes are *p*-bromo-l-phenylalanine (BRF) bound to aaRS (PDB ID: 2AG6) (Turner et al., 2006); tRNA^{Phe} with 3,4-dihydroxy-l-phenylalanine (DHF) (PDB ID: 3TEG) (Moor et al., 2011); evolved PylRS charged with *o*-methyl-l-tyrosine (OMY) (PDB ID: 3QTC) (Takimoto et al., 2011); tyrosine-tRNA ligase mutant complexed with 3-methyl-tyrosine (MEY) (PDB ID: 4HPW); and 3,5-difluoro-l-tyrosine (DFY) incorporated into

TABLE 3 | Averaged heavy-atom RMSDs (Å) with standard errors of the mean for the UAAs and corresponding proteins in isolated systems.

UAA	RMSD (Å)			PDB ID
	UAA fit to crystal UAA	UAA fit to MD UAA	Protein fit to backbone	
ACF131	1.31 ± 0.11	0.39 ± 0.39	1.94 ± 0.24	3HWL
AZF108	0.75 ± 0.11	0.28 ± 0.13	3.44 ± 0.31	6HCS
BFA11	0.22 ± 0.06	0.23 ± 0.07	1.75 ± 0.15	4S0I
AMY249	0.42 ± 0.12	0.23 ± 0.09	4.36 ± 0.48	3MBB
NIY5	0.72 ± 0.29	0.66 ± 0.30	2.43 ± 0.29	4B9R
NIY66	0.95 ± 0.20	0.40 ± 0.25	2.43 ± 0.29	4B9R
NIY83	0.30 ± 0.09	0.23 ± 0.08	2.43 ± 0.29	4B9R
NIY150	0.28 ± 0.07	0.31 ± 0.09	2.43 ± 0.29	4B9R
CHY16	0.26 ± 0.01	0.24 ± 0.10	2.80 ± 0.25	5D82
IOY111	0.15 ± 0.04	0.14 ± 0.04	2.02 ± 0.20	2Z10



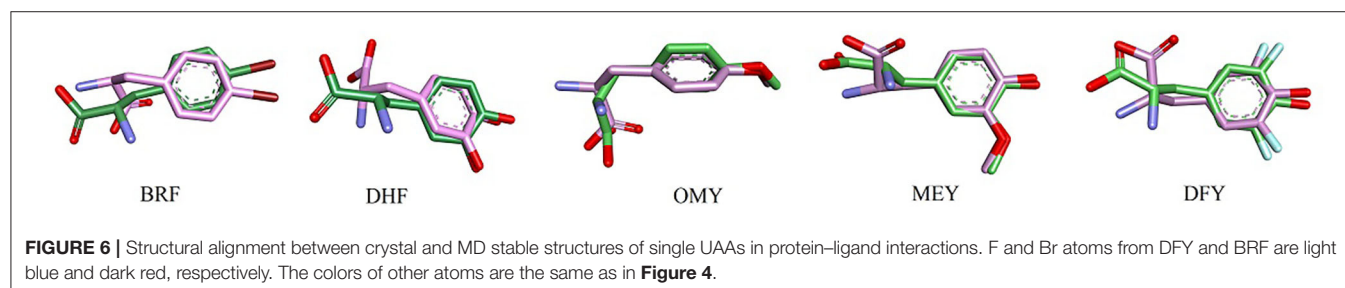
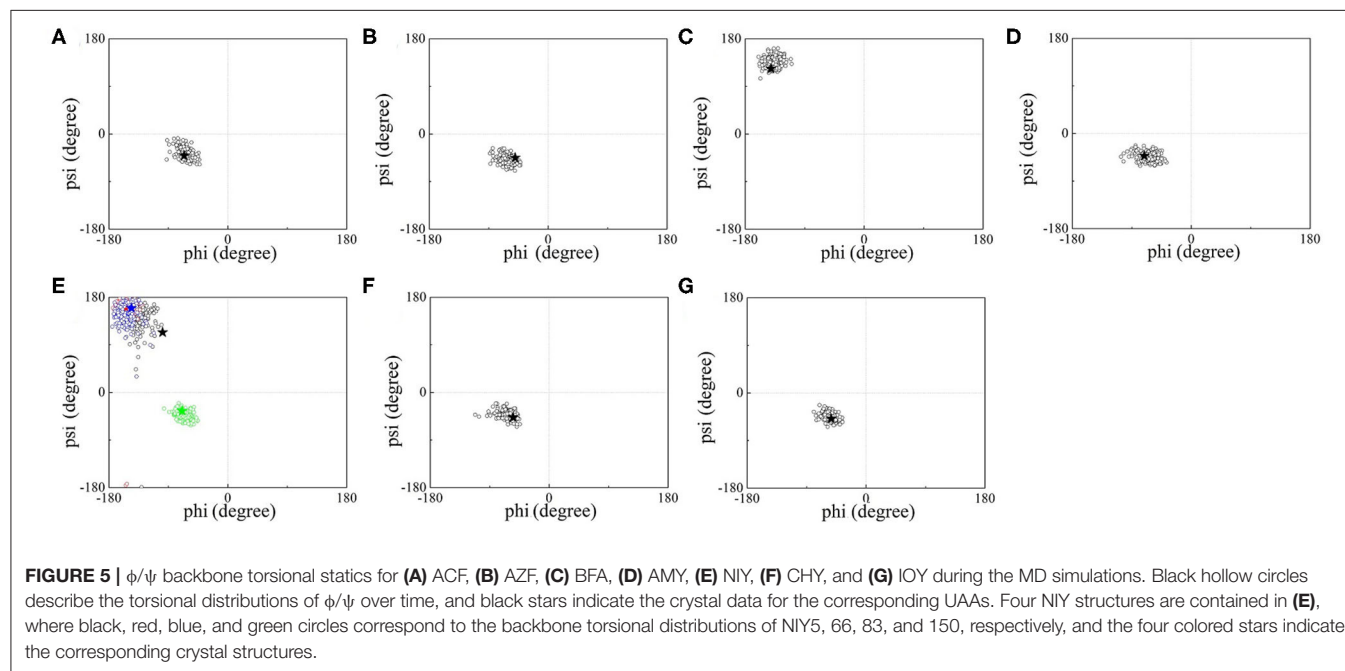
tyrosine phosphorylation (PDB ID: 4HJX) (Li et al., 2013). In addition, we added H and OH groups to the -NH and -C=O termini, respectively, to achieve neutral UAA ligands. We made minor modifications to the charge parameters of the terminal H and OH groups; the modified terminal charges for the H and OH groups of the five ligands BRF, DHF, OMY, MEY, and DFY are listed in **Supplementary Table 3**. These charge parameters should be more appropriate for UAAs when they are treated as ligands.

Compared with the UAAs incorporated into isolated proteins, the UAAs involved as substrates in protein–ligand interactions seem to shift more obviously, particularly the backbone structures (**Figure 6**). This may be due to the flexible UAA structures acting as ligands to bind with the proteins. The average RMSD values were also calculated to be larger, around 1.3 Å, as listed in **Table 4**. After choosing one equilibrium MD structure as the reference, the averaged RMSDs for all UAA ligands decreased to below 1.0 Å, suggesting good stability in the simulation process. **Supplementary Figure 6** plots the RMSD distributions as a function of time for the five ligands BRF, DHF, OMY, MEY, and DFY during the MD simulation starting from the experimental structure set. As shown, the

DHF, OMY, MEY, and DFY ligands were well-balanced after 3 ns, whereas BRF reached another stable state after 10 ns. The RMSD values of all the UAA ligands are in the vicinity of 1.5 Å, showing stable movements over the initial structures. The final whole structures also overlap well with the crystal structures, as depicted in **Supplementary Figures 7H–L**, indicating that our new parameters can reproduce the experimental structures of these protein–UAA interactions.

TABLE 4 | Averaged heavy-atom RMSDs (Å) with standard errors of the mean for UAAs and corresponding proteins in protein–ligand complexes.

UAA	RMSD			PDB ID
	UAA fit to crystal UAA	UAA fit to MD UAA	Protein fit to backbone	
BRF	0.96 ± 0.37	0.83 ± 0.51	2.23 ± 0.15	2AG6
DHF	1.33 ± 0.16	0.81 ± 0.49	2.42 ± 0.15	3TEG
OMY	1.33 ± 0.27	0.87 ± 0.46	2.38 ± 0.20	3QTC
MEY	1.31 ± 0.38	0.61 ± 0.44	2.44 ± 0.22	4HPW
DFY	1.11 ± 0.27	0.47 ± 0.31	3.29 ± 0.30	4HJX



Further, we used the MD structures obtained from the new parameters to calculate the five protein-UAA complexes. **Table 5** shows the binding free energies of aaRS-BRF (PDB ID: 2AG6), tRNA^{Phe}-DHF (PDB ID: 3TEG), PylRS-OMY (PDB ID: 3QTC), tRNA^{Tyr}-MEY (PDB ID: 4HPW), and tyrosine phosphorylation (F2YRS)-DFY (PDB ID: 4HJX) based on MM/PBSA analysis. The binding free energy of 3TEG is the highest at -21.3 kcal/mol, while the weakest binding affinity of -6.4 kcal/mol corresponds to 2AG6. As mentioned above, the RMSD of BRF reaches another local equilibrium within a period of 12–20 ns. To check the convergence of the MM/PBSA calculation, the results were usually estimated from different time intervals (Spiliotopoulos et al., 2012). As shown in **Supplementary Table 4**, the binding free energy of 2AG6 estimated using the 12–20 ns trajectory is -7.1 kcal/mol, which is largely consistent with the one obtained using entire trajectory and the one in an early stage. We also predicted the binding free energies of tRNA^{Tyr}-MEY (4HPW) and tyrosine phosphorylation-DFY (4HJX) as -17.3 and -13.6 kcal/mol, respectively. The energy decomposition analysis also indicates that vdW and electrostatic interactions are the dominate factors contributing to the total binding free energy. The polar energies contribute positively to the solvation. Overall, the binding free energies for the five systems were well stabilized by the MM contributions.

In addition, the binding affinities from experimental data for two of the complexes are available. One is the reported Michaelis constant K_M for the system tRNA^{Phe}-DHF (3TEG) of $380 \pm 40 \mu\text{M}$ (Moor et al., 2011), which determines the performance of the catalytic reaction and positively correlates with the dissociation constant K_d (Johnson and Goody, 2011). The other is for OMY as one of the compstatin variants reported as $K_d = 118$ nM and $\Delta G = -9.5 \pm 1.2$ kcal/mol (Magotti et al., 2009). The binding free energy of PylRS and OMY interaction is predicted to be -15.7 ± 0.7 kcal/mol by MM/PBSA, which is in satisfying agreement with the experimental one. No experimental

data of the binding affinity is available to date for the other three protein-UAA systems (PDB IDs: 2AG6, 4HPW, and 4HJX). Nevertheless, MM/PBSA analysis has been demonstrated to be an effective approach to estimate qualitatively the relative binding free energy of protein-ligand interaction (Homeyer and Gohlke, 2012; Kumari et al., 2014; Genheden and Ryde, 2015; Wang et al., 2019).

Per-Residue Energy Decomposition Analysis of Protein-UAA Interactions

The structural interaction modes between UAAs and proteins have been established by experimental reports (Turner et al., 2006; Moor et al., 2011; Takimoto et al., 2011; Li et al., 2013). We show the interaction details of the UAAs BRF, DHF, OMY, MEY, and DFY as substrates bound to the respective proteins in **Figure 7**. The per-residue binding free energies of the major contacts involved in the interactions are provided in **Table 6**. The interactions of the UAAs as substrates are discussed in the following sections.

aaRS and BRF interactions

In the 2AG6 system, our parameters predicted several direct connections of C-halogen-bonding interactions, which are consistent with the experimental results (**Figure 7A**). For example, the bromine of BRF forms a C-Br $\cdots \pi$ interaction with WT H160, which has been extensively reported in the crystal structures of protein-small molecules (Saraogi et al., 2003; Turner et al., 2006). One crystal structure report showed that the mutant L32 is a key mutant residue providing binding room for the bromine without vdW contributions (Turner et al., 2006). Here, the small contribution is -0.45 kcal/mol of free energy as predicted by our new parameters (see **Table 6**). In addition, we did not observe obvious contact between WT Y161 and BRF, and the predicted binding free energy was 0.47 kcal/mol, with weak contributions from MM, polar, and nonpolar interactions of -0.84 , 1.47 , and -0.17 kcal/mol, respectively. This is consistent with the experimental finding that the O atom of Y161 is too far (4.6 Å) to form H-bonded contact with the Br-atom of BRF in the active loop (Turner et al., 2006). In addition, two potential H-bonded contacts that have not been anticipated experimentally are predicted here. In particular, WT E36 and WT Q173 of 2AG6 use side chains to combine with the amide group of BRF in the form of H-bonds. As shown in **Table 6**, Q173 produces stronger polar interactions than the MM component, leading to a positive contribution of 5.46 kcal/mol. Strong electrostatic and vdW interactions of -6.29 and -4.31 kcal/mol were calculated for both E36 and Q173, respectively, which provide important conditions for H-bond formation (Li et al., 2014; Hao and Wang, 2015).

tRNA^{Phe} and DHF interactions

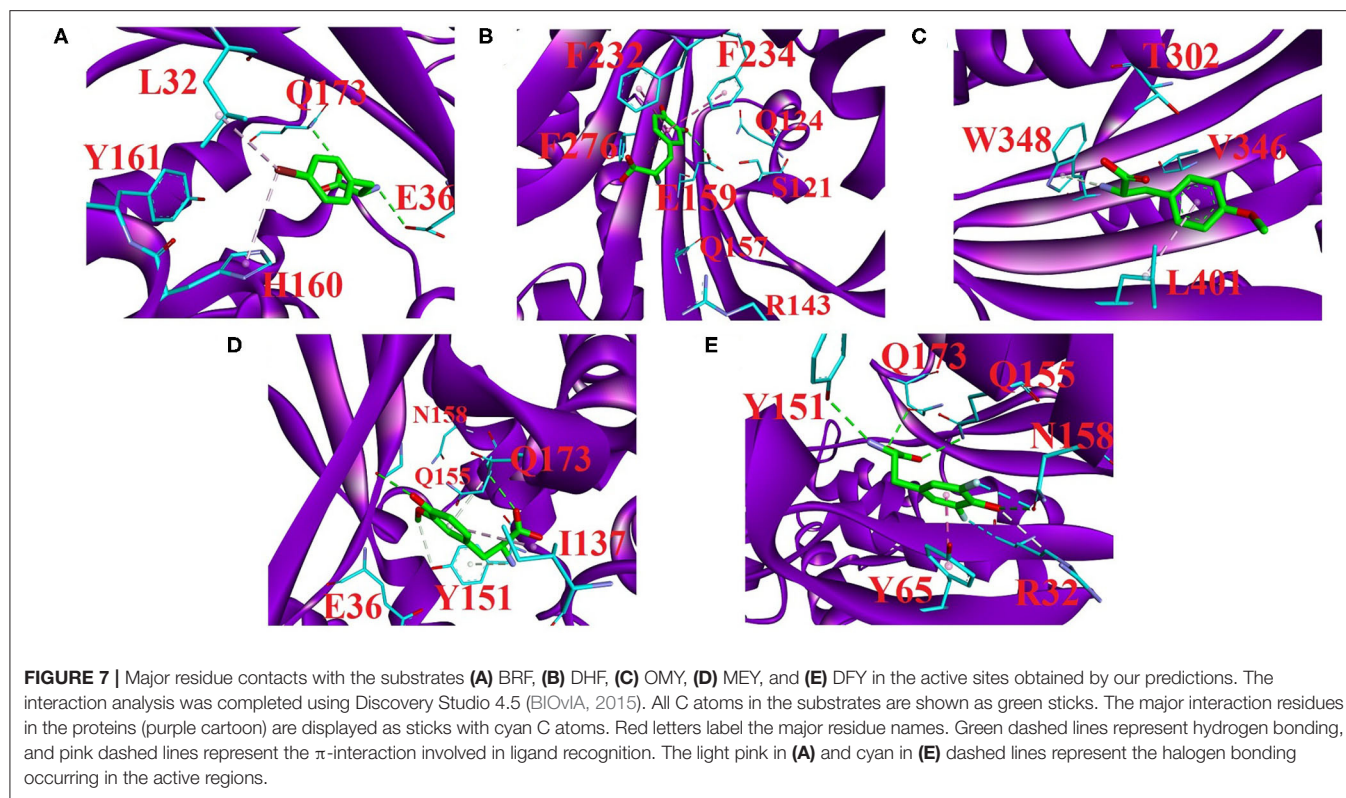
We provide the structural basis of the reported 3TEG (tRNA^{Phe} binding with DHF) in **Figure 7B**. F232 and F234 located at the FPF loop maintain major contacts with the phenyl ring of the ligand DHF (Moor et al., 2011). This was also observed from our predictions between F232/F234 and DHF in the form of $\pi \cdots \pi$ interactions. Simultaneously, F276 has a novel predicted role

TABLE 5 | Binding free energies (kcal/mol) with standard deviation^a for the systems 2AG6, 3TEG, 3QTC, 4HPW, and 4HJX obtained from MM/PBSA calculations and various energy components.

Component	2AG6	3TEG	3QTC	4HPW	4HJX
ΔE_{vdW}	-25.2 (0.4)	-24.3 (0.5)	-20.6 (1.4)	-25.4 (0.9)	-26.4 (0.6)
ΔE_{ele}	-11.8 (1.5)	-35.6 (1.7)	-12.7 (1.5)	-20.2 (1.9)	-18.8 (1.5)
$\Delta G_{pb/solv}$	33.6 (1.5)	41.4 (1.4)	20.0 (1.1)	31.1 (1.6)	34.7 (1.5)
$\Delta G_{np/solv}$	-3.0 (0.0)	-2.7 (0.0)	-2.4 (0.1)	-2.9 (0.0)	-3.1 (0.0)
ΔG_{pb}	21.8 (0.0)	5.8 (0.2)	7.3 (0.2)	10.9 (0.2)	15.9 (0.0)
ΔG_{np}	-28.2 (0.2)	-27.0 (0.3)	-23.0 (0.8)	-28.3 (0.6)	-29.5 (0.3)
ΔE_{MM}	-37.0 (0.7)	-59.9 (0.8)	-33.3 (0.0)	-45.6 (0.6)	-45.2 (0.5)
ΔG_{solv}	30.6 (0.9)	38.7 (0.8)	17.6 (0.6)	28.2 (0.9)	31.6 (0.9)
ΔG_{bind}	-6.4 (0.8)	-21.2 (1.1)	-15.7 (1.0)	-17.4 (1.1)	-13.6 (0.8)

^aThe standard deviations are calculated by the equation $SD = \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 / N}$.

Here, $\Delta G_{pb} = \Delta G_{pb/solv} + \Delta E_{ele}$; $\Delta G_{np} = \Delta G_{np/solv} + \Delta E_{vdW}$; $\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele}$; $\Delta G_{solv} = \Delta G_{pb/solv} + \Delta G_{np/solv}$.



involved in DHF binding through an amide $\cdots \pi$ interaction (see **Supplementary Figure 8A**). These π -interaction modes formed by F232, F234, and F276 are similar to the reported “edge-to-face” contact (Fishman et al., 2001), an interaction network formed by three phenylalanine in tRNA^{Phe} binding with the phenyl moiety of the DHF ligand. The $\pi \cdots \pi$ and amide $\cdots \pi$ interactions mainly originate from vdW contributions (Gao et al., 2017). As shown in **Table 6**, the total vdW and electrostatic contributions of F232, F234, and F276 are all more negative than -4.0 kcal/mol. Furthermore, the binding free energy of E159 is a remarkable -16.57 kcal/mol, with surprisingly large non-bonded and polar contributions of -55.39 and 39.20 kcal/mol, respectively. As shown in **Figure 7B**, one hydrogen bonding connection occurs through the side-chain O atom of E159 with the negative charge binding to the OH group of DHF. Additionally, H-bonded connections have been reported between S121, Q124, R143, and Q157 in the protein and DHF shown in **Supplementary Figure 8B** (Moor et al., 2011). However, we failed to observe these hydrogen bonding contacts. Per-residue energy decomposition analysis further indicates that only R143 and Q157 provide dispensable non-bonded interactions of -3.66 and -8.86 kcal/mol, respectively. The contributions of S121 and Q124 are almost too weak for binding.

PylRS and OMY interactions

The four residue mutations in PylRS are A302T, N346V, C348W, and V401L, which play a vital role in the OMY selectivity (Takimoto et al., 2011). We also predicted these four important

residue contacts with OMY based on the new parameters and per-residue binding free energy analysis. **Figure 7C** shows the interaction modes between OMY and the four residues T302, V346, W348, and LV401, and the binding free energy of each residue (PDB ID: 3QTC) is listed in **Table 6**. As shown, W348 uses a side-chain 5-membered ring as a π -donor to form hydrogen bonds with the N-atom in the amide group of OMY. This results in one quadrupole–dipole interaction formed by the indole plane of W348 being vertical to the O-methyl moiety of OMY (Takimoto et al., 2011). The binding free energy of W348 is -2.99 kcal/mol, providing strong vdW and electrostatic interactions of -4.53 kcal/mol. In the activation region, alkyl $\cdots \pi$ interactions occur by the methylene group of L401 binding with OMY, with the highest binding affinity contribution of -3.91 kcal/mol. Even though no direct connection forms between T302 and OMY, a moderate impact with a -2.11 kcal/mol binding free energy was evaluated, which also provides strong electrostatic and vdW interactions of -7.32 kcal/mol. In addition, the binding contribution of V346 is mainly derived from electrostatic and vdW contributions at -2.03 kcal/mol, but we did not observe hydrogen bonding between them. This is in agreement with the experimental observation that the H-bonds formed by WT N346 and OMY are abolished after the N346V mutation in PylRS (Takimoto et al., 2011).

tRNA^{Tyr} and MEY interactions

The structural basis for MEY recognition to tRNA^{Tyr} has not been reported to date, but the binding modes of the

TABLE 6 | Energy decomposition analysis of 2AG6, 3TEG, 3QTC, 4HPW, and 4HJX for major residues.

PDB ID	#Residue	Energy component (kcal/mol)				Standard deviation (kcal/mol)			
		MM	Polar	Non polar	Total	MM	Polar	Non polar	Total
2AG6	L32	-1.26	1.15	-0.34	-0.45	0.38	0.31	0.03	0.48
	E36	-6.29	5.77	-0.33	-0.78	2.98	2.70	0.05	0.99
	H160	-2.24	4.85	-0.21	2.39	0.32	0.83	0.04	0.68
	Y161	-0.84	1.47	-0.17	0.47	0.12	0.20	0.02	0.19
	Q173	-4.31	10.22	-0.44	5.46	0.73	1.43	0.05	0.82
3TEG	S121	0.10	-0.40	-0.07	-0.36	0.21	0.38	0.02	0.25
	Q124	1.96	-0.39	-0.02	1.53	0.38	0.58	0.01	0.30
	R143	-3.66	7.75	-0.08	3.91	1.74	1.60	0.03	0.69
	L157	-8.86	10.66	-0.58	1.18	1.12	0.85	0.04	0.75
	E159	-55.39	39.20	-0.38	-16.57	1.94	2.80	0.02	1.69
	F232	-4.07	2.03	-0.47	-2.53	0.60	0.38	0.03	0.35
	F234	-4.48	3.39	-0.55	-1.66	0.35	0.24	0.04	0.23
	F276	-6.78	3.94	-0.09	-2.91	0.40	0.40	0.02	0.45
3QTC	T302	-7.32	5.55	-0.39	-2.11	1.31	0.94	0.06	0.47
	V346	-2.03	0.26	-0.08	-1.84	0.34	0.09	0.02	0.38
	W348	-4.53	1.68	-0.15	-2.99	1.16	0.44	0.04	0.75
	L401	-4.10	0.68	-0.50	-3.91	0.77	0.19	0.07	0.67
4HPW	E36	-14.75	9.85	-0.52	-5.34	1.79	1.69	0.07	1.21
	I137	-4.56	0.75	-0.62	-4.43	0.37	0.35	0.04	0.25
	Y151	-13.06	7.57	-0.68	-6.17	0.95	0.78	0.06	1.09
	Q155	-9.69	8.13	-0.32	-1.90	1.19	0.96	0.05	0.45
	N158	-0.17	0.49	-0.02	0.31	0.21	0.49	0.02	0.34
	Q173	-17.58	16.48	-0.81	-1.97	1.55	0.87	0.06	1.02
4HJX	R32	-7.32	8.48	-0.27	0.82	0.71	1.19	0.03	0.98
	Y65	-7.78	4.41	-0.58	-3.91	0.68	0.47	0.05	0.56
	Y151	-6.95	6.62	-0.45	-0.76	0.65	0.62	0.05	0.50
	Q155	-23.04	21.54	-1.07	-2.59	1.59	1.62	0.06	0.67
	N158	-7.46	5.45	-0.28	-2.30	1.41	0.68	0.02	0.96
	Q173	-12.13	12.83	-0.50	0.20	1.49	1.17	0.03	0.71

Bold letters represent residue contacts from experimental reports.

tRNA^{Tyr}-MEY interaction can be analyzed and determined using the Mol* tool provided in the PDB (Sehna et al., 2018). Accordingly, E36, Y151, Q155, N158, and Q173 are the main residue contacts with MEY. **Table 6** shows that the binding free energies of these residues provide negative contributions of -1.90 to -6.17 kcal/mol, except for N158 with 0.31 kcal/mol. **Figure 7D** displays the hydrogen bonding and alkyl $\cdots \pi$ interaction network between Y151, Q155, Q173, and MEY. Even though E36 does not form hydrogen bonding with MEY, a -5.34 kcal/mol strong affinity is derived from vdW and electrostatic attractions of -14.75 kcal/mol. Furthermore, I137 shows a new potential contact with MEY via an alkyl $\cdots \pi$ interaction with a -4.34 kcal/mol binding free energy.

F2YRS and DFY interactions

F2YRS shares approximately identical sequences with tRNA^{Tyr} except for the asparagine and cysteine at positions 108 and 109, respectively, corresponding to F108 and G109 in tRNA^{Tyr}. The complex of F2YRS-DFY was obtained after Y32R, L65Y, H70G, F108N, Q109C, D158N, and L162S mutations by an experimental technique (Li et al., 2013). We assumed DFY to be in a neutral state due to the reported pKa value close to 7.0 (Seyedsayamdost et al., 2006). **Figure 7E** shows the six key residues binding to the DFY substrate. R32 and N158 form halogen bonding with the two different fluorine atoms of DFY; meanwhile, hydrogen bonding of R32 and N158 occurs with the OH group of DFY. This is consistent with experimental findings (Li et al., 2013). Experiments have also shown that there are strong dipolar interactions between the fluorine atoms and amide/guanidine groups. Notable polar contributions of 8.48 and 5.45 kcal/mol are estimated by our predictions for R32 and N158, respectively. In addition, Y65 forms $\pi \cdots \pi$ stacking interactions with the phenyl group of DFY, and Y151, Q155, and Q173 form hydrogen bonds with the amide and carbonyl groups of DFY. Among them, Q155 provides the largest MM contribution of -23.04 kcal/mol with 21.54 kcal/mol of polar energy. Y65 and Q155 with -3.91 and -2.59 kcal/mol free energies, respectively, contribute moderately to the observed binding.

CONCLUSION

This work presents the charge parameters of 18 UAAs related to phenylalanine and tyrosine that are compatible with the use of the Amber ff14SB force field included in the GROMACS package. The newly derived charge parameters initially fitted by the RESP protocol were tested on structural optimizations and relative energies of the 18 UAAs in α -/ β -backbone conformations, with an RMS deviation of 0.33 kcal/mol compared with the QM dataset, whereas the M06-2X method produces an RMS deviation of 1.08 kcal/mol. After the parameters were determined, the energy function was further applied to MD simulations of the UAA-mutated proteins and protein-UAA complexes. The motifs containing UAAs and their respective backbone torsions generally overlapped well with the initial coordinates, with an average RMSD of approximately 1.5 Å. The MM/PBSA approach showed that the binding free energy of tRNA^{Phe}-DHF is higher than that of PyIRS-OMY, which is consistent with experimental data. Comparisons with crystal residue contacts and satisfactory treatments for the interaction modes between proteins and UAAs by substrate binding are presented from the analysis of the per-residue energy decomposition.

Nevertheless, the development of force field is too far from only the development of charge parameters. To increase the transferability and compatibility to the standard Amber force field, the atoms in the common structure of these UAAs should be optimized to be of identical partial charges by applying restraints/constraints in the fitting to RESP in a future study. The bonded parameters, especially the torsional terms related to the gas-phase QM conformational potential energy scan, require further adjustment. The current testing concentrated

on conformational and energetic investigations is also limited, and thus more extensive studies focusing on the dynamic and thermodynamic properties of polypeptides and proteins should be explored.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

WL proposed the core idea of the study. XW designed the project, performed the research, analyzed the data

and drafted the manuscript. Both authors critically reviewed the manuscript.

FUNDING

This work was financially supported by the National Natural Science Foundation of China under grant number 31770777, Start-up Foundation for Peacock Talents (827-000365), and the Start-up Grant for Young Scientists (860-000002110384), Shenzhen University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.608931/full#supplementary-material>

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001
- Ackaert, C., Kofler, S., Horejs-Hoek, J., Zulehner, N., Asam, C., von Grafenstein, S., et al. (2014). The impact of nitration on the structure and immunogenicity of the major birch pollen allergen Bet v 1.0101. *PLoS ONE* 9:e104520. doi: 10.1371/journal.pone.0104520
- Andersen, H. C. (1983). Rattle: a “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52, 24–34. doi: 10.1016/0021-9991(83)90014-1
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ 1 and χ 2 dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x
- BIOvIA, D. S. (2015). *Discovery Studio Modeling Environment*. San Diego: Dassault Systemes. Release. 4.
- Bourquin, F., Riezman, H., Capitani, G., and Grütter, M. G. (2010). Structure and function of sphingosine-1-phosphate lyase, a key enzyme of sphingolipid metabolism. *Structure* 18, 1054–1065. doi: 10.1016/j.str.2010.05.011
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101. doi: 10.1063/1.2408420
- Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., et al. (2020). *AMBER 2020*. San Francisco: University of California.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002
- Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993). Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* 115, 9620–9631. doi: 10.1021/ja00074a030
- Creon, A., Josts, I., Niebling, S., Huse, N., and Tidow, H. (2018). Conformation-specific detection of calmodulin binding using the unnatural amino acid p-azido-phenylalanine (AzF) as an IR-sensor. *Struct. Dyn.* 5:064701. doi: 10.1063/1.5053466
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Dennington, R., Keith, T. A., and Millam, J. M. (2016). *GaussView 6*. Shawnee Mission, KS Semichem Inc.
- Fishman, R., Ankilova, V., Moor, N., and Safro, M. (2001). Structure at 2.6 Å resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese. *Acta Crystallogr. D* 57, 1534–1544. doi: 10.1107/S090744490101321X
- Fleissner, M. R., Brustad, E. M., Kálai, T., Altenbach, C., Cascio, D., Peters, F. B., et al. (2009). Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl. Acad. Sci.* 106:21637. doi: 10.1073/pnas.0912009106
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16 Rev. C.01*. Wallingford, CT: Gaussian, Inc.
- Gao, X.-C., Hao, Q., and Wang, C.-S. (2017). Improved polarizable dipole-dipole interaction model for hydrogen bonding, stacking, T-shaped, and X-H... π interactions. *J. Chem. Theory Comput.* 13, 2730–2741. doi: 10.1021/acs.jctc.6b00936
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461. doi: 10.1517/17460441.2015.1032936
- Hao, J.-J., and Wang, C.-S. (2015). Rapid evaluation of the interaction energies for carbohydrate-containing hydrogen-bonded complexes via the polarizable dipole-dipole interaction model combined with NBO or AM1 charge. *RSC Adv.* 5, 6452–6461. doi: 10.1039/C4RA12814A
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12, 281–296. doi: 10.1021/acs.jctc.5b00864
- Homeyer, N., and Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Mol. Inform.* 31, 114–122. doi: 10.1002/minf.201100135
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* 268:1144. doi: 10.1126/science.7761829
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123
- Johnson, K. A., and Goody, R. S. (2011). The original michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* 50, 8264–8269. doi: 10.1021/bi201284u
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Khoury, G. A., Bhatia, N., and Floudas, C. A. (2014a). Hydration free energies calculated using the AMBER ff03 charge model for natural and unnatural amino acids and multiple water models. *Comput. Chem. Eng.* 71, 745–752. doi: 10.1016/j.compchemeng.2014.07.017

- Khoury, G. A., Smadbeck, J., Tamamis, P., Vandris, A. C., Kieslich, C. A., and Floudas, C. A. (2014b). Forcefield_NCAA: Ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synth. Biol.* 3, 855–869. doi: 10.1021/sb400168u
- Kumari, R., Kumar, R., and Lynn, A. (2014). g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54, 1951–1962. doi: 10.1021/ci500020m
- Leaverfay, A., Tyka, M. D., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487, 545–574. doi: 10.1016/B978-0-12-381270-4.00019-6
- Li, F., Shi, P., Li, J., Yang, F., Wang, T., Zhang, W., et al. (2013). A genetically encoded 19F NMR probe for tyrosine phosphorylation. *Angew. Chem. Int. Ed.* 52, 3958–3962. doi: 10.1002/anie.201300463
- Li, S.-S., Huang, C.-Y., Hao, J.-J., and Wang, C.-S. (2014). A polarizable dipole-dipole interaction model for evaluation of the interaction energies for N-H...O=C and C-H...O=C hydrogen-bonded complexes. *J. Comput. Chem.* 35, 415–426. doi: 10.1002/jcc.23473
- Liu, C. C., and Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444. doi: 10.1146/annurev.biochem.052308.105824
- Magotti, P., Ricklin, D., Qu, H., Wu, Y.-Q., Kaznessis, Y. N., and Lambris, J. D. (2009). Structure-kinetic relationship analysis of the therapeutic complement inhibitor compstatin. *J. Mol. Recognit.* 22, 495–505. doi: 10.1002/jmr.972
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Minnihan, E. C., Young, D. D., Schultz, P. G., and Stubbe, J. (2011). Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* 133, 15942–15945. doi: 10.1021/ja207719f
- Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi: 10.1002/jcc.540130805
- Mondal, S., Chowdhuri, D. S., Ghosh, S., Misra, A., and Dalai, S. (2007). Conformational study on dipeptides containing phenylalanine: a DFT approach. *J. Mol. Struct. THEOCHEM* 810, 81–89. doi: 10.1016/j.theochem.2007.02.006
- Moor, N., Klipcan, L., and Saforo, Mark, G. (2011). Bacterial and eukaryotic phenylalanyl-tRNA synthetases catalyze misaminoacylation of tRNA^{Phe} with 3,4-dihydroxy-L-phenylalanine. *Chem. Biol.* 18, 1221–1229. doi: 10.1016/j.chembiol.2011.08.008
- Nosé, S., and Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50, 1055–1076. doi: 10.1080/00268978300102851
- Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi: 10.1063/1.328693
- Pauling, L. (1939). *The Nature of the Chemical Bond*. Ithaca, NY: Cornell University Press.
- Pearson, A. D., Mills, J. H., Song, Y., Nasertorabi, F., Han, G. W., Baker, D., et al. (2015). Trapping a transition state in a computationally designed protein bottle. *Science* 347:863. doi: 10.1126/science.aaa2424
- Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C., and Zagrovic, B. (2013). A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* 9:e1003154. doi: 10.1371/journal.pcbi.1003154
- Qin, M., Li, F., Huang, Y., Ran, W., Han, D., and Song, Y. (2015). Twenty natural amino acids identification by a photochromic sensor chip. *Anal. Chem.* 87, 837–842. doi: 10.1021/ac504121d
- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99. doi: 10.1016/S0022-2836(63)80023-6
- Renfrew, P. D., Choi, E. J., Bonneau, R., and Kuhlman, B. (2012). Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS ONE* 7:e32637. doi: 10.1371/journal.pone.0032637
- Robertson, M. J., Tirado-Rives, J., and Jorgensen, W. L. (2015). Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* 11, 3499–3509. doi: 10.1021/acs.jctc.5b00356
- Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., et al. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 30, 4692–4699. doi: 10.1093/nar/gkf589
- Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., et al. (2009). Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* 17, 335–344. doi: 10.1016/j.str.2009.01.008
- Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20, 1044–1048. doi: 10.1038/nbt742
- Saraogi, I., Vijay, V. G., Das, S., Sekar, K., and Guru Row, T. N. (2003). C-halogen... π interactions in proteins: a database study. *Cryst. Eng.* 6, 69–77. doi: 10.1016/S1463-0184(03)00068-6
- Sehnal, D., Rose, A. S., Koča, J., Burley, S. K., and Velankar, S. (2018). “Mol*: towards a common library and tools for web molecular graphics,” in *Paper Presented at the Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data, Brno, Czech Republic*.
- Seyedsayamdost, M. R., Reece, S. Y., Nocera, D. G., and Stubbe, J. (2006). Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* 128, 1569–1579. doi: 10.1021/ja055926r
- Si, L., Xu, H., Zhou, X., Zhang, Z., Tian, Z., Wang, Y., et al. (2016). Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* 354:1170. doi: 10.1126/science.aah5869
- Sitkoff, D., Sharp, K. A., and Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98, 1978–1988. doi: 10.1021/j100058a043
- Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* 5:367. doi: 10.1186/1756-0500-5-367
- Spiliotopoulos, D., Spitaleri, A., and Musco, G. (2012). Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS ONE* 7:e46902. doi: 10.1371/journal.pone.0046902
- Takimoto, J. K., Dellas, N., Noel, J. P., and Wang, L. (2011). Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient *in vivo* incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* 6, 733–743. doi: 10.1021/cb200057a
- Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2006). Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc. Natl. Acad. Sci.* 103:6483. doi: 10.1073/pnas.0601756103
- Wang, C., Greene, D. A., Xiao, L., Qi, R., and Luo, R. (2018). Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* 4:87. doi: 10.3389/fmolb.2017.00087
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* 119, 9478–9508. doi: 10.1021/acs.chemrev.9b00055
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21, 1049–1074. doi: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., et al. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765–784. doi: 10.1021/ja00315a051
- Wu, Y., Fried, S. D., and Boxer, S. G. (2015). Dissecting proton delocalization in an enzyme's hydrogen bond network with unnatural amino acids. *Biochemistry* 54, 7110–7119. doi: 10.1021/acs.biochem.5b00958
- Xiao, H., Nasertorabi, F., Choi, S.-H., Han, G. W., Reed, S. A., Stevens, R. C., et al. (2015). Exploring the potential impact of an expanded genetic code on protein function. *Proc. Natl. Acad. Sci.* 112:6961. doi: 10.1073/pnas.1507741112

- Yang, T., Wu, J. C., Yan, C., Wang, Y., Luo, R., Gonzales, M. B., et al. (2011). Virtual screening using molecular simulations. *Proteins* 79, 1940–1951. doi: 10.1002/prot.23018
- Young, D. D., and Schultz, P. G. (2018). Playing with the molecules of life. *ACS Chem. Biol.* 13, 854–870. doi: 10.1021/acscchembio.7b00974
- Yuet, K. P., Doma, M. K., Ngo, J. T., Sweredoski, M. J., Graham, R. L. J., Moradian, A., et al. (2015). Cell-specific proteomic analysis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* 112:2705. doi: 10.1073/pnas.1421567112
- Yurieva, A. G., Poleshchuk, O. K., and Filimonov, V. D. (2008). Comparative analysis of a full-electron basis set and pseudopotential for the iodine atom in DFT quantum-chemical calculations of iodine-containing compounds. *J. Struct. Chem.* 49, 548–552. doi: 10.1007/s10947-008-0073-9
- Zhao, J., Burke, A. J., and Green, A. P. (2020). Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* 55, 136–144. doi: 10.1016/j.cbpa.2020.01.006
- Zou, H., Li, L., Zhang, T., Shi, M., Zhang, N., Huang, J., et al. (2018). Biosynthesis and biotechnological application of non-canonical amino acids: complex and unclear. *Biotechnol. Adv.* 36, 1917–1927. doi: 10.1016/j.biotechadv.2018.07.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.