

# Optimising the $\beta$ -lactam Parameters Using the Force Field Toolkit

Qiyang Wu<sup>1,+</sup>, Songyan Xia<sup>1,+</sup>, Frank Otto<sup>2</sup>, Tzong-Yi Lee<sup>3</sup>, Hsien-Da Huang<sup>3</sup>, and Ying-Chih Chiang<sup>1,\*</sup>

<sup>1</sup>Kobilka Institute of Innovative Drug Discovery, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Longgang District, Shenzhen, 518172, China

<sup>2</sup>Department of Chemistry, University College London, London, WC1H 0AJ, United Kingdom

<sup>3</sup>Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Longgang District, Shenzhen, 518172, China

\*Corresponding author: Ying-Chih Chiang, [chiangyc@cuhk.edu.cn](mailto:chiangyc@cuhk.edu.cn)

+these authors contributed equally to this work

## ABSTRACT

In this work we performed force field parameter optimisation for various  $\beta$ -lactams using the Force Field Toolkit (FFTK) and Gaussian calculations. Two problems that users are likely to encounter when optimising other drug molecules are identified. First, inappropriate CGenFF parameters without penalty prediction could cause a difficulty in optimisation. Second, multiple dihedral parameter sets may produce the same molecular mechanics (MM) potentials of similar quality, raising the question on how to choose the right parameter set as the best solution. A systematic protocol incorporating molecular dynamics simulations and a principle for selecting the dihedral phase shifts are introduced. Using the protocol, we then successfully optimised both neutral and anionic forms of penam and of cephem. Our results highlight the importance of selecting proper phase shifts during the dihedral optimisation, and the protocol proposed in this work is beneficial to other users.

## Introduction

Benefiting from the growth of computing power and from the development of highly accurate force fields, e.g. CHARMM<sup>1,2</sup>, AMBER<sup>3</sup>, OPLS-AA<sup>4,5</sup>, RSFF<sup>6,7</sup>, etc., molecular dynamics (MD) simulations are now a standard tool employed during computer-aided drug design<sup>8,9</sup>. From generating the ensemble protein conformations for virtual screening<sup>10,11</sup> to sampling microstates for binding affinity calculations<sup>12,13</sup>, (enhanced) MD simulations have been carried out in a broad range of scenarios. Apart from these examples, (enhanced) MD is also commonly used to reveal the interrelation between a protein's conformations and its function<sup>14,15</sup> as well as for exploring the ligand-protein interactions<sup>16</sup>.

For MD results to be reliable, accurate force fields must be employed. While accurate parameters for amino acids are provided in the standard force fields like CHARMM36<sup>1,2</sup> or AMBER ff19SB<sup>3</sup>, parameters are also available for a huge range of ligands, e.g. in CGenFF<sup>17</sup> for CHARMM or GAFF<sup>18</sup> for AMBER. However, owing to the complexity of the chemical space, optimised parameters can not be provided for all drug-like molecules. Instead, web services<sup>19–21</sup> and software tools<sup>22–24</sup> are available to help users predict and optimise parameters for the ligands of interest. For instance, if one would like to conduct a protein-ligand simulation using the CHARMM force field, one can obtain the ligand parameters from the ParamChem web server<sup>25</sup>, which will report the atom types, atom charges, and bonded parameters based on analogy with the CGenFF force fields. Penalties are then reported for these parameters based on the level of analogy. If the analogy is good, a low penalty is expected. Otherwise, a large penalty will be reported. The user should then further optimise those parameters with large penalties, using software such as the Force Field Toolkit (FFTK)<sup>22</sup>.

FFTK<sup>22</sup> is a VMD<sup>26</sup> plugin designed to assist a normal user to carry out the cumbersome ligand parameterization using CHARMM compatible ligand force fields. It provides a graphical user interface to generate input files for quantum calculations, to analyze the quantum calculation outputs, and to carry out the fitting of the molecular mechanics (MM) parameters to the quantum data. Currently, input/output to two quantum chemistry packages, Gaussian<sup>27</sup> and ORCA<sup>28</sup>, are provided. While FFTK and other software have greatly simplified the procedure of ligand parameterization, it can still be very challenging to perform this task for drug-like molecules. For instance, Pavlova and Gumbart employed FFTK to optimize the macrolides<sup>29</sup>, a group of antibiotics that target the protein synthesis in bacteria. They reported the problem that for the same dihedral angles, different multiplicities can reproduce the quantum potentials equally well. However, only the multiplicities from their own prediction, but not those from ParamChem, can reproduce the desired ligand conformation during MD simulations. The same

group recently carried out the parameterization of the drug molecule AT130<sup>30</sup>, and revealed many details that one should pay attention to during the optimisation process. The third example is the optimisation of norfloxacin, performed by Morado et al.<sup>23</sup> using their own software ParaMol<sup>31</sup>. For this molecule, an extensive 2-dimensional potential energy surface scan was performed for optimising the dihedral parameters. All these examples demonstrate that the optimisation of dihedral angle parameters is not as straight-forward as the other bonded parameters for bonds or angles, and therefore extra care must be taken in order to obtain reliable parameters.

From the chemical complexity, heterocyclic compounds are often considered difficult to parameterize. If a compound contains fused rings or spiro rings, this process can be even more challenging. Unfortunately, many drug molecules fall into this category, and hence require extensive force field parameterization. Apart from drug molecules, custom force fields for non-natural amino acids<sup>32,33</sup> and bacterial components<sup>34</sup> are not included in the standard force field databases and were constructed by experts additionally. While some of the parameters are available, users are still required to optimise many of the parameters by themselves. However, ligand parameterization remains challenging and requires a lot of experience, especially for those containing heterocyclic structures or fused rings. Hence, sharing protocols and lessons learned from optimising new ligands is still beneficial to other users.

In this work, we optimise the force field parameters for  $\beta$ -lactams, an important class of antibiotics that are composed of heterocyclic fused rings. We discuss the difficulties that one may encounter during the parameterization, and show that there is a systematic protocol to address these issues.

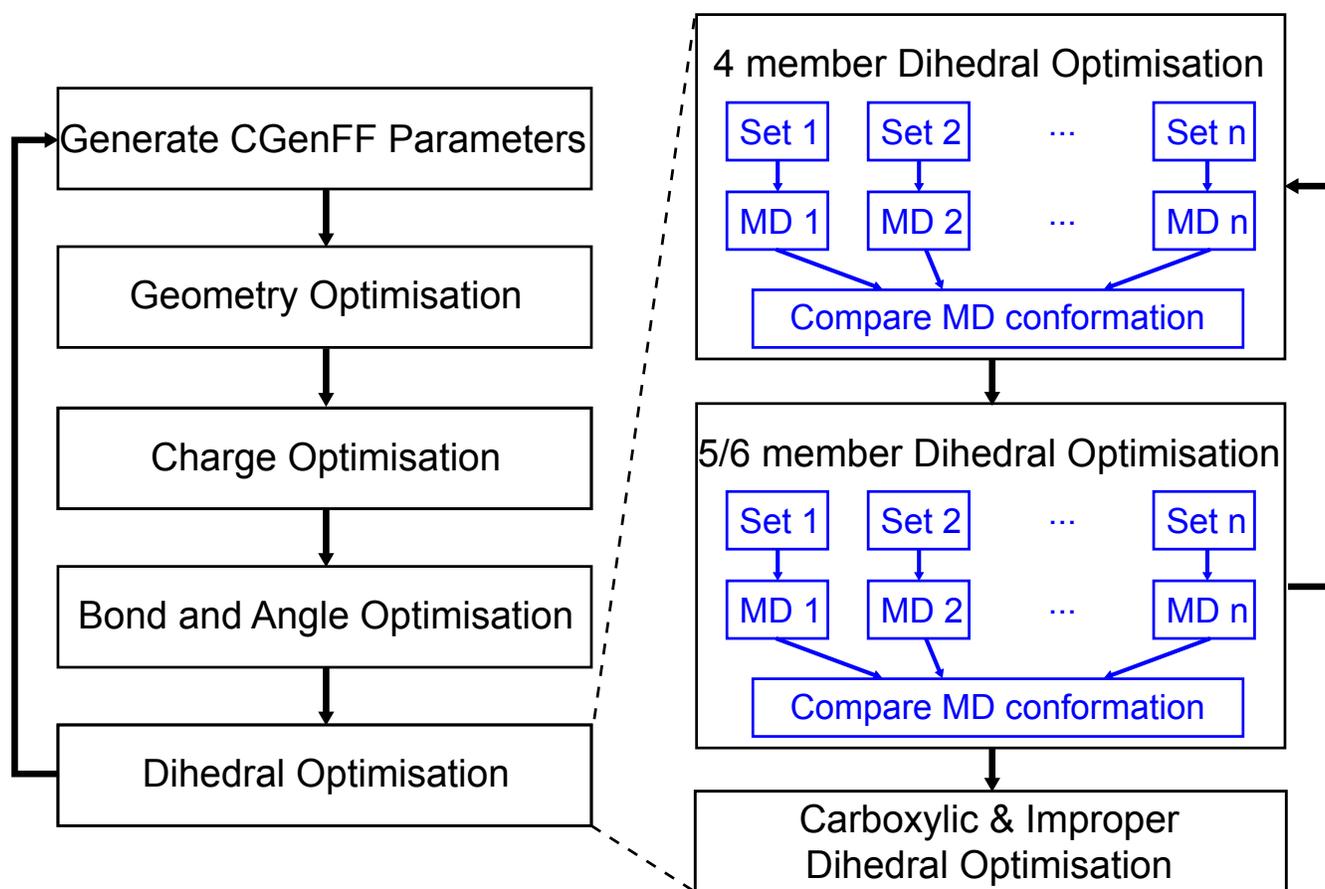
## Method

Our workflow follows the standard FTK parameterization scheme<sup>22</sup>, with a more rigorous protocol that incorporates information obtained from MD simulations to determine proper phase shifts for the dihedral angle parameters. Fig. 1 shows the entire workflow.

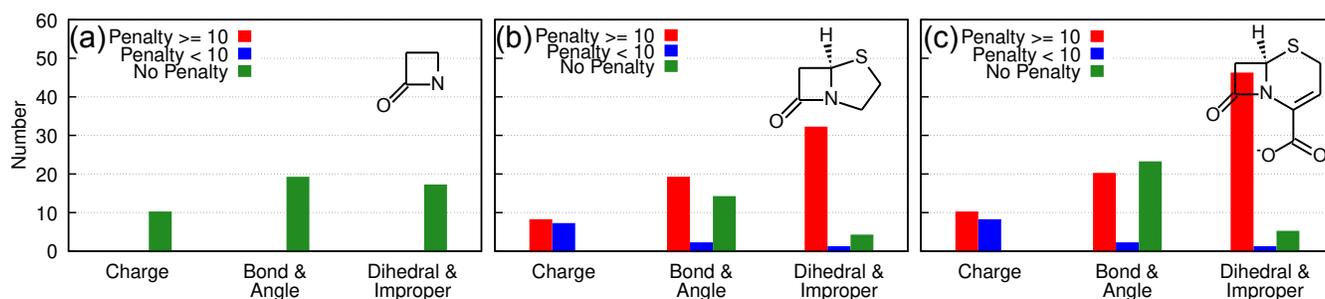
The parameterization begins with using ParamChem<sup>25</sup> to obtain CGenFF parameters for the ligand, for which the initial geometry can be taken from PubChem<sup>35</sup> or RCSB PDB<sup>36</sup>. ParamChem is a web server that automates the atom type mapping and determines the associated CGenFF force field parameters and their penalties based on analogy. The output, including the atom types, charges, bond and angle parameters, as well as dihedral and improper dihedral parameters, are listed in a *str* file. In general, a penalty score is listed for each parameter. Ranging from 0 to 10, the penalty shows that analogy is fair and no extra validation is required. Ranging from 10 to 50, some basic validation is required. Penalties larger than 50 indicate that the analogy is poor and further optimisation is mandatory. In this work, we only distinguish the low penalty ( $< 10$ ) that requires no parameterization and the high penalty ( $\geq 10$ ) that requires parameterization. Apart from these groups, some parameters are reported with no penalty prediction (i.e. penalty is reported as zero). These are parameters that are provided in the original CGenFF database<sup>17</sup>, i.e. they were optimised carefully with respect to a set of molecules. If other molecules share a similar structural analogy with molecules from this original set, their parameters will be taken directly from this database, and no penalties will be assigned. For instance, monobactam (also known as 2-azetidinone) is the simplest  $\beta$ -lactam that contains only a 4-member heterocyclic ring. Being the core of an important class of antibiotics, it was included and optimised carefully in the original CGenFF database. Consequently, analysis of the parameter penalties for monobactam shows zero parameters with penalties, see Fig. 2(a). In contrast, penam and cepem are not included in the original CGenFF database, so most of their parameters show a large penalty and require an update, as shown in Fig. 2(b) and (c). Only a few parameters from penam and cepem (mostly among the bond and angles parameters) have no penalty reported, because these are parameters taken from monobactam in the original CGenFF. However, no penalty does not mean that these parameters are actually good choices and do not require any optimisation. Instead, one should carefully check the validity of these parameters and include the bad ones for parameterisation. This is the first pitfall that could trap an user.

After the initial ligand geometry and CGenFF parameters are prepared, we conduct the geometry optimisation using Gaussian 16<sup>27</sup>. Dihedral rotations are employed to search for the global minimum, and the candidates will be further submitted to geometry optimisation. The global minimum, referred to as the quantum geometry hereafter, will be used as the reference equilibrium geometry in all later steps of the parameterization.

Charge optimisation is achieved by fitting the MM atom charges to reproduce the QM interactions with water. Additionally, the distance (hydrogen bond length) and the dipole moment are taken into consideration during the fitting<sup>22</sup>. Following the original CGenFF protocol<sup>17</sup>, the interaction energy is scaled by a factor of 1.16 for a neutral polar molecule, and the quantum hydrogen bond length is offset by  $-0.2 \text{ \AA}$  to account for the hydrogen bond length in solution. For carbons that can interact with water as both a hydrogen bond donor and acceptor, a weighting factor of 0.5 is applied during the fitting to avoid giving it double weight. Ideally, one aims for a fitting error within 0.2 kcal/mol of the quantum interaction energy, a dipole moment within 1.2 to 1.5 times of the quantum dipole moment, and a distance error below  $0.1 \text{ \AA}$ <sup>17</sup>. In practice, a larger interaction energy error ( $< 3 \text{ kcal/mol}$ ) is acceptable, as long as the dipole moment is still within the acceptable range<sup>30</sup>. The distance error is often considered less relevant.



**Figure 1.** The parameterization workflow. Generate CGenFF parameters: The initial CGenFF parameters are predicted using the ParamChem server. Geometry Optimisation: The equilibrium geometry is determined via quantum calculations. Charge Optimisation: The atom charges are fitted by reproducing the quantum interaction energy with water, the quantum equilibrium distance, and the quantum dipole moment. Bond and Angle Optimisation: Force constants and equilibrium bond lengths/angles are determined by fitting distortion energies calculated from the quantum Hessian matrix. Dihedral Optimisation: Force constants and phase shifts are determined by fitting MM potentials to quantum dihedral potentials. Sometimes extra multiplicities are introduced. To apply the protocol to fused rings, we further separate the dihedrals into 3 groups: dihedrals belonging to the 4-member ring, dihedrals belonging to the 5- or 6-member ring, and the remaining dihedral and improper angles. Optimisation is performed per group, and several sets (with different choices of phase shifts) are tested using MD simulations. Parameters from the set with a good MD conformation, good potential fitting, and consistent choice of multiplicity over various  $\beta$ -lactam classes will be taken.



**Figure 2.** Penalty analysis for monobactam (a), penam (b), and cephem (c). For each molecule, we show how many parameters fall into each class of penalty severity (see legend), and we group the parameters into atom charges, bond and angle parameters, as well as dihedral and improper dihedral parameters. Monobactam's parameters, as well as other parameters based on monobactam analogy, are originally provided in the CGenFF database and thus they are reported with no penalty.

CGenFF models the bonds and angles as harmonic oscillators, i.e. ligands are assumed to vibrate only around the optimised quantum geometry. Because FFTK determines the force constants and the associated equilibrium values of bonds and angles by fitting distortion energies calculated via the Hessian matrix, wrong parameters not updated during the fitting could lead to a convergence problem, e.g. equilibrium angles constantly jump to a wrong value. Thus, it is recommended to take all angles and bonds for parameterization. If one wishes to keep using the default CGenFF parameters, one at least should check carefully whether the equilibrium reported by the CGenFF parameters matches the true equilibrium found by the quantum geometry optimisation. Generally acceptable error bounds for bonds and angles are 0.03 Å and 3°<sup>17</sup>, respectively.

The last step of our protocol is to determine the dihedral and improper dihedral parameters. Relaxed 1-dimensional potential energy surface (1D PES) scans along all dihedral rotations are performed with Gaussian 16. Note that FFTK removes redundant single bond rotations and those with hydrogens, but we include all of them during our optimisation. CGenFF models the dihedral potentials as<sup>17</sup>,

$$V_{\text{dihedrals}} = \sum_{\text{dihedrals}} k_{\phi} [1 + \cos(n\phi - \delta)] \quad , \quad (1)$$

where  $k_{\phi}$  is the force constant for a dihedral angle  $\phi$ ,  $n$  is the multiplicity ( $n = 1, 2, 3, 4, \text{ or } 6$ ), and  $\delta$  is the associated phase shift ( $\delta = 0^{\circ}$  or  $180^{\circ}$ ). The dihedral parameterization aims to reproduce the quantum potentials using these MM parameters. Important modifications such as introducing extra multiplicities for some dihedrals, avoiding over-fitting, correcting the default CGenFF multiplicity, and validating the outcome by molecular dynamics (MD) simulations have been reported<sup>29,30</sup>. Here we stress that predicting the phase shifts in advance can increase the fitting quality, especially when the search space is so large that getting correct phase shifts for all dihedrals becomes difficult. Our principle is to choose a  $\delta$  that places the equilibrium dihedral angle at the valley of the cos function for the associated  $n$ , or at least as close to this valley as possible. For instance, for a dihedral angle of multiplicity  $n = 1$  with an equilibrium angle at  $0^{\circ}$ , we will choose the phase shift to be  $\delta = 180^{\circ}$ , because the minimum of  $\cos(\phi - \delta)$  appears at  $\phi = 0^{\circ}$  if  $\delta = 180^{\circ}$ .

To further validate the prediction obtained in this way, we introduce a search protocol to systematically explore the dihedral parameter space. The parameters of the 4-member ring and of the 5- or 6-member ring are updated separately. When necessary, we perform another iteration to ensure convergence of the parameters, see the dihedral optimisation protocol in Fig. 1. This is only needed when the maximum force constant from the 5- or 6-member ring (which is updated later) is much larger than the one from the 4-member ring. Several sets of optimised parameters using different choices of phase shifts are generated and tested using MD simulations. The best set is selected for further optimisation, based on three criteria: how well the fitted MM PES compares to the QM PES, how well the MD conformation matches the quantum geometry (measured via root mean squared deviation, RMSD), and how well the choice (of multiplicity and phase shifts) agrees with those selected for other  $\beta$ -lactams. Through these selection criteria, information from MD simulations is included into our parameterization protocol and helps us navigate through the search of dihedral parameter space (in particular the space of phase shifts) efficiently. Using this protocol, we successfully optimise the dihedral parameters for the fused rings to reproduce the quantum geometry using only default CGenFF multiplicities, i.e. no extra multiplicities are needed during the fitting.

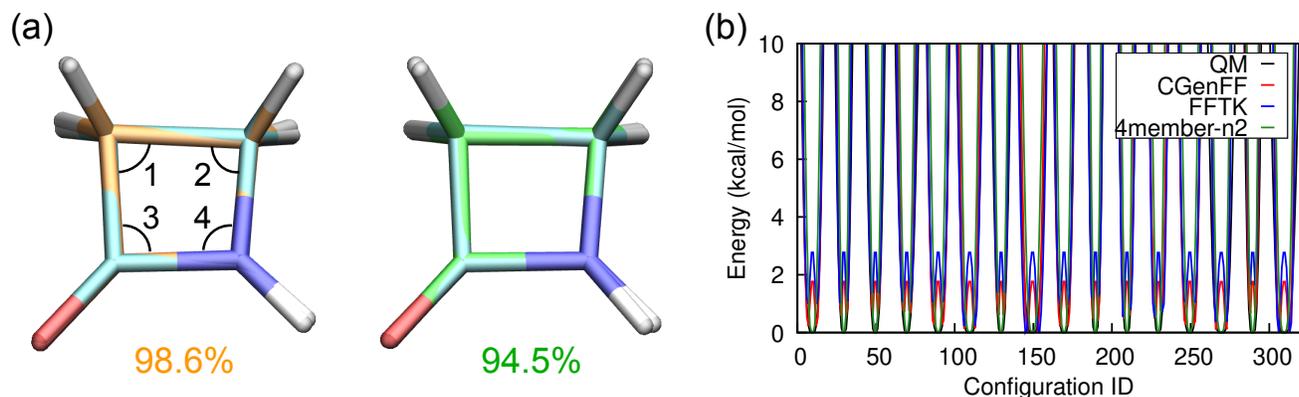
In this process, the MD conformation is obtained from the major cluster centroid of a 50 ns NPT simulation, using Gromacs<sup>37-45</sup>. The ligand is solvated in a cubic box of size 30 Å, and only for the anionic ligand a counter ion  $\text{Na}^+$  is introduced to make the system neutral. Following the CHARMM standard, the van der Waals force switching is turned on at 10 Å and the force is completely off beyond 12 Å. The electrostatic interaction is calculated using the Particle Mesh Ewald (PME) method<sup>46</sup>, with a spline order of 4 and a grid spacing of 1.25 Å. The temperature and pressure are controlled at 300 K and 1 bar using the V-rescale thermostat<sup>47</sup> and Parrinello-Rahman pressure coupling<sup>48,49</sup>, respectively. Bonds involving hydrogens are constrained using lincs<sup>50</sup>. Prior to the MD simulation, the system is first subjected to a 50,000 step energy minimization to remove any high energy clashes, a 100 ps NVT equilibration to heat up the system, and a 100 ps NPT equilibration to equilibrate the pressure. Conformations sampled in the 50 ns trajectory are clustered based on the fused ring structure using the GROMOS method<sup>51</sup> with a 0.2 Å RMSD cutoff.

Last but not least, all quantum calculations are performed using Gaussian 16<sup>27</sup> at MP2/6-31G(d) level for neutral molecules and MP2/6-31+G(d) for anions, except for the water interactions required for the charge optimisation. For the latter, HF/6-31G(d) is employed in the quantum calculations to stay consistent with the CHARMM water model, which was originally parameterized at the HF/6-31G(d) level.

## Results

### Monobactam

Monobactam is the simplest  $\beta$ -lactam. Its parameters are provided in the original CGenFF, and hence no penalties are predicted by ParamChem. Furthermore, MD simulation reveals a stable ligand structure that compares well with the quantum geometry, see Fig. 3(a). Therefore, there hardly seems to be a reason to re-optimize the parameters. However, on careful inspection,



**Figure 3.** Monobactam. (a) Comparison between the quantum geometry (cyan), the MD conformation obtained using CGenFF parameters (orange), and the MD conformation obtained using re-optimised parameters (green). Both sets of parameters reproduce the quantum geometry very well. Associated cluster sizes reach 98.6% and 94.5%, respectively. (b) Dihedral fitting results. The quantum potential is depicted in black, while potentials fitted with the CGenFF phase shifts (labeled as CGenFF) and with phase shifts initially from CGenFF but allowed to vary during optimisation (labeled as FFK) are depicted in red and blue, respectively. In contrast to the quantum dihedral potentials, these two MM potentials show strange barriers where local minima should be. Only the MM potential fitted using phase shifts from our prediction (green) agrees well with the quantum data.

we find that the MM parameters have equilibrium angles that deviate strongly from the quantum geometry, especially for the essential four angles that are labeled in panel (a). Table 1 lists the equilibrium values for QM and MM and shows that they differ by up to  $20^\circ$ . While in monobactam these four angles can compensate each other and lead to a good match with the quantum geometry during MD simulation, this feature is not reproducible when we try to optimise penam and cephem, where the fused ring structure will distort the 4-member ring, making it less planar than in monobactam. Considering how the parameters for angles and bonds are determined, wrong equilibrium angles that are not updated during the fitting can cause a convergence problem. Therefore, we will include all bond and angle parameters when optimising penam and cephem, and we will also carry out the parameterization for monobactam for validation of our approach. All parameters, except for the charges of the hydrogen atoms, are re-optimized.

Surprisingly, even for this simple molecule, finding the correct phase shifts via fitting the dihedral potentials is challenging. Fig. 3(b) shows the resulting PES obtained from using the CGenFF phase shifts (red curve;  $\delta$ s locked during fitting), from FFK (blue curve;  $\delta$ s allowed to vary during fitting), and from our prediction (green curve;  $\delta$ s chosen as described in the Methods section and locked during fitting). Among these three sets, only our prediction produces a harmonic-like potential, similar to the quantum potentials (black curve), while the other two results show strange barriers across all potentials. Consequently, only our prediction yields an MD conformation comparable to the quantum geometry, see Fig. 3(a) for the major MD conformation depicted in green. This motivates us to carefully select the phase shifts when optimising penam and cephem, and see if this boosts the quality of the dihedral parameterization.

Angle index	CGenFF equilibrium ( $^\circ$ )	QM optimised geometry ( $^\circ$ )
1	106.00	86.08
2	104.50	87.36
3	104.50	90.76
4	111.50	95.81

**Table 1.** List of the monobactam equilibrium angles reported in the CGenFF parameters and calculated via QM optimisation. The angles are indexed according to Fig. 3(a).

## Penam

Penam is the core structure of penicillin. Owing to its importance, penicillin's force field parameters are long available in the literature<sup>52</sup>. However, those reported parameters contain many multiplicities different from the ParamChem prediction. For consistency, we re-parameterize the ligand based on the multiplicities predicted by ParamChem. As mentioned previously, parameters provided in the original CGenFF have no penalties, but we have learned from the monobactam example that at least four essential angles from those parameters require a proper update. Furthermore, Fig. 2(b) shows that more than half of the parameters require optimisation due to high penalties. Thus, for simplicity we optimise all parameters except for charges from hydrogens and charges with low penalties.

Penam comprises the  $\beta$ -lactam ring, a 5-member ring with sulfur, and a carboxylic acid group. For simplicity, our discussion will focus on the basic core structure shown in Fig. 4(a) (penam 1, atom names labeled). Results for more complicated structures (penam 2 and penam 3) will be addressed at the end of this section. These three forms are optimised individually using the aforementioned protocol. For the dihedral parameterization, we first optimise several sets of parameters for the 4-member ring, then we conduct MD simulations to select the best set and repeat the same operation for the parameters of the 5-member ring. Once the parameterization of the core structure is completed, we optimise the dihedrals of the carboxylic group and finally the improper dihedral angles. The tested sets of parameters for the 4-member ring and a comparison of their MD conformation are listed in Table 2. The sets labeled CGenFF and FFTK both employ the default phase shifts from ParamChem. The former keeps these phase shifts locked during the fitting, while the latter allows them to vary. In addition to these two common choices, we introduce three sets with phase shifts predicted as described in the Method section. That is, we choose a  $\delta$  that places the quantum equilibrium close to the potential valley of the given multiplicity  $n$ , and we keep the predicted phase shifts locked during the fitting. The associated sets are labeled as 4member-n2, 4member-n2+n3, and 4member-n3, where the suffix indicates the choice of multiplicities for the dihedral angles C1-C3-N-C2, C1-C3-N-C0, O-C3-N-C2 and O-C3-N-C0. As suggested by the name, 4member-n2 utilizes the multiplicity  $n = 2$ , which is the default multiplicity from ParamChem. This choice agrees with the corresponding dihedrals' multiplicity in monobactam. We also introduced the sets 4member-n2+n3 (multiplicity  $n = 2$  and  $n = 3$ ) and 4member-n3 (only multiplicity  $n = 3$ ), to see if extra or different multiplicities are required to describe the non-planar 4-member ring in penam.

Fig. 4(b) shows the potential fitting results for the 4-member ring parameters of penam 1. To make the visualisation more clear, data from the sets 4member-n2+n3 and 4member-n3 are not plotted. Like in monobactam, the 4member-n2 parameter set (green) fits the quantum potential (black) better than the CGenFF (red) and the FFTK (blue) parameters. This finding is also reflected in the associated MD conformations: MD simulation using the 4member-n2 parameters reproduces the quantum geometry best, with a RMSD of 0.9 Å. See Table 2 for the MD results of penam 1. Notably, although the 4member-n2+n3 parameter set can also reproduce the quantum geometry during MD simulation, it requires extra multiplicity ( $n = 3$ ) for four dihedral angles. Since the default multiplicities predicted by ParamChem (4member-n2) can already reproduce both the quantum potentials and geometry very well, we continue optimising the 5-member parameters using the parameters of 4member-n2.

Results of the 5-member parameter optimisation are plotted in Fig. 4(c). Again we optimised the dihedral parameters for four scenarios: using the default CGenFF phase shifts but locked during fitting (CGenFF), allowing FFTK to vary the phase shifts during fitting (FFTK), using our predicted phase shifts (5member-locked), and allowing our predicted phase shifts to vary during fitting (5member-unlock). Panel (c) shows that the MM potentials obtained from these four scenarios agree well with the quantum potentials, with the 5member-locked set slightly worse than the others. The RMSD listed in Table 2 reveals that locking the phase shifts according to the aforementioned method reproduces the quantum geometry best during MD simulations, with the RMSD being a mere 0.4 Å. Hence, through carefully choosing the phase shifts and optimising only the force constants, one can reproduce both the quantum potentials and geometry relatively well for penam 1, without any extra multiplicities.

The MD simulation using the final parameters (all updated) finds 75.9% of the conformations in the MD trajectory reproducing the quantum geometry closely, while another 17.5% show a flipped 5-member ring, see the structures of penam 1 depicted in in panel (d). This flipped conformation is an outcome of two dihedral angles, C0-S-C-C2 and N-C2-C-S, rotating in concert. Fig. 4(e) shows the 2D PES of C0-S-C-C2 and N-C2-C-S, calculated using Gaussian relaxed dihedral scans for the associated single bond rotations. The positions of the two major MD conformations are highlighted using the yellow dots in the plot. Along the diagonal of the plot, one finds a low energy valley that connects these two conformations. Consequently, this conformational change is energetically accessible only when the two dihedrals rotate together. A mere single dihedral rotation can not lead to such a conformational change. Interestingly, although the current dihedral optimisation protocol does not support fitting 2D dihedral PES, one still finds related fingerprints in the 1D potential scans: Each QM potential in panel (c) has a second dip to a local minimum. Such a potential shape suggest the existence of another stable conformation, which will be discovered during MD simulations.

Following the same protocol, penam 2 and penam 3 are also parameterized using the predicted phase shifts. No extra multiplicities are required to reproduce the potentials for the 4- and 5-member rings. The associated RMSDs are listed in

Ring	Label	Choice of $\delta$	Penam 1	Penam 2	Penam 3
4-member	CGenFF	CGenFF (locked)	6.7	1.8	7.7
	FFTK	CGenFF (unlocked)	6.5	1.1	8.0
	4member-n2	Self predicted (locked)	0.9	0.4	1.0
	4member-n2+n3	Self predicted (locked)	1.0	0.7	1.3
	4member-n3	Self predicted (locked)	2.9	1.8	2.7
5-member	CGenFF	CGenFF (locked)	0.8	0.5	1.5 <sup>†</sup>
	FFTK	CGenFF (unlocked)	0.8	0.5	5.1 <sup>†</sup>
	5member-locked	Self predicted (locked)	0.4	0.6	0.8 <sup>†</sup>
	5member-unlock	Self predicted (unlocked)	0.5	0.4	1.0 <sup>†</sup>

**Table 2.** Comparison between MD conformations obtained from different sets of penam 4-member ring and 5-member ring parameters. The table lists the RMSD (in Å) of the major MD cluster conformation from the quantum geometry. For the 4-member ring, CGenFF predicts a multiplicity of  $n = 2$  for the dihedral angles associated with the C3-N single bond rotation. Parameter sets with extra multiplicity  $n = 3$  (4member-n2+n3) and with the multiplicity changed to  $n = 3$  (4member-n3) are also tested. <sup>†</sup> Reported for penam 3 is the RMSD of the second major cluster.

Table 2, while the major two MD conformations are depicted in Fig. 4(d). For penam 2, again the MD simulation successfully finds a conformation closely resembling the quantum geometry in 86.5% of the trajectory, except for the carboxylic group that can rotate freely in water. The second major conformation is similar to the one in penam 1. Both are the consequence of the C0-S-C-C2 and N-C2-C-S dihedrals rotating in concert. Interestingly, populations of these two conformations invert in the MD simulations for penam 3. This is a general feature of penam 3 that can be observed in all MD simulations conducted using all parameter sets listed in Table 2. This could be related to the potential valley that connects the two conformations being more flat in penam 3, leading to an easier conformational change. See Fig.S1 in the supplementary information.

## Cephem

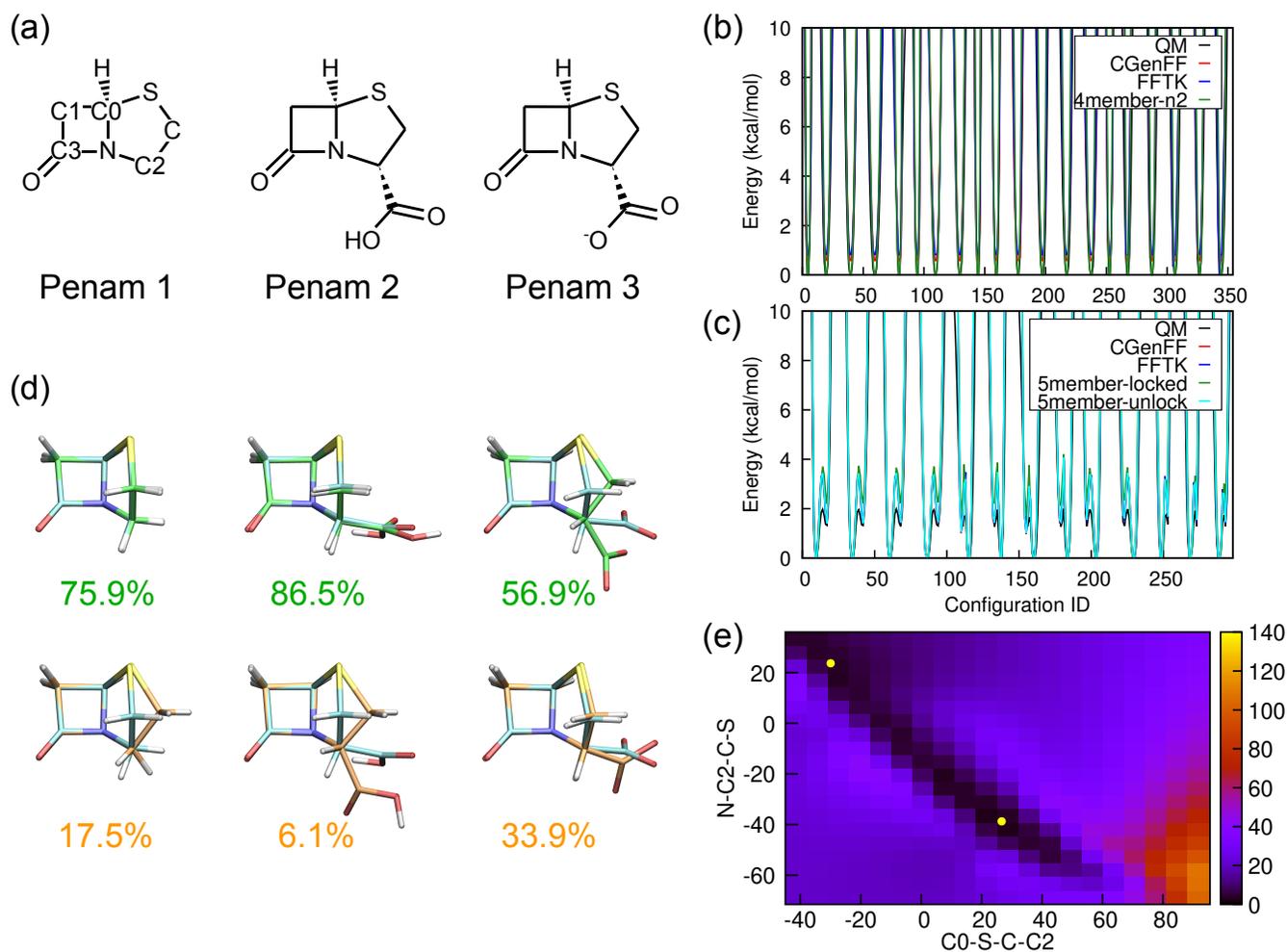
Cephem is the core structure of the latest  $\beta$ -lactam antibiotics in clinical use, i.e. the cephalosporins, which are less susceptible to  $\beta$ -lactamases. Similar to penam, cephem requires a proper parameterization since ParamChem predicts many parameters with high penalty, see the red bars in Fig. 2(c).

In this work, two forms of cephem are optimised: the neutral cephem 1 and the anionic cephem 2, see Fig. 5(a). This time we will focus on the anionic cephem 2 to show how our protocol can work for anions. Again, several choices of phase shifts are tested. For the 4-member ring, these are CGenFF, FFTK, 4member-n2, 4member-n2+n3, and 4member-n3, i.e. the same as in the penam case. Different from all other cases, anionic cephem has three dihedral angles for which the prediction of phase shifts is ambiguous, i.e. the quantum geometry lies exactly in between the minima from the two choices of the phase shifts so that none of the choices is evidently better. These dihedrals are C10-C9-N5-C8, C21-C9-N5-C8, and C10-C11-S1-C7, see cephem 2 in Fig. 5(a) for the atom names. We took the phase shifts from the associated dihedrals in the neutral cephem 1 instead. The associated MM potentials are depicted in Fig. 5(b).

This time, the MM potentials fail to reproduce the quantum dihedral potentials with acceptable quality, see Fig. 5(b). (MM potentials of 4member-n2+n3 and 4member-n3 are again omitted for a clearer visualization.) Although the 4member-n2 set (green curve) reproduces some features of the QM data, it still fails to reproduce some potentials, e.g. the MM energy is still too high for some potential wells. This bad potential fitting is caused by the fact that parameters of the 6-member ring are updated after the 4-member ring. Consequently, any 6-member ring parameters with large force constants but wrong phase shifts could lead to a bad potential fitting in 4-member ring optimisation. This issue can be resolved easily by carrying out the second iteration of optimisation, as one shall see later in panel (c).

MD simulations reveal that an acceptable (i.e. matching the quantum geometry) molecular skeleton without the carboxylic group can be obtained using the 4member-n2 parameter set. As listed in Table 3, the lowest RMSD is 1.4 Å. Nevertheless, this value is already much lower compared to the CGenFF or FFTK phase shifts. Among the three sets that yield the lowest RMSD, the 4member-n2 parameters utilize only the default ParamChem multiplicities, and therefore will be employed for the subsequent optimisation.

Fig. 5(b) also shows the MM potentials for the 6-member ring, obtained in the first iteration. In contrast to the 4-member ring optimisation, all MM potentials now seem to reproduce the QM data very well, and the associated MD simulations of cephem 2 show a similar RMSD for the four sets. While this may pose a selection difficulty, our strategy in searching for a consistent choice suggests to select the set of 6member-locked, because it is the only set that also captures the quantum geometry for cephem 1 during MD simulation, see the RMSDs in Table 3.



**Figure 4.** (a) The three penam structures that are optimised in this work. Penam 1 contains merely the core penam structure; penam 2 is a neutral molecule with a carboxylic acid; penam 3 is anionic. Important atom names are labeled in penam 1. (b) The 4-member ring dihedral potentials, obtained using quantum calculations (black) and various MM data. Plotted here are the MM potentials for the CGenFF, FFTK, and 4member-n2 parameter sets (see text) using red, blue, and green curves, respectively. (c) Dihedral potentials of the 5-member rings. Shown are data from the quantum calculations (QM, black), from CGenFF (red), from FFTK (blue), from 5member-locked (green), and from 5member-unlock (cyan). (d) The MD conformations obtained using the optimised parameters. The reference quantum geometry is always depicted in cyan, the major MD conformation in green, and the second major MD conformation in orange. The cluster size is given as the ratio of the trajectory length. (e) The 2D dihedral PES for penam 1, obtained from Gaussian calculations. Axes are labeled in degrees. The optimised quantum geometry is at (26.66, -38.75) and the second major MD conformation found is at (-29.95, 23.70). Both points are highlighted with yellow dots. A relatively low energy valley connecting the two dots is observed, suggesting that the conformational change occurs through a simultaneous rotation of two dihedral angles.

Ring	Label	Choice of $\delta$	Cephem 1	Cephem 2
4-member	CGenFF	CGenFF (locked)	2.3	2.7
	FFTK	CGenFF (unlocked)	2.8	3.4
	4member-n2	Self predicted (locked)	1.0	1.4
	4member-n2+n3	Self predicted (locked)	1.1	1.4
	4member-n3	Self predicted (locked)	1.0	1.4
6-member	CGenFF	CGenFF (locked)	5.5	0.9
	FFTK	CGenFF (unlocked)	5.2	1.0
	6member-locked	Self predicted (locked)	0.5	1.0
	6member-unlock	Self predicted (unlocked)	5.3	0.9

**Table 3.** Comparison between MD conformations obtained from different sets of cephem 4-member ring and 6-member ring parameters. The table lists the RMSD (in Å) from the quantum geometry.

The final potentials obtained from a second iteration for dihedral optimisation, which is followed by another iteration of the entire parametrisation, are plotted in panel (c). Although there is still some minor mismatch in potentials of the 6-member ring dihedrals, our MM potentials in principle grasp the QM features very well. In particular, we note that the QM potentials of the 4-member ring dihedral angles are well reproduced, proving that the extra iteration can indeed improve the fitting. The major MD conformation obtained from a 50 ns NPT simulation is depicted in Fig. 5(d). The MD conformation of the neutral cephem 1 almost perfectly reproduces its quantum geometry (RMSD 0.5 Å), while the anionic cephem 2 has a larger deviation in the skeleton (RMSD 1.4 Å). The slightly increased RMSD after the second iteration is due to the updated improper angle parameters, which in the first iteration had to be turned off to improve the dihedral fitting results.

## Discussion

Using the Force Field Toolkit (FFTK) software and a protocol for predicting phase shifts, we systematically optimised CGenFF parameters for two important  $\beta$ -lactam antibiotic classes, viz. penam and cephem. Two important lessons have been learned:

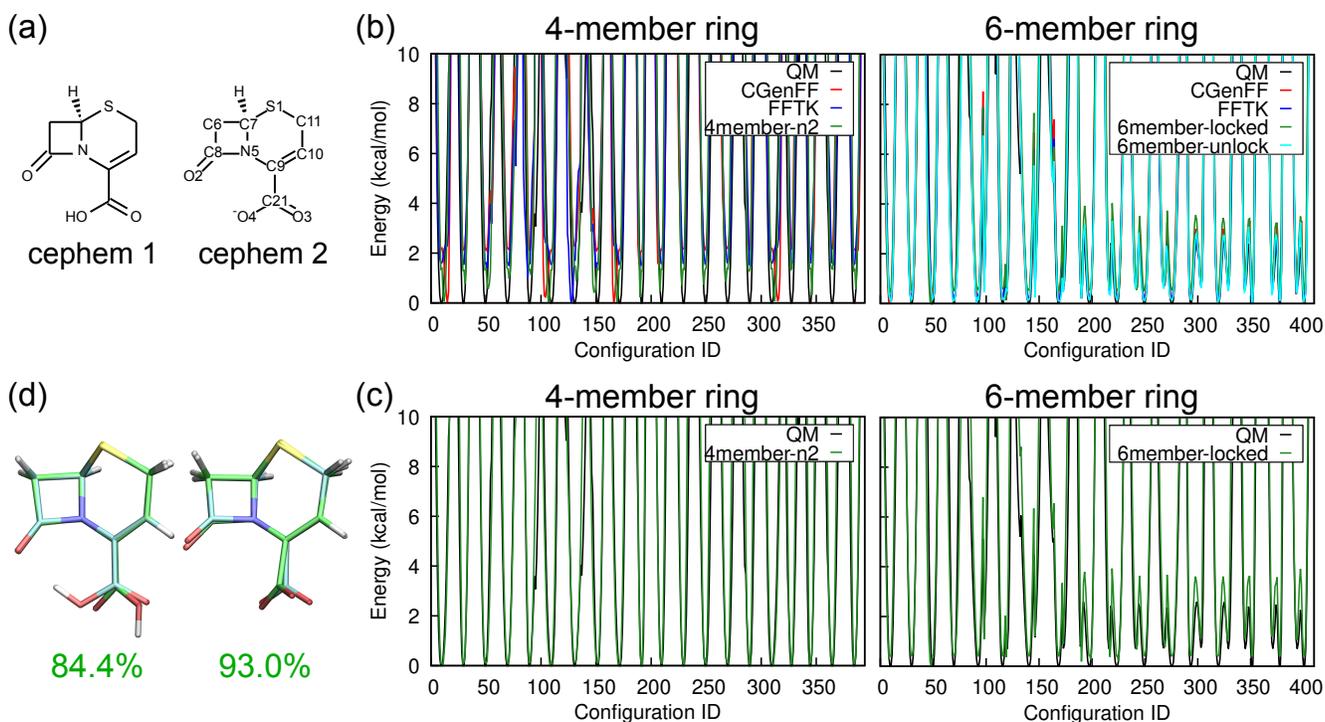
First, the original CGenFF utilises equilibrium angles that deviate from the QM equilibrium for the 4-member  $\beta$ -lactam ring, leading to a difficulty in optimising the bond and angle parameters. Although ParamChem by default does not produce any penalty warning for these parameters, users should at least double check the MM equilibrium against the optimised quantum geometry. If any large deviation is found, the associated parameter should be included into the optimisation, regardless of the predicted penalty value. If too many essential parameters are different from the quantum geometry, users should include all bond and angle parameters for optimisation.

Second, multiple sets of dihedral parameters can reproduce the quantum potential energy surfaces relatively well. This is a known problem<sup>29</sup> that can only be solved by testing the parameters' performance using MD simulations. Thus, we combine the MD simulation results with dihedral potential fitting to determine the correct combination of phase shifts. Additionally, we propose to choose the phase shifts that place the dihedral angles of the quantum geometry as close as possible to the energy minima of the associated dihedral functions. Based on the presented cases of  $\beta$ -lactams, predicting the phase shifts in advance to the dihedral parameterization can reduce the search difficulty during optimising the dihedral parameters. If the phase shifts cannot be determined easily, one can refer to the corresponding phase shifts from a similar molecule, such as using the neutral cephem's phase shifts for the anionic cephem. Following this protocol, we successfully parameterized several penam and cephem molecules, using only the default dihedral multiplicities taken from ParamChem. No extra multiplicity is required to fit the MM dihedral potentials to the QM data. However, this does not mean that extra dihedral multiplicities are never required during parameterization. Rather, we suggest to optimise the phase shifts first and then include extra multiplicities when necessary, to avoid the over-fitting problem.

Finally, MD simulations using optimised parameters reproduce the quantum geometry well in most of the cases, except for the anionic penam. In this case, the molecule can undergo a simultaneous rotation of two dihedral angles during the MD simulation, leading to a major MD conformation with a flipped 5-member ring.

## References

- Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\varphi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273, DOI: [10.1021/ct300400x](https://doi.org/10.1021/ct300400x) (2012). <https://doi.org/10.1021/ct300400x>.



**Figure 5.** (a) Neutral and anionic cephem structures. Names for the relevant atoms are shown for cephem 2. (b) Dihedral potentials of the 4-member and 6-member rings obtained during the first iteration. Depicted are potentials from quantum calculations (black) and various MM data. Note that none of the MM potentials reproduce the 4-member ring quantum potentials well, but fitting of the 6-member ring seems to be good. (c) Dihedral potentials obtained in the second iteration. Plotted are the quantum (black) and the MM potentials (green) optimised using self-predicted phase shifts. (d) The major MD conformations (green) obtained using optimised parameters. As a reference, the optimised quantum geometries are depicted in cyan. For the neutral molecule, the major MD conformation resembles the quantum geometry well. The molecular skeleton (i.e. excluding the carboxylic group) has a RMSD of 0.5 Å from the quantum geometry. For the anionic cephem 2, the major MD conformation yields an acceptable conformation, with a slightly larger RMSD of 1.4 Å.

2. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73, DOI: [10.1038/nmeth.4067](https://doi.org/10.1038/nmeth.4067) (2017).
3. Tian, C. *et al.* ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552, DOI: [10.1021/acs.jctc.9b00591](https://doi.org/10.1021/acs.jctc.9b00591) (2020). <https://doi.org/10.1021/acs.jctc.9b00591>.
4. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236, DOI: [10.1021/ja9621760](https://doi.org/10.1021/ja9621760) (1996). <https://doi.org/10.1021/ja9621760>.
5. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. 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](https://doi.org/10.1021/acs.jctc.5b00356) (2015). <https://doi.org/10.1021/acs.jctc.5b00356>.
6. Jiang, F., Zhou, C.-Y. & Wu, Y.-D. Residue-specific force field based on the protein coil library. RSFF1: Modification of OPLS-AA/L. *The J. Phys. Chem. B* **118**, 6983–6998, DOI: [10.1021/jp5017449](https://doi.org/10.1021/jp5017449) (2014). <https://doi.org/10.1021/jp5017449>.
7. Zhou, C.-Y., Jiang, F. & Wu, Y.-D. Residue-specific force field based on protein coil library. RSFF2: Modification of AMBER ff99SB. *The J. Phys. Chem. B* **119**, 1035–1047, DOI: [10.1021/jp5064676](https://doi.org/10.1021/jp5064676) (2015). <https://doi.org/10.1021/jp5064676>.
8. Yu, W. & MacKerell Jr., A. D. Computer-aided drug design methods. In Sass, P. (ed.) *Antibiotics. Methods in Molecular Biology*, vol. 1520, 85–106, DOI: [10.1007/978-1-4939-6634-9\\_5](https://doi.org/10.1007/978-1-4939-6634-9_5) (Humana Press, New York, NY, 2017).
9. Cournia, Z., Allen, B. & Sherman, W. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *J. Chem. Inf. Model.* **57**, 2911–2937, DOI: [10.1021/acs.jcim.7b00564](https://doi.org/10.1021/acs.jcim.7b00564) (2017). <https://doi.org/10.1021/acs.jcim.7b00564>.
10. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229, DOI: [10.1038/s41586-019-0917-9](https://doi.org/10.1038/s41586-019-0917-9) (2019).
11. Bender, B. J. *et al.* A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799–4832, DOI: [10.1038/s41596-021-00597-z](https://doi.org/10.1038/s41596-021-00597-z) (2021).
12. Raniolo, S. & Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **15**, 2837–2866, DOI: [10.1038/s41596-020-0342-4](https://doi.org/10.1038/s41596-020-0342-4) (2020).
13. Wan, S., Bhati, A. P., Zasada, S. J. & Coveney, P. V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus*. **10**, 20200007, DOI: [10.1098/rsfs.2020.0007](https://doi.org/10.1098/rsfs.2020.0007) (2020). <https://royalsocietypublishing.org/doi/pdf/10.1098/rsfs.2020.0007>.
14. Latorraca, N. R. *et al.* Molecular mechanism of gpcr-mediated arrestin activation. *Nature* **557**, 452–456, DOI: [10.1038/s41586-018-0077-3](https://doi.org/10.1038/s41586-018-0077-3) (2018).
15. Renault, P., Louet, M., Marie, J., Labesse, G. & Floquet, N. Molecular dynamics simulations of the allosteric modulation of the adenosine a2a receptor by a mini-g protein. *Sci. Reports* **9**, 5495, DOI: [10.1038/s41598-019-41980-x](https://doi.org/10.1038/s41598-019-41980-x) (2019).
16. Araki, M. *et al.* Exploring ligand binding pathways on proteins using hypersound-accelerated molecular dynamics. *Nat. Commun.* **12**, 2793, DOI: [10.1038/s41467-021-23157-1](https://doi.org/10.1038/s41467-021-23157-1) (2021).
17. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690, DOI: [10.1002/jcc.21367](https://doi.org/10.1002/jcc.21367) (2010). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21367>.
18. 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.* **25**, 1157–1174, DOI: [10.1002/jcc.20035](https://doi.org/10.1002/jcc.20035) (2004). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20035>.
19. Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J. & Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **45**, W331–W336, DOI: [10.1093/nar/gkx312](https://doi.org/10.1093/nar/gkx312) (2017). <https://academic.oup.com/nar/article-pdf/45/W1/W331/23740908/gkx312.pdf>.
20. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154, DOI: [10.1021/ci300363c](https://doi.org/10.1021/ci300363c) (2012). <https://doi.org/10.1021/ci300363c>.
21. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52**, 3155–3168, DOI: [10.1021/ci3003649](https://doi.org/10.1021/ci3003649) (2012). <https://doi.org/10.1021/ci3003649>.

22. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E. & Gumbart, J. C. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comput. Chem.* **34**, 2757–2770, DOI: [10.1002/jcc.23422](https://doi.org/10.1002/jcc.23422) (2013). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23422>.
23. Morado, J. *et al.* ParaMol: A package for automatic parameterization of molecular mechanics force fields. *J. Chem. Inf. Model.* **61**, 2026–2047, DOI: [10.1021/acs.jcim.0c01444](https://doi.org/10.1021/acs.jcim.0c01444) (2021). <https://doi.org/10.1021/acs.jcim.0c01444>.
24. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260, DOI: [10.1016/j.jmgl.2005.12.005](https://doi.org/10.1016/j.jmgl.2005.12.005) (2006).
25. CGenFF ParamChem web interface. <https://cgenff.umaryland.edu>.
26. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, DOI: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) (1996).
27. Frisch, M. J. *et al.* Gaussian 16 Rev. C.01 (2016). Gaussian Inc. Wallingford CT.
28. Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2**, 73–78, DOI: [10.1002/wcms.81](https://doi.org/10.1002/wcms.81) (2012). <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.81>.
29. Pavlova, A. & Gumbart, J. C. Parametrization of macrolide antibiotics using the Force Field Toolkit. *J. Comput. Chem.* **36**, 2052–2063, DOI: [10.1002/jcc.24043](https://doi.org/10.1002/jcc.24043) (2015). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.24043>.
30. Pang, Y. T., Pavlova, A., Tajkhorshid, E. & Gumbart, J. C. Parameterization of a drug molecule with a halogen  $\sigma$ -hole particle using ffTK: Implementation, testing, and comparison. *The J. Chem. Phys.* **153**, 164104, DOI: [10.1063/5.0022802](https://doi.org/10.1063/5.0022802) (2020). <https://doi.org/10.1063/5.0022802>.
31. Morado, J. ParaMol. <https://github.com/JMorado/ParaMol> (2020–2021).
32. Croitoru, A. *et al.* Additive charmm36 force field for nonstandard amino acids. *J. Chem. Theory Comput.* **17**, 3554–3570, DOI: [10.1021/acs.jctc.1c00254](https://doi.org/10.1021/acs.jctc.1c00254) (2021). PMID: 34009984, <https://doi.org/10.1021/acs.jctc.1c00254>.
33. Wang, X. & Li, W. Development and testing of force field parameters for phenylalanine and tyrosine derivatives. *Front. Mol. Biosci.* **7**, DOI: [10.3389/fmolb.2020.608931](https://doi.org/10.3389/fmolb.2020.608931) (2020).
34. Gumbart, J. C., Beeby, M., Jensen, G. J. & Roux, B. Escherichia coli peptidoglycan structure and mechanics as predicted by atomic-scale simulations. *PLOS Comput. Biol.* **10**, 1–10, DOI: [10.1371/journal.pcbi.1003475](https://doi.org/10.1371/journal.pcbi.1003475) (2014).
35. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395, DOI: [10.1093/nar/gkaa971](https://doi.org/10.1093/nar/gkaa971) (2020). <https://academic.oup.com/nar/article-pdf/49/D1/D1388/35363961/gkaa971.pdf>.
36. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242, DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235) (2000). <https://academic.oup.com/nar/article-pdf/28/1/235/9895144/280235.pdf>.
37. Bekker, H. *et al.* Gromacs: A parallel computer for molecular dynamics simulations. In de Groot, R. A. & Nadrchal, J. (eds.) *Physics Computing '92*, 252–256, DOI: [10.1142/9789814536295](https://doi.org/10.1142/9789814536295) (World Scientific, Singapore, 1993).
38. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56, DOI: [10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E) (1995).
39. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol. modeling annual* **7**, 306–317, DOI: [10.1007/s008940100045](https://doi.org/10.1007/s008940100045) (2001).
40. van der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718, DOI: [10.1002/jcc.20291](https://doi.org/10.1002/jcc.20291) (2005). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20291>.
41. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447, DOI: [10.1021/ct700301q](https://doi.org/10.1021/ct700301q) (2008). <https://doi.org/10.1021/ct700301q>.
42. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854, DOI: [10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055) (2013). <https://academic.oup.com/bioinformatics/article-pdf/29/7/845/17343875/btt055.pdf>.
43. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In Markidis, S. & Laure, E. (eds.) *EASC 2014: Solving Software Challenges for Exascale*, vol. 8759 of *Lecture Notes in Computer Science*, chap. 1, 3–27, DOI: [10.1007/978-3-319-15976-8\\_1](https://doi.org/10.1007/978-3-319-15976-8_1) (Springer, Cham, Switzerland, 2015).
44. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25, DOI: <https://doi.org/10.1016/j.softx.2015.06.001> (2015).

45. Lindahl, E., Abraham, M. J., Hess, B. & van der Spoel, D. GROMACS source code (2021.3). <https://doi.org/10.5281/zenodo.5053201>, DOI: 10.5281/zenodo.5053201 (2021).
46. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The J. Chem. Phys.* **103**, 8577–8593, DOI: 10.1063/1.470117 (1995). <https://doi.org/10.1063/1.470117>.
47. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The J. Chem. Phys.* **126**, 014101, DOI: 10.1063/1.2408420 (2007). <https://doi.org/10.1063/1.2408420>.
48. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190, DOI: 10.1063/1.328693 (1981). <https://doi.org/10.1063/1.328693>.
49. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076, DOI: 10.1080/00268978300102851 (1983). <https://doi.org/10.1080/00268978300102851>.
50. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122, DOI: 10.1021/ct700200b (2008). <https://doi.org/10.1021/ct700200b>.
51. Daura, X. *et al.* Peptide folding: When simulation meets experiment. *Angewandte Chemie Int. Ed.* **38**, 236–240, DOI: [https://doi.org/10.1002/\(SICI\)1521-3773\(19990115\)38:1/2<236::AID-ANIE236>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M) (1999).
52. Fiset, O., Gagné, S. & Lagüe, P. Molecular dynamics of class a  $\beta$ -lactamases – effects of substrate binding. *Biophys. J.* **103**, 1790–1801, DOI: 10.1016/j.bpj.2012.09.009 (2012).

## Acknowledgments

YCC thanks Prof. Jonathan W. Essex for helpful discussions, the Chinese University of Hong Kong, Shenzhen (CUHK-SZ) and the Kobilka Institute of Innovative Drug Discovery (KIIDD) for funding, as well as The Royal Society for travel support (AL\201047). Computational resources from KIIDD and from ITSO at CUHK-SZ are gratefully acknowledged.

## Author contributions

Q.W., S.X., F.O., Y.-C.C. performed the quantum calculations and parameterization. Y.-C.C. organized the work and oversaw the entire project. T.-Y.L., H.-D.H., Y.-C.C. wrote the manuscript together.

## Data availability

The optimised parameters are provided in the supplementary information. Other data are available from the corresponding author on reasonable request.

## Competing interests

The authors declare no competing interests.