Comparison of Charge Derivation Methods Applied to Amino Acid Parameterization

Pietro G. A. Aronica,*†‡, Stephen J. Fox,*∥ and Chandra S. Verma*§

†Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, 138671, Singapore
‡Department of Biological Sciences, National University of Singapore, 16 Science Drive, 117558, Singapore
§School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551, Singapore

ABSTRACT: When using non-natural amino acids in computational simulations of proteins, it is necessary to ensure appropriate parameterization of the new amino acids toward the creation of appropriate input files. In particular, the charges on the atoms may have to be derived de novo and ad hoc for the new species. As there are many variables in the charge derivation process, an investigation was devised to compare different approaches and determine their effect on simulations. This was done with the purpose to identify the methods which produced results compatible with the existing parameters. It was found in this study that all analyzed charge derivation methods reproduce with sufficient accuracy the literature values and can be used with confidence when parameterizing novel species.

INTRODUCTION

The reliability of molecular dynamics (MD) simulations is dependent on the underlying force field, which is a collection of parameters including the strengths of various interactions, the charges on the atoms, the lengths of bonds, and so forth. Therefore, obtaining high-quality parameters is one of the first steps in ensuring that simulations are reliable. MD packages such as AMBER, CHARMM, GROMACS, NAMD, and many others use different varieties of the basic force field, some of which may be specific to classes of biological molecules or to the computational program being used. These force fields may have different parameters that result from the differences in the methods employed for their derivation.

Parameters have been derived for the standard sets of amino acids and nucleic acids, sugars and lipids, and also for many kinds of functional groups in small molecule ligands. Indeed, general algorithms have been developed to parameterize new molecules that enable speed in high-throughput studies. However, while the parameters for the covalent interactions such as bond lengths, angles, dihedrals, and noncovalent van der Waals parameters are transferable from the general datasets that already exist, the same is not the case for charges, which are very much dependent on the context and hence need to be calculated for every new molecule. Although some general rules have been derived for their transferability, the paucity of experimental data to benchmark the charge datasets (either directly or through the prediction of properties dependent on charges) can make this issue problematic. The matter is further complicated because a wide variety of methods have been used for the purpose of charge derivation, regularly producing subtly different results. Therefore, we have decided to review and compare the existing methods in representative test cases and establish how charges derived with different variables measure with those reported in the literature. We hope this can help understand which of the current approaches is best suited for parameterizing nonstandard amino acids.

The charges for a series of representative natural amino acids were derived following as much as possible the procedure reported in the literature; the values that we used as the benchmark are taken from the AMBER force field FF14SB (which are the same as in FF12SB and FF99SB9). Two methods were used for the derivation of charges: the restrained electrostatic potential (RESP) fitting procedure, which was originally employed in the derivation of the AMBER force field charges, and AM1-BCC,12 which is used by the antechamber tool part of AMBERTools. MD simulations were then carried out separately using these derived charges and compared to simulations carried out using the original charges from AMBER.

METHODS

Charge Derivation Methods. Perhaps the most commonly used technique in the derivation of charges is RESP‎13 fitting, where point charges are assigned to atoms based on a grid of electrostatic potential points derived through quantum mechanical (QM) methods. It is an extension of the ESP
method and is more powerful, creating charges of higher quality because it is less dependent on the molecular conformation.\textsuperscript{14} Although it is an obvious approximation to treat charges as fixed, immutable, and located in a single point, it is necessary within the context of MD. There are other methods for deriving charges such as from Mulliken analysis\textsuperscript{15} or NBO analysis,\textsuperscript{16} but for our purposes, we only considered RESP as it is the one used in the charge derivation of FF99SB and its successors,\textsuperscript{11} and our aim was to be as compatible as possible with the established procedures. One final remark must be made about polarization and the effects that arise from it, which are generally not accounted for in the RESP approach. Some force fields, such as Amoeba,\textsuperscript{17} do try to incorporate these effects in their parameters, but this was not done for the FF14SB force field and its predecessors. Our investigation aims to follow the protocols laid out in the literature and therefore does not explicitly take into account polarization. This could be an interesting future venue of exploration to further improve our understanding on how to parameterize amino acids.

We used the programs NWChem\textsuperscript{18} and Red Server\textsuperscript{19} to perform the calculations necessary for RESP, each of which present different approaches to charge derivation. NWChem is a suite of tools that allows for many types of QM calculations; for the purposes of this study, we used the geometry optimization and the ESP charge fitting protocols to generate atomic charges for our systems. NWChem can be downloaded and run as an executable on common desktop machines and is relatively quick and easy to use. However, only a single conformation can be used as an input at any one time. It has been shown that charges vary with molecular conformation,\textsuperscript{19} thus necessitating the use of multiple conformations to obtain robust results.\textsuperscript{20} To investigate the effect of conformation on the derived charges, the Red Server package was also used; it is an online server and it is a little more complicated to use than NWChem. Both NWChem and Red Server allow for the use of several different basis sets and QM methods for the RESP fitting procedure. For the derivation of charges, we focused on perhaps the most commonly used basis set, 6-31G*, and employed both Hartree–Fock (HF) and the density functional theory exchange correlation function, B3LYP. Although the former method is the one most traditionally used for this task, because of a fortuitous cancellation of errors,\textsuperscript{22} B3LYP is also employed.\textsuperscript{23} The other charge derivation procedure we tested is the semiempirical QM approach using the AM1-BCC charge fitting procedure. For the derivation of charges, we focused on perhaps the most commonly used basis set, 6-31G*, and employed both Hartree–Fock (HF) and the density functional theory exchange correlation function, B3LYP. Although the former method is the one most traditionally used for this task, because of a fortuitous cancellation of errors,\textsuperscript{22} B3LYP is also employed.\textsuperscript{23} The other charge derivation procedure we tested is the semiempirical QM approach using the AM1-BCC charge fitting method,\textsuperscript{12} as used by programs such as MOPAC.\textsuperscript{24} It is meant to emulate RESP fitting and provide a faster alternative; conceptually, it is similar to RESP but the implementation is much quicker. It is an approximation which can satisfactorily describe the underlying electronic features and although not as accurate as RESP,\textsuperscript{16} it does so in a fraction of the time. This method is also available as part of AMBER and is employed to derive charges of small molecule ligands.\textsuperscript{26}

As previously mentioned, the conformation of the molecule has a great effect on the derived electrostatic potential;\textsuperscript{20} the same molecule in different orientations may produce different charge distributions depending on local polarization, steric clashes, and other similar effects. In order to obviate this problem, the literature protocol uses multiple conformations to describe each amino acid.\textsuperscript{11,27} These conformations, differing in backbone dihedrals, were obtained through a statistical analysis of amino acids in crystallized proteins. It was possible to do this for all natural amino acids, but it is, for obvious reasons, much harder to apply to non-natural ones, for which there would not be a large enough database to mine statistically relevant conformations. In our investigation, as we have only examined natural amino acids, we used the conformations reported in the literature with the intention of determining how this approach can then be applied to residues for which there is less data available.

**Test Case Methodology.** The amino acids isoleucine (Ile, I), lysine (Lys, K), glutamic acid (Glu, E), glutamine (Gln, Q), tyrosine (Tyr, Y), and phenylalanine (Phe, F) were chosen as they represent nonpolar, polar, positive, negative, and aromatic amino acids. Each amino acid (X) was capped using an acetyl group (Ace) at the N-terminus and an N-methylamide (Nme) at the C-terminus, resulting in Ace-X-Nme. Following the spirit of the derivation of charges for FF14SB,\textsuperscript{11} different combinations of $\phi$, $\psi$, $\chi^1$, and $\chi^2$ dihedrals were used to generate $\alpha_R$ or $C_5$ conformations, as shown in Figure 1.

![Figure 1. $\alpha_R$ (left) and $C_5$ (right) conformations of isoleucine.](image)

The RESP and AM1-BCC charges were generated separately for each of the two different conformations and were also subsequently averaged. As a result, 11 sets of derived charges were obtained: AM1-BCC $\alpha_R$, AM1-BCC $C_5$, the average of these two sets of charges, RESP $\alpha_R$ with HF, RESP $\alpha_R$ with B3LYP, RESP $C_5$ with HF, RESP $C_5$ with B3LYP, the average of the single conformation HF charges, the average of the single conformation B3LYP charges, the multiconformation HF charges, and the multiconformation B3LYP charges.

In each case, if possible, the backbone atoms of N, H, C, and O were restrained to the literature values\textsuperscript{11} (for neutral residues: $N = -0.4157$, $H = 0.2719$, $C = 0.5973$, and $O = -0.5679$; for positively charged residues: $N = -0.3479$, $H = 0.2747$, $C = 0.7341$, and $O = -0.5894$; and for negatively charged residues: $N = -0.5163$, $H = 0.2936$, $C = 0.5366$, and $O = -0.5819$), as the philosophy is to have a charge derivation that is as modular as possible. For AM1-BCC, which does not do this automatically, this step was not performed, which leads to those four atoms having different charges from the other methods. All chemically equivalent atoms, such as hydrogens attached to the same carbon or carbons occupying the same relative position in a ring were constrained, using symmetry, to have the same charge.

The obtained charges were all very similar to each other across individual amino acids. This means that, for example, the charges of each atom of the glutamic acid residue were consistent and differed usually only in the second significant digit onward; the carboxylic carbon had a strongly positive charge compensated by the negative charges of the oxygens, whereas the $\beta$-carbon and $\gamma$-carbon and their hydrogens were weakly charged, and so on. Similar patterns could be seen for all the other amino acids, with functional groups carrying the expected charge distributions, that is, similar to those reported...
in the literature. This is perhaps a testament to the robustness and versatility of both RESP and AM1-BCC.

All of the charges were obtained easily and automatically; by far, the fastest method was AM1-BCC, which operated within seconds, although the inability to restrain backbone charges is a drawback. NWChem and Red Server both took similar amounts of time, with the expected advantages and disadvantages inherent to obtaining and running a program on a machine and accessing a web server, respectively. The only exception is the multiconformational charge derivation of lysine, carried out on the Red Server, which did not work with any combination of parameters (neither with B3LYP or HF) nor by changing the convergence criteria. The program never managed to reach a minimized structure, and thus it was not considered in this study.

Once the charges were obtained, it was a matter of testing how the properties of the system were affected by the different charge models. Free energy was chosen as a method to measure this, as the computations of the free energies of binding between two molecules are highly dependent on the nonbonded parts of the force field equation, that is, the Coulombic and van der Waals terms. In order to evaluate the effect of the differing charges from the different models, two approaches have been used: the molecular mechanics Poisson–Boltzmann (generalized born) surface area [MMPB(GB)SA] approach and the multistate Bennett’s acceptance ratio (MBAR). MMPBSA is a widely used approach which is fast and has been shown to produce trends in binding affinities that are consistent with experimentally observed trends; in contrast, MBAR is much more accurate, but is also much more computationally expensive.

The system we have chosen to use as a test case is oligopeptide permease A (OppA; Figure 2). This protein transports oligopeptides into the cells, and the crystal structures of this protein in complex with tripeptides of the sequence KXX, where X is any of the 20 natural amino acids, were available. This makes it an ideal test case, as it allows us to have bound structures of proteins and peptides where the amino acid is at the interface and is therefore presumed to be important for the interaction. For each of the six selected amino acids, the relevant OppA complex was obtained from the pdb repository (PDB reference codes: KIK, 1b3g.pdb; KKK, 2olb.pdb; KEK, 1jeu.pdb; KFK, 1b40.pdb, KQK, 1b5j.pdb; KYK, 1b58.pdb).

In each complex, the crystallographic waters were retained, and then the whole system was solvated with an octahedral box which was truncated to ensure that no part of the system was less than 8 Å from the edge of the box; the system was neutralized with Cl\(^-\) or Na\(^+\) ions where necessary. The parameter file was modified to change the charge of the residue X to the values calculated in this study; all other amino acids were assigned the original FF14SB charges. The resulting system was minimized for 1500 steps of steepest descent followed by a further 3500 steps of conjugate gradient. Equilibration consisted of heating from 0 to 300 K in the NVT ensemble for 2.5 ns followed by a further 20 ns at 300 K in the NPT ensemble. Production simulations were carried out for 100 ns in the NPT ensemble at 300 K and 1 atm pressure, using the Langevin thermostat and Berendsen barostat, respectively. Periodic boundary conditions were employed using an 8 Å radius; inside the radius, the nonbonded terms were calculated explicitly, while the particle mesh Ewald method was used outside that radius. The interaction energy between the tripeptide and the receptor was then measured using MMPBSA with all the differing charge models, using the PB4 method. This was done to remove any intrinsic variation of the binding energy due to simulation variability.

We are aware that the interaction in this system is mediated by water, and although water at the interface between the ligand and protein can be of fundamental importance when calculating the binding free energy within this system, in this case, our focus is on only the difference between the binding free energies due to the differing charge models and hence not on the absolute values. The presence of interfacial waters is less important in this situation as we are not attempting to replicate the experimental values or obtain water-mediated dynamics; any error that derives from neglecting water molecules will be consistent across different systems.

Another venue of analysis was the calculation of the solvation energies, which are dependent on the nonbonded terms (as well as on the conformation of the molecule). A highly accurate free energy perturbation approach for the calculation of the free energy of binding differences is MBAR. These calculations were carried out using 19 lambda windows from 0.05 to 0.95, extrapolating to the end points. Production simulations were 1 ns in the NPT ensemble at 1 atm and 300 K using the Berendsen barostat and the Langevin thermostat, respectively. The system setup was the same as that used above for MMGBSA calculations.

Finally, some structural features that depend on charges were also analyzed and compared across different derivation methods.

## RESULTS AND DISCUSSION

### Comparison of Charges from the Different Charge Models.
Each charge derivation method generates, at the atomic positions, point charges that recreate the QM (or semiempirical) electrostatic potential. The charges for the six different amino acids are shown in the supplemental information for the different methods used.

The largest difference in single point charges between the methods is in the sidechain carbons, most notably at CB, which consistently has the largest standard deviation between the methods. The maximum difference in partial charge on any single atom type between the methods was 0.51. Charges can
also differ in sign; an example is the β-hydrogen of isoleucine, whose charge ranges from −0.0817 to 0.0696. In general, there is a relatively small difference between the partial charges obtained from the different approaches used here; an example is shown in Table 1 for isoleucine. More details on the actual values from each method for each amino acid are shown in the Supporting Information.

### Structural Comparisons. Salt Bridge Analysis

In addition to energetic differences, we have also looked at two structural features dependent on charges. The first is a salt bridge interaction that is present in the OppA–KEK system, where the central glutamate of the peptide interacts with Arg404. As this is an interaction that is directly dependent on the charges of the positively and negatively charged residues, it is reasonable to suggest that it might be susceptible to variations in atomic partial charges due to different methods of charge derivation. We therefore compared the distance between the carboxylic and guanidinium groups of these two residues to see if there are any significant differences in the salt bridge. It is to be noted that in this interaction the charges of the arginine residue do not change between different systems, as they remain those of the FF14SB literature values; it is the glutamate charges that change. The salt bridge is shown in Figure 3, while the averages are summarized in Table 2.

The distance between the center of mass of the carboxylic oxygen of the glutamate and the center of mass of the nitrogen atoms of the guanidine of the arginine over the simulation for each charge model is shown in Figure 4. The data follows a Boltzmann distribution, with an average value between 2.86 (for FF14SB) and 2.96 Å (for RESP, multiconformational, HF). As it can be seen in the table, the standard deviation between the charge methods is an order of magnitude smaller than that from each ensemble, showing no noticeable discrepancies between the different simulations.

This shows that there is very little difference in the salt bridge in terms of its length or interaction due to differences in the charges.

### Root Mean Square Fluctuation Analysis

Another type of structural analysis that we have carried was the measurement of the mobility of the residues as measured using the root mean square fluctuation (RMSF) as a metric. RMSF values were averaged over the set of three simulations. Because the ligand is buried, conventional MD was considered sufficient for sampling of the ligand bound in the cavity. The data is shown in Table 3. The RMSF values are small and without significant variations across the different charge derivation methods.

---

**Table 1. Average Charges for Each Atom in Isoleucine**

| Atom | Average | Standard Deviation | Range | Max | Min |
|------|---------|--------------------|-------|-----|-----|
| N    | −0.4157* | 0.0000*            | 0.0000* | −0.4157* | −0.4157* |
| H    | 0.2719*  | 0.0000*            | 0.0000* | 0.2719*  | 0.2719*  |
| CA   | −0.0886  | 0.0952             | 0.2748  | 0.0397   | −0.2351  |
| HA   | 0.1104   | 0.0410             | 0.1335  | 0.1839   | 0.0504   |
| CB   | 0.1717   | 0.1721             | 0.4370  | 0.3723   | −0.0647  |
| HB   | 0.0118   | 0.0498             | 0.1446  | 0.0817   | −0.0629  |
| CG2  | −0.1854  | 0.0745             | 0.2263  | −0.0941  | −0.3204  |
| HG21 | 0.0436   | 0.0150             | 0.0594  | 0.0882   | 0.0288   |
| HG22 | 0.0436   | 0.0150             | 0.0594  | 0.0882   | 0.0288   |
| HG23 | 0.0436   | 0.0150             | 0.0594  | 0.0882   | 0.0288   |
| CG1  | 0.0283   | 0.0902             | 0.2290  | 0.1436   | −0.0854  |
| HG12 | −0.0103  | 0.0433             | 0.1110  | 0.0482   | −0.0628  |
| HG13 | −0.0103  | 0.0433             | 0.1110  | 0.0482   | −0.0628  |
| CD1  | −0.0863  | 0.0359             | 0.1260  | −0.0427  | −0.1687  |
| HD11 | 0.0203   | 0.0134             | 0.0399  | 0.0381   | −0.0018  |
| HD12 | 0.0203   | 0.0134             | 0.0399  | 0.0381   | −0.0018  |
| HD13 | 0.0203   | 0.0134             | 0.0399  | 0.0381   | −0.0018  |
| C    | 0.5973*  | 0.0000*            | 0.0000* | 0.5973*  | 0.5973*  |
| O    | −0.5679* | 0.0000*            | 0.0000* | −0.5679* | −0.5679* |

*The exact partial charges for all the amino acids derived using the different charge derivation methods are listed in the Supporting Information. The backbone charges (N, H, C, O, marked with an asterisk) include data only from methods that can restrain charges, therefore excluding AM1-BCC methods. The other atoms include AM1-BCC charges.

**Table 2. Average Distance Across All Methods of the Carboxylic Group in KEK and the Guanidinium Group in Arginine 404**

| Method | Average | Standard Deviation (within method) |
|--------|---------|-----------------------------------|
| FF14SB | 2.86    | 0.17                              |
| AM1-BCC, single conformation αR | 2.83 | 0.17 |
| AM1-BCC average of conformations | 2.85 | 0.16 |
| RESP, single conformation αR, B3LYP | 2.95 | 0.19 |
| RESP, single conformation αR, B3LYP | 2.91 | 0.20 |
| RESP, average of B3LYP conformations | 2.93 | 0.21 |
| RESP, single conformation αR, HF | 2.93 | 0.23 |
| RESP, single conformation αR, HF | 2.83 | 0.14 |
| RESP, average of HF conformations | 2.84 | 0.13 |
| RESP, multiconformation, B3LYP | 2.96 | 0.24 |
| RESP, multiconformation, HF | 2.90 | 0.20 |
| Standard deviation (across methods) | 0.06 |

*The comparison includes the standard deviation for all measurements within the method, and the standard deviation across all methods. All values are in angstrom.

---
With these analyses, we believe we have shown that structural features that are likely to be dependent on charges appear to be agnostic to different charge derivation methods. Regardless of which variables are used, all produce values that are similar to those reported in the literature. Our next step in the investigation was to study what may be of more interest in a comparison with the experiment: free energies of binding, which are related to the interaction energy between the protein and the ligand. When employing non-natural amino acids, it is often done with the aim of producing a better binder, and thus it is crucial to ensure that new parameters do not skew the free energy calculations.

Free Energy Calculations. MMPBSA Results. The MD simulations carried out using the original and derived charges were processed to compute the free energies of interactions between tripeptides carrying each of the six amino acids chosen using MMPBSA calculations. Entropy was not considered when using MMPBSA, as it is notoriously complex and resource-intensive to address. The $\Delta G$ value describes the difference between each derivation method and that data obtained by using the original FF14SB values. The closer that $\Delta G$ is to zero, the more it can be assumed that small differences in charge have a minimal impact on the overall free energy. When calculating the averages of the $\Delta G$ values to obtain a mean for a particular residue or across a particular residue, it was carried out on the absolute values, ignoring the sign. By and large, the results in Table 4 seem to suggest that there is little difference in using one method over another for charge derivation, and all methods reproduce values similar to those obtained using literature charges. The $\Delta G$ ranges from 0.25 to approximately 1 kcal mol$^{-1}$ in terms of the absolute average, which is well within the error obtained for a typical MMPBSA or MMGBSA calculation. Although this difference is theoretically not negligible (as 2 kcal mol$^{-1}$ would imply a difference of about one order of magnitude in terms of $K_d$), it is too small to conclusively attribute it to the differences in charge derivation, and likely arises due to the inherent imprecision of MMPBSA.

Thus, we conclude that as long as MMPBSA is used as the tool for the analysis of free energy changes, all the examined charge derivation methods should give results that are comparable with each other and similar to the literature values.

Another level of analysis we have done is comparing MMPBSA results between different residues, as MMPBSA is often used to obtain trends between ligands. To examine the effect of the charge model for this purpose, we have calculated the $\Delta G$ between the peptide with isoleucine and the other ligands in turn. From Figure 5, we observe that there is very little difference between the charge models, all yielding the same trends for the ligands. The greatest deviation is between isoleucine and glutamic acid, with a standard deviation of 1.2 kcal mol$^{-1}$. All other $\Delta G$’s have a standard deviation less than 0.65 kcal mol$^{-1}$ between the different charge models (full values in the Supporting Information), which is comparable to thermal fluctuations; the overall trend clearly remains unchanged.

Additionally, we have investigated the sidechain contribution to the MMPBSA binding energy, specifically the electrostatic component, which depends the most on charges. The other components (polar solvation and van der Waals) are not shown here. The results are shown below in Table 5.

The size of the error bars and the lack of obvious overarching trends complicated the task of interpreting the data, but some conclusions could be drawn from the analysis. One observation that can be made is that the values obtained with AM1-BCC calculations are those that mirror the FF14SB ones most closely, although only for the polar residues: for glutamine and glutamic acid, with a standard deviation of 1.2 kcal mol$^{-1}$. All other $\Delta G$’s have a standard deviation less than 0.65 kcal mol$^{-1}$ between the different charge models (full values in the Supporting Information), which is comparable to thermal fluctuations; the overall trend clearly remains unchanged.

Additionally, we have investigated the sidechain contribution to the MMPBSA binding energy, specifically the electrostatic component, which depends the most on charges. The other components (polar solvation and van der Waals) are not shown here. The results are shown below in Table 5.

The size of the error bars and the lack of obvious overarching trends complicated the task of interpreting the data, but some conclusions could be drawn from the analysis. One observation that can be made is that the values obtained with AM1-BCC calculations are those that mirror the FF14SB ones most closely, although only for the polar residues: for glutamine and

Table 3. RMSF of the Central Residue of the KKK Peptide Averaged Across the Entire Simulation and all Runs, Shown for Each Method$^a$

|      | GLU | GLN | TYR | PHE | ILE | LYS |
|------|-----|-----|-----|-----|-----|-----|
| FF14SB | 0.96 | 1.22 | 1.53 | 0.93 | 1.29 | 1.85 |
| AM1-BCC, single conformation $\alpha$, B3LYP | 0.85 | 1.21 | 1.30 | 0.92 | 1.95 | 1.65 |
| AM1-BCC single conformation $\alpha$, B3LYP | 0.93 | 1.53 | 1.51 | 1.23 | 1.51 | 1.98 |
| AM1-BCC average of conformations | 0.79 | 1.28 | 1.92 | 1.10 | 1.53 | 1.92 |
| RESP, single conformation $\alpha$, B3LYP | 0.85 | 1.46 | 1.72 | 1.25 | 1.55 | 1.87 |
| RESP, single conformation $\alpha$, B3LYP | 0.77 | 1.26 | 1.53 | 1.09 | 1.19 | 1.94 |
| RESP, average of B3LYP conformations | 0.78 | 1.17 | 1.60 | 1.08 | 1.33 | 1.97 |
| RESP, single conformation $\alpha$, HF | 1.00 | 1.35 | 1.43 | 1.20 | 1.55 | 1.91 |
| RESP, single conformation $\alpha$, HF | 0.88 | 1.06 | 1.59 | 1.01 | 1.59 | 2.01 |
| RESP, average of HF conformations | 0.82 | 1.26 | 1.74 | 1.08 | 1.66 | 1.79 |
| RESP, multiconformation, B3LYP | 0.99 | 1.35 | 1.59 | 0.99 | 1.36 | n/a |
| RESP, multiconformation, HF | 0.87 | 1.55 | 2.02 | 0.98 | 1.54 | n/a |
| Average and standard deviation | 0.88 | 1.31 | 1.62 | 1.07 | 1.51 | 1.89 |

$^a$All values are in angstrom.
Table 4. ΔΔMMPBSA Data, Obtained by Subtracting from the FF14SB MMPBSA Free Energy the MMPBSA Free Energy Calculated Using a Different Method Parameters on the FF14SB Runs.

| Residue | ΔΔG (kcal mol⁻¹) | Standard Deviation (kcal mol⁻¹) |
|---------|------------------|---------------------------------|
| GLU     | -570 ± 10.37     | 0.91                            |
| GLN     | -292 ± 9.55      | 0.23                            |
| TYR     | -31.3 ± 6.30     | 0.77                            |
| PHE     | -31.8 ± 5.98     | 1.06                            |
| ILE     | -30.1 ± 5.3     | 0.77                            |
| LYS     | -30.1 ± 5.3     | 0.64                            |
| GLU     | -292 ± 9.55      | 0.23                            |
| GLN     | -31.8 ± 5.98     | 1.06                            |
| TYR     | -31.8 ± 5.98     | 1.06                            |
| PHE     | -30.1 ± 5.3     | 0.77                            |
| ILE     | -31.3 ± 6.30     | 0.77                            |
| LYS     | -31.3 ± 6.30     | 0.77                            |

*All values are in kcal mol⁻¹.

glutamic acid, the electrostatic contribution of those methods is very similar to the one obtained using literature charges. To a lesser extent, this is also true of lysine and tyrosine, whereas this is not really the case for isoleucine and phenylalanine. It can perhaps mean that AM1-BCC most closely replicates the charge distribution of more strongly charged residues. However, it should be noticed that the correlation between the electrostatic contribution to the binding energy and the binding energy itself is mediocre at best. The highest $R^2$ value for isoleucine (0.77), but it is much lower for the other amino acids: it goes from poor (glutamic acid, 0.34 and lysine, 0.26) to very poor (tyrosine, 0.11; phenylalanine, 0.02; and glutamine, 0.01). In other words, the electrostatic component is not necessarily predictive of the overall binding energy, and there are other factors that affect the free energy. Furthermore, the free energy results, as it can be seen above in Table 4, are very consistent across different charge models.

A further level of analysis that we have performed is correlating the obtained MMPBSA results with the sum of the absolute charge values for each derivation method. For each method and each amino acid, the absolute value of the charge was summed to obtain an overall value, which was averaged across each method and across each residue. The obtained number can be a property of the charge derivation method; even if the overall sum of the charges is 0, as it is the case for neutral amino acids, the sum of absolute values is a measure of the magnitude of individual charges. The average values for each amino acid across different methods yield unsurprising results: isoleucine and phenylalanine, which contain no heteroatoms in the sidechain, have lower sums than other residues, which have greater sums in proportion to the number of oxygens and nitrogens they include. The standard deviation is also fairly narrow, indicating that the results are consistent for each residue. This data is shown in Table 6.

This sort of analysis can reveal important information regarding the trends and correlations within different methods. For example, when averaging only the values of those methods that involve HF and averaging only those that use B3LYP, the former are generally of larger magnitude and have smaller standard deviations, as shown in Table 7. This is due to the well-known overpolarization of HF.38

It is however more interesting to consider the average for one charge derivation method across different residues and comparing that to the binding affinity data to obtain correlation values, which is shown in Tables 8aa and 8bb. Although it may seem strange to consider together charged, polar, and nonpolar residues alike, focusing on each method across different amino acids can reveal if there are trends in how the charges affect the interaction of the peptide. The average values across the method were correlated with the ΔG MMPBSA results obtained by using different charges on the FF14SB runs. As shown by the $R^2$, the correlation goes from poor (glutamine and phenylalanine) to very poor (tyrosine and lysine). We believe that this supports our observation that there is very little difference between charge derivation methods: even in the case of glutamine, the change in the sum of charges given by the differences in charge derivation methods correlates poorly with the obtained values, and all other residues have even weaker correlations. Thus, although they may yield charges of different magnitude and of different distributions, all charge derivation methods end up giving similar results when analyzed with MMPBSA. Even when considering known differences within...
methods, this does not translate to significantly different MMPBSA results.

**MBAR Results.** The other free energy approach we used to evaluate the effect of charges was MBAR. We calculated the hydration energy of the Ace-X-Nme system. Table 9 presents the average of the absolute $\Delta\Delta G$ values from MBAR. The

| method | average | standard deviation |
|--------|---------|-------------------|
| FF14SB | 4.23    | 0.87              |
| AM1-BCC, single conformation $\alpha_R$ | 4.62 | 1.03 |
| AM1-BCC single conformation $C_5$ | 4.66 | 1.02 |
| AM1-BCC average of conformations | 4.64 | 1.02 |
| RESP, single conformation $\alpha_R$, B3LYP | 4.64 | 1.07 |
| RESP, single conformation $C_5$, B3LYP | 4.11 | 0.81 |
| RESP, average of B3LYP conformations | 4.33 | 0.90 |
| RESP, single conformation $\alpha_R$, HF | 4.79 | 1.08 |
| RESP, single conformation $C_5$, HF | 4.68 | 1.14 |
| RESP, average of HF conformations | 4.72 | 1.09 |
| RESP, multiconformation, B3LYP | 3.75 | 0.87 |
| RESP, multiconformation, HF | 4.15 | 1.07 |
average ΔΔG for the MBAR results across all the residues and charge models is 0.28 kcal mol⁻¹ with a standard deviation of 0.18 kcal mol⁻¹. Tyrosine has the largest differences, with a max ΔΔG of 1.06 kcal mol⁻¹ (RESP, average of B3LYP conformations) and an average of 0.63 kcal mol⁻¹. The next highest difference is for phenylalanine with a max ΔΔG of 0.81 kcal mol⁻¹ (AM1-BCC, single conformation αR) and an average of 0.28 kcal mol⁻¹, which is much lower than that for tyrosine. Charged residues, and those containing a phenyl ring, have larger average differences and standard deviations. Of the various methods examined, no single approach produced the smallest differences across the amino acids. Overall, the average differences between the different methods are quite comparable, however, AM1-BCC single conformation Cα and RESP multiconformational HF from the Red Server performed the best, and the averaged B3LYP approach performed the worst. The experimental uncertainty for binding free energies is approximately 0.6 kcal mol⁻¹, setting the upper limit for in silico approaches. Differences in predicted binding free energy results within this range could therefore be considered indistinguishable. On the basis of this conclusion, all charge methods produce fairly comparable results, although some appear to be better.

### CONCLUSIONS

Given the evidence that we have presented above, we can conclude that any of the approaches that we have used to derive charges can reproduce, within error, the values that are observed when using the FF14SB charges. In other words, if we were to use these methods to obtain charges for an unknown, non-natural amino acid, we would do so with the confidence that our results would be comparable and compatible with parameters of existing residues.

That is not to say that there are not any differences between the methods. For example, from the analysis of the electrostatic contribution to the binding energy, the AM1-BCC methods seem to replicate the values obtained with FF14SB the most closely for polar residues. From the MBAR data, the methods that use HF instead of B3LYP have greater agreement with FF14SB results and therefore appear to be superior for charge derivation.

However, all methods generally produce results that are within error. This is because the general approach, which includes steps such as restraining the backbone and maintain equivalence between chemically identical atoms, is robust enough to account for minor variations due to differing initial structures, or differing basis sets and quantum methods. The resulting charges may vary, and this may have an impact on the simulation, but such small differences are eclipsed by other interactions which do not depend on charges, by the inherent variability of MD runs, and by the intrinsic inaccuracy of the methods commonly used to analyze the simulations. It is our opinion that an accurate assigning of the atom types that determine the internal parameters is far more important than the minute differences in charge, as the former control bonds, angles, torsions, and van der Waals interactions which will have a larger effect on the overall simulation.

It should be noted however that we have only tested relatively easy systems that involved natural amino acids, which are well-known and lack problematic features. The largely aliphatic and well-documented nature of the residues did not include functional groups which may have been described by our approaches in a less robust manner. Future efforts should focus on determining what is the best approach to parametrize systems involving highly polarizable or highly charged species, metals, or halides, which may all present unique challenges.

In conclusion, the current study provides compelling evidence that for the investigated class of chemical species, all examined charge derivation approaches produce equivalent results.

### ASSOCIATED CONTENT

#### Supporting information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b00438.

Tables for all the derived atomic charges for each residue and the statistical analysis of these charges; the binding data and the ΔΔG of the calculations based on FF14SB

---

Table 9. MBAR Data for the Transformation between FF14SB Charges and the Charges Derived Using Different Methods

| Method                          | GLU  | GLN  | TYR  | PHE  | ILE  | LYS  | average |
|---------------------------------|------|------|------|------|------|------|---------|
| AM1-BCC, single conformation αR | 0.26 | 0.26 | 0.04 | 0.81 | 0.41 | 0.20 | 0.33    |
| AM1-BCC single conformation Cα | 0.27 | 0.02 | 0.23 | 0.27 | 0.11 | 0.16 | 0.18    |
| AM1-BCC average of conformations| 0.17 | 0.20 | 0.37 | 0.59 | 0.05 | 0.22 | 0.27    |
| RESP, single conformation αR, B3LYP | 0.30 | 0.00 | 0.32 | 0.46 | 0.22 | 0.09 | 0.23    |
| RESP, single conformation Cα, B3LYP | 0.28 | 0.37 | 0.57 | 0.24 | 0.30 | 0.12 | 0.31    |
| RESP, average of B3LYP conformations | 0.31 | 0.09 | 1.06 | 0.08 | 0.06 | 0.70 | 0.38    |
| RESP, single conformation αR, HF | 0.28 | 0.37 | 0.57 | 0.24 | 0.30 | 0.12 | 0.31    |
| RESP, single conformation Cα, HF | 0.08 | 0.29 | 0.91 | 0.44 | 0.01 | 0.38 | 0.35    |
| RESP, average of HF conformations | 0.40 | 0.06 | 0.73 | 0.48 | 0.06 | 0.01 | 0.29    |
| RESP, multiconformation, B3LYP | 0.08 | 0.04 | 0.70 | 0.26 | 0.24 | n/a  | 0.26    |
| RESP, multiconformation, HF | 0.29 | 0.22 | 0.33 | 0.04 | 0.01 | n/a  | 0.18    |
| average                         | 0.23 | 0.15 | 0.63 | 0.28 | 0.17 | 0.21 | 0.28    |
| standard deviation              | 0.12 | 0.14 | 0.26 | 0.21 | 0.12 | 0.21 | 0.18    |
runs; tables containing sums of the absolute values of the charges, per residues and across methods, and their correlation with binding data; and the decomposition of sidechain contribution to the binding energy (PDF)

■ AUTHOR INFORMATION
Corresponding Authors
*E-mail: pietroa@bita-star.edu.sg (P.G.A.).
*E-mail: chandra@bita-star.edu.sg (C.S.V.).

ORCID
Pietro G. A. Aronica: 0000-0002-2927-8378
Chandra S. Verma: 0000-0003-0733-9798

Author Contributions
$Equal contributions.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS
We gratefully acknowledge A*STAR and an A*STAR-JCO grant for funding, support, and expertise.

■ REFERENCES
(1) Liang, G.; Fox, P. C.; Bowen, J. P. Parameter analysis and refinement toolkit system and its application in MM3 parameterization for phosphine and its derivatives. J. Comput. Chem. 1996, 17, 940–953.
(2) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2013, 3, 198–210.
(3) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Pettrella, R. J.; Roux, B.; et al. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 2009, 30, 1545–1614.
(4) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 1995, 91, 43–56.
(5) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; et al. Scalable molecular dynamics with NAMD. J. Comput. Chem. 2005, 26, 1781–1802.
(6) Levitt, M. The birth of computational structural biology. Nat. Struct. Mol. Biol. 2001, 8, 392–393.
(7) Bowen, J. P.; Allinger, N. L. Molecular Mechanics: The Art and Science of Parameterization. Reviews in Computational Chemistry; John Wiley & Sons, Inc., 1991; pp 81–97.
(8) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; 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. 2010, 31, 671–690.
(9) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Summerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 2015, 11, 3696–3713.
(10) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Struct., Funct., Bioinf. 2010, 78, 1950–1958.
(11) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. J. Comput. Chem. 1995, 16, 1357–1377.
(12) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model II: Parameterization and validation. J. Comput. Chem. 2002, 23, 1623–1641.
(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. J. Am. Chem. Soc. 1993, 115, 9620–9631.
(14) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. J. Phys. Chem. 1993, 97, 10269–10280.
(15) Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. J. Chem. Phys. 1955, 23, 1833–1840.
(16) Foster, J. P.; Weinhold, F. Natural hybrid orbitals. J. Am. Chem. Soc. 1980, 102, 7211–7218.
(17) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; et al. Current Status of the AMOEBA Polarizable Force Field. J. Phys. Chem. B 2010, 114, 2549–2564.
(18) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. Comput. Phys. Commun. 2010, 181, 1477–1489.
(19) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. Phys. Chem. Chem. Phys. 2010, 12, 7821–7839.
(20) Stouch, T. R.; Williams, D. E. Conformational dependence of electrostatic potential-derived charges: Studies of the fitting procedure. J. Comput. Chem. 1993, 14, 858–866.
(21) Reynolds, C. A.; Essex, J. W.; Richards, W. G. Atomic charges for variable molecular conformations. J. Am. Chem. Soc. 1992, 114, 9075–9079.
(22) Sherrill, C. D.; Merz, K. M. Quantum Mechanical Methods for Quantifying and Analyzing Non-Covalent Interactions and for Force-Field Development. In Many-Body Effects and Electrostatics in Biomolecules; Cui, Q., Meuwly, M., Ren, P., Eds.; CRC Press: New York, 2016; p 73.
(23) Kang, Y. K.; Scheraga, H. A. An efficient method for calculating atomic charges of peptides and proteins from electronic populations. J. Phys. Chem. B 2008, 112, 5470–5478.
(24) Stewart, J. J. P. MOPAC: A semiepiirical molecular orbital program. Proteins: Struct., Funct., Bioinf. 1990, 4, 1–103.
(25) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. J. Mol. Graphics Modell. 2006, 25, 247–260.
(26) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. J. Am. Chem. Soc. 2001, 222, U403.
(27) Reynoldson, C. A.; Essex, J. W.; Richards, W. G. Atomic charges for variable molecular conformations. J. Am. Chem. Soc. 1992, 114, 9075–9079.
(28) Lee, M. R.; Duan, Y.; Kollman, P. A. Use of MM-PBSA in estimating the free energies of proteins: Application to native, intermediates, and unfolded villin headpiece. Proteins: Struct., Funct., Bioinf. 2000, 39, 309–316.
(29) Goghen, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discovery 2015, 10, 449–461.
(30) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 2008, 129, 124105.
(31) Suenaga, A.; Okimoto, N.; Hirano, Y.; Fukui, K. An Efficient Computational Method for Calculating Ligand Binding Affinities. PLoS One 2012, 7, No. e42846.
(32) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, R. J. H. Crystallographic and calorimetric analysis of peptide binding to OppA protein. J. Mol. Biol. 1999, 291, 393–415.
(33) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Isaiah, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Merz, K. M.; Monard, G.; York, D. M.; Kollman, P. A. Amber 2017 Reference Manual; University of California, 2017.
(34) Huggins, D. J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein–ligand interactions. Protein Eng., Des. Sel. 2011, 24, 777–789.
(35) Tame, J. R. H.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* 1996, 3, 998.
(36) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 2011, 51, 69–82.
(37) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson–Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* 2005, 48, 7796–7807.
(38) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* 2002, 23, 1623–1641.
(39) Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A. The experimental uncertainty of heterogeneous public Ki data. *J. Med. Chem.* 2012, 55, 5165–5173.
(40) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* 2012, 8, 4405–4412.
(41) Lin, F.; Wang, R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.* 2010, 6, 1852–1870.
(42) Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* 2011, 32, 2564–2574.