Recent Progress in General Force Fields of Small Molecules

Xibing He,[1] Brandon Walker,[2] Viet H. Man,[1] Pengyu Ren,[2*] and Junmei Wang[1*]

[1]Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA

[2]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, USA

*Corresponding authors:

Pengyu Ren (pren@utexas.edu)

Junmei Wang (juw79@pitt.edu)

Abstract

Recent advances in computational hardware and free energy algorithms enable a broader application of molecular simulation of binding interactions between receptors and small molecule ligands. The underlying molecular mechanics force fields (FFs) for small molecules have also achieved advancements in accuracy, user friendliness and speed during the past several years (2018-2020). Besides expansion of chemical space coverage of ligand-like molecules among major popular classical additive FFs and polarizable FFs, new charge models have been proposed for better accuracy and transferability, new chemical perception of avoiding predefined atom types have been applied, and new automated parameterization toolkits including machine learning approaches have been developed for users' convenience.

Introduction

Molecular simulations have played essential roles in biochemical and biophysical sciences [1-4]. They capture the structural characteristics and dynamic behaviors of biomolecules (proteins, nucleic acids, carbohydrates, etc.) interacting with ligands, solvents, co-solvents, and other molecules at an atomistic level and fine temporal resolution. Especially in the field of modern drug discovery, *in silico* simulations are commonly used to virtually screen potential compounds active to specific drug targets from vast compound pools, typically by calculating the protein-ligand binding free energies for drug lead identifications and optimizations. Such a vital application of molecular simulations is highly desired but still holds many challenges currently. A key to the success of molecular simulation studies and structure-based rational drug design is the quality of the utilized molecular mechanics force field (MMFF). A MMFF consists of a set of simple

functions for bonded terms (bonds, angles, dihedrals, improper dihedrals) and nonbonded Coulombic and van der Waals (VDW) interactions and associated parameters, which enable the calculation of potential energies and forces.

Among the numerous MMFFs, the following four families are the most popular ones for atomistic molecular dynamics (MD) simulations of biological systems: Assisted Model Building with Energy Refinement (AMBER) [5,6], Chemistry at HARvard Macromolecular Mechanics (CHARMM) [7,8], Optimized Potentials for Liquid Simulations (OPLS) [9,10], and GROningen MOlecular Simulation (GROMOS) [11]. These force field (FF) families usually include specific FFs for proteins, nucleic acids, lipids, and carbohydrates. They also include general FFs for various small molecules, such as the General AMBER Force Field (GAFF) [12,13], the CHARMM General Force Field (CGenFF) [14-16], and expanded parameters in recent OPLS [17] and GROMOS series [18]. It is a tradition to group a MMFF to a family mainly based on its nonbonded terms, especially VDW parameters. Thus, even a general FF carries different name from the FF from which it is originated, it still belongs to the same family according to this tradition. The recent review by Nerenberg and Head-Gordon [19] summarizes the developments before 2018 in all-atom biomolecular FFs for proteins, nucleic acids, and small molecules. In this review, we focus on the progresses of general FFs for small molecules from 2018 to 2020. We summarize classical additive FFs first, and then polarizable FFs (Figure 1).

Major upgrade of popular small-molecule force fields.

The most notable update among the widely used small-molecule FFs during the past two years is the release of OPLS3e [17]. Building upon the previous OPLS series, OPLS3e further expanded with extensive parameters for bonds, angles and torsional terms related to drug-like compounds.

[17]. It also integrated a ligand-specific approach to assign atomic charges on the fly [17]. The charge model included off-atom centered virtual sites for charge distributions of lone pairs and sigma holes as CGenFF had done [16]. The improved parameter transferability lead to an improved performance on small molecule conformational energy profiles, solvation free energies, and receptor−ligand binding free energies [17]. This FF is implemented in the commercial Schrodinger software suite.

Since 2015, the second generation of GAFF (GAFF2) has steadily expanded its parameters, which can be obtained for free through AmberTools. Originally, GAFF and GAFF2 were developed utilizing the RESP (restrained electrostatic potential) [12] method to assign charges for atoms, which fits the electrostatic potential from *ab initio* HF/6-31G* calculation, the same protocol as for the AMBER protein FFs. But in practice, the users of GAFF/GAFF2 prefer the fast semi-empirical AM1-BCC model [20,21] to obtain the atomic partial charges because *ab initio* calculations are avoided. Very recently He et al. adjusted the bond charge correction (BCC) terms in the original AM1-BCC model for GAFF2, and achieved mean unsigned error (MUE) of only 0.37 kcal/mol for the hydration free energy of more than 400 organic solutes [13]. This new version of charge model (named ABCG2) combined with GAFF2 parameters also works well on the calculations of solvation free energy for nearly 900 pairs of various organic solutes in various organic solvents, whose dielectric constants range from 1.8 to 37.2 (Figure 2), and the corresponding MUE is only 0.51 kcal/mol [13]. This result demonstrated the capability of dealing different dielectric environments by the newly derived ABCG2 model, which is important for quantitatively predicting transferring free energies and binding free energies.

Although CGenFF and GROMOS did not announce major updates on small molecule parameters, new toolkits (described below) emerged to facilitate the FF parameterization.

Automated toolkits to facilitate parameterizations.

FF parameterization is a tedious and time-consuming process which often involves decisions from human experts. In order to reduce the burden of developing missing parameters for non-expert users, automated toolkits have been developed to generate specific parameters (usually bonded terms or partial charges) from quantum mechanics (QM) data, such as Parmscan and Antechamber for GAFF/GAFF2, Paramfit for AMBER, ffTK for CGenFF, ATB for GROMOS, LigParGen for OPLS-AA, and Poltype for polarizable FF AMOEBA [22-25]. Recently Nash et al. presented a Java tool, ForceGen [26], to extract the force constants and equilibrium values of bonds and angles via performing a vibrational frequency analysis, and the output is formatted with the Gromacs topology. The MacKerell lab developed the FFParam package [22] to facilitate the parametrization process for CGenFF and CHARMM Drude polarizable FF. Horton et al. developed a toolkit, QUBEKit (QUantum mechanical BEspoke Kit), which derives FF parameters directly from quantum mechanics for specific small molecules [27].

Machine-learning methods in force field parameterization

Recently, machine learning (ML) methods were adopted in FF parameterization for efficiency. Galvelis et al. combined a general FF and several neural network potentials (NNPs) to improve dihedral parameters [23]. They demonstrated that small molecules can be parameterized in much shorter time with torchani-ANI-1x NNP compared to equivalent procedure using density

functional theory (DFT) calculations. However, extensive evaluation is necessary on applying NNPs in MMFF parameterization.

Martin et al. used ML algorithms to fast assign partial charges for screening molecules which are encoded as a cyclic undirected graph with atoms corresponding to vertices, and bonds to edges [28]. A web tool, ContraDRG [28], was developed to assign partial charges using random forest models. Similarly, Wang et al. recently developed a deep learning algorithm to predict QM-based atomic charges via a graph convolutional network [29].

Wilkins et al. [30] calculated the static dipole polarizability tensors of over 7,000 small organic molecules (containing up to seven heavy atoms) using highly accurate QM LR-CCSD theory, and used a ML approach named SA-GPR to accurately predict the LR-CCSD molecular polarizabilities with a negligible computational cost. The molecular dipole polarizability is an important parameter in some polarizable FFs discussed below.

New chemical perception without relying on atom types

Usually, the process of assigning appropriate FF parameters for a molecule is done based on a set of pre-defined atom types. The Open Force Field Consortium worked on an approach to automatically recognize the moieties and assign each type of parameters via standard chemical substructure queries [31,32]. This approach was previously used by Poltype [25] to facilitate automated parameter assignment for AMOEBA FF. The identification of specific atoms inside a chemical pattern was carried out via an industry-standard SMARTS language and its SMIRKS extensions [31]. Starting from GAFF atom types and parameters, a new FF format was

implemented and named as the SMIRKS Native Open Force Field (SMIRNOFF)[31]. The obtained definition file of parameters contains only approximately 300 lines but can cover five million drug-like molecules. The obtained FF (named SMIRNOFF99Frosst) showed similar accuracy on free energies of hydration for small molecules and tested properties of organic bulk liquids compared with GAFF [31].

Another recently developed method of determining FF torsional parameters without relying on the atom type concept is H-TEQ (Hyperconjugation for Torsional Energy Quantification) [24,33,34]. It is based on the following chemical principle and hypothesis: (1) the torsional interaction is controlled by a combination of hyperconjugation, electrostatic and steric effect; (2) a hyperconjugation term can replace the conventional torsion form to improve torsional profiles; (3) the hyperconjugation term can be derived from a simple chemical property – electronegativity of the atoms along the torsion. The authors of H-TEQ found that the parameters of their proposed hyperconjugation term can be conveniently obtained from the electronegativity values of central and neighboring atoms with a few correlation rules, and do not need any atom types. With this method, atomic charge, electronegativity, and hence hyperconjugation terms can be derived on the fly for molecules of interest [33]. The authors showed that H-TEQ performed comparably well to GAFF in reproducing QM torsional profiles for diverse organic molecules [24,33], and better torsional profiles for conjugated drug-like molecules [34].

Polarizable force fields.

The classical additive FF models remain problematic when the same set of fixed partial charges is applied to different environments, in which the charge distribution is expected to change, such as gas to aqueous solution, solvent to protein cavity, cell membrane permeation, and heterogeneous interfaces. Growing effort has been made in improving the electrostatic model and to address the lack of polarization in additive models. Inakollu et al. recently reviewed the polarizable FFs and discussed in depth the key interactions and special cases that are difficult to model well with a purely additive FF [35]. A more in-depth review of polarizable FF algorithms and recent successes for a variety of biological systems is discussed elsewhere [36]. Currently, empirical polarizable models that have been applied to common organic molecules can be divided into four classes: fluctuating charge, Drude oscillator, gaussian electrostatic, and induced dipole. However, the fluctuating charge model has not been actively developed during the past years. Hence, it is not discussed in this review.

The Drude oscillator method introduces explicit polarization by attaching a charged virtual particle (the Drude oscillator) to every polarizable atom via a harmonic spring. The atomic dipoles can change according to the surrounding electric field by optimizing the positions of Drude particles with respect to the fixed atomic cores. The CHARMM Drude oscillator model for small molecules, including alkanes, ethers, alcohols, amides, aromatics, sulfur containing compounds, ketones, aldehydes and halogenated ethanes and benzenes, is reviewed in more depth elsewhere [37]. Ions in particular are susceptible to polarization effects and monovalent ions have been previously parameterized to with CHARMM, as well as molecular ions such as ammonium and derivatives, imidazolium, guanidinium, acetate, methanethiolate, ethanethiolate, phenolate [38]. Parameters for the polarizable Drude ions were optimized to be consistent with experimental hydration free energies of the neutral salts while generating accurate energies and geometries for monovalent ions [38]. This was an improvement over previous non-polarizable ion model for CHARMM.

The Gaussian Electrostatic Model (GEM) and Gaussian Multipole Model (GMM) utilizes a continuous representation of charge density with gaussian functions to model polarization and charge penetration effects. The AMBER Force Field Consortium developed polarizable Gaussian Multipole (pGM) model in which a Gaussian function or its derivatives are adopted to represent the atomic multipoles for more efficient electrostatics [39]. Another unique ML based many-body potential is the MB-pol model, which has shown excellent results for water and its interactions with ions and CH_4 etc. [40-42].

The AMOEBA polarizable FF employs induced dipoles on atomic sites to model polarization. [43,44]. Additionally, an automated procedure known as Poltype, has been described to generate small molecule parameters for AMOEBA [25]. A later version Poltype 2, is currently in development and includes additional features such as fragmentation for torsion fitting, ring puckering (for torsion parameterization), as well as automated VDW parameterization and 2D torsion-torsion correction maps. A molecular fragmenter for torsion parameterization is also implemented to significantly reduce the computation time of necessary *ab initio* calculations. AMOEBA has been successfully applied to protein-ligand binding free energy computations, typically with ligands containing many rings such as host-guest systems [45]. Poltype has been used to derive parameters for organochlorine compounds [46], IN17 derivatives to inhibit MELK (many ring containing compounds) [47], ALDOLASE inhibitors containing phosphate groups and fluorinated groups [48], ATP and ADP [49]. A range of small organic molecules have also been validated in implicit solvents, producing high quality experimental hydration free energy [46]. Recent improvements have been made to AMOEBA functional forms. This new generation FF, AMOEBA+, includes the new charge penetration, charge transfer and geometry-dependent charge flux terms. Charge flux parameters have been derived for alkane, alkyne, amine, alcohol, Di-oxyl, sulfide, halide, ketone, carboxylate ester, amide, anhydride, oxalic acid oxalate, malonic acid, imide, azide, nitro, amino acid, sulfoxide, thiocyanate, phosphate, benzene derivatives [50].

Potential problems in general MMFF development

Accuracy, efficiency, transferability, integrity of deriving force field parameters in a consistent fashion, and compatibility between FFs when multiple FFs are employed to describe a system are the major factors measuring the success of a MMFF for small molecules. Unlike the first two metrics, transferability, integrity, compatibility to the accompanying FFs (biomolecular FFs, lipid FFs, water and ion models etc.) are sometimes neglected. Critical evaluation on a MMFF and its variants require a great deal of effort. A FF variant which is likely to increase accuracy or efficiency by re-parameterizing a subset of parameters might adversely affect the integrity, transferability and compatibility of the FF as a whole. This "unwanted" outcome may occur when nonbonded terms are changed and the training set data are limited. Thus, systematic evaluation a FF variant is critical to maintain or improve a general MMFF performance measured by the five metrics.

Conclusions and outlook

In the last two years, we have witnessed a significant advance in general force field (FF) development for organic molecules, which in turn enhances the accuracy of thermodynamic prediction. Machine learning algorithms have been increasingly applied in FF development, especially on high-quality partial charge assignments. The artificial neural network-based potentials represented by ANI-1x, may be applied to detect bad FF parameters efficiently, even though currently they may not be able to take the role of high-level *ab initio* models to generate reference data for MMFF parameterization.

Advanced FFs such as polarizable FFs for small molecules have also shown encouraging success in various applications. Parameterization of advanced FF benefits from direct use of quantum mechanics molecular properties in gas-phase due to explicit treatment of polarization. Nonetheless, additional physical terms and parameters require more data and more sophisticated parameterization approaches based on big data. Robust and automated tools for parameterization is essential for increasing application of advanced FFs to small molecules.

Notes Conflict of interest statement

Pengyu Ren is a co-founder of qubit pharmaceuticals. The other authors declare no competing financial interest.

Acknowledgements

The authors gratefully acknowledge the funding support from the National Science Foundation (NSF) and National Institutes of Health (NIH) to J.W. (NSF 1955260, NIH R01GM079383), and to P.R. (National Institutes of Health (NIH R01GM106137, NIH R01GM114237 and NSF CHE-1856173).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- **• of special interest**
- **•• of outstanding interest**

1. Hollingsworth SA, Dror RO: **Molecular Dynamics Simulation for All**. *Neuron* 2018, **99**:1129-1143.

2. Hossain S, Kabedev A, Parrow A, Bergstrom CAS, Larsson P: **Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning**. *Eur J Pharm Biopharm* 2019, **137**:46-55.

3. Sledz P, Caflisch A: **Protein structure-based drug design: from docking to molecular dynamics**. *Curr Opin Struct Biol* 2018, **48**:93–102.

4. Aminpour M, Montemagno C, Tuszynski JA: **An overview of molecular modeling for drug discovery with specific illustrative examples of applications**. *Molecules* 2019, **24**: 1693.

5. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz Jr, KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA: **A second generation force field for the simulation of proteins, nucleic acids, and organic molecules**. *J Am Chem Soc* 1995, **117**:5179-5197.

6. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C: **ff19SB: amino-acid-specific protein backbone**

parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* 2020, **16**:528−552.

7. MacKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher III WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, M. Karplus M: **All-atom empirical potential for molecular modeling and dynamics studies of proteins**. *J Phys Chem B* 1998, **102**:3586-3616.

8. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell Jr AD: **CHARMM 36m: an improved force field for folded and intrinsically disordered proteins**. *Nat Methods* 2017, **14**:71-73.

9. Jorgensen WL, Maxwell DS, 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* 1996, **118**:11225-11236.

10. Robertson MJ, Qian Y, Robinson MC, Tirado-Rives J, Jorgensen WL: **Development and testing of the OPLS-AA/M force field for RNA**. *J Chem Theory Comput* 2019, **15**:2734-2742.

11. Oostenbrink C, Villa A, Mark AE, van Gunsteren W: **A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6**. *J Comput Chem* 2004, **25**:1656–1676.

12. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA: **Development and testing of a general amber force field**. *J Comput Chem* 2004, **25**:1157-1174.

13. He X, Man VH, Yang W, Lee TS, Wang J: **A fast and high-quality charge model for the next generation general AMBER force field**. *J Chem Phys* 2020, **153**:114502.

•• A new set of charge model named ABCG2 which demonstrates better accuracy and transferability in solvation free energy calculation of various organic solutes in a diversity of organic solvents compared to the original AM1-BCC charge model when combined with GAFF2.

14. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell Jr AD: **CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field**. *J Comput Chem* 2010, **31**: 671-690.

15. Yu W, He X, Vanommeslaeghe K, MacKerell Jr AD: **Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations**. *J Comput Chem* 2012, 33:2451-2468.

16. Soteras Gutierrez I, Lin FY, Vanommeslaeghe K, Lemkul JA, Armacost KA, Brooks III CI, MacKerell Jr AD: **Parametrization of halogen bonds in the CHARMM general force field: improved treatment of ligand-protein interactions**. *Bioorg Med Chem* 2016, **24**: 4812-4825.

17. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED: **OPLS3e: extending force field coverage for drug-like small molecules**. *J Chem Theory Comput* 2019, **15**:1863−1874.

•• The most recent version of OPLS general force field for drug-like molecules.

18. Horta BAC, Merz PT, Fuchs PFJ, Dolenc J, Riniker S, Hünenberger PH: **A GROMOS-compatible force field for small organic molecules in the condensed phase: the 2016h66 parameter set**. *J Chem Theory Comput* 2016, **12**:3825−3850.

19. Nerenberg PS, Head-Gordon T: **New developments in force fields for biomolecular simulations**. *Curr Opin Struct Biol* 2018, **49**:129–138.

•• This review summarizes the recent developments before 2018 in all-atom biomolecular force fields for proteins, nucleic acids, and small molecules.

20. Jakalian A, Bush BL, Jack DB, Bayly CI: **Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method**. *J Comput Chem* 2000, **21**:132-146.

21. Jakalian A, Jack DB, Bayly CI: **Fast, efficient generation of highquality atomic charges. AM1-BCC model: II. Parameterization and validation**. *J Comput Chem* 2002, **23**:1623-1641.

22. Kumar A, Yoluk O, MacKerell Jr AD: **FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules**. *J Comput Chem* 2020, **41**:958–970.

• Most recent toolkit for parameterization of small molecules compatible to CHARMM.

23. Galvelis R, Doerr S, Damas JM, Harvey MJ, De Fabritiis G: **A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning**. *J Chem Inf Model* 2019, **59**:3485−3493.

24. Champion C, Barigye SJ, Wei W, Liu Z, Labute P, Moitessier N. **Atom type independent modeling of the conformational energy of benzylic, allylic, and other bonds adjacent to conjugated systems**. *J Chem Inf Model* 2019, **59**:4750−4763.

25. Wu JC, Chattree G, Ren P: **Automation of AMOEBA polarizable force field parameterization for small molecules**. *Theor Chem Acc* 2012, **131**:1138.

26. Nash A, Collier T, Birch HL, de Leeuw NH: **ForceGen: atomic covalent bond value derivation for Gromacs**. *J Mol Model* 2018, **24**:5.

27. Horton JT, Allen AEA, Dodda LS, Cole DJ: **QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics**. *J Chem Inf Model* 2019, **59**:1366−1381.

28. Martin R, Heider D: **ContraDRG: automatic partial charge prediction by machine learning**. *Front Genet* 2019, **10**:990.

29. Wang J, Cao D, Tang C, Xu L, He Q, Yang B, Chen X, Sun H, Hou T: **DeepAtomicCharge: a new graph convolutional network-based architecture for accurate prediction of atomic charges**. *Briefings Bioinf* 2021, **22**:bbaa183. https://doi.org/10.1093/bib/bbaa183

30. Wilkins DM, Grisafi A, Yang Y, Lao KU, DiStasio Jr RA, Ceriotti M. **Accurate molecular polarizabilities with coupled cluster theory and machine learning**. *Proc Natl Acad Sci USA* 2019, **116**:3401–3406.

31. Mobley DL, Bannan CC, Rizzi A, Bayly CI, Chodera JD, Lim VT, Lim NM, Beauchamp KA, Slochower DR, Shirts MR, Gilson MK, Eastman PK: **Escaping atom types in force fields using direct chemical perception**. *J Chem Theory Comput* 2018, **14**:6076−6092.

• Traditional pre-defined graph of atom types is replaced by automatic chemical substructure queries via the industry-standard SMARTS chemical perception language, using SMIRKS extensions that permit labeling of specific atoms within a chemical pattern.

32. Zanette C, Bannann CC, Bayly CI, Fass J, MK, Shirts MR, Chodera JD, Mobley DL. **Toward learned chemical perception of force field typing rules**. *J Chem Theory Comput* 2019, **15**:402−423.

• Further description by the Open Force Field Initiative about their approach to automatically specify atom types and more general fragment types typically used in biomolecular force fields.

33 Liu ZM, Pottel J, Shahamat M, Tomberg A, Labute P, Moitessier N: **Elucidating Hyperconjugation from Electronegativity to Predict Drug Conformational Energy in a High Throughput Manner**. *J Chem Inf Model* 2016, **56**:788−801.

34. Wei W, Champion C, Barigye SJ, Liu Z, Labute P, Moitessier N: **Use of extended-Hückel descriptors for rapid and accurate predictions of conjugated torsional energy barriers**. *J Chem Inf Model* 2020, **60**:3534−3545.

35. Inakollu VSS, Geerke DP, Rowley CN, Yu H**: Polarisable force fields: what do they add in biomolecular simulations?** *Curr Opin Struct Biol* 2020, **61**:182–190.

•• This review summarizes the latest developments in accounting for key biomolecular interactions with polarizable force fields and applications to address challenging biological questions.

36. Jing Z, Liu C, Cheng SY, Qi R, Walker BD, Piquemal JP, Ren P: **Polarizable force fields for biomolecular simulations: recent advances and applications**. *Annu Rev Biophys* 2019, **48**:371-94.

•• This review describes the recent advances and applications, as well as future directions, of polarizable force fields in biomolecular simulations.

37. Lin F-Y, MacKerell AD. **Force fields for small molecules. Methods in molecular biology** (Clifton, NJ). 2019, 2022:21-54.

• This review discusses the importance of small molecule force fields and their evolution from additive to polarizable. Emphasis is placed on the additive CHARMM Force Field and the classical Drude oscillator.

38. Lin FY, Lopes PEM, Harder E, Roux B, MacKerell AD: **Polarizable Force Field for Molecular Ions Based on the Classical Drude Oscillator**. *J Chem Inf Model* 2018, **58**:993-1004.

39. Wei H, Qi R, Wang J, Cieplak, Duan Y, Luo R**: Efficient formulation of polarizable Gaussian multipole electrostatics for biomolecular simulations**. *J Chem Phys* 2020, **153**:114116.

• The most recent work on polarizable Gaussian Multipole (pGM) model defined with respect to a local frame formed with a set of covalent basis vectors.

40. Arismendi-Arrieta D, Riera M, Bajaj P, Prosmiti R, Paesani F: **i-TTM model for ab initio-based ion–water interaction potentials. 1. halide–water potential energy functions**. *J Phys Chem B* 2016, **120**:1822-1832.

41. Riera M, Hirales A, Ghosh R, Paesani F: **Data-driven many-body models with chemical accuracy for CH4/H2O mixtures**. *J Phys Chem B* 2020, 124, **49**:11207–11221.

42. Muniz MC, Gartner TE, Riera M, Knight C, Yue S, Paesani F, Panagiotopoulos AZ: **Vapor–liquid equilibrium of water with the MB-pol many-body potential**. *J Chem Phys* 2021, **154**:21103.

43. Ren P, Wu C, Ponder JW: **Polarizable atomic multipole-based molecular mechanics for organic molecules**. *J Chem Theory Comput* 2011, **7**:3143-3161.

44. Mu X, Wang Q, Wang LP, Fried SD, Piquemal JP, Dalby KN, Ren P: **Modeling organochlorine compounds and the sigma-hole effect using a polarizable multipole force field**. *J Phys Chem B* 2014, **118**:6456-6465.

45. Laury ML, Wang Z, Gordon AS, Ponder JW: **Absolute binding free energies for the SAMPL6 Cucurbit[8]uril host-guest challenge via the AMOEBA polarizable force field**. *J Comput-Aided Mol Des* 2018, **32**:1087-95.

46. Corrigan RA, Qi G, Thiel AC, Lynn JR, Walker BD, Casavant TL, Lagardere L, Piquemal J-P, Ponder JW, Ren P, Schnieders MJ: **Implicit solvents for the polarizable atomic multipole AMOEBA force field**. *J Chem Theory Comput* 2021, **17**:2323-2341.

47. Harger M, Lee J-H, Walker B, Taliaferro JM, Edupuganti R, Dalby KN, Ren P: **Computational insights into the binding of IN17 inhibitors to MELK**. *J Mol Model* 2019, **25**:151.

• This article details an assessment of the current capabilities of the AMOEBA force field, to serve as a molecular dynamics-based free energy perturbation platform for lead optimization on the MELK-IN17 cancer target.

48. Qi R, Walker B, Jing Z, Yu M, Stancu G, Edupuganti R, Dalby KN, Ren P: **Computational and experimental studies of inhibitor design for Aldolase A**. *J Phys Chem B* 2019, **123**:6034-6041.

49. Walker B, Jing Z, Ren P: **Molecular dynamics free energy simulations of ATP:Mg2+ and ADP:Mg2+ using the polarisable force field AMOEBA**. *Mol Simul* 2020, 0:1-10.

50. Yang X, Liu C, Walker BD, Ren P: **Accurate description of molecular dipole surface with charge flux implemented for molecular mechanics**. *J Chem Phys* 2020, **153**:064103.

Figure Captions

Figure 1. The themes of small molecule force fields reviewed in this article.

Figure 2. The calculated solvation free energies of 895 pairs of various neutral organic solutes in diverse neutral organic solvents versus their experimental data. The dielectric constants (epsilon) of the solvents range from 1.8 to 37.2: A, epsilon < 2.0; B, epsilon between 2.0 and 3.0; C, epsilon between 3.0 and 10.0; D, epsilon between 10.0 and 40.0. The calculations were done with the thermodynamic integration (TI) method, and the solvent and solute molecules were described with the GAFF2 parameters with the ABCG2 charge model. The data were taken from Ref. 13 and were re-analyzed. RMSE: root mean square error; MUE: mean unsigned error; MSE: mean signed error; PI: predictive index; R: Pearson's correlation coefficient; n: the number of data.