

AUTOMATED FITTING OF TRANSITION STATE FORCE FIELDS FOR BIOMOLECULAR SIMULATIONS

Taylor R. Quinn,^{1,2} Himani N. Patel,² Kevin J. Koh,² Brandon E. Haines,³ Per-Ola Norrby,⁴ Paul Helquist,²
and Olaf Wiest^{2,5}

¹Early TDE Discovery, Early Oncology, Oncology R&D, AstraZeneca, Boston, United States

² Department of Chemistry and Biochemistry, University of Notre Dame, IN 46556, USA

³Department of Chemistry, Westmont College, Santa Barbara, CA 93108, USA

⁴Data Science and Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, SE-43183
Mölndal, Sweden

⁵Lab of Computational Chemistry and Drug Design, School of Chemical Biology and Biotechnology,
Peking University, Shenzhen Graduate School, Shenzhen, China

ABSTRACT

The generation potential energy functions (PEF) that are orders of magnitude faster to compute but as accurate as the underlying training data from high-level electronic structure methods is one of the most promising applications of machine learning (ML) in chemistry. In contrast to such studies in materials and small molecules, which parameterizes the entire system without constraints on the functional form of the PEF, the simulation of biomolecular systems requires that the PEF is compatible with one of the extensively validated biomolecular force fields. Here, we describe the application of the quantum guided molecular mechanics (Q2MM) method to transition states of enzymatic reactions to generate a transition state force field (TSFF) with the functional form of the well-established AMBER force field. The differences to fitting small molecule TSFFs and the similarities of the approach to transfer learning are discussed. Finally, the application of the to the transition state of the second hydride transfer in HMGCoA Reductase from *Pseudomonas mevalonii* is demonstrated.

Introduction

Understanding how enzymes achieve their catalytic function is one of the grand challenges of chemistry and biology. Studying enzymes using computational methods has produced highly impactful work, as highlighted by the award of the Nobel Prize in 2014¹ for the development of multiscale methods such as the Quantum Mechanics/Molecular Mechanics (QM/MM) method.² Because enzymes consist of tens of thousands of atoms, using even low level electronic structure methods is cost prohibitive for the full system. Furthermore, extensive sampling of the conformational space, e.g. by molecular dynamics simulation at the microsecond time scale for the enzyme, possible ligands, and the surrounding water molecules, is necessary to obtain physically meaningful results. To enable such simulations, a range of classical force fields that approximate atoms and bonds as masses connected by springs have been developed.^{3,4} The accuracy of these simulations is dependent on the quality of the force field used.⁵ As a result, extensive validation studies of the force field functional form as well as the parameters themselves have been performed.

The use of machine learning (ML) methods in science and technology has expanded exponentially in recent years, in part due to the rapid expansion in computational power and available datasets. In chemistry, applications of ML range from basic research through material research to drug discovery.⁶ More pertinent to the topic of the present study, ML has been applied to force field and PEF parameterization given its strengths in pattern recognition.⁷⁻¹⁰ There are numerous examples in materials chemistry, where the accurate description of large systems to predict material properties demanded a cheap method at high accuracy.¹¹ Another well-recognized example is the ANI-1 potentials¹² that use active learning and neural network algorithms to take high-level QM data to create transferable ML potentials.¹³ Even though the development of ML methods for the treatment of enzymatic reactions provides an alternative to the computationally expensive QM/MM methods, there have been comparatively few ML applications for force field development reactions and/or biomolecular systems. One reason is that in most cases, the new potential energy surfaces created break away from the restrictions of a predefined functional form.

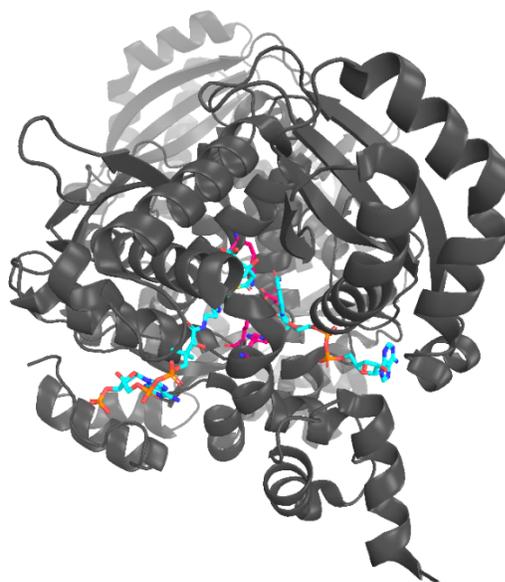


Figure 1. Structure of *PmHMGR* (pdb code 1QAX 2.80 Å resolution) with substrate, cofactor and reparameterized residues shown in colored stick representation while the remainder of the protein is shown in gray cartoon representation.

This is less likely to be successful for the case of the study of reactions in biomolecular systems, as exemplified by *Pseudomonas mevalonii* 3-hydroxy-3-methylglutaryl-CoA Reductase (*PmHMGR*) shown in Figure 1. Here, the vast majority of the system (shown in grey) is well described by extensively validated classical force fields. However, these cannot describe the substrate, cofactor, or residues involved in the transition state the reaction (shown in color). The large dataset needed for training of an ML PEF for the reactive center is not available from experimental data and cannot be generated from high-level electronic structure calculations due to their high computational cost. Here, we propose an alternative approach that is reminiscent of transfer learning where the functional form and extensively validated parameters of a classical force field (in the present case, AMBER) are used and retrained for a subset of the structure that includes the bond breaking and making atoms as well as key active site residues and cofactors (shown in color in Figure 1) using the quantum-guided molecular mechanics

(Q2MM) method that was originally developed for the parameterization of small molecule force fields, especially TSFFs.^{14, 15}

As mentioned before, most of the work done on the use of ML for all-atom force fields has been focused on small molecules or solvents using functional forms determined by e.g. a neural network.^{16, 17} There are a few examples of the use of ML for fitting predefined functional forms using both linear and non-linear regression algorithms in the literature to reproduce training data from appropriate electronic structure methods. ML in the form of a genetic algorithm was used to optimize a polarizable force field from ab initio QM data¹⁸ as well as the parameterization of reactive force fields.¹⁹ The Parsely force field for small molecules uses QM data for parameterization of an AMBER-lineage with SMIRNOFF atom specification.²⁰ Similarly, the AMBER-15 Force Balance force field²¹ for use with the TIP3P-Force Balance water model²² is fitted using a weighted least-squares method. The AMOEBA-2013 force field also was optimized using automated techniques to obtain a general polarizable protein force field.²³ However, these studies concern ground state (GS) force fields that are not able to describe bond breaking and making steps of an enzymatic reaction where a TSFF is needed.

One of the best established²⁴ automated fitting procedures for the parameterization of both ground state and transition state force fields (TSFF) is the Quantum Guided Molecule Mechanics (Q2MM) approach that has been used extensively for the development of TSFF for the prediction of stereoselectivity of small molecule reactions.²⁵⁻²⁸ To the best of our knowledge, the only application of Q2MM to biomolecular systems is a TSFF for transition-state docking of small molecular drugs to P450 enzymes to identify potential sites of hydroxylation.^{29, 30} However, the code used for this fitting procedure is to the best of our knowledge not widely available.

Q2MM uses training data from electronic structure (usually density functional theory) reference calculations to automatically parameterize molecular mechanics TSFF based on the MM3* PEF. The details of this process for asymmetric catalysis by small molecules have been covered elsewhere^{14, 15, 31} and will not be elaborated on here. Here, we will describe the application of the Q2MM method to derive TSFFs of a predefined functional form compatible with the AMBER-family force fields with particular

attention to the differences to the fitting of small molecule TSFFs. We will also discuss the interfacing of the Q2MM tools to the AMBER suite of molecular dynamics programs and demonstrate this workflow for the case of a TSFF for the second hydride transfer of *PmHMGR*.

Fitting Methods

Q2MM fits the FF parameters by minimizing the value of objective or loss function,

$$\chi^2 = \sum_i w_i^2 (x_i^0 - x_i)^2 \quad (1)$$

where x_i^0 is the reference data point, x_i is the FF data point, and w_i is the weight for that data point. The minimization step in the parameter space is calculated using gradient-based method such as the Newton-Raphson technique and simplex method.³² The gradient-based method is general and utilizes the Jacobian matrix J where

$$J_{ij} = \frac{\partial x_i}{\partial p_j} \quad (2)$$

and p_j is j-th parameter, which is calculated in many programs using numerical differentiation and therefore the rate-determining step. Thus, the simplex method is often used to avoid the high cost of numerical derivatives.³³ The simplex method in Q2MM is modified to move toward the best point(s) in the parameter space using a bias of reflection point.³² The modified simplex method has shown to have faster convergence than the Raphson-type methods up to ca. 40 parameters.³¹ Thus, it is used to optimize a medium-sized parameter set or a subset of the larger parameter set.

Q2MM, unlike most traditional methods for fitting system-specific FF parameters,^{22, 34, 35} uses the Hessian Matrix for the fitting of force constants of bonded parameters with geometric data for reference structures.^{32, 36, 37} The Hessian matrix is the second partial derivative of the energy with respect to the xyz coordinates of atoms, which gives the matrix size of $3N \times 3N$ where N is the number of atoms. It can be obtained by appropriate electronic structure calculations of suitable model systems including, in the case of the large biomolecular systems discussed here, QM/MM calculations. In the later case, the calculation

of the Hessian Matrix usually needs to be limited to a subsystem due to the memory demands of such calculations. The Hessian matrix's eigenvalues and eigenvectors provides information on the vibrational frequencies and normal modes, respectively. Normally, eigenvalues of the Hessian matrix are positive, but at the transition state geometry, the eigenvalues contain one significantly negative value with its eigenvector representing the reaction vector. By providing Hessian matrix information in the objective function, Q2MM uses information on both the transition state geometry and the shape of the potential energy surface around it when fitting the FF parameters. However, because Q2MM fits these parameters to represent the transition state, which is a saddle point, as a minimum on the potential energy surface, the matrix element that corresponds to the negative eigenvalue is, inevitably, altered during the fitting process. This leads to an increase of the objective function value.

To address this and the fact that the algorithms in most molecular force field-based programs^{38, 39} are designed to optimize towards minima rather than transition states, a small modification to Q2MM is made. Traditionally, in the Cartesian Hessian fitting method, all indices of the Hessian matrix are accounted for in the objective function with respect to the reference values. However, different weights are assigned to each element of the Hessian matrix to correctly represent the transition state as a minimum. The indices of Hessian matrix are given a weight of 0.0 to 1-1 interactions, 0.031 to 1-2 and 1-3 interactions, 0.31 to 1-4 interactions and 0.031 to all other interactions.⁴⁰ The Cartesian Hessian matrix fitting method is used for large molecule systems such as an enzyme, where only one reference structure is used to fit the parameters.

Alternatively, users can use the eigenmode fitting method in Q2MM. In this method,⁴¹ the reference Hessian matrix $H = V^T E V$ is decomposed into eigenvector V and eigenvalue E . Then the objective function includes the calculated eigenvalue matrix E' where $E' = V H' V^T$ and H' is the Hessian matrix of the FF calculated Hessian matrix. By preserving the original eigenvector V , all of the originally positive eigenvalues are preserved and only the negative eigenvalue is converted into a positive value by zeroing the weight of the eigenvalue to represent a transition state as a

minimum. This method has yielded an FF that is stable to unnatural distortions and is used for small-molecule systems such as metal-ligand-substrates, where multiple reference data are used to fit the parameters. It should be noted that this inversion of the potential energy surface in the reaction coordinate is done to allow the use of simple energy minimization techniques available in all force field packages to locate the stationary point. However, it is not absolutely required and alternative approaches have been developed.⁴²

The Q2MM Flow Scheme

The following parameterization scheme is specific towards the implementation of the AMBER20³⁹ interface of Q2MM and its use for large large biomolecular systems. Details of the method regarding parameterization of TSFF for asymmetric catalysis using other programs such as Macromodel have been documented elsewhere.¹⁵ As an example of using Q2MM for a large biomolecular system, the development of a TSFF for the second hydride transfer transition state of *Pm*HMGR,⁴³⁻⁴⁵ shown in Figure 1, will be discussed. Examples of the files discussed in this section as well as the final TSFF are given in the Supporting Information. The Q2MM code itself, which contains the interface to the AMBER Suite of programs, and several published TSFFs are freely available on the Q2MM/CatVS github repository (github.com/q2mm).

In order to develop a TSFF for an enzyme, the first step is to define a model system that includes the reactive species and the relevant parts of the protein involved in catalysis to generate the training data for the TS of this model system. For the example discussed here, the QM/MM or theozyme⁴⁶ model incorporated the relevant residues in the QM region derived from our previous studies^{43, 44} of the mechanism of HMGR and shown in Figure 2, though other model systems were also explored.⁴³ Since this model system is derived from electronic structure calculations, only the most essential atoms should be included for efficiency of the fitting procedure even though the methodology is equally applicable to larger numbers of refitted atoms. A `fixedatoms.txt` file is created to include any atoms frozen in the electronic structure calculation (Figure 2, green atoms). Because the frozen atoms create unphysical

Hessian elements, the weight of the Hessian values associated with these atoms are set to zero during the parameterization. Results of transition state optimizations, in a .log file, contain the energetic and geometric data that are used by Q2MM in the parametrization and are thereby included in the Q2MM input as reference. Currently, Q2MM supports interfaces to Gaussian⁴⁷ and Jaguar³⁸ .log files as training data for the parameterization process. The .log file is also used to create a .mol2 file of the model system using the RESP protocol in AMBER. The .mol2 file contains updated partial charges of all the atoms in the model system at the TS and is for used throughout the parametrization.

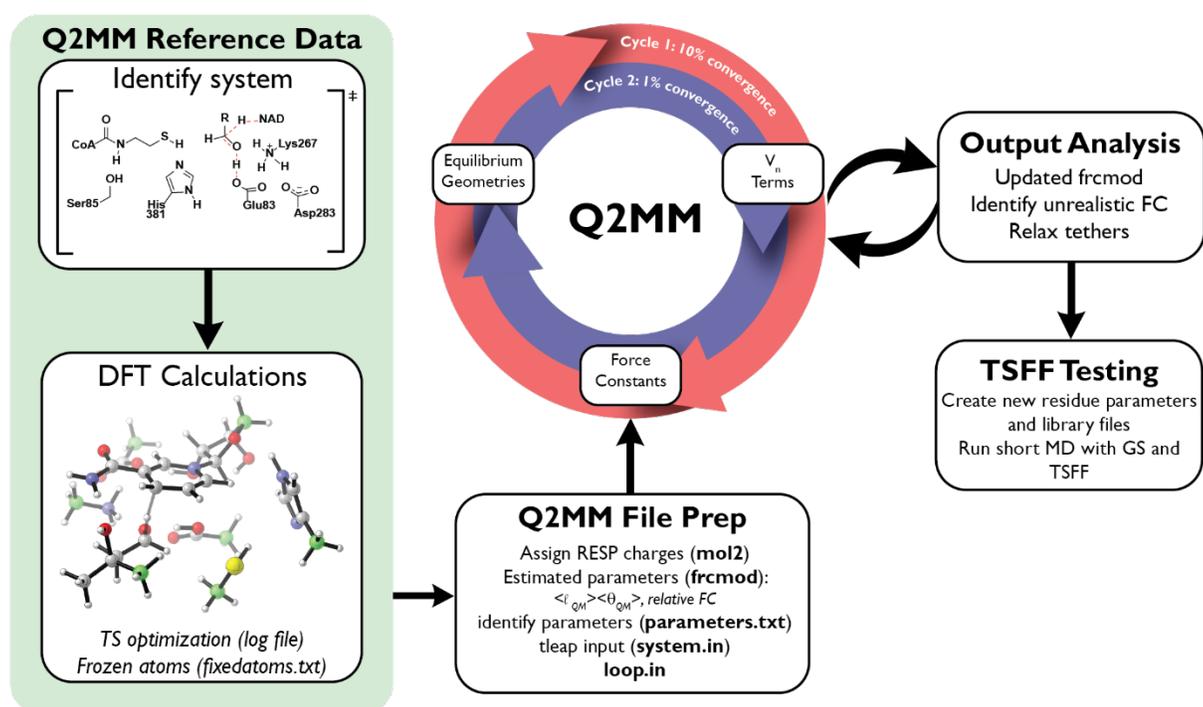


Figure 2. Flow scheme of the Q2MM method for the parameterization of TSFFs for enzymatic reactions using the AMBER interface

At this point, new atom types should be assigned to the atoms directly involved in the reaction, as their properties will be sufficiently different from that of the parent force field. This allows for the parameters defined by the TSFF to be restricted to a specific atom in the entire system. The atoms to be reparametrized in the case discussed here are shown in Fig. 3A. It should be noted that this procedure is analogous to transfer learning in that parameters trained to a much larger dataset (standard parameters of

the Amber force field) and extensively validated in the literature are used as a starting point for retraining a much smaller subset for which smaller training data sets are available. It is a key difference from the development of TSFFs for transition metal catalyzed reactions^{14, 15, 25, 26} where there are usually no parameters available for the transition metal environment. As a result, a much larger training set is needed in those cases to achieve a reliable TSFF. Even though the number of atoms to be retrained is usually larger for the case of enzyme catalyzed reactions, the use of a transfer learning approach makes the fitting procedure much more effective because the vast majority of atoms only undergoes minor perturbations in proceeding from the ground state to the transition state of the reaction.

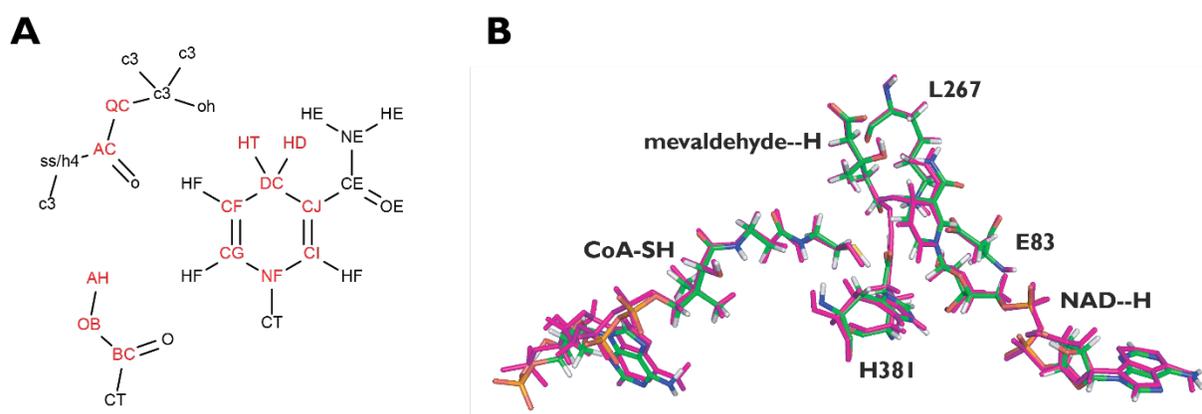


Figure 3. (A) reassigned atom types and (B) overlay of geometries of substrate, cofactor, and reparametrized protein atoms from electronic structure calculations (purple) and from the TSFF (atom colored)

The .mol2 file should also be used to generate the force field modification (frcmod) file, using antechamber program of AMBER.³⁹ The .frcmod file needs to be updated accordingly to be used in Q2MM, examples of which can be found in the documentation on github. All parameters such as bonds, angles, and dihedrals for atoms directly involved in the reaction should be included in the .frcmod file. Transition state parameters are different from the ground state ones, so initial guesses of the bond lengths and angles should be for the system at the TS as described by the QM reference data. The estimation of

the parameters prevents optimization to local nonphysical minima of the objective function and decreases the number of iterations required for parameterization. Force constants are initially set to standard values based on the Generalized Amber Force Field (GAFF),⁴⁸ and initial estimations for dihedrals should in our experience be avoided to prevent over-parameterization. The parameter.py module of Q2MM generates a list of a specified parameter type to be optimized that references the .frcmod file line and includes the range of values acceptable for that parameter type.

The input file, loop.in, for Q2MM files should contain all of the relevant information for an optimization cycle. The FFLD being read every cycle should be the given AMBER .frcmod file and the RDAT being read should be the Gaussian or Jaguar .log file. For CDAT, a tleap input file that calls the mol2 file and frcmod file of the model system and relevant Amber force fields should be created to generate a prmtop and inpcrd file that is used during the parameterization process. The optimization criteria of the penalty function are set in the loop.in file under the LOOP flag. Initially the penalty function can be set to a 10% convergence criterion. The loop.in file can be submitted by >python loop.py loop.in .

Partial charges should remain unchanged throughout the course of the parameterization. The order of parameterization (Figure 2) is largely the same as discussed earlier.¹⁵ The force constants should be optimized first while ensuring that the optimized value stays above $32.2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for bond distances and angles and $3.2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for dihedrals. Subsequently, the bond length parameter can be refined to reflect the reference data. Bond angles can be optimized after the bond lengths while ensuring that the optimized values are within reasonable ranges. If the optimized angles deviate towards unreasonable values, then this angle parameters value should be “tethered” to the reference data to prevent major deviations during optimization. The tether is defined as a weight value associated that would thereby control the deviation of the parameter being optimized. A higher tether weight should be used in the first round of optimization, then slowly decreased to zero in subsequent optimizations cycles. Finally, the V_n terms for the torsional potentials are fit to the Hessian data first before being further refined. A second round of optimizations should be performed with a 1% convergence criterion for the

penalty function to allow parameter refinement to be closer to the reference data parameters. Additional optimization cycles can be performed as needed until a working transition state force field has been obtained. For enzymatic systems, a working TSFF is obtained when an optimization step changes the objective function by less than a 1% and the values and parameters are deemed realistic by the given user.

Additionally, the resulting force field should be tested in a large-scale molecular dynamics simulations in conjunction with the ground state force field to describe the remainder of the protein (shown in gray in Figure 1). The TSFF will have to be parsed to generate new residue types that contain reparametrized atoms and new library files will need to be created to read into the leap module of AMBER20. This could also involve setting conditions that allow the reacting atoms to have more than the standard amount of bonds in a system. Other important considerations are adjusting the time step of the simulation to account for the vibrations of the reacting atoms and potentially removing the SHAKE algorithm for hydrogens in the TSFF. A short MD simulation should then be performed to ensure that the total energy of the system remains stable with the TSFF in combination with the ground state FF that would be used for the rest of the enzyme.

Application to *Pm*HMGR

This method described above was employed for the second hydride transfer TS of *Pm*HMGR. Here, the reference data for the training of the TSFF were obtained from QM/MM calculations where the atoms indicated in Figure 2 were treated at the ONIOM-(B3LYP/6-31G(d,p):AMBER) level of theory.^{43, 44} This includes the side chains of H381, K267, D283 and E83 as well as the substrates and cofactor as shown in Figure 3 and the hmgrqm.log example file in the Supporting Information. As the functional form of the underlying force file to which to fit the TSFF to, AMBER99SB and GAFF for atoms on residues and substrates were used, respectively, as seen in the ts2.frcmod file. During parameterization, the full size of the substrates and cofactor, along with the backbone and sidechains of the residues mentioned above, were included while calculating the MM data (Figure 3B). As discussed earlier, the bonding character and partial charges of the atoms directly involved in the TS change in going from the ground to the transition state. Furthermore, the nicotinamide ring of the cofactor is dearomatized. To describe these perturbations, new atom types were introduced as indicated in Figure 3A. It is worth reemphasizing that the initial parameters for these new atom types were derived from the standard ground state AMBER99SB parameters and then trained to reproduce the electronic structure results in the

training data. In this specific case, only parameters directly associated with these atoms (within 3 bonds) were reparametrized for the TSFF.

As shown in Figure 3 B, the TSFF successfully reproduced the geometries around the reacting center of the active site and could successfully be incorporated into the rest of the enzyme that is treated with a traditional ground state force field. Using this, the enzyme could be simulated at the transition state on the microsecond timescale. The results of these studies will be discussed elsewhere.

Conclusions

In this contribution, we have discussed an automated workflow that combines the Q2MM method with transfer learning-type approaches for the generation of fast and accurate TSFFs for large biomolecular systems. Application of the workflow to the second hydride transfer of HMGR, an enzyme of high biomedical importance, shows that the transition state of this reaction can be accurately reproduced by the TSFF derived by this workflow.

The use of machine learning to generate potential energy functions that are orders of magnitude faster to compute than their training data, which often are derived from accurate but slow electronic structure calculations, is a promising application of ML in chemistry. The work presented here uses the philosophy of transfer learning and applies it to the parametrization of TSFF by retraining of well validated existing force fields as oppose to creating completely new atom types and parameters, as is done in the generation of small molecule TSFF that cover transition metal catalyzed reactions. The results are an early example for using only electronic structure reference data and a much larger number of parameters adjusted in the biomolecular TSFF than in the earlier cases of small molecule TSFFs. They show that idea derived from ML can be used to parameterize a TSFF to simulate enzymes at the transition state $\sim 10^4$ times faster than the underlying electronic structure methods, allowing for molecular dynamics simulation for system sizes and timescales well beyond the accessibility of DFT-based methods.

Acknowledgments

This work was supported by the National Institutes of Health through grant R01GM111645 to OW and PH. BEH, HNP and TRQ are fellows of the Chemistry-Biochemistry-Biology Interface (CBBI) Program at the University of Notre Dame, supported by training grant T32GM075762 from NIGMS.

Additional Supporting Information containing the example files for the Q2MM fitting discussed in the manuscript may be found in the online version of this article. The full Q2MM code is available at github.com/q2mm.

Correspondence to: Olaf Wiest (E-mail: owiest@nd.edu)

FUNDING

This work supported by the National Institutes of Health through CBBI Fellowships to TRQ, BEH, and HP (T32GM075762) and grant R01GM111645 to OW and PH.

References

1. Warshel, A. Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture). *Angew. Chem. Intl. Ed Engl.* **2014**, 53, 10020-10031.
2. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol* **1976**, 103, 227-249.
3. Dauber-Osguthorpe, P.; Hagler, A. T. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *J. Comp. Aid. Des.* **2019**, 33, 133-203.
4. Hagler, A. T. Force field development phase II: Relaxation of physics-based criteria... or inclusion of more rigorous physics into the representation of molecular energetics. *J. Comp. Aid. Des.* **2019**, 33, 205-264.
5. Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. USA* **2018**, 115, E4758-E4766.
6. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Mod.* **2017**, 57, 942-957.
7. Jordan, M. I.; Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, 349, 255-260.
8. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, 145, 170901.
9. Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Intl. Ed.* **2017**, 56, 12828-12840.
10. Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett* **2007**, 98, 146401.
11. Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *npj Comp. Mat.* **2017**, 3, 1-8.
12. Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, 8, 3192-3203.

13. Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Comm.* **2019**, *10*, 1-8.
14. Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O. Prediction of Stereochemistry using Q2MM. *Acc. Chem. Res.* **2016**, *49*, 996-1005.
15. Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. Application of Q2MM to predictions in stereoselective synthesis. *Chem. Comm.* **2018**, *54*, 8294-8311.
16. Yao, K.; Herr, J. E.; Toth, D. W.; McKintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: A neural network augmented with long-range physics. *Chem. Sci.* **2018**, *9*, 2261-2269.
17. Schran, C.; Uhl, F.; Behler, J.; Marx, D. High-dimensional neural network potentials for solvation: The case of protonated water clusters in helium. *J. Chem. Phys.* **2018**, *148*, 102310.
18. Li, Y.; Li, H.; Pickard, F. C.; Narayanan, B.; Sen, F. G.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B. Machine Learning Force Field Parameters from Ab Initio Data. *J. Chem. Theor. Comp.* **2017**, *13*, 4492-4503.
19. Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396-9409.
20. Lim, V. T.; Mobley, D. Benchmark Assessment of Molecular Geometries and Energies from Small Molecule Force Fields. *ChemRxiv* **2020**, <https://doi.org/10.26434/chemrxiv.12551867.v1>.
21. Wang, L. P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023-4039.
22. Wang, L. P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885-1891.
23. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theor. Comp.* **2013**, *9*, 4046-4063.

24. Norrby, P.-O.; Åkermark, B.; Haeffner, F.; Hansson, S.; Blomberg, M. Molecular Mechanics (MM2) Parameters for the (eta-3-allyl) palladium Moiety. *J. Am. Chem. Soc.* **1993**, 115, 4859-4867.
25. Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O. Rapid virtual screening of enantioselective catalysts using CatVS. *Nature Catalysis* **2019**, 2, 41-45.
26. Rosales, A.; Ross, S. P.; Helquist, P.; Norrby, P.-O.; Sigman, M. S.; Wiest, O. Transition State Force Field for the Asymmetric Redox Relay Heck Reaction. *J. Am. Chem. Soc.* **2020**, 142, 9700–9707.
27. Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Prediction of enantioselectivity in rhodium catalyzed hydrogenations. *J. Am. Chem. Soc.* **2009**, 131, 410-411.
28. Limé, E.; Lundholm, M. D.; Forbes, A.; Wiest, O.; Helquist, P.; Norrby, P.-O. Stereoselectivity in asymmetric catalysis: The case of ruthenium-catalyzed ketone hydrogenation. *J. Chem. Theor. Comp.* **2014**, 10, 2427-2435.
29. Rydberg, P.; Hansen, S. M.; Kongsted, J.; Norrby, P.-O.; Olsen, L.; Ryde, U. Transition-state docking of flunitrazepam and progesterone in cytochrome P450. *J. Chem. Theor. Comp.* **2008**, 4, 673-681.
30. Rydberg, P.; Olsen, L.; Norrby, P.-O.; Ryde, U. General transition-state force field for cytochrome P450 hydroxylation. *J. Chem. Theor. Comp.* **2007**, 3, 1765-1773.
31. Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Development of a Q2MM force field for the asymmetric rhodium catalyzed hydrogenation of enamides. *J. Chem. Theor. Comp.* **2008**, 4, 1313-1323.
32. Norrby, P.-O.; Liljefors, T. Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J. Comp. Chem.* **1998**, 19, 1146-1166.
33. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Downhill Simplex Method in Multidimensions*. Cambridge University Press: New York, 1992.
34. Huang, L.; Roux, B. Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data. *J. Chem. Theor. Comp.* **2013**, 9, 3543-3556.

35. Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, 131, 1138.
36. Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comp. Chem.* **1994**, 15, 162-182.
37. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.* **1996**, 17, 490-519.
38. Schrödinger Release 2020-1: Schrödinger, LLC, New York, NY, 2020.
39. Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E.;; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Kazavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pantano, S.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skynnikov, N.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, V.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 20*, University of California, San Francisco, CA, USA, 2020.
40. Hagler, A. T.; Ewig, C. S. On the use of quantum energy surfaces in the derivation of molecular force fields. *Comp. Phys. Comp.* **1994**, 84, 131-155.
41. Limé, E.; Norrby, P.-O. Improving the Q2MM method for transition state force field modeling. *J. Comp. Chem.* **2015**, 36, 244-250.
42. Madarász, A.; Berta, D.; Paton, R. S. Development of a true transition state force field from quantum mechanical calculations. *J. Chem. Theor. Comp.* **2016**, 12, 1833-1844.
43. Haines, B. E. Computational Studies on the Mechanism of HMG-CoA Reductase and the Grignard $S_{RN}1$ reaction. PhD Thesis. PhD Thesis Notre Dame 2014.
44. Haines, B. E., Steussy, C. N., Stauffacher, C. V., Wiest, O. Molecular Modeling of the Reaction Pathway and Hydride Transfer Reactions of HMG-CoA Reductase. *Biochemistry* **2012**, 51, 7983-7995.

45. Haines, B. E., Wiest, O., Stauffacher, C. V. The Increasingly Complex Mechanism of HMG CoA Reductase. *Acc. Chem. Res.* **2013**, 46, 2416-2426.
46. Tantillo, D.; Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **1998**, 2, 743-750.
47. Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Revision C.01*, Wallingford, CT, 2016.
48. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comp. Chem.* **2005**, 26, 1668-1688.