Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Structural Biology

Computer-aided drug design, quantum-mechanical methods for biological problems

Madushanka Manathunga[1], Andreas W. Götz[2] and Kenneth M. Merz Jr. [1]

Abstract

Quantum chemistry enables to study systems with chemical accuracy (<1 kcal/mol from experiment) but is restricted to a handful of atoms due to its computational expense. This has led to ongoing interest to optimize and simplify these methods while retaining accuracy. Implementing quantum mechanical (QM) methods on modern hardware such as multiple-GPUs is one example of how the field is optimizing performance. Multiscale approaches like the so-called QM/molecular mechanical method are gaining popularity in drug discovery because they focus the application of QM methods on the region of choice (*e.g.*, the binding site), while using efficient MM models to represent less relevant areas. The creation of simplified QM methods is another example, including the use of machine learning to create ultra-fast and accurate QM models. Herein, we summarize recent advancements in the development of optimized QM methods that enhance our ability to use these methods in computer aided drug discovery.

Addresses
[1] Department of Chemistry and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, United States
[2] San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, United States

Corresponding author: Merz Jr., Kenneth M. (merz@chemistry.msu.edu)
(Manathunga M.), (Götz A.W.), (Merz K.M.)

Current Opinion in Structural Biology 2022, **75:**102417

This review comes from a themed issue on **Biophysical Methods**

Edited by **David Fushman** and **Dagmar Ringe**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2022.102417

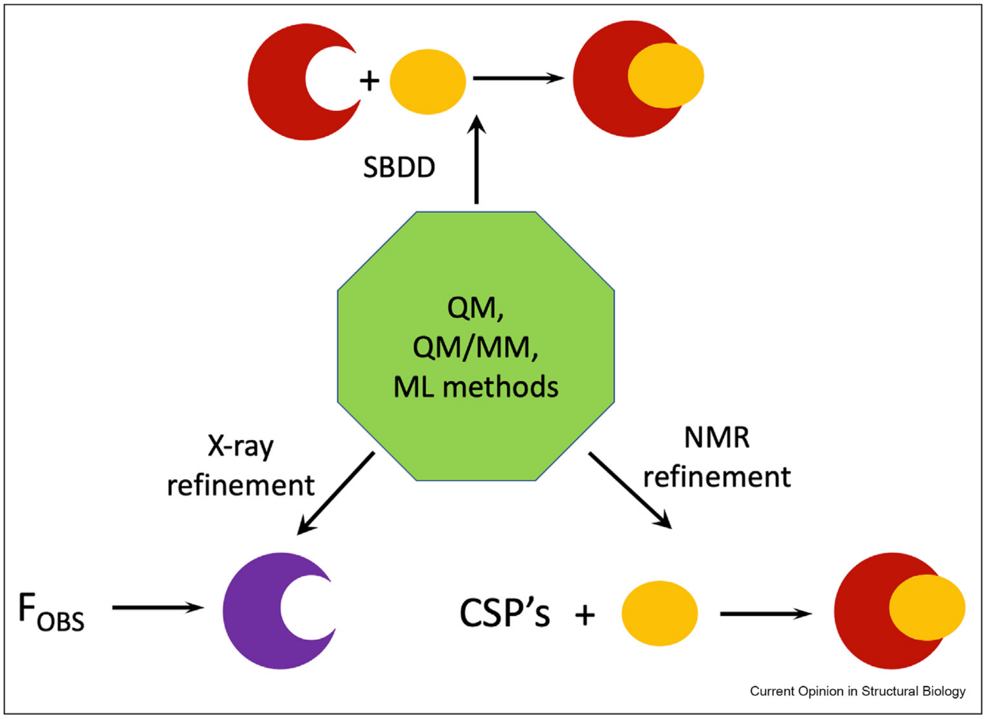
Introduction

The use of QM in computer aided drug design [1] (CADD) has been evolving over the past several decades [2,3]. The early focus in quantum mechanics (QM) was on small molecules due to the computational expense, but with a range of algorithmic and computer hardware innovations the ability to realistically use QM in CADD has emerged. In what follows we will summarize relevant advances in the past couple of years, but for those interested in digging deeper into the field they could do no worse than consulting a recent book on QM in Drug Discovery, which provides a good overview of the field with contributed chapters covering QM methods, protocols, and applications [4]. Here, we will review recent developments in QM methods, hybrid quantum mechanical/molecular mechanical (QM/MM) methods, and newly emerging quantum machine learning (QML) methods of relevance to CADD (Scheme 1).

Advances in QM methods for structure-based drug design

QM methods have had a long tradition in CADD campaigns and continue to have impact in the form of tools/workflows based on QM methods and novel computational implementations of traditional quantum chemical (QC) methods. Among the recent QC method developments, extended tight binding methods developed by Grimme et al. [5] are particularly interesting. Their most recent method, GFN2-xTB, which is based on a revised version of tight binding and on the D4 dispersion correction, has become popular due to its applicability across the periodic table in combination with good accuracy and low-computational cost. Several workflows centered around GFN2-xTB and GFNFF (a partially polarizable forcefield) have appeared recently [6–9]. Meanwhile, traditional semi-empirical methods, Hartree-Fock (HF), and density functional theory (DFT) methods with improved exchange correlation functionals continue to be useful in various branches of CADD, while approaches using highly accurate correlated wave function methods also start to appear. Useful benchmark datasets consisting of 547 protein fragment interaction energies (PLF547) and 15 active site-ligand models (PLA15) were established and used to assess the accuracy of several semi-empirical QM methods, demonstrating that PM6 and DFTB3 can achieve results nearly on par with dispersion corrected DFT when

Scheme 1



Diverse use of QM, QM/MM and ML methods in CADD applications, including structure-based drug design (SBDD), X-ray refinement based on differences between computed and observed structure factors (F_{OBS}), and NMR refinement using differences in chemical shifts of free and bound ligands (chemical shift perturbations, CSP's).

combined with noncovalent dispersion and hydrogen bonding corrections [10]. An example for state-of-the-art *ab initio* methods is the linear scaling domain-based local pair natural orbital coupled cluster method, DLPNO-CCSD(T), which was recently applied to quantify ligand—residue interactions in nicotinic acetylcholine receptor agonist binding through a local energy decomposition analysis [11]. While computationally tractable, DLPNO-CCSD(T) is still computationally demanding, and such highly accurate calculations thus remain limited to single point energy calculations of moderately sized cluster models for the foreseeable future.

In the area of conformer generation, Udvarhely et al. [12] reported the ReSCoSS workflow for generating conformers of druglike molecules in solution. In this semiautomated workflow, the initial conformer generation is performed using traditional methods, but the subsequent ranking of conformers based on the relative free energies uses geometry optimizations using the DFT-D method with the COSMO-RS solvent model. Meanwhile, Grimme's group has reported on the CREST tool [6], where the sampling in conformational space is performed using an iterative procedure and the subsequent energy calculation is performed using the GFN2-xTB method. This tool has been shown to provide reasonable conformers with a sufficient accuracy at a lower computational cost [13]. A related software tool has been reported that is aimed at generating torsional conformers of small acyclic molecules [14]. In this approach, the torsional space is explored based on chemical knowledge and a random search and geometry optimizations and Hessian calculations are carried out at a low (usually HF/3-21G) and high level of theory (M08HX/MG3S).

High-throughput docking is one of the most widely used tools in drug lead discovery, holding the potential for triaging potential drug candidates much faster and cheaper than experimentally. In this context, Cavasotto and Aucar have introduced a PM7 based scoring function [15]. Using enrichment factors computed for a set of 10 proteins representing different families and binding site characteristics, the authors reported that their function outperforms a traditional docking method. More recently, full DFT calculations were employed to successfully rank order a scaffold-diverse set of ligands to a SARS-CoV-2 main protease (Mpro) model with nearly 3000 atoms based on absolute ligand binding energies [16]. DFT calculations for such large models are significantly more expensive than semi-empirical scoring functions, however, this work demonstrates the practical possibility to employ more rigorous and large-scale QM approaches with reasonable turnaround times where required in the drug discovery pipeline.

In the area of pKa prediction, new QM based workflows have been reported by Haslak et al. and the Grimme group [7,17]. The former, which exclusively focuses on carboxylic acids, is based on a linear relationship between computed atomic charges of the anionic carboxylic fragment and experimental pKa values. In contrast, Grimme's is a widely applicable protocol based on a new cubic free energy relationship equation. In the area of C13 and proton NMR spectra prediction, MOSS-DFT for organic molecules in solution, and the MIM2-NMR workflow for large biomolecules have been reported [18,19]. Das and co-workers have reported a QM based NMR [20] and collisional cross section (CCS) workflow [21] for metabolites. A new hybrid density functional aimed at C13 chemical shift prediction has also been introduced [22]. This functional called xOPBE, was constructed by mixing HF exchange into the previously reported OPBE functional. The authors report that the predicted C13 chemical shifts using this functional are more accurate than its parent functional. Among other new tools/workflows, Grimme's group incorporated collision-induced dissociation spectra calculations into their QCEIMS software tool [8]. Another significant development from the same group is a new workflow to compute the optical rotation of organic molecules in solution [9].

Among the recent work related to dipole moment calculation, Zapata and McKemmish studied the accuracy and performance of different basis sets using different quantum chemical methods and documented a few useful recommendations [23]. The authors used a dataset consisting of over a hundred small molecules containing second and third row elements. They reported that DFT calculations with a hybrid functional such as ωB97X−V or SOGGA11-X and augmented double zeta basis sets from Jensen (aug-pc-1 or augpcseg-1) and Dunning (aug-cc-pvDZ) families can provide accurate dipole moments at an affordable computational cost. Furthermore, the authors concluded that single zeta quality and Pople style basis sets are unsuitable for dipole moment calculation.

There has been a number of recent efforts to increase the performance of traditional quantum chemistry methods. This mainly includes implementing algorithms to harness the power of new computer architectures, particularly GPUs. Among these, SCF implementation by Gordon group [24], XC quadrature scheme reported by NWChemX group [25] and work related to the QUICK code in the AMBERtools package are notable [26,27]. Finally, it is important to highlight that semi-empirical methods, particularly xTB methods, are becoming increasingly popular. Given the fast evolution of computer hardware and the pace of method development, it is likely that these methods and QML potentials will either replace classical force fields (FFs) in drug development workflows or will be used interoperably with *ab inito* QM calculation, QML potentials and FFs.

Multiscale QM/MM models

Multiscale simulation methods divide a large molecular system into different regions that are modeled at appropriate levels of theory to focus the computational effort on regions where it is required. Of particular interest are QM/MM approaches that embed a QM region into a molecular mechanics (MM) description of the environment, opening the door for simulations of entire proteins in their native environment without making gross approximations. A fairly recent editorial and collection of articles discuss developments and applications of QM/MM approaches with a focus on biomolecular systems [28] and a recent perspective highlights current developments and challenges in the field [29]. QM/MM approaches are often employed to investigate enzymatic reaction mechanisms and the interested reader is pointed to recent reviews [30,31]. In the context of CADD, an understanding of enzymatic reaction mechanisms enables the design of scaffolds based on transition state analogs for the development of enzyme inhibitors with improved affinity and specificity if enzymes are known as therapeutic targets. QM/MM methods are also successfully employed in X-ray crystal structure refinement. Recent work employed a PM6 semi-empirical Hamiltonian in combination with the AMBER force field for the protein environment to the set of structures from the Community Structure Activity Resource data set and demonstrated improved protein-ligand geometries in comparison to conventional refinement methods [32].

Due to the computational cost of *ab initio* and DFT methods, QM/MM studies of reaction mechanism are often restricted to geometry optimizations of reactant, product, and transition states or reaction paths while QM/MM molecular dynamics (MD) simulations often employ semi-empirical QM methods. However, even full scale DFT based MD simulations of entire small proteins have become feasible exploiting the latest GPU accelerators, although these are still limited to small basis sets and short time scales on the order of 10 ps [33]. A recent publication on the other hand computed the potential of mean force (PMF) along predetermined reaction paths of enzymatic reactions from biased QM/MM MD simulations with the dispersion corrected hybrid exchange-correlation functional B3LYP-D3 and a relatively large aug-cc-pVDZ basis set [34]. The development of widely available and open-source implementations such as the efficient GPU accelerated QUICK code in AmberTools [27] mentioned above hold the promise to make such investigations routine.

A challenge that any practitioner of QM/MM approaches faces is the appropriate selection of the QM region. A promising approach that aims to alleviate this burden by automating the construction of QM/MM models has recently presented [35]. In this approach the size and composition of the QM region are evaluated in an automated manner such that the hybrid model describes the atomic forces in the center of the QM region accurately. Furthermore, a good FF potential to describe the MM region must be chosen. For biomolecular simulations there are a range of well-established and highly successful pairwise additive force fields with fixed point charges that are readily available. In some cases, however, it may be important to explicitly account for electronic polarization in the MM region. Progress has been made in developing QM/MM methods and software implementations towards this end using different methods to represent polarization, for example polarizable point dipoles [36−38]. Whether these methods will become useful for CADD applications will depend on further improvements of the underlying polarizable FFs and wider availability in user-friendly software packages.

Another aspect that is of critical importance for QM/MM simulations in condensed phase under periodic boundary conditions (PBC), is the treatment of long-range electrostatic interactions between the QM and MM regions. It is common practice to truncate electrostatics beyond a cutoff distance in *ab initio* and DFT based QM/MM simulations, thus introducing uncontrolled errors into the simulation. A notable method that properly accounts for long-range electrostatics is the ambient potential composite Ewald method that performs a direct interaction of the QM charge density both with the real space and reciprocal space electrostatic potential, which was shown to remedy numerical instabilities and artifacts in DFT based QM/MM simulations [39]. An alternative approach that was recently developed projects the long-range electrostatic potential in form of augmentary charges onto MM atoms within the cutoff region and is applicable both to large finite clusters and PBC simulations [40]. With these approaches rigorous QM/MM simulations with bounded errors are now possible.

Finally, it is worth to mention recent developments in QM “bookending” methods, which are commonly known as reference potential methods which were originally pioneered 30 years ago [41,42]. These methods enable highly efficient QM/MM free energy simulations via an indirect route using free energy perturbation, thermodynamic integration, or employing Bennett’s acceptance ratio (BAR). The trick of bookending is to invoke a thermodynamic cycle in which the bulk of the free energy simulation, for instance an alchemical transformation, is performed using a series of computationally cheap reference potentials, while only the end points (bookends so to say) need to be sampled with the expensive QM/MM potential [43]. This opens the door to compute solvation free energies and ligand binding free energies with accurate QM/MM methods that otherwise would be computationally too demanding. In a similar vein, mapping to reference potentials has been developed for the targeted free energy perturbation approach (TFEP) using neural networks, resulting in an approach that requires only a small number of single point calculations with the expensive target potential, thus improving the accuracy and convergence of absolute and relative free energies and free energy surfaces [44].

Emergence of quantum machine learning models

The use of ML and NN models to develop potential energy functions of relevance to chemistry have been undergoing rapid development. Reflecting the interest in this field several reviews have appeared in 2021 for those interested in further details of the advances of the last decade [45,46]. An exciting area for CADD is to take large databases of QM derived information and then create a QML model that reproduces this data set with the expectation that the resultant model would be extensible to molecules not represented by the training data set [45]. In CADD applications these *ab* initio ML potential energy functions should be superior to MM potentials, but have a similar computational cost. However, while many of the potentials described aim at reproducing *ab initio* or DFT data for small molecules the use of these models in condensed phase simulation are just emerging. In particular, interoperability or the ability to mix and match ML force field, *ab initio*, *etc*. potentials is an area ripe for exploitation and validation. Combined together these models can be used to mitigate deficiencies in each other: for example, force fields aren’t great at conformational energies, but QM based ML models can improve this situation. Overall, the use of QML potentials in drug design has the potential to greatly improve computed binding or relative free energies using modern free energy methods [47]. Many advances along these lines will be seen in the coming years and could speed up free energy calculations while improving their accuracy.

The evolution of QML potentials is readily described by a series of generations [48]. In the first generation rudimentary potentials that can be used to model simple systems were described [49]. To overcome this deficiency and to generate models that can be used in 3-dimensional space second generation potentials were developed. Examples of these so-called second generation models like ANI [50−55] or AP-net [56] among many others [45] are based off of high-dimensional neural network potentials (HDNNPs) described by Behler and Parrinello [57], which while effective to study large systems lack long-range electrostatics and dispersion interactions thereby reducing their effectiveness to model intermolecular interactions in particular in inhomogeneous systems [48,56,58]. A major drawback of many of these models is the lack of atomic charges and the use of fixed charges in the third generation addressed these issues [59–61]. Finally, fourth generation (4G) models have begun to appear combining many of the earlier generation models with atomic charges determined on the fly using, for example, charge equilibration (Ceq) [62] models [48,60,63]. In these 4G models the intramolecular degrees of freedom are modeled using accurate 2G-HDNNP coupled with a charge equilibration model that updates the charges on the fly allowing the model to respond to the local environment and effectively deal with multiple charge states and long-range charge transfer. Ko et al. [48] reported an initial attempt at a 4G model and showed that it had the proper behavior on a series of test systems. These 4G models have the potential to have an impact on CADD because they have all the required features: an accurate intermolecular potential combined with an accounting for polarization and charge transfer. Combining 4G models with classical force fields incorporating long-range electrostatic interactions could usher in a new generation of accurate relative free energy calculations. While the field is not quite there yet it is a very exciting avenue to pursue that will challenge standard paradigms where many potentials can be layered on top of one another to create hybrid models to solve problems in CADD.

Applications of QML potentials, in the context of chemical reactivity, have been reported in the past year and we highlight just a few here. Böselt et al. [64], Zeng et al. [65], Gastegger et al. [66] and Pan and coworkers [67], building on earlier work of the Yang lab [68–70], have reported the use of QML potentials in combination with QM/MM methods. The studies of Böselt et al. and Pan and co-workers use Δ-machine learning (Δ-ML) approaches [71] where the differences between a lower level of QM theory *versus* that of a higher level of theory is learned for a specific reaction and then applied to correct the cheaper QM models to achieve high quality results. The work of Zeng et al. [65] introduces range-corrections to the ML potentials to also improve short-range QM/MM interactions affecting the MM atoms within a relevant cutoff of the QM region. These application studies are interesting and point to the need for improvements in the treatment of long-range interactions and in building standalone 4G potentials that can deliver high-quality results without resorting to Δ-ML approaches. That said these latter models may ultimately prove to be the best way forward. While these studies are of interest in studying chemical reactivity how these methods can be applied to drug discovery applications will need to be more fully fleshed out in the coming years.

Conclusions and outlook

QM has been a key field in theoretical and computational chemistry for decades and will continue to be of importance going forward since it is the only method available in the computational armamentarium that guarantees to asymptotically reach the correct answer via the variational principle. That said, improvement in the performance of traditional QM methods by exploiting modern hardware and novel algorithms continues to be an area worth pursuing. The development of new QM/MM methods that take advantage of the advancements has already begun to appear in CADD applications. The ML/AI field is advancing rapidly to provide the CADD field with fast and accurate potentials that directly mimic QM models that can be applied to a myriad of problems in CADD. The most exciting aspect of the use of QM in CADD has been the emergence of many new directions that all promise to deliver chemical accuracy to several aspects of CADD in the coming years.

CRediT author statement

Madushanka Manathunga: Writing - Original Draft, Writing - Review & Editing, **Andreas W. Götz**: Writing - Original Draft, Writing - Review & Editing, Funding acquisition, **Kenneth M. Merz, Jr.**: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

Conflict of interest statement

Nothing to declare

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation (NSF) through grant OAC-1835144. KMM gratefully thanks the National Institutes of Health for financial support (Grant Numbers GM130641 and U2CES030167).

References

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

* of special interest
* * of outstanding interest

1. Merz KM, Ringe D, Reynolds CH: *Drug design: structure- and ligand-based approaches*. New York: Cambridge University Press: Cambridge [U.K.]; 2010.
2. Merz Jr KM: **Using quantum mechanical approaches to study biological systems**. *Acc Chem Res* 2014, **47**:2804–2811.
3. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz Jr KM: **The role of quantum mechanics in structure-based drug design**. *Drug Discov Today* 2007, **12**:725–731.
4. Heifetz A: *Quantum mechanics in drug discovery*. New York, NY: Humana; 2020.
5. Bannwarth C, Caldeweyher E, Ehlert S, Hansen A, Pracht P, Seibert J, Spicher S, Grimme S: **Extended tight-binding quantum chemistry methods**. *WIREs Comput Mol Sci* 2021, **11**: e1493.
6. Pracht P, Bohle F, Grimme S: **Automated exploration of the low-energy chemical space with fast quantum chemical methods**. *Phys Chem Chem Phys* 2020, **22**:7169–7192.

7. Pracht P, Grimme S: **Efficient quantum-chemical calculations of acid dissociation constants from free-energy relationships**. *J Phys Chem* 2021, **125**:5681–5692.

8. Koopman J, Grimme S: **From QCEIMS to QCxMS: a tool to routinely calculate CID mass spectra using molecular dynamics**. *J Am Soc Mass Spectrom* 2021, **32**:1735–1751.

9. Bohle F, Seibert J, Grimme S: **Automated quantum chemistry-based calculation of optical rotation for large flexible molecules**. *J Org Chem* 2021, **86**:15522–15531.

10. Kriz K, Rezac J: **Benchmarking of semiempirical quantum-mechanical methods on systems relevant to computer-aided drug design**. *J Chem Inf Model* 2020, **60**:1453–1460.

11. Beck ME, Riplinger C, Neese F, Bistoni G: **Unraveling individual host-guest interactions in molecular recognition from first principles quantum mechanics: insights into the nature of nicotinic acetylcholine receptor agonist binding**. *J Comput Chem* 2021, **42**:293–302.

12. Udvarhelyi A, Rodde S, Wilcken R: **ReSCoSS: a flexible quantum chemistry workflow identifying relevant solution conformers of drug-like molecules**. *J Comput Aided Mol Des* 2021, **35**:399–415.

13. Boz E, Stein M: **Accurate receptor-ligand binding free energies from fast QM conformational chemical space sampling**. *Int J Mol Sci* 2021, **22**:3078.

14. Ferro-Costas D, Mosquera-Lois I, Fernández-Ramos A: **TorsiFlex: an automatic generator of torsional conformers. Application to the twenty proteinogenic amino acids**. *J Cheminf* 2021, **13**:100.

15. Cavasotto CN, Aucar MG: **High-throughput docking using quantum mechanical scoring**. *Front Chem* 2020, **8**.

16. Wang Y, Murlidaran S, Pearlman DA: **Quantum simulations of SARS-CoV-2 main protease M(pro) enable high-quality scoring of diverse ligands**. *J Comput Aided Mol Des* 2021, **35**: 963–971.

17. * Haslak ZP, Zareb S, Dogan I, Aviyente V, Monard G: **Using atomic charges to describe the pKa of carboxylic acids**. *J Chem Inf Model* 2021, **61**:2733–2743.
This article presents a new broadly applicable protocol, based on DFT and GFN2-XTB methods, to compute pKa of drug-like molecules in water

18. * Hoffmann F, Li D-W, Sebastiani D, Brüschweiler R: **Improved quantum chemical NMR chemical shift prediction of metabolites in aqueous solution toward the validation of unknowns**. *J Phys Chem* 2017, **121**:3071–3078.
A paper describing a DFT and ML based efficient novel protocol for computing NMR chemical shifts of metabolites.

19. Chandy SK, Thapa B, Raghavachari K: **Accurate and cost-effective NMR chemical shift predictions for proteins using a molecules-in-molecules fragmentation-based method**. *Phys Chem Chem Phys* 2020, **22**:27781–27799.

20. Das S, Edison AS, Merz Jr KM: **Metabolite structure assignment using in silico NMR techniques**. *Anal Chem* 2020, **92**: 10412–10419.

21. Das S, Tanemura KA, Dinpazhoh L, Keng M, Schumm C, Leahy L, Asef CK, Rainey M, Edison AS, Fernandez FM, Merz Jr KM: **In Silico collision cross section calculations to aid metabolite annotation**. *J Am Soc Mass Spectrom* 2022.

22. Zhang J, Ye Q, Yin C, Wu A, Xu X: **xOPBE: a specialized functional for accurate prediction of 13C chemical shifts**. *J Phys Chem* 2020, **124**:5824–5831.

23. ** Zapata JC, McKemmish LK: **Computation of dipole moments: a recommendation on the choice of the basis set and the level of theory**. *J Phys Chem* 2020, **124**:7538–7548.
This paper reports an extensive study and provides useful recommendations on selecting basis sets for dipole moment prediction using QM methods.

24. Barca GMJ, Alkan M, Galvez-Vallejo JL, Poole DL, Rendell AP, Gordon MS: **Faster self-consistent field (SCF) calculations on GPU clusters**. *J Chem Theor Comput* 2021, **17**:7486–7503.

25. Williams-Young DB, Bagusetty A, de Jong WA, Doerfler D, van Dam HJJ: **Vázquez-Mayagoitia, Á.; Windus, T. L.; Yang, C., Achieving performance portability in Gaussian basis set density functional theory on accelerator based architectures in NWChemEx**. *Parallel Comput* 2021, **108**:102829.

26. Manathunga M, Jin C, Cruzeiro VWD, Miao Y, Mu D, Arumugam K, Keipert K, Aktulga HM, Merz KM, Götz AW: **Harnessing the power of multi-GPU acceleration into the quantum interaction computational kernel program**. *J Chem Theor Comput* 2021, **17**:3955–3966.

27. Cruzeiro VWD, Manathunga M, Merz KM, Götz AW: **Open-source multi-GPU-accelerated QM/MM simulations with AMBER and QUICK**. *J Chem Inf Model* 2021, **61**:2109–2115.

28. Hofer TS, de Visser SP: **Quantum mechanical/molecular mechanical approaches for the investigation of chemical systems - recent developments and advanced applications**. *Front Chem* 2018, **6**:357.

29. Cui Q, Pal T, Xie L: **Biomolecular QM/MM simulations: what are some of the "burning issues**. *J Phys Chem B* 2021, **125**:689–702.

30. Quesne MG, Borowski T, de Visser SP: **Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs**. *Chemistry* 2016, **22**:2562–2581.

31. * Magalhães RP, Fernandes HS, Sousa SF: **Modelling enzymatic mechanisms with QM/MM approaches: current status and future challenges**. *Isr J Chem* 2020, **60**:655–666.
Excellent concise contemporary review of the current status of QM/MM approaches for biomolecular systems including an overview of different application fields and discussion of future challenges.

32. Borbulevych OY, Martin RI, Westerhoff LM: **The critical role of QM/MM X-ray refinement and accurate tautomer/protomer determination in structure-based drug design**. *J Comput Aided Mol Des* 2021, **35**:433–451.

33. Vennelakanti V, Nazemi A, Mehmood R, Steeves AH, Kulik HJ: **Harder, better, faster, stronger: large-scale QM and QM/MM for predictive modeling in enzymes and proteins**. *Curr Opin Struct Biol* 2022, **72**:9–17.

34. Yagi K, Ito S, Sugita Y: **Exploring the minimum-energy pathways and free-energy profiles of enzymatic reactions with QM/MM calculations**. *J Phys Chem B* 2021, **125**:4701–4713.

35. Brunken C, Reiher M: **Automated construction of quantum-classical hybrid models**. *J Chem Theor Comput* 2021, **17**: 3797–3813.

36. Bondanza M, Nottoli M, Cupellini L, Lipparini F, Mennucci B: **Polarizable embedding QM/MM: the future gold standard for complex (bio)systems?** *Phys Chem Chem Phys* 2020, **22**: 14433–14448.

37. Loco D, Lagardere L, Cisneros GA, Scalmani G, Frisch M, Lipparini F, Mennucci B, Piquemal JP: **Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings**. *Chem Sci* 2019, **10**: 7200–7211.

38. Nochebuena J, Naseem-Khan S, Cisneros GA: **Development and application of quantum mechanics/molecular mechanics methods with advanced polarizable potentials**. *Wiley Interdiscip Rev Comput Mol Sci* 2021, **11**.

39. Giese TJ, York DM: **Ambient-potential composite Ewald method for ab initio quantum mechanical/molecular mechanical molecular dynamics simulation**. *J Chem Theor Comput* 2016, **12**:2611–2632.

40. * Pan X, Nam K, Epifanovsky E, Simmonett AC, Rosta E, Shao Y: **A simplified charge projection scheme for long-range electrostatics in ab initio QM/MM calculations**. *J Chem Phys* 2021, **154**. 024115.
This paper introduces a new approach to treat long-range electrostatics in QM/MM simulations and contains a nice write-up and comparison of alternative methods. Table 1 lists the most relevant methods including references to the original papers.

41. Gao JL: **Absolute free-energy of solvation from monte-carlo simulations using combined quantum and molecular mechanical potentials**. *J Phys Chem Us* 1992, **96**:537–540.

42. Luzhkov V, Warshel A: **Microscopic models for quantum-mechanical calculations of chemical processes in solutions - Ld/ampac and scaas/ampac calculations of solvation energies**. *J Comput Chem* 1992, **13**:199–213.

43. Giese TJ, York DM: **Development of a robust indirect approach for MM–> QM free energy calculations that combines force-matched reference potential and Bennett's acceptance ratio methods**. *J Chem Theor Comput* 2019, **15**: 5543–5562.
* *

This paper introduces a robust approach to obtain DFT-based QM/MM free energies at minimal computational cost via reference potentials through careful analysis of solvation free energies and ligand binding relative free energies with different approaches.

44. Rizzi A, Carloni P, Parrinello M: **Targeted free energy perturbation revisited: accurate free energies from mapped reference potentials**. *J Phys Chem Lett* 2021, **12**:9449–9454.

45. Huang B, von Lilienfeld OA: **Ab initio machine learning in chemical compound space**. *Chem Rev* 2021, **121**: 10001–10036.
*

Excellent contemporary review describing the use of QML models in chemistry. Table 1 in this article gives a list of the various synthetic QM datasets available for organic, transition metal and materials systems

46. Manzhos S, Carrington Jr T: **Neural network potential energy surfaces for small molecules and reactions**. *Chem Rev* 2021, **121**:10187–10217.
*

Excellent contemporary review describing the use of neural networks (NN) to represent reactive potential surfaces of molecules

47. Song LF, Merz Jr KM: **Evolution of alchemical free energy methods in drug discovery**. *J Chem Inf Model* 2020, **60**: 5308–5318.

48. Ko TW, Finkler JA, Goedecker S, Behler J: **A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer**. *Nat Commun* 2021, **12**:398.
*

A lovely paper giving a nice introduction to the field of QML models as we all a nice demonstration of the use of a fourth generation HDNNPs QML model using a charge equilibration scheme. A series of test calculations are reported showing the overall performance of the model.

49. Blank TB, Brown SD, Calhoun AW, Doren DJ: **Neural network models of potential energy surfaces**. *J Chem Phys* 1995, **103**: 4129–4137.

50. Devereux C, Smith JS, Davis KK, Barros K, Zubatyuk R, Isayev O, Roitberg AE: **Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens**. *J Chem Theor Comput* 2020, **16**:4192–4202.

51. Gao X, Ramezanghorbani F, Isayev O, Smith JS, Roitberg AE: **TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials**. *J Chem Inf Model* 2020, **60**:3408–3415.

52. Smith JS, Isayev O, Roitberg AE: **ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost**. *Chem Sci* 2017, **8**:3192–3203.

53. Smith JS, Isayev O, Roitberg AE: **ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules**. *Sci Data* 2017, **4**:170193.

54. Smith JS, Nebgen BT, Zubatyuk R, Lubbers N, Devereux C, Barros K, Tretiak S, Isayev O, Roitberg AE: **Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning**. *Nat Commun* 2019, **10**:2903.

55. Smith JS, Zubatyuk R, Nebgen B, Lubbers N, Barros K, Roitberg AE, Isayev O, Tretiak S: **The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules**. *Sci Data* 2020, **7**:134.

56. Glick ZL, Metcalf DP, Koutsoukas A, Spronk SA, Cheney DL, Sherrill CD: **AP-Net: an atomic-pairwise neural network for smooth and transferable interaction potentials**. *J Chem Phys* 2020, **153**. 044112.

57. Behler J, Parrinello M: **Generalized neural-network representation of high-dimensional potential-energy surfaces**. *Phys Rev Lett* 2007, **98**:146401.

58. Yue S, Muniz MC, Andrade Calegari, F M, Zhang L, Car R, Panagiotopoulos AZ: **When do short-range atomistic machine-learning models fall short?** *J Chem Phys* 2021, **154**. 034111.

59. Yao K, Herr JE, Toth DW, McKintyre R, Parkhill J: **The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics**. *Chem Sci* 2018, **9**:2261–2269.

60. Ko TW, Finkler JA, Goedecker S, Behler J: **General-Purpose machine learning potentials capturing nonlocal charge transfer**. *Acc Chem Res* 2021, **54**:808–817.

61. Bereau T, Andrienko D, von Lilienfeld OA: **Transferable atomic multipole machine learning models for small organic molecules**. *J Chem Theor Comput* 2015, **11**:3225–3233.

62. Leven I, Hao H, Tan S, Guan X, Penrod KA, Akbarian D, Evangelisti B, Hossain MJ, Islam MM, Koski JP, Moore S, Aktulga HM, van Duin ACT, Head-Gordon T: **Recent advances for improving the accuracy, transferability, and efficiency of reactive force fields**. *J Chem Theor Comput* 2021, **17**: 3237–3251.

63. Xie X, Persson KA, Small DW: **Incorporating electronic information into machine learning potential energy surfaces via approaching the ground-state electronic energy as a function of atom-based electronic populations**. *J Chem Theor Comput* 2020, **16**:4256–4270.

64. Böselt L, Thurlemann M, Riniker S: **Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems**. *J Chem Theor Comput* 2021, **17**:2641–2658.

65. Zeng J, Giese TJ, Ekesan S, York DM: **Development of range-corrected deep learning potentials for fast, accurate quantum mechanical/molecular mechanical simulations of chemical reactions in solution**. *J Chem Theor Comput* 2021, **17**: 6993–7009.

66. Gastegger M, Schutt KT, Muller KR: **Machine learning of solvent effects on molecular spectra and reactions**. *Chem Sci* 2021, **12**:11473–11483.

67. Pan X, Yang J, Van R, Epifanovsky E, Ho J, Huang J, Pu J, Mei Y, Nam K, Shao Y: **Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions**. *J Chem Theor Comput* 2021, **17**:5745–5758.

68. Shen L, Wu J, Yang W: **Multiscale quantum mechanics/molecular mechanics simulations with neural networks**. *J Chem Theor Comput* 2016, **12**:4934–4946.

69. Wu J, Shen L, Yang W: **Internal force corrections with machine learning for quantum mechanics/molecular mechanics simulations**. *J Chem Phys* 2017, **147**:161732.

70. Shen L, Yang W: **Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks**. *J Chem Theor Comput* 2018, **14**:1442–1455.

71. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA: **Big data meets quantum chemistry approximations: the delta-machine learning approach**. *J Chem Theor Comput* 2015, **11**: 2087–2096.