ELSEVIER



# The role of quantum mechanics in structure-based drug design

**Kaushik Raha[1], Martin B. Peters, Bing Wang, Ning Yu[2], Andrew M. Wollacott[3], Lance M. Westerhoff[4] and Kenneth M. Merz Jr**

Department of Chemistry, Quantum Theory Project, University of Florida, 2328 New Physics Building, P.O. Box 118435, Gainesville, FL 32611-8435, United States

**Herein we will focus on the use of quantum mechanics (QM) in drug design (DD) to solve disparate problems from scoring protein–ligand poses to building QM QSAR models. Through the variational principle of QM we know that we can obtain a more accurate representation of molecular systems than classical models, and while this is not a matter of debate, it still has not been shown that the expense of QM approaches is offset by improved accuracy in DD applications. Objectively validating the improved applicability and performance of QM over classical-based models in DD will be the focus of research in the coming years along with research on the conformational sampling problem as it relates to protein–ligand complexes.**

## Introduction

The routine use of quantum mechanics (QM) in all phases of *in silico* DD is the logical next step in the evolution of this field. The first principles nature of QM allows it to improve systematically the accuracy of the description of the nature of the interactions between molecules. Moreover, the systematic way in which one can approach the use of QM methods to solve chemical and biological problems is quite appealing, but the practical use of many of the attractive features of QM in *in silico* DD applications is still to be realized due, in large part, to computational limitations. In recent years it has become clear that classical potential functions are being pushed to their limits and as many pitfalls of using them are coming to light, one is tempted to explore the use of QM procedures. This is a somewhat naïve view, however, since one of the main observations of a large body of computational work has shown that sampling of relevant conformational states can be as important as providing an accurate representation of an inter- or intramolecular interaction. Hence, even as QM becomes a routine tool used to calculate the energy of individual states of a biological system, one still faces the daunting task of sampling relevant conformational space, which, in our view, will for the near term be largely confined to classical models.

The last couple of years have seen significant advances with respect to the use of QM in all aspects of DD. This has, in part, been fueled by the extraordinary increase in computational power and the plummeting cost of CPU time and storage space, which has, in turn, accelerated the development and validation of more sophisticated algorithms for calculating wave functions of macromolecular systems. Hand-in-hand with CPU performance increases has been the equally impressive improvement in algorithms and software that allows researchers to address large-scale biological questions using QM models. In the following sections, we will highlight the evolving role played by QM in all aspects of *in silico* DD and we describe what, in our view, are significant recent advances. The focus of this review is on the use of QM in DD, but QM has found broad application, for example, in the study of enzyme catalysis [1–5]. The latter will not be discussed here, but the interested reader is directed to many of the recent reviews on QM studies of enzyme catalysis.

*Corresponding author:* Merz, K.M. Jr (merz@qtp.ufl.edu)

[1] Current address: Computational and Structural Chemistry, GlaxoSmithKline, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989, United States.

[2] Current address: Simulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93534, United States.

[3] Current address: Health Sciences Building Box #357350, University of Washington, Seattle, WA 98195, United States.

[4] Current address: QuantumBio Inc., 200 Innovation Boulevard Suite 261, State College, PA 16803, United States.

1359-6446/06/$ - see front matter  doi:10.1016/j.drudis.2007.07.006

The use of QM in *in silico* DD can be divided into two broad categories, structure- and ligand-based methods (see Figure 1). Structure-based drug design (SBDD) methods involve the explicit treatment of the receptor, as well as its associated ligands, and includes scoring protein–ligand poses using QM or QM/MM methods, homology modeling of the receptor (before docking studies, e.g.), and energy decomposition methods like COMBINE which is based on a quantitative structure–activity relationship (QSAR) of pairwise interaction energies between a receptor and a series of ligands. SBDD requires either an X-ray or NMR structure of the ligand in complex with the receptor and this information is shown as inputs in Figure 1. An important aspect of the structure determination process is the refinement process, which can be impacted by QM-based methods as well, while ligand-based drug design (LBDD) methods include various QSAR methods, which rely on the knowledge of the ligand structure. QSAR can be carried out using 2D, 2.5D (structures generated from 2D), or 3D structures (from NMR or X-ray studies), but they are generally obtained from purely computational means. However, one has to utilize 3D structures when using QM because of the need to have an all-atom description of the nuclei and associated electrons.

## Structure-based drug design

### *Qualitative insights into protein and protein–ligand structure*

The ability to characterize a macromolecule, such as a protein, using QM opens up a whole new range of descriptors or molecular representations that can aid drug discovery. Many of these descriptors are beyond the reach of classical potentials and by their very nature can be used to gain a qualitative understanding of protein–ligand interactions and then be used in the rational design of drug molecules. Linear-scaling QM methods have made therapeutically important protein targets routinely accessible to qualitative analysis. For example, new QM-derived descriptor classes, such as molecular electrostatic potential (ESP) maps, local hardness and softness, Fukui indices, frontier orbital analysis, density of states, *etc.* can be used to probe protein–ligand complexes.

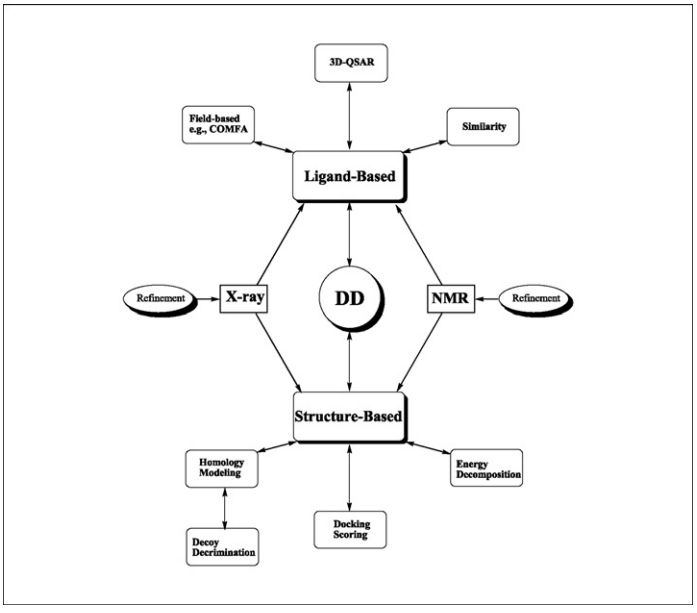


**FIGURE 1**

Hierarchy of QM methods used in *in silico* drug design.

ESP maps have been widely used as a tool for characterizing protein or DNA binding sites. However, these maps have traditionally been derived from classical point charge models that were used to compute the electrostatic potential on the surface of proteins by solving the linear or nonlinear Poisson–Boltzmann equation [6]. With the advent of linear-scaling QM algorithms, combined with selfconsistent reaction field methods to model solvation, ESP maps can now be computed quantum mechanically [7]. Khandogin and York, using linear-scaling QM technology to generate ESP maps, have probed properties of therapeutically important protein targets, such as HIV-1 nucleocapsid (NC) protein [8]. These authors have clearly demonstrated the advantage of using the PM3/COSMO computed MEP map over a classical point charge (PARSE/PB) map, in discerning between the electronegativity of the C-terminal and N-terminal zinc finger region of NC. These results agree with earlier experimental work that arrived at a similar conclusion [9]. Furthermore, they used the relative proton potential as a descriptor to predict proton affinity of titratable sites of the ovomucoid third domain (OMTYK3). The agreement between experimental p$K_a$ and relative proton potentials of these residues is very encouraging, with a linear correlation coefficient ($R$) of −0.996. There is a wealth of experimental p$K_a$ data and high-resolution X-ray crystallographic data available for other therapeutically important protein targets. A systematic study of all these targets to confirm the predictive ability of relative proton potential is in order.

In related studies, Rajamani and Reynolds have also used linear-scaling QM [10], implemented in the computer program DivCon, to model protonation states of catalytic aspartates in β-secretase [11]. These studies suggest that the aspartates prefer the mono-protonated state in the presence of the inhibitor, whereas in its absence they favor the di-deprotonated state. A more recent study came to a different conclusion about the protonation pattern, through the use of QM-based X-ray refinement followed by relative energy ordering using semiempirical QM methods [12]. Raha and Merz, again using DivCon, have also formulated a scheme to calculate the proton affinity of the catalytic aspartates of HIV-1 protease in the presence and absence of inhibitors bound to the proteases and discussed the results in light of their binding affinity calculations [13].

Polarization and charge transfer (CT) have been documented to be important at some level in macromolecular interactions [14–17]; but it has only recently been quantified in ways relevant to SBDD. Hensen and coworkers, using QM/MM methods, have studied the interaction of HIV-1 protease with three high affinity inhibitors, nelfinavir, mozenavir, and tripnavir [18]. They find that polarization of the ligand by the enzyme environment contributes to up to 39% of the total electrostatic interaction energy. Based on their analysis, they propose modifications to one of the inhibitors that can possibly lead to increase in binding affinity. In a similar study, Garcia-Viloca *et al.* have investigated the role of polarization of the substrate tetrahydrofolate, and the cofactor NADPH, at various stages of the dihydrofolate reductase-catalyzed hydride transfer reaction [19]. The authors find that polarization contributes 4% of the total electrostatic interaction and stabilizes the transition state by 9 kcal/mol over the reactants.

Charge transfer in receptor–ligand interaction in the context of SBDD has been discussed by Raha and Merz [13]. In a study of 165

noncovalent protein–ligand complexes, they find that in 11% of the complexes more than 0.1 electron units of charge is transferred from the protein to ligand. In the 49 metalloenzyme complexes, there is, on average, up to 0.6 electron units of charge transferred between the protein and the ligand. The direction of CT depends on the protein–ligand complex. For example, in matrix metalloproteases (MMP), charge is transferred from the protein to the ligand, whereas in human carbonic anhydrase (HCA) and carboxypeptidases (CPA) charge is transferred from the ligand to protein. The role of CT in biological systems has been questioned, in particular the magnitude of its contribution to the overall interaction energy [17]. However, recent work, using Car-Parrinello (CP) techniques [20] and Fragment Molecular Orbital (FMO) methods [21], shows that the distribution of charge and electrons in a system is strongly affected by CT and polarization effects. All these studies highlight the fact that QM effects are important in biological systems and, in particular, protein–ligand systems and cannot be ignored in accurate *in silico* drug design efforts.

QM-derived atomic point charges have recently been shown to be important for the study of protein–ligand complexes. For example, a database (ZINC) of commercially available drug-like molecules prepared with QM charges and desolvation penalties has been developed [22]. Irwin *et al.*, using ZINC, have successfully enriched known ligands that bind to metalloenzymes, over non-binders in retrospective docking screens [23]. QM-derived charges and the resulting desolvation penalties clearly contributed to the success of this approach.

Further evidence of the importance of QM-derived charges comes from another study by Cho *et al.*, in which ligand charges only were calculated using QM/MM methods. The resulting QM-derived ligand charges led to significant improvement in the ability of docking studies to obtain the correct binding mode of the inhibitor [24]. The docking method that employed QM charges performed decisively better than force field-based charges in ranking native binding modes as the best pose. The difference was more pronounced for poses that were predicted within 0.5–1.0 Å RMSD of the native pose. Raha and Merz have also designed a classical scoring function – the molecular recognition model – that used CM2 charges calculated using semiempirical QM for modeling electrostatic and solvation effects during binding [13]. It is noteworthy that charges, in this case, were computed for the entire protein–ligand complex using linear-scaling methods, thus accounting for polarization and charge transfer. The molecular recognition model was able to calculate p$K_i$s that agreed with experimental p$K_i$ (correlation coefficient $R^2$ of 0.78) for 33 inhibitors modeled in the active site of HIV-1 protease.

In a recent report [25], we demonstrated the use of semiempirical quantum mechanics (QM) and molecular dynamics simulations (MD) in conjunction with the Frohlich–Kirkwood theory to calculate the dielectric permittivity of proteins. The proteins Staphylococcus nuclease and T4 lysozyme were examined in order to investigate the structural basis of the macroscopic dielectric permittivity from microscopic simulations. The use of QM allowed a realistic representation of electronic polarizability of the proteins, which is otherwise inaccessible because of the use of fixed-point charge models in the classical force fields which are typically used to study proteins. The results from the MD simulation of T4 lysozyme followed by single point QM calculations are shown in Figure 2. The key findings of this study included the confirmation of earlier reports that the dielectric permittivity is not a constant, but varies with region of the protein, and its structural and electronic features; the dielectric permittivity is highest on the surface and boundary regions and drops off sharply towards the interior of the protein and the main new observation that electronic polarization or charge transfer, whether due to solvent, or to the protein environment, significantly influences permittivity.

### *Quantitative characterization of protein and protein–ligand structure*

While QM can provide valuable insights and a different perspective regarding the interaction between receptor and ligand in structure-based drug design, the holy grail of computational drug discovery still remains the ability to calculate accurately the free energy of binding to allow the routine discovery of new inhibitors using *in silico* techniques. Part of this problem involves the prediction of the correct binding mode or ''pose'' of the inhibitor when bound to a protein target. Several docking programs have been reasonably successful in obtaining the correct binding mode [26]. However, calculating the binding free energy or the correct score has proven to be challenging [27–29]. This is not surprising, considering that the free energy of binding between two molecular systems depends on a complex interplay of interactions between them. Computational methods that strive to calculate the free energy of binding, usually use an energy function also known as a ''scoring function'' that computes a score directly or indirectly, related to the binding free energy. Scoring functions have traditionally been either simplistic empirical or statistical potentials that relate observables to the free energy of binding by using statistical methods, or they are extremely detailed in nature and use physics-based descriptions of the molecular energetics and extensive sampling of receptor–ligand conformations via molecular simulation. We have reviewed all categories of scoring functions and discussed their pros and cons with respect to SBDD in an earlier review [30].

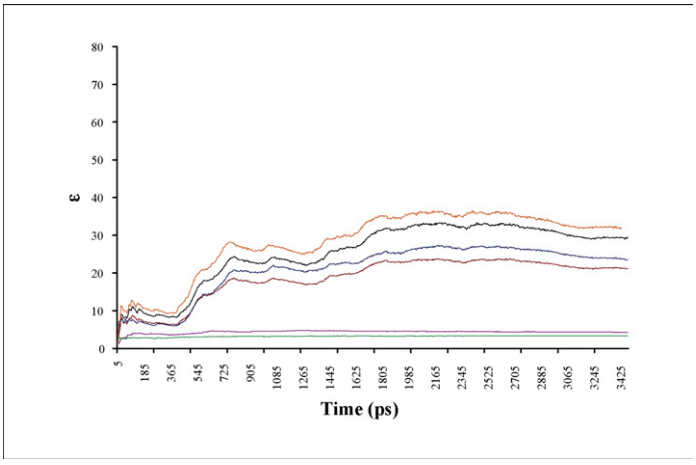


**FIGURE 2**

Dielectric permittivity of T4 lysozyme calculated from MD simulation and QM calculations. Color code: classical AMBER charges (black); QM-derived gas-phase charge model 2 (CM2) charges (red); sidechain residues (blue); uncharged residues (purple); backbone residues (green); solvated QM-derived CM2 charges (orange).

QM/MM methods are generally used to study mechanistic aspects of enzyme catalysis [1–3], but they are beginning to be employed to compute protein–ligand binding affinities. Khandelwal *et al.* used a four-tier approach that involves docking, QM/MM optimization, MD simulation, and QM/MM interaction energy calculation to predict binding affinity [31]. The authors used a modified version of extended linear response theory (ELR), where the van der Waals and electrostatic terms are replaced by the QM/MM interaction energy. The authors calculated the binding affinity of 28 hydroxamate-based inhibitors of matrix metalloprotease (MMP-9) using this approach, with impressive accuracy. The agreement between the calculated and experimental p$K_i$ is excellent ($R^2 = 0.9$ and crossvalidated $R^2$ ranging from 0.77 to 0.88). What is also noteworthy is that the authors clearly demonstrated an improvement in predictive accuracy with every step of their four-tier approach. This study demonstrates the importance of a quantum mechanical treatment and the sampling of active conformations in accurate binding affinity prediction. Specifically, the QM/MM treatment of the active site is very important (step 4) because it was shown that a proton is transferred from the hydroxamate hydroxyl to the active site glutamate. This observation could have been missed using an approach based on a classical force field.

Grater *et al.* used QM/MM with Poisson–Boltzmann/Surface Area approach to calculate binding free energy of trypsin and FKBP inhibitors [32]. The unbound ligand free energy, the unbound protein free energy, and the complex free energy were calculated using the QM/MM-PB/SA formalism. The ligand free energy and polarization was accounted for using QM/MM at the AM1 level of theory. Experimentally determined structures were available for the FKBP inhibitors, whereas the binding modes of the trypsin inhibitors were obtained by docking. The accuracy of prediction was higher with FKBP inhibitors in the set (correlation coefficient = 0.56) as opposed to only trypsin inhibitors (correlation coefficient = 0.20). The authors suggested that this was because of the certainty in the binding mode of the FKBP inhibitors, which had been determined experimentally.

The QM/MM approach clearly shows promise for calculating the binding affinity of protein–ligand interaction. However, it is obvious from the above discussion that firstly these approaches still require extensive sampling of ligand–receptor conformations through molecular simulation and are very time consuming and secondly in many of the QM/MM studies reported to date, only the ligand is treated quantum mechanically (excepting cases where a metal ion and its ligand environment is necessary in the modeling [31,33]), because including even small portions of the protein is computationally too expensive. Thirdly, if the protein-ligand complex is to be divided into QM and MM regions across covalent bonds, then there are well-documented computational difficulties associated with the treatment of this so-called boundary region between the QM and MM atoms, which could affect the reliability of a binding affinity calculation using QM/MM methods [34].

These problems have, to some extent, been surmounted by the development of linear-scaling QM technology in the past decade. Semiempirical Hamiltonians such as AM1 and PM3 can now be employed to calculate the molecular wavefunction for proteins with thousands of atoms [10,35,36]. The first application of linear-scaling methods to the computation of protein–ligand binding free energies was reported by Raha and Merz, where they calculated the binding affinity of ligands bound to the metalloenzyme HCA with reasonable accuracy [37] (see Figure 3). The free energy of binding in solution was calculated using the following set of equations:

$$\Delta G^{\mathrm{sol}}_{\mathrm{bind}} = \Delta G^{\mathrm{g}}_{\mathrm{b}} + \Delta G^{\mathrm{PL}}_{\mathrm{solv}} - \Delta G^{\mathrm{P}}_{\mathrm{solv}} - \Delta G^{\mathrm{L}}_{\mathrm{solv}},$$
$$\Delta G^{\mathrm{g}}_{\mathrm{b}} = \Delta H^{\mathrm{g}}_{\mathrm{b}} - T\Delta S^{\mathrm{g}}_{\mathrm{b}} \quad (1)$$

Here the free energy of binding in solution was calculated as the sum of the gas-phase interaction energy and a solvation correction. The gas-phase interaction energy consisted of enthalpic and entropic components. The electrostatic part of the enthalpic component was calculated with the program DivCon, using semiempirical Hamiltonians. The solvation correction was calculated as a difference between the solvation free energies of the protein–ligand complex (PL) with the protein (P) and the ligand (L) free in solution. The solvation free energy was calculated using a Poisson–Boltzmann-based selfconsistent reaction field (PB/SCRF) method in which the polarization of the solute electron density due to the presence of the solvent reaction field is calculated selfconsistently using a QM Hamiltonian [7]. This is a major advantage of using a QM-based solvation method, wherein the dielectric relaxation (or the internal dielectric) of the protein in response to a solvent reaction field is not preset.

This study was followed by a very large-scale and detailed validation of a fully QM-based scoring function, termed QMScore. Interaction energies for a diverse range of protein–ligand complexes comprising 165 noncovalent complexes and 49 metalloenzyme complexes [13] were calculated. For the 165 noncovalent complexes, the interaction energies without any fitting agreed with experimental binding affinity within 2.5 kcal/mol. When different parts of the scoring function were fitted to experimental free energies of binding using regression methods, the agreement was within 2.0 kcal/mol. For metalloenzymes, the agreement with experiments without fitting was within 1.7 kcal/mol and with fitting was within 1.4 kcal/mol. The authors thus demonstrated the inherent predictive ability of this first generation full QM-based scoring function that takes into account all aspects of binding.

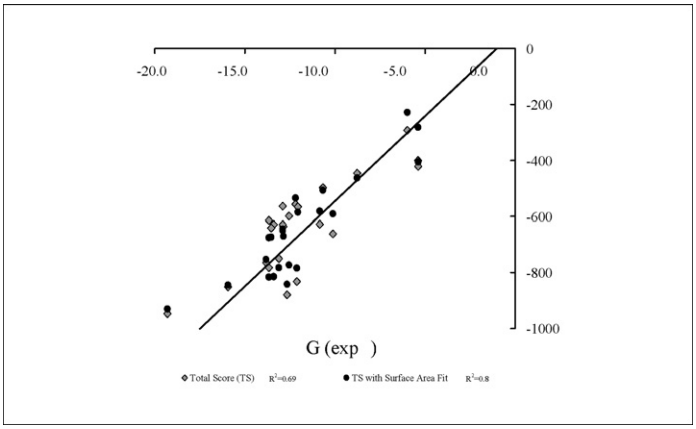


**FIGURE 3**

Plot of calculated QMScore (labeled as total score) versus the Δ*G* (exp) for the set of 23 complexes. (Gray squares) sum of the individual contributions from Eq. (1). The square of the correlation coefficient ($R^2$) is 0.69. (Solid circles) surface areas fitted against Δ*G* (exp) for the set of 23 complexes. The square of the correlation coefficient ($R^2$) is 0.8.

In a related study, Nikitina *et al.*, using linear-scaling QM methods, calculated the binding enthalpy of eight ligands bound to protein conformations from the PDB [38]. The authors chose enthalpy to examine the ability of the semiempirical Hamiltonian PM3 to calculate the enthalpy of binding. The choice of the enthalpy of binding, instead of the free energy of binding, was wise because the computation of entropy is far more challenging. Another important aspect of the study was inclusion of water molecules in the calculation of enthalpy. The structural water molecules were included in the computation of reference state enthalpies of the protein and ligand. They tried two different schemes where water molecules that were hydrogen bonded to both the protein and the ligand in the complex were considered in both reference state calculations of the protein and the ligand. One drawback of the study is the exclusion of solvation effects, or the solvation correction to the enthalpy of binding. However, the authors argue that solvation effects are modeled enthalpically by including explicit water molecules. The calculated enthalpies agreed with the experimental enthalpies within 2 kcal/mol.

Other recent examples of using of linear-scaling QM in SBDD include a study by Vasilyev and Bliznyuk, where the semiempirical computer program MOZYME was used to rescore the top 100 predicted ligands from another docking program. The authors evaluated the feasibility of using a linear-scaling QM program for such a task [39]. In another application of MOZYME, Ohno *et al.* studied the affinity maturation of an antibody by calculating the binding free energy of the hapten bound to a germline antibody and the mature form [40]. The authors emphasize the importance of polarization and charge transfer in the maturation process.

The recent development of linear-scaling technology has focused on higher levels of theory, such as Hartree–Fock or Density Functional Theory (DFT) to calculate the wavefunctions of macromolecules. Gao *et al.* have described the development and application of a density matrix (DM) scheme based on Molecular Fractionation with Conjugate Caps (MFCC) [41]. Using this method, the density matrix is calculated for capped fragments of a macromolecule at high levels of theory. The total energy is then calculated from the full DM that is assembled from the fragment DMs. In an application of this method, Chen and Zhang calculated the ligand–DNA/RNA interaction at high levels of theory [42]. Although further validation is needed for evaluating the ability of such a method to calculate binding free energies, it clearly has potential.

Fukuzawa *et al.* have used another approach – *ab initio* Fragment Molecular Orbital (FMO) – to calculate the interaction energy of ligands that bind to the human estrogen receptor [43]. While the agreement between the calculated and observed binding affinity is modest, they have examined the feasibility of modeling the receptor using only a few of the residues surrounding the ligand. They found no significant difference in the computed interaction energy between the complete receptor and the pruned receptor that had residues surrounding only the ligand. This hints toward a strategy to reduce even further the time taken for such calculations; however, a more thorough validation study is still needed.

Experimental measures of binding affinity give very little insight into the relationship of the binding pose of an active inhibitor and its interaction with the receptor. Such insights can be very useful for the process of going from a lead to a drug. Computational methods, in general, provide access to the decomposition of the interaction energy between the ligand and the receptor. However, with the application of QM to SBDD, these insights are more grounded theoretically and can often be validated by experiments. These insights can be utilized in design cycles comprising prediction and testing for increasing the potency of submicromolar leads in drug discovery.

Both QM/MM and linear-scaling QM methods have been used to dissect the interaction of a ligand with its receptor. Hensen *et al.* used MD and QM/MM to dissect the interaction of inhibitors bound to the HIV-1 protease [18]. They demonstrated that a 4-hydroxy-dihydropyrone substructure of the most potent inhibitor, tripnavir, made favorable interactions with the catalytic aspartates and isoluecine residues of the HIV-1 protease. He *et al.* used the linear-scaling DM-MFCC approach to dissect the interaction between the HIV-1 reverse transcriptase (RT) and its drug resistant mutants with the inhibitor nevirapine. The authors calculated a QM interaction spectrum that sheds light on crucial aspects of resistance to RT [44].

Raha *et al.*, using linear-scaling QM and a pairwise energy decomposition (PWD) scheme, dissected the interaction of a series of fluorine substituted ligands (*N*-(4-sulfamylbenzoyl)benzylamine or SBB) with HCA [45]. They divided the enzyme and inhibitors into subsystems and calculated the exchange energy that consisted of the off diagonal elements of the density matrix and the one-electron matrix elements between subsystems:

$$E_{AB} = \sum_{\mu}^{A}\sum_{\nu}^{B} P_{\mu\nu}^{AB}\left(2H_{\mu\nu}^{AB} - \frac{1}{2}\sum_{\lambda}^{B}\sum_{\sigma}^{A} P_{\lambda\sigma}^{BA}\left(\mu^{A}\sigma^{A}|\lambda^{B}\nu\right)^{B}\right) \quad (2)$$

Here $A$ and $B$ are residue subsystems, and $P$ and $H$ are the density matrix and the one-electron matrix, respectively. Using this PWD scheme, the authors investigated the effect of substitution of fluorines on the distal aromatic rings of SBB inhibitors, on its interaction with HCA. The authors probed at the relationship of various pairwise interactions with the free energy of binding of the inhibitors. It was found that the substitution of fluorine at the distal group did not directly affect the free energy of binding. Rather, it geometrically influenced the strongest interaction between the sulfonamide group of the inhibitor and the Thr199 residue of the protein. This strong interaction, which was chemically identical in each of the inhibitors, was directly correlated with the binding affinity of the ligand. Such insights can be valuable in designing new and potent inhibitors.

The PWD scheme was also incorporated into the Comparative Binding Energy Analysis (COMBINE) [46] methodology of Ortiz and coworkers to create SE-COMBINE by Peters and Merz [47]. This method elucidated the most important interactions between trypsin and a series of trypsin inhibitors. Protein–ligand interaction energies are decomposed to find the most or least stabilizing interactions as well as providing a means to identify regions of significant variation (thereby targeting areas that could benefit from more optimization). The multivariate statistical tools, PCA and PLS, were used to mine the interactions between the receptor residues and the ligand fragments to generate QSAR models. The authors introduced the so-called Intermolecular Interaction Maps (IMMs), an example of which is given in Figure 4, which enable the

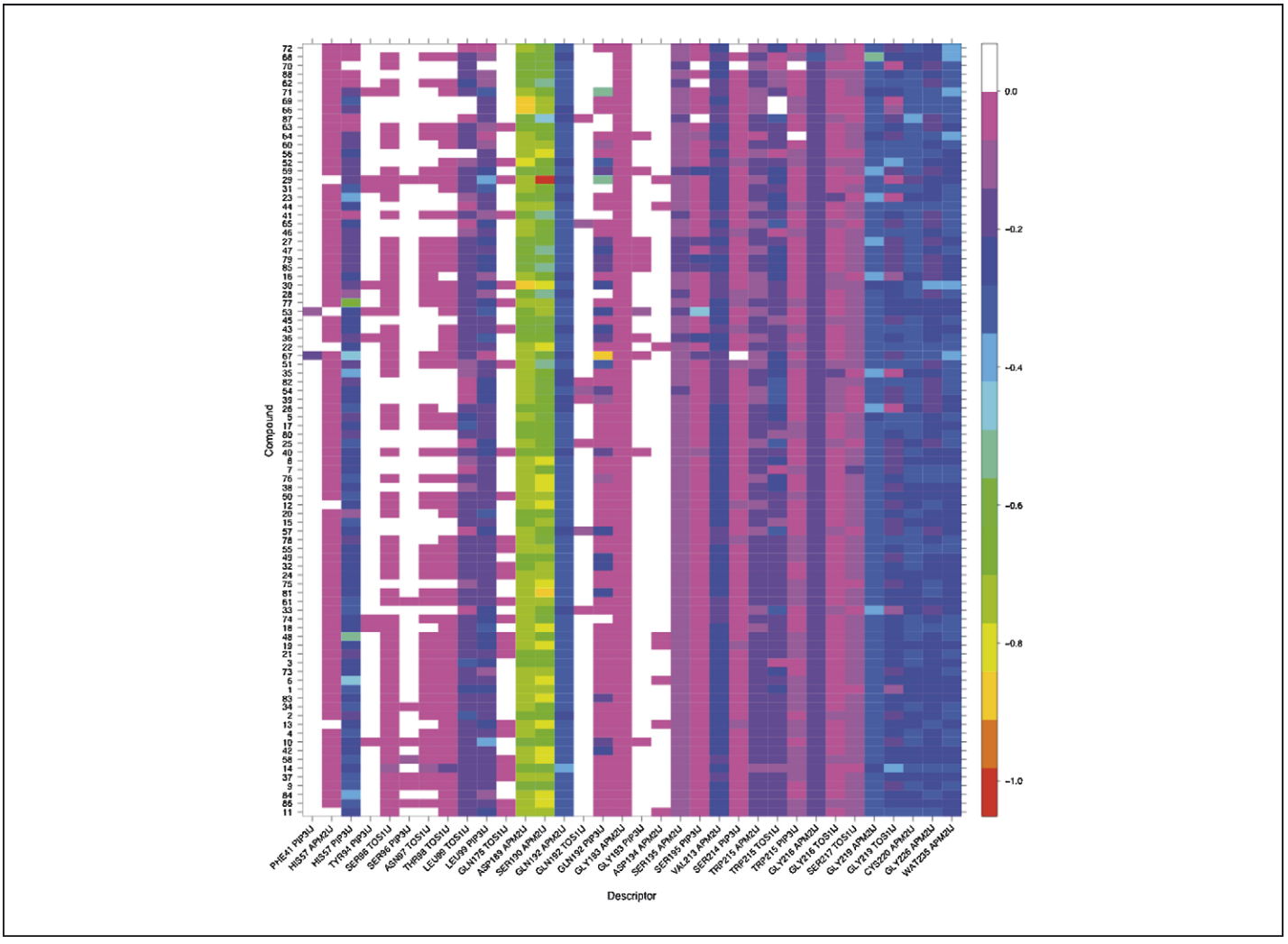


**FIGURE 4**

Model Lig3C Intermolecular Interaction Map (IMM) of the important $E_{AB}$ descriptors. The key residues of trypsin that interact with the triple fragment ligand (APM, TOS, and PIP; see Figure 5) label the *x*-axis. The compounds on the *y*-axis are ordered with respect to activity. The activity decreases from top to bottom. The legend indicates the magnitude of the unscaled descriptor in eV.

researcher to view graphically where a candidate drug could be modified or optimized.

## Outlook

QM-based methods can impact many aspects of SBDD as indicated by Figure 1 and this review has touched on a few examples of the application of QM-derived methods to DD. As with any brief review, it is difficult to catalog all the most recent advances; however, the use of quantum mechanical approaches in drug design problems, using both ligand- and structure-based drug design applications, will certainly experience tremendous growth in the coming years. The ability, through the variational principle, for QM to give chemically accurate interaction energies between a receptor and ligand and its ability to generate novel descriptor classes should attract even more attention to the use of QM in drug design in the coming years. The transformation is likely to happen slowly and almost imperceptibly, because even faster QM methodologies are required and careful validation studies to demonstrate improved performance over classical methodologies will multiply in the coming decade. Entropic and dynamical effects still are major hurdles to effective ligand- and structure-based drug design methods in both the classical and QM realms. Perhaps a marriage of classical methods with QM will provide a way to solve these vexing fundamental chemistry problems that transcend many fields where conformational dynamics is important in biomolecular function.

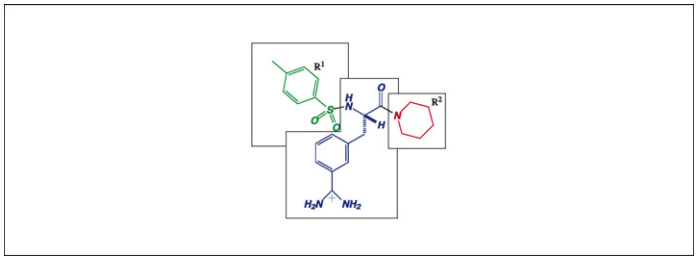

**FIGURE 5**

Schematic diagram of a trypsin inhibitor fragmentation. The structure in blue is the 3-amidino-phenylalanine moiety (APM). The TOS group is colored green while the PIP group is shown in red.

## References


1 Friesner, R.A. and Gullar, V. (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* 56, 389–427

2 Monard, G. and Merz, K.M. (1999) Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.* 32, 904–911

3 Gao, J.L. and Truhlar, D.G. (2002) Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* 53, 467–505

4 Gao, J.L. *et al.* (2006) Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* 106, 3188–3209

5 Cavalli, A. *et al.* (2006) Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* 106, 3497–3519

6 Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science* 268, 1144–1149

7 Gogonea, V. *et al.* (2001) New developments in applying quantum mechanics to proteins. *Curr. Opin. Struct. Biol.* 11, 217–223

8 Khandogin, J. and York, D.M. (2004) Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins: Struct. Funct. Bioinform.* 56, 724–737

9 Bombarda, E. *et al.* (2001) Determination of the pK(a) of the four $Zn^{2+}$-coordinating residues of the distal finger motif of the HIV-1 nucleocapsid protein: consequences on the binding of $Zn^{2+}$. *J. Mol. Biol.* 310, 659–672

10 Dixon, S.L. and Merz, K.M., Jr (1996) Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* 104, 6643–6649

11 Rajamani, R. and Reynolds, C.H. (2004) Modeling the protonation states of catalytic aspartates in β-secretase. *J. Med. Chem.* 47, 5159–5166

12 Yu, N. *et al.* (2006) Assigning the protonation states of the key aspartates in beta-secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.* 2, 1057–1069

13 Raha, K. and Merz, K.M., Jr (2005) Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes. *J. Med. Chem.* 48, 4558–4575

14 Gao, J.L. and Xia, X.F. (1992) A priori evaluation of aqueous polarization effects through Monte-Carlo Qm–Mm simulations. *Science* 258, 631–635

15 van der Vaart, A. and Merz, K.M., Jr (1999) The role of polarization and charge transfer in the solvation of biomolecules. *J. Am. Chem. Soc.* 121, 9182–9190

16 van der Vaart, A. and Merz, K.M., Jr (2000) Charge transfer in biologically important molecules: comparison of high-level ab initio and semiempirical methods. *Int. J. Quantum Chem.* 77, 27–43

17 Mo, Y.R. *et al.* (2000) Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* 112, 5530–5538

18 Hensen, C. *et al.* (2004) A combined QM/MM approach to protein–ligand interaction: polarization effects of HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* 47, 6673–6680

19 Garcia-Viloca, M. *et al.* (2003) Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J. Mol. Biol.* 372, 549–560

20 Dal Peraro, M. *et al.* (2005) Solute–solvent charge transfer in aqueous solution. *Chemphyschem* 6, 1715–1718

21 Komeiji, Y. *et al.* (2007) Change in a protein's electronic structure induced by an explicit solvent: an *ab initio* fragment molecular orbital study of ubiquitin. *J. Comput. Chem.*, Early view

22 Irwin, J.J. and Shoichet, B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model* 45, 177–182

23 Irwin, J.J. *et al.* (2005) Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry* 44, 12316–12328

24 Cho, A.E. *et al.* (2005) Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical approach. *J. Comp. Chem.* 29, 917–930

25 Raha, K., and Merz, J.K.M. Structural basis of dielectric permittivity of proteins: insights from quantum mechanics. *Proceedings of The Fermi School* (in press)

26 Taylor, R. *et al.* (2002) A review of protein–small molecule docking methods. *J. Comput. Aided Mol. Des.* 16, 151–166

27 Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931

28 Warren, G.L. *et al.* (2004) Critical assessment of docking programs and scoring functions. *Abstr. Pap. Am. Chem. Soc.* 228, U513–U514

29 Schneidman-Duhovny, D. *et al.* (2004) Predicting molecular interactions in silico. II. Protein–protein and protein–drug docking. *Curr. Med. Chem.* 11, 91–107

30 Raha, K. and Merz, K.M., Jr (2005) Calculating binding free energy in protein–ligand interaction. *Ann. Rep. Comput. Chem.* 1, 113–130

31 Khandelwal, A. *et al.* (2005) A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* 48, 5437–5447

32 Grater, F. *et al.* (2005) Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J. Phys. Chem. B* 109, 10474–10483

33 Merz, K.M., Jr and Banci, L. (1997) Binding of bicarbonate to human carbonic anhydrase II: a continuum of binding states. *J. Am. Chem. Soc.* 119, 863–871

34 Klahn, M. *et al.* (2005) On possible pitfalls of ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J. Phys. Chem. B* 109, 15645–15650

35 Dixon, S.L. and Merz, K.M., Jr (1997) Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* 107, 879–893

36 Lee, T.S. *et al.* (1996) Linear-scaling semiempirical quantum calculations for macromolecules. *J. Chem. Phys.* 105, 2744–2750

37 Raha, K. and Merz, K.M., Jr (2004) A quantum mechanics based scoring function: study of zinc-ion mediated ligand binding. *J. Am. Chem. Soc.* 126, 1020–1021

38 Nikitina, E. *et al.* (2004) Semiempirical calculations of binding enthalpy for protein–ligand complexes. *Int. J. Quantum Chem.* 97, 747–763

39 Vasilyev, V. and Bliznyuk, A.A. (2004) Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor. Chem. Acc.* 112, 313–317

40 Ohno, K. *et al.* (2005) Quantum chemical study of the affinity maturation of 48G7 antibody. *Theor. Chem. Acc.* 722, 203–211

41 Gao, A.M. *et al.* (2004) An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* 394, 293–297

42 Chen, X.H. and Zhang, J.Z.H. (2004) Theoretical method for full ab initio calculation of DNA/RNA–ligand interaction energy. *J. Chem. Phys.* 120, 11386–11391

43 Fukuzawa, K. *et al.* (2005) Ab initio quantum mechanical study of the binding energies of human estrogen receptor alpha with its ligands: an application of fragment molecular orbital method. *J. Comput. Chem.* 26, 1–10

44 He, X. *et al.* (2005) Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations. *Proteins: Struct. Funct. Bioinform.* 61, 423–432

45 Raha, K. *et al.* (2005) Pairwise decomposition of residue interaction energies using semiempirical quantum mechanical methods in studies of protein–ligand interaction. *J. Am. Chem. Soc.* 127, 6583–6594

46 Ortiz, A.R. *et al.* (1995) Prediction of drug-binding affinities by comparative binding-energy analysis. *J. Med. Chem.* 38, 2681–2691

47 Peters, M.B. and Merz, K.M., Jr (2006) Semiempirical comparative binding energy analysis (SE-COMBINE) of a series of trypsin inhibitors. *J. Chem. Theory Comput.* 2, 383–399