Quantum Mechanical Methods for Drug Design

Ting Zhou,* Danzhi Huang, and Amedeo Caflisch*

Department of Biochemistry,University of Zurich,

Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

E-mail: t.zhou@bioc.uzh.ch; Caflisch@bioc.uzh.ch

Phone: +41 44 635 55 21. Fax: +41 44 635 68 62

Abstract

Quantum mechanical (QM) methods are becoming popular in computational drug design and development mainly because high accuracy is required to estimate (relative) binding affinities. For low- to medium-throughput in silico screening, (e.g., scoring and prioritizing a series of inhibitors sharing the same molecular scaffold) efficient approximations have been developed in the past decade, like linear scaling QM in which the computation time scales almost linearly with the number of basis functions. Furthermore, QM-based procedures have been used recently for determining protonation states of ionizable groups, evaluating energies, and optimizing molecular structures. For high-throughput in silico screening QM approaches have been employed to derive robust quantitative structure-activity relationship models. It is expected that the use of QM methods will keep growing in all phases of computer-aided drug design and development. However, extensive sampling of conformational space and treatment of solution of macromolecules are still limiting factors for the broad application of QM in drug design.

*To whom correspondence should be addressed

Introduction

Accurate models for computing the binding free energy between small molecules and proteins are needed for drug discovery and design.[1] The increasing popularity of quantum mechanical methods in computer-aided drug design (CADD) is not just a consequence of ever growing computing power but is also due to the first principle nature of QM, which should provide the highest accuracy.[2, 3] Because of their first principle nature, both the time-consuming ab initio methods[4] and fast semi-empirical approaches[5, 6] do not suffer from the limitation inherent to the ball and spring description and the fixed-charge approximation used in the force fields (FFs). In a recent review, it has been suggested that the application of QM methods in all phases of CADD is likely to become reality.[3] At the same time, interest for QM in CADD has spurred further methodological development of QM methods and in particular QM approaches for docking, scoring, improvement of known lead compounds, and unraveling the reaction mechanism. As an example, QM calculations were performed to investigate significant differences in binding affinities upon modification of a $-CH_2-$ linker into a carbonyl.[7]

In the following sections, we will classify the QM methods into two broad classes according to their functionalities: the first class includes the methods used for quantifying energies and optimizing structures, while the second contains the techniques employed for calculating molecular properties. The methods in the first class are the conventional and straightforward applications of QM. They can be directly exploited for interpreting the reactivities of biologically active molecules, which are always accompanied by the transfer of energy and transformation of molecular structures. However, the currently available computating power is not enough for the direct ab initio QM calculations of macromolecules with accuracy similar to that of in vitro experiments. Therefore this section will unavoidably give the prominence to the acceleration of QM methods for macromolecules, including linear scaling algorithms and hybrid quantum-mechanics/molecular-mechanics (QM/MM). Apart from the methods for calculating the energies and optimizing structures of biomolecules, we also illustrate two typical applications of QM relevant to structure-based drug design: the analysis of the protonation states of titratable side chains[8, 9] and the evaluation, with high structural and energetic

accuracy, of cation-π and π-π interactions that are beyond the limits of classical FF methods. The methods in the second class are mainly used for calculating specific properties of molecules, such as partial charges, bond strength, and torsion angles which can be applied in the parameterization of FFs, and other descriptors that can be used in building quantitative structure-activity relationship (QSAR) models or quantitative structure-property relationship (QSPR) models. We also discuss recent advances in two emerging topics: molecular quantum similarity and variational particle number approach for molecular design.

Using QM to calculate energies and optimize structures

QM is preferable to classical FF based methods for accurate energies and electronic structure calculations,[2, 3] and even for examining the potential energy hypersurface of small molecules.[10] Recently, Butler and coworkers used QM-based methods to describe both the internal energy of the ligand and the solvation effect.[11] Their analysis indicates that two thirds of the bioactive conformations of small-molecule inhibitors lie within $0.5\,\mathrm{kcal{\cdot}mol^{-1}}$ of a local minimum, and conformations with penalties above $2.0\,\mathrm{kcal{\cdot}mol^{-1}}$ are generally attributable to inaccuracies in structure determination. However QM can only be applied to molecular systems of limited size, so that QM has to be simplified to adapt to the available computational power, e.g., apply pure QM to a small subset of atoms and polarizable continuum model to emulate the protein and the solvent,[12] or use accelerated techniques which will be mentioned in the following sections.

Linear scaling QM methods

The computational time of QM ranges from N^3 (semi-empirical) to N^5 (second order Møller-Plesset perturbation theory (MP2) and other post-Hartree-Fock (HF) methods), where N is the number of basis functions.[13] Linear scaling quantum mechanics (LSQM) has been applied extensively for the evaluation of binding enthalpy between small molecules and proteins.[2, 3] In LSQM, the computing time scales with N^2 or even N if the local character of chemical interactions is fully

exploited.[14–18] In the divide-and-conquer (D&C) approach, one of the typical LSQM techniques, a large system is decomposed into many subsystems, and the density matrix of each subsystem is determined separately. Finally contributions of individual subsystems are summed to obtain the total density matrix and energy of the system (Figure 1).[15–17, 19] Raha and Merz developed a semi-empirical D&C-based scoring function[18] and studied the ion-mediated ligand binding process. Their study shows that QM is needed for metal-containing system, because the atom types and parameters of metal atoms in most classical FFs are not accurate enough to describe the nature of the interactions between a small molecule and a metal ion in the active site.[20]

We have suggested the use of a semi-empirical D&C strategy as an improvement of the linear interaction energy model with continuum electrostatic solvation (LIECE).[21] The new method QMLIECE was compared to LIECE by application to three enzymes belonging to different classes: the West Nile virus NS3 serine protease (WNV PR) , the aspartic protease of human immunodeficiency virus (HIV-1 PR), and the human cyclin-dependent protein kinase 2 (CDK2).[19] Our results indicate that QMLIECE is superior to LIECE when the inhibitor/protein complexes have highly variable charge–charge interactions, as in the case of 44 peptidic inhibitors of WNV PR, because of the variable polarization effects (Figure 2) which are captured only by QMLIECE (Figure 3).

Localized molecular orbital (LMO) theory is an LSQM approach in which occupied-virtual interactions involving distant LMOs are neglected, i.e., only density matrix and energies of LMOs that belong to a limited number of atoms need to be calculated.[28] Used as a scoring function in virtual screening, however, LMO theory is still not fast enough for evaluating all poses of small molecules generated by docking. Therefore, Vasilyev and Bliznyuk, performed QM implemented in MOZYME[28] based on LMOs only for the $10-100$ top binders predicted by simplified scoring functions.[29] By comparing the results with and without solvation they pointed out that although QM was able to provide more accurate enthalpy values, the solvation model needed to be improved. Anikin and coworker developed another linear-scaling semi-empirical algorithm based on LMOs named LocalSCF.[30] The method resolves the SCF task through the finite atomic expansion of weakly nonorthogonal localized molecular orbitals. The inverse overlap matrix arising from the

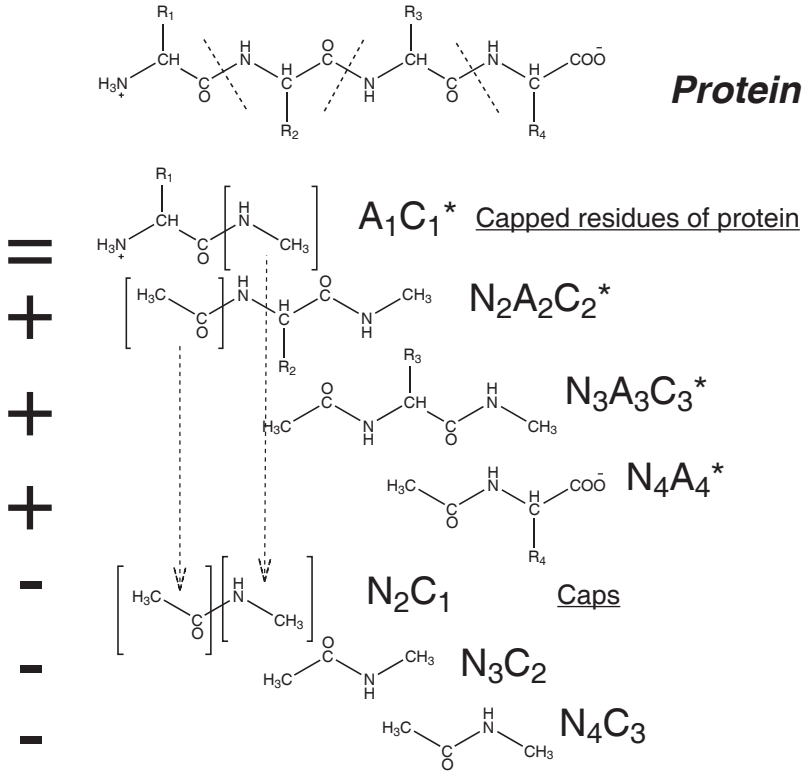


Figure 1: D&C protocol[15] for calculation of QM interaction energy between a protein and a small molecule (ligand). The interaction energy between a protein with m residues and the ligand is decomposed into

$$E_{\text{ligand-protein}} = E_{\text{ligand-A}_1\text{C}_1} + E_{\text{ligand-N}_2\text{A}_2\text{C}_2} + \ldots + E_{\text{ligand-N}_{m-1}\text{A}_{m-1}\text{C}_{m-1}} + E_{\text{ligand-N}_m\text{A}_m} - E_{\text{ligand-N}_2\text{C}_1} - E_{\text{ligand-N}_3\text{C}_2} - \ldots - E_{\text{ligand-N}_m\text{C}_{m-1}}$$

where N_i and C_i are N terminal and C terminal cap, respectively, of residue A_i. The fragments with blue names are protein residues with conjugate caps,[17] while the ones with red names are pure "caps" that have to be subtracted to remove the duplication in energy calculation.[19] This figure is reprinted from Ref.19 with permission of ACS.

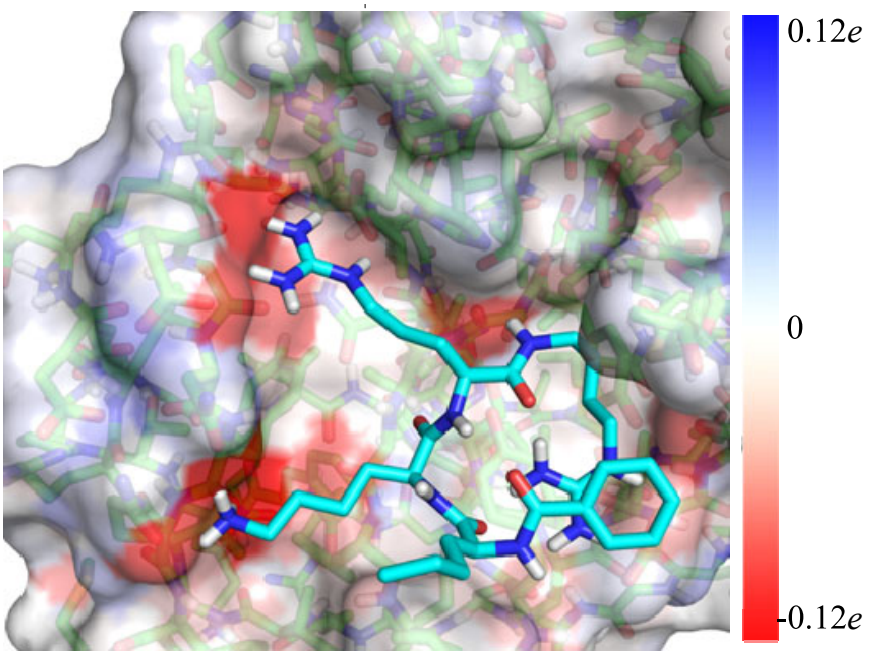


Figure 2: Polarization of protein atoms due to inhibitor binding: WNV PR, whose carbons are in green, in complex with a peptidic inhibitor, whose carbons are in cyan. The polarized charges were calculated by subtracting self-consistent field (SCF) atomic charges before binding from that after binding, using the D&C protocol (see Figure 1).[15] The protein surface was rendered with the blue-white-red spectrum according to polarized charges of atoms. The blue color on the surface denotes atomic partial charges that become more positive upon binding, while red color means more negative atomic charges upon binding, and white color indicates atomic charges which do not change upon binding.

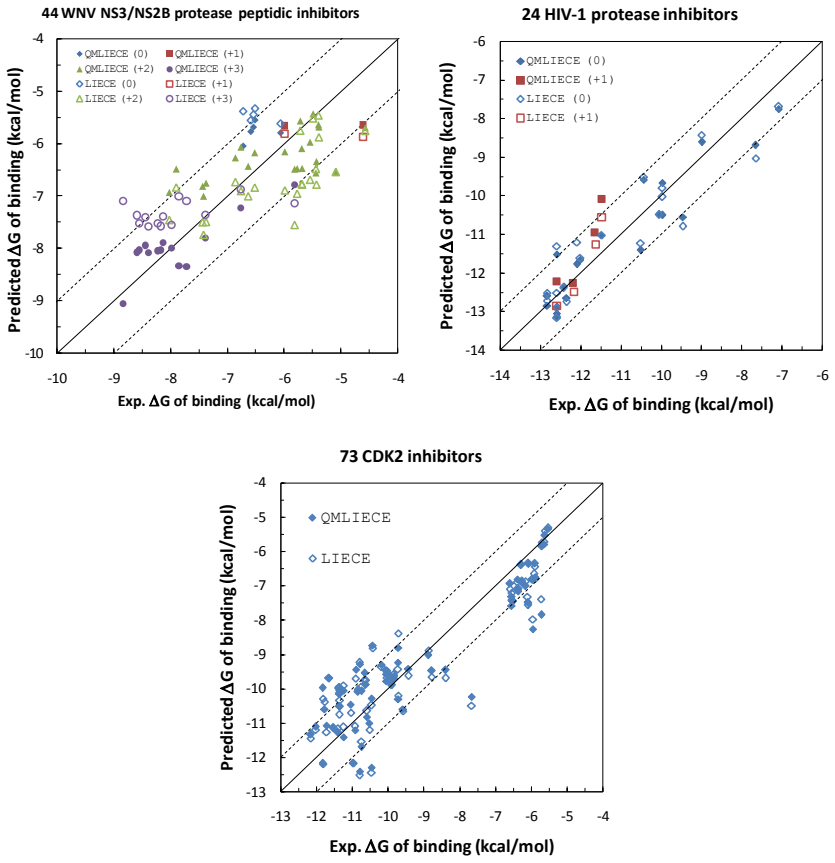


Figure 3: Comparison of the calculated (QMLIECE filled symbols, LIECE empty symbols) versus experimental binding free energies for 44 WNV PR[22–24] (top left), 24 HIV-1 PR[25] (top right), and 73 CDK2[26, 27] (bottom) inhibitors. The experimental data are fitted with two-parameter models for WNV PR, three-parameter models for HIV-1 PR, and two-parameter models for CDK2. Digit in parentheses is the total charge of the inhibitor. This figure is reprinted from Ref.19 with permission of ACS.

nonorthogonality of the localized orbitals is approximated by preserving the first-order perturbation term and applying the second-order correction by means of a penalty function. Furthermore, for very large systems ($>$ 104 atoms) the performance of LocalSCF is more efficient in CPU time and memory consumption than MOZYME with the help of the fast multipole method.[31] Based on the semiempirical variational finite localized molecular orbital approximation, Anisimov and Bugaenko developed the novel semiempirical QM/QM method in which a part of the system, including the ligand and protein active site, are treated self-consistently, while the protein bulk is considered as carrying a frozen electronic density matrix.[32]

QM/MM

The computational procedures based on QM/MM (see also the review article by J. Gascon at page nnn) combine the strengths of both QM (accuracy) and molecular mechanics (MM) (efficiency) methods, and are widely employed to model chemical reactions and other electronic processes in biomolecular systems.[33–43] QM/MM can be used for preparing the structures of small molecules and proteins, such as optimizing the binding poses obtained from docking,[44] and refining the geometries of enzyme active sites obtained from a harmonically restrained minimization with MM,[45] or X-ray structures.[46] It has been suggested that within the drug discovery process QM/MM is valuable for (1) helping the interpretation of poorly resolved electron density,[46] (2) probing the details of the interactions within enzymes active sites,[47] and (3) investigating the effects of different substituents on the binding mode or in the assessment of alternate scaffolds.[48]

QM/MM is also very useful in describing the process of charge polarization and electron transfer, which is not possible by classical FF methods. Anisimov and coworkers used a QM/MM docking approach based on variational finite localized molecular orbital approximation to speed up conventional QM,[30, 31] which took explicitly into account the effects of charge polarization and intermolecular charge transfer. Gentilucci et al. carried out QM/MM calculation to investigate the binding mode into the M-opioid receptor and the electronic properties of an atypical agonist, the cyclic peptide c[YpwFG] which contains aromatic side chains.[44] In their research the highly

favorable dipole-dipole interaction between the protein and the peptide agonist indicates that ligand polarization induced by the protein environment contribute noteworthily to the overall binding energy. Gao and coworkers used docking, molecular dynamics (MD), and QM/MM methods to study the reaction dynamics between pyrimidine nucleoside phosphorylase and a substrate.[49] Their results show that catalysis involved residues stabilize the uridine in a high-energy conformation by electrostatic interactions and the activation of phosphorolytic catalysis stems from polarization effects. As mentioned above, QM is also appropriate to describe electron transfer. Blumberger and coworkers applied QM/MM to calculate the free energy profile for peptide bond cleavage.[50] Zheng and coworkers developed a QM/MM based approach for in silico screening of transition states of the enzymatic reaction and calculated the activation energy.[51] By this approach they designed a human butyrylcholinesterase mutant with a 2000-fold improved catalytic efficiency for therapeutic use as an exogenous enzyme in humans to treat cocaine overdose and addiction. Wallrapp and coworkers presented docking and QM/MM studies for the electron transfer pathway between cytochrome P-450 camphor and putidaredoxin.[52]

QM/MM is a powerful instrument of parameterization of FFs for the system containing structural motifs not adequately described by empirical FFs, such as diverse drug-like molecules,[53] and metal-containing system.[54] QM/MM can also be used for incorporating polarization effects into a FF, which enables the qualitative improvement in constructing patterns of hydrogen bonds of the docked ligand, water structures and dynamics.[55]

QM Simulation

QM simulation is a useful tool for unraveling the mechanism of reactions.[56] In the drug design field which involves biological macromolecules, QM simulation is often working with the classical MD simulation.[41, 57] To explain the catalytic pathway of metalloenzyme farnesyltransferase (FTase), Ho and coworkers exploited the Car-Parrinello MD[58] version of QM(B3LYP density functional theory (DFT))/MM(Amber FF[59]) dynamics. Their results might be helpful in designing selective inhibitors of FTase, given the proposed mechanism of the FTase reaction and the inhibition

by fluorine substituents of farnesyl diphosphate substrate.[60]

QM combined with classical MD is also useful for improving accuracy of interaction energy and sampling of conformational space. Feenstra and coworkers used semi-empirical QM to calculate activation energy barriers, and compare substrate activation barriers at different locations from MD simulations in the enzyme.[61] Alves and coworkers explained the viral resistance of diketo acids (DKAs) to the integrase of human immunodeficiency virus (HIV-1 IN) N155S mutant by QM/MM MD simulation.[62] Their decomposition analysis of energy terms shows that there is a strong interaction between the Lys159, Lys156, Asn155, and Mg^{2+} cation and the DKA inhibitor with complex electrostatic interactions. QM/MM can be used in free energy perturbation (FEP) method. QM/MM FEP was applied to calculate the relative solvation free energies for a diverse set of small molecules (root mean square deviation (RMSD)) from experimental data $< 1.02\,\mathrm{kJ{\cdot}mol^{-1}}$).[53] Using the same method, the $>$ 2000-fold decrease in the affinity for fructose-1,6-bisphosphatase of an adenosine monophosphate (AMP) analogue (phosphonate 4) compared with AMP was explained by the absence of hydrogen bonds and the loss of the electrostatic interactions,[63] which were well described by the QM method.[64] Similarly Khandelwal and coworkers reported that QM/MM calculated energies for the time averaged structures from MD simulations were able to distinguish subtle differences in binding affinities of only one order of magnitude (Figure 4).[65, 66] These methods, however, are very time-consuming, and therefore are not applicable for in silico high throughput screening at present.

Protonation states

The rapid growth of the number of protein structures determined by X-ray crystallography calls for robust methods for determining hydrogen positions, in particular for active site residues in enzymes.[36, 37, 67] Explicit hydrogen atoms are required for most of the structure based drug design methods,[68] e.g., all-atom MM, MD, docking, and electrostatic calculations. A recent study reaffirms that the protonation state in the active site influences the ability of scoring methods to determine the native binding pose.[69] Although other classical methods, e.g., MD[67] and

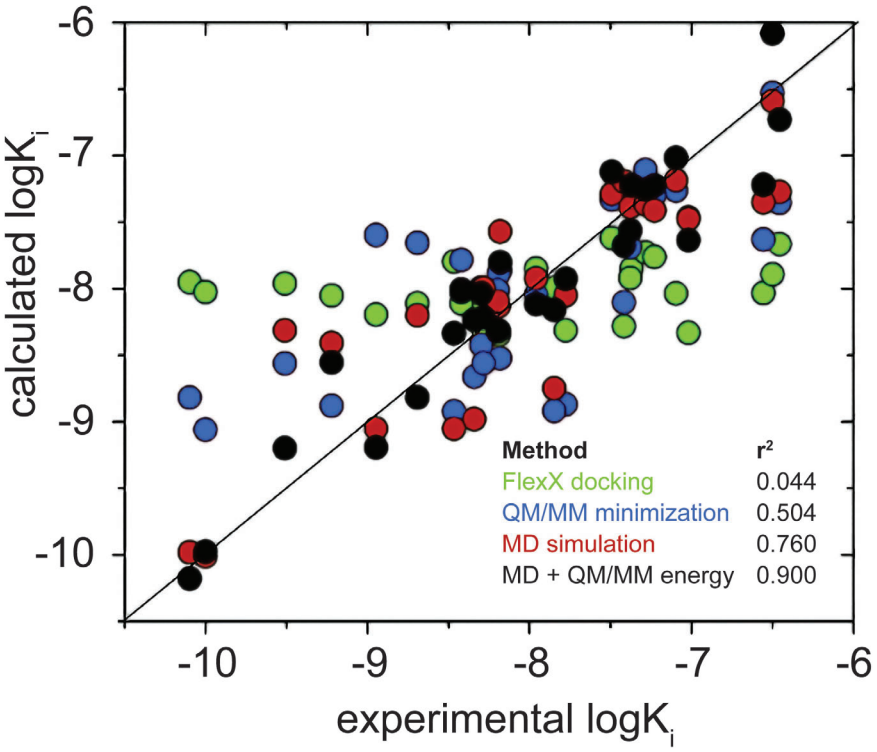


Figure 4: Correlations between experimental and calculated inhibition potencies of hydroxamates vs MMP-9 as obtained by FlexX docking with the zinc-binding-based selection of modes (green), QM/MM minimization (blue), MD simulation with constrained zinc bonds (red), and by QM/MM energy calculations for the time-averaged structures from MD simulation (black).[65, 66] This figure is reprinted from Ref65 with permission of ACS.

MM/Poisson-Boltzmann (PB) surface area,[70] can be used for determining the position of hydrogen atoms, the prediction of protonation states should be more robust by means of QM because protonation is related to the formation of the covalent bond between the hydrogen and heavy atom.

There are several studies on the determination of protonation states of protease, e.g., β-secretase (BACE),[71–74] plasmepsin,[67] and HIV-1 PR.[69, 75] Which of the two aspartates in the catalytic dyad of BACE (Asp32 or Asp228) is protonated is likely to depend on the presence and type of inhibitor.[71, 72, 76] Rajamani and coworkers used a LSQM method and the finite-difference PB method to determine the protonation state and proton location in the presence and absence of an inhibitor.[72] They performed structural optimization in the region surrounding the catalytic dyad. Their calculation favors the monoprotonated state of Asp228 in presence of the hydroxyethylene based inhibitor and di-deprotonated state for the apo enzyme. Yu and coworkers applied the QM/MM to further refine the X-ray structure of BACE, and observed an energetically favored monoprotonated configuration of Asp32 by fitting eight refined structures of BACE and an inhibitor to the observed electron density.[74]

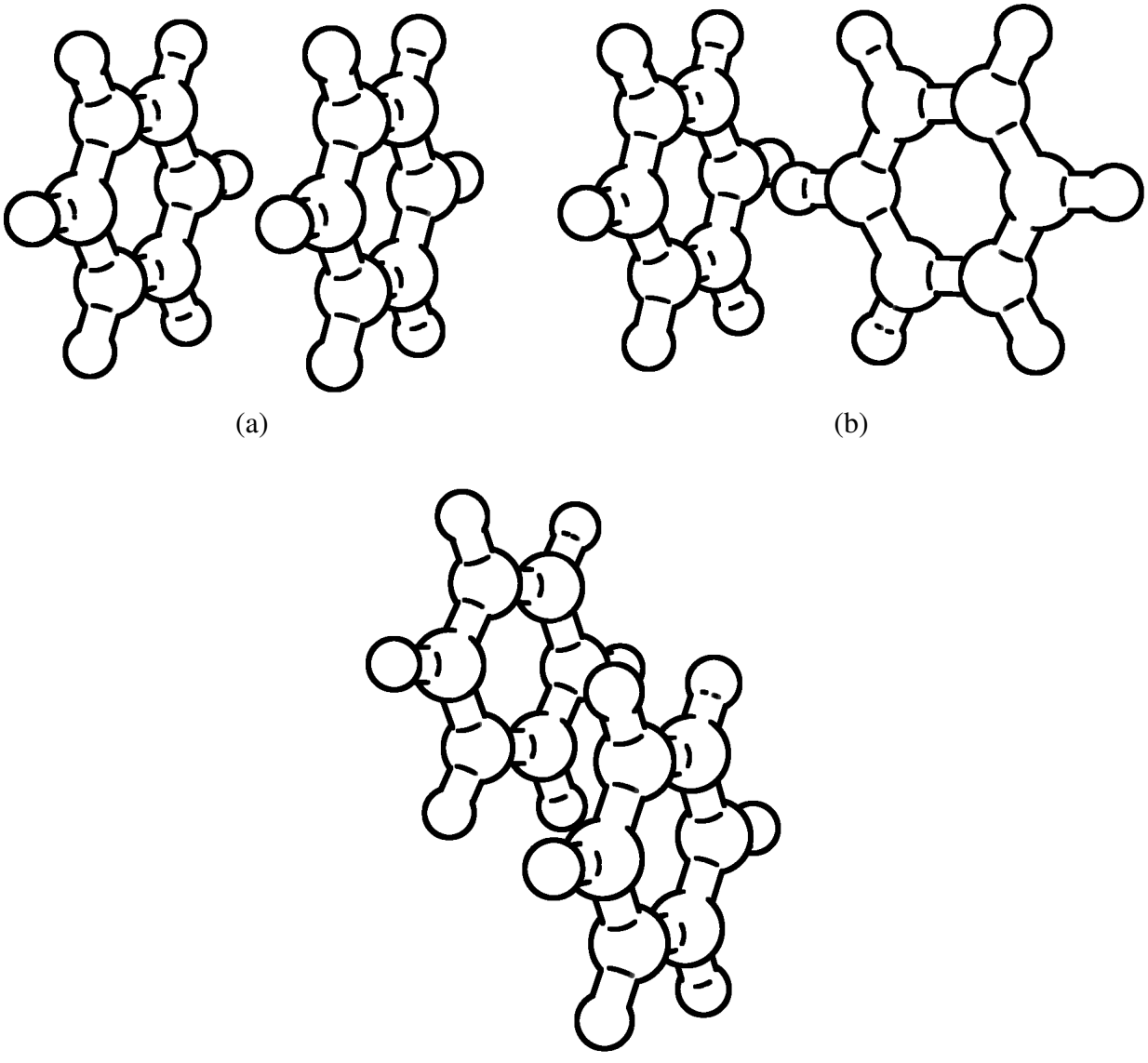
Determination of protonation states of metal-binding sites poses challenges on classical methods.[77, 78] Lin and Lim used a combination of QM and continuum dielectric methods to compute the free energies for deprotonating a Zn-bound imidazole/water in various zinc complexes.[79] They found that the protonation state of the His in the Zn-binding site depends on the solvent accessibility of metal-binding site and Lewis acid ability of the zinc atom. They also suggested that it is critical for the QM region to include not only the metal's first-shell interactions, but also the second shell in QM/MM modeling of metal-binding sites of metalloproteins.[79, 80] A comprehensive review of Kamerlin and coworkers summarizes the progresses in ab initio QM/MM free-energy simulations of electrostatic energies in proteins.[41] Their accelerated QM/MM method, which uses an updated mean charge distribution and a classical reference potential, was benchmarked on the pK_a of titratable side chains. For Asp3 in the bovine pancreatic trypsin inhibitor they obtained the deviation of ~ 1 pK_a unit ($1\,\mathrm{kcal\cdot mol^{-1}}$). For Lys102 in T4-lysozyme mutant the deviation was 2.4 pK_a unit ($\sim 3\,\mathrm{kcal\cdot mol^{-1}}$). The protonation state of Lys102 may affect the conformation of the protein, since

it is deeply buried in the hydrophobic surface. Therefore there is much larger likelihood to attain significant errors in calculation of pK_a of its side chain.[81] Compared to the 7 kcal·mol^{-1} energy difference required for catalysis, an error of 3 kcal·mol^{-1} may be acceptable to determine the main energetic contribution to the reaction.

Cation-π and π-π interactions

Cation-π and π-π stacking interactions play a fundamental role in chemical and biological recognition.[82] Classical FFs sometimes fail to describe these interactions because of the lack of charge delocalization in fixed-charge models or the particular FF parameters. Even HF methods have limitations in capturing π-interactions because of incompleteness of electronic correlation.[83, 84] Villar and coworkers analyzed whether ligand-protein binding enthalpies evaluated by semi-empirical Austin Model 1 (AM1) are sufficient for use in the rational design of new drugs by comparing with B3LYP DFT, and MP2 method.[85] They pointed out that with the exception of cation-π interactions the enthalpies calculated by AM1 correlated well with that by counterpoise-corrected MP2/6-31G(d). However, the structures calculated by AM1 and DFT do not correlated with that calculated by MP2 consistently. Wu and McMahon applied DFT and MP2 to optimize the structures of the most stable isomers of protonated Tyr and ammonia or methylamine and to calculate the enhancement of binding energies due to cation-π interactions. Møller-Plesset perturbation and coupled-cluster methods show that dispersive forces and electrostatic and exchange-repulsion forces play the primary stabilizing role in π-stacked complexes.[84, 86, 87] High-level ab initio calculations, including extrapolation to the MP2 basis set limit and inclusion of a CCSD(T) correction, show that T-shaped and parallel-displaced configurations are virtually isoenergetic in gas phase, with binding energies of -11.46 and -11.63 kJ·mol^{-1} respectively, whereas the sandwich structure is less stable at -7.57 kJ·mol^{-1} (Figure 5),[84] and substituted benzene dimers bind more strongly than unsubstituted benzene.[88] Hobza and coworkers suggested to model the π-π stacking interactions by MP2 with a medium-sized basis set with a more diffuse polarization function, i.e., MP2/6 31G*(0.25) where exponents of *d* polarization functions are changed into more diffuse 0.25 from 0.8 used in the standard 6-31G*

basis.[89–92] Because of its computational efficiency, DFT has been used by several groups to describe π-stacking interactions.[93–97] To attain predicting power similar to high-level ab initio methods, some researchers have combined HF theory and DFT, and using modest basis sets have reproduced the potential energy surface of higher level calculations for a number of instances of π-stacking.[98–100]



(a) (b) (c)

Figure 5: (a) Sandwich, (b) T-shaped, and (c) parallel-displaced configurations of the benzene dimer.[84, 88]

Using QM to calculate molecular properties

It has long been recognized that if one could accurately evaluate the standard free energy change of complexation of biologically active molecules, it would be possible both to gain a deeper understanding of molecular recognition in biology, and to shed light onto the first principles design of pharmaceuticals and other compounds.[101] The currently available computer power does not allow highly accurate QM calculations of free energies, particularly for proteins and ligands in solution. Moreover, usage of QM methods in high-throughput docking is prohibitive. Therefore QM is more suitable to derive models for prediction rather than for the direct evaluation of binding free energies. Classical FFs and QSAR are examples of compromises between accuracy and efficiency.

QM derived FFs

Due to the large chemical space of molecules, FFs do not include all the parameters required for describing drug-like molecules.[102] QM is being used routinely in optimizing geometries, fitting the torsion parameters, and deriving atomic charges for proteins,[103] DNA,[104] and in particular small molecules.[105–108] Spiegel and coworkers developed a new set of FF parameters of platinated moiety via a force matching procedure of the classical forces to ab initio forces obtained from QM/MM trajectories, and extended the classical MD simulation to describe slow converging rearrangement of dinuclear Pt compounds and DNA duplex.[109] Sugiyama and coworkers used DFT calculated partial charges and FF parameters for the atoms near the active site, which are usually significantly polarized, and metal atoms for which FF parameters are not available.[110]

Multipole expansion (ME) is often used in the representation of the molecular electrostatic potential.[30, 111–113] To account for the effects of charge penetration, the point charges, dipoles, quadrupoles, and octupoles in ME model need to be damped. The damping strategies are particular crucial for short-range energies. For example, damping strategies have to be used when ME is applied on calculation of electrostatic potentials or electric fields on van der Waals (vdW) or solvent accessible surface of a molecule.[114] Werneck and coworkers suggested a general methodology

to optimize the damping functions with the ab initio (HF/6-31G** and 6-31G**+) electrostatic potential.[115]

QM-derived partial charges

In the molecular simulations with fixed charge models, the method used to derive partial charges influences the computed physical properties and subsequent docking and scoring significantly.[116, 117] Mobley and coworker compared the hydration free energies of small molecules, whose partial charges are assigned according to different levels of QM, including AM1, HF, DFT, and MP2, by explicit water MD simulations.[118] They found that AM1 bond charge correction method[119] for computing charges works almost as well as any of the more computationally expensive ab initio method. Fischer et al. compared FF-based scoring functions with QM-based scoring functions by computing binding free energies of eleven ligands to the human estrogen receptor subtype α (ERα) and four ligands to the human retinoic acid receptor of isotype γ.[120] They found that the improvement for the complexes with the ERα receptor stemmed from applying classical electrostatic models partial charges derived by fragment molecular orbital (FMO).[121] Illingworth and coworkers implemented QM/MM derived induced charges into a classical framework, redocked 12 difficult protein-ligand complexes with AutoDock,[122] and found that there was no significant improvement in RMSD of the lowest energy structure against the crystal structure but an increment of the largest cluster size.[123] Pasquini and coworkers explained different binding affinities of similar compounds to HIV-1 IN by calculating the partial charges of these compounds and attributed the difference to a poor interaction of the molecules with the divalent metal ions of the active site due to the electron-withdrawing effect.[7] Instead of using fixed and point-charge model, Wang and coworkers calculated solvation free energies of 31 small neutral molecules from QM charge density and continuum dielectric theory (finite-difference PB equation).[124] The QM and PB equations were solved self-consistently until both the charge and reaction field converged. The calculations took into account polarized electronic wave function asymmetric distortion, and spreading out of the electron cloud. In particular, when the solute is treated by QM, part of its electron density penetrates

into the solvent. The experimentally measured solvation free energies of these molecules spanned a range of 25 kcal·mol^{-1}. The authors reported root mean square error of only 1.3 kcal·mol^{-1} upon tuning a single parameter to shift the calculated values.

QM descriptors in QSAR/QSPR models

The information provided by QM is more accurate than FFs, therefore more robust QSAR models and/or QSPR models are expected with QM descriptors.[125] Partial charges are the most common descriptors in QSAR/QSPR models due to their simplicity and informative content. Occhiato and coworkers employed atomic partial charges derived from DFT electrostatic potential in a CoMFA model, with which they designed new 5α-reductase 1 inhibitors.[126] Lepp and Chuman applied LocalSCF calculated atomic charges to build a QSAR model to predict Michaelis-Menten constants, and attained better correlation than classical QSAR descriptors.[127] Wan et al. found that the net charge of the atoms and polarizability correlate with biological activity.[128] Furthermore, they reported that the predictive power of QSAR models derived from DFT charges is higher than from semi-empirical PM3 charges. Besides partial charges, other QM descriptors are commonly used to build QSAR/QSPR models. Yamagami and coworkers used various quantum chemical descriptors, e.g., frontier energy and frontier electron density, which are powerful for describing chemical reactivity.[129] Their CoMFA method shows that the antimutagenic activities are increased by electron-withdrawing substituents and also by hydrogen-bonding between 2-hydroxy group and the receptor. Singh and coworkers developed a QSAR model of derivatives of testosterone with several QM parameters, e.g., absolute hardness and electronegativity.[130] Pasha and coworkers derived QSAR models utilizing various QM descriptors to analyze the factors affecting inhibitory potency for a series of analogues of the MK886 inhibitor of microsomal prostaglandin E2 synthase-1.[131] These QM models indicate that the steric properties, as well as electrostatic and hydrophobic interactions are relevant to the inhibitory potency.

Molecular quantum similarity

Molecular similarity measures have been used in CADD since more than 15 years.[132] Malde and coworkers have investigated boron analogs of natural peptides by QM to find the secondary structural preferences and the impact on stability of different substitutions on boron.[133–135] Recently they have shown that the B(OH)−NH isostere is an interesting surrogate for the peptide bond because of the similar geometry and barrier for rotation around the backbone dihedral angle ω, as well as stability to proteolytic enzymes.[136] Carbó et al. measured the similarity of electron density calculated by QM, and developed a novel QSAR descriptor named molecular quantum similarity measures (MQSM).[137–144] The MQSM relies on the first order electronic density function as molecular descriptor. Before comparing the similarities of electronic density functions, approximated functions[145, 146] and a maximization algorithm are needed to obtain optimal molecular superposition.[147] The MQSM was then used to predict the toxicity[148, 149] and to describe the substituent effect in an aromatic series traditionally described by empirical Hammett equation.[144, 150] MQSM also can be applied for classification of molecules using dendrograms.[151] A further development of MQSM is quantum topological molecular similarity (QTMS),[152] which is based on the definition of distances between molecules in bond critical point space.[153, 154] QTMS is very useful to describe pK_a of molecules as QSAR descriptors.[155, 156] Singh et al. suggested the connection between QTMS and relative bond dissociation enthalpies, and attained good QSAR.[157] Hemmateenejad and Mohajeri used QTMS indices for describing the quantitative effects of molecular electronic environments on the O-methylation kinetic of substituted phenols.[158] Their results revealed that the rate constant of esterification of phenols is highly influenced by the electronic properties of the C_2-C_1-O-H fragment of the parent molecule, which can be considered as frontier bonds in the O-methylation reaction. As shown in these examples, the effects of substitutions are related to the electron density of the bond connecting the scaffolds and the substituents, such that molecular quantum similarity is suitable for studying substitutive effects.

Variational particle number approach for molecular design

Chemical space is the high-dimensional molecular space spanned by the astronomical number of accessible chemical structures. How to sample the chemical space efficiently is always a difficult problem in de novo drug design. In general terms, compound design efforts usually attempt a mapping of a given molecular system to the observable of interest. However in structure-based drug design, the inverse question applies, i.e., which modification of a given compound will result in a desired molecular property. Two independent research groups have recently addressed this question. Lilienfeld and coworkers developed an approach that can explore chemical space in a less heuristic manner by extending the conceptual DFT by the chemical potential for nuclei (alchemical potential).[159] With their approach, they modified a peptidic inhibitor of an anticancer target (human X-chromosome linked inhibitor-of-apoptosis-proteins) into a nonpeptidic inhibitor by optimizing the interaction energy between the inhibitor and the target. Almost at the same time, Wang and coworkers optimized molecular polarizability and hyperpolarizability using a similar method. The main idea of these methods is mapping the discrete chemical structures onto a continuous hypersurface (Figure 6). In this case, enumerating the astronomical number of discrete chemical structures can be avoided by a systematic optimization of parameters introduced in the mapping procedure. Up to now, these methods are merely applied for optimizing several molecular properties e.g., polarizability and hyperpolarizability, which can be calculated by QM straightforwardly.[160–166] QSAR/QSPR models may bridge the gap between the properties calculated directly by QM and those that are useful for drug discovery, e.g., high binding affinity and selectivity to the target, good pharmacokinetics and pharmacodynamics, and low toxicity, so that the variational particle number approach might become a routine of structure-based drug design.

Conclusion and outlook

Several QM-based methods already play an important role in many phases of CADD, and will have a stronger impact in the future because of the ever growing computing power and development

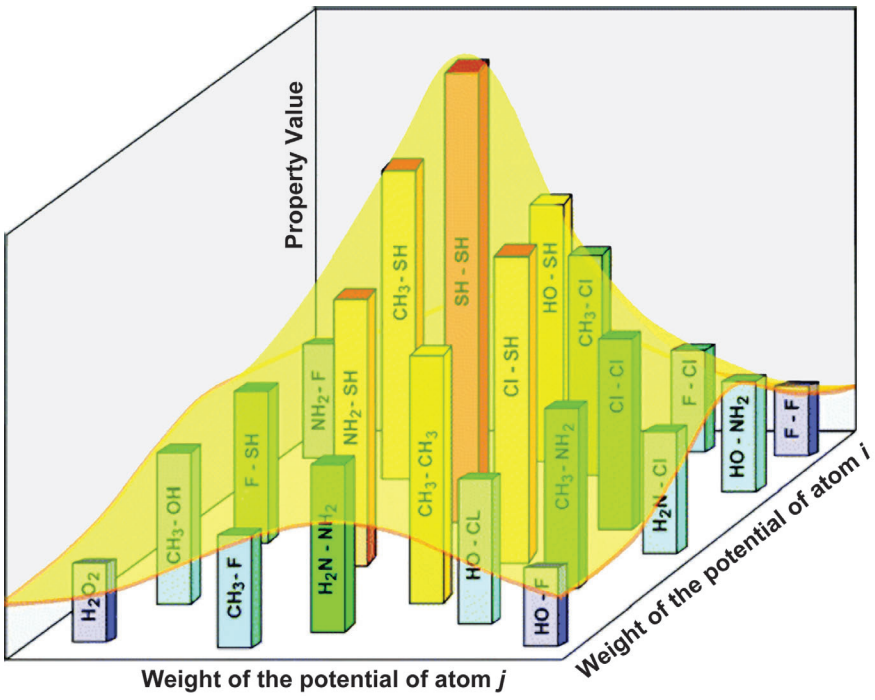


Figure 6: Schematic representation of optimization of molecular properties by a linear combination of atomic potentials. Bar heights represent electronic polarizabilities for candidate structures. The optimization of the property is performed on the smooth (hyper)surface (only two degrees of freedom are denoted). Establishing a well behaved property surface that interpolates among the realizable molecules is a key aspect of the variational particle number approach.[167] This figure is reprinted from Ref.167 with permission of ACS.

of efficient algorithms. The compromise between accuracy and efficiency is a perpetual issue in the applications of QM methods. It is important to select the most appropriate technique at each phase of drug development, and QM methods should be selected only if there is a real advantage with respect to the classical approaches. The initial phase of CADD, e.g., high-throughput docking, which is useful to identify hit compounds,[168] requires full sampling of conformations of the small molecules within the protein binding site. Such extensive sampling calls for approximated energy functions and predicted properties thereof, which are usually calculated by classical FF methods or fast semi-empirical QM methods. In the subsequent phase, hits have to be optimized to leads which does not require extensive sampling but high accuracy because of the small differences in the binding free energy. Therefore QM methods should be applied on the hits to shed light on the energetics of binding. The QM methods are particularly important to capture charge transfer and polarization effects, which are usually pronounced in systems containing metal atoms or charged groups, and/or dispersion forces which play a significant role in the interactions of conjugated π systems. Importantly, before starting CADD it is necessary to evaluate the status of the project, which in turn dictates the number and diversity of molecules to be evaluated and the demand of accuracy, and to select the most appropriate approach accordingly.

Acknowledgement

This work was supported by a grant of the Swiss National Science Foundation to A.C.

References

[1] Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.

[2] Cavalli, A.; Carloni, P.; Recanatini, M. Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* **2006**, *106*, 3497–3519.

[3] Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; WollaCott, A. M.; Westerhoff, L. M.; Merz, K. M.

The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* **2007**, *12*, 725–731.

[4] Peters, M. B.; Raha, K.; Merz, K. M. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 370–379.

[5] Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RMI: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.

[6] Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

[7] Pasquini, S.; Mugnaini, C.; Tintori, C.; Botta, M.; Trejos, A.; Arvela, R. K.; Larhed, M.; Witvrouw, M.; Michiels, M.; Christ, F.; Debyser, Z.; Corelli, F. Investigations on the 4-quinolone-3-carboxylic acid motif. 1. Synthesis-activity relationship of a class of human immunodeficiency virus type 1 integrase inhibitors. *J. Med. Chem.* **2008**, *51*, 5125–5129.

[8] Vorobjev, Y. N.; Scheraga, H. A.; Honig, B. Theoretical modeling of electrostatic effects of titratable side-chain groups on protein conformation in a polar ionic solution. 2. pH-induced helix-coil transition of poly-(L-lysine) in water and methanol ionic solutions. *J. Phys. Chem.* **1995**, *99*, 7180–7187.

[9] Vorobjev, Y. N.; Scheraga, H. A.; Hitz, B.; Honig, B. Theoretical modeling of electrostatic effects of titratable side-chain groups on protein conformation in a polar ionic solution. 1. Potential of mean force between charged lysine residues and titration of poly-(L-lysine) in 95% methanol solution. *J. Phys. Chem.* **1994**, *98*, 10940–10948.

[10] Bombasaro, J. A.; Masman, M. F.; Santagata, L. N.; Freile, M. L.; Rodriguez, A. M.; Enriz, R. D. A comprehensive conformational analysis of bullacin B, a potent inhibitor of complex I. Molecular dynamics simulations and ab initio calculations. *J. Phys. Chem. A* **2008**, *112*, 7426–7438.

[11] Butler, K. T.; Luque, F. J.; Barril, X. Toward accurate relative energy predictions of the bioactive conformation of drugs. *J. Comput. Chem.* **2009**, *30*, 601–610.

[12] Buback, V.; Mladenovic, M.; Engels, B.; Schirmeister, T. Rational design of improved aziridine-based inhibitors of cysteine proteases. *J. Phys. Chem. B* **2009**, *113*, 5282–5289.

[13] Van der Vaart, A.; Gogonea, V.; Dixon, S. L.; Merz, K. M. Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method. *J. Comput. Chem.* **2000**, *21*, 1494–1504.

[14] Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165–9169.

[15] Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643–6649.

[16] Lee, T. S.; Lewis, J. P.; Yang, W. T. Linear-scaling quantum mechanical calculations of biological molecules: The divide-and-conquer approach. *Comp. Mat. Sci.* **1998**, *12*, 259–277.

[17] Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.

[18] Raha, K.; Merz, K. M. 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.* **2005**, *48*, 4558–4575.

[19] Zhou, T.; Huang, D.; Caflisch, A. Is quantum mechanics necessary for predicting binding free energy? *J. Med. Chem.* **2008**, *51*, 4280–4288.

[20] Raha, K.; Merz, K. M. A quantum mechanics-based scoring function: Study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020–1021.

[21] Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.

[22] Knox, J. E.; Ma, N. L.; Yin, Z.; Patel, S. J.; Wang, W. L.; Chan, W. L.; Rao, K. R. R.; Wang, G.; Ngew, X.; Patel, V.; Beer, D.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors. *J. Med. Chem.* **2006**, *49*, 6585–6590.

[23] Yin, Z.; Patel, S. J.; Wang, W. L.; Chan, W. L.; Rao, K. R. R.; Wang, G.; Ngew, X.; Patel, V.; Beer, D.; Knox, J. E.; Ma, N. L.; Ehrhardt, C.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 40–43.

[24] Yin, Z.; Patel, S. J.; Wang, W. L.; Wang, G.; Chan, W. L.; Rao, K. R. R.; Alam, J.; Jeyaraj, D. A.; Ngew, X.; Patel, V.; Beer, D.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of dengue virus NS3 protease. Part 1: Warhead. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 36–39.

[25] Dreyer, G. B.; Lambert, D. M.; Meek, T. D.; Carr, T. J.; Tomaszek, T. A.; Fernandez, A. V.; Bartus, H.; Cacciavillani, E.; Hassell, A. M.; Minnich, M.; Petteway, S. R.; Metcalf, B. W.; Lewis, M. Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease - structure activity analysis using enzyme-kinetics, X-ray crystallography, and infected T-cell assays. *Biochemistry (Mosc.)* **1992**, *31*, 6646–6659.

[26] Bramson, H. N. et al. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.

[27] Gibson, A. E. et al. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O6 substituted guanine derivatives. *J. Med. Chem.* **2002**, *45*, 3381–3393.

[28] Stewart, J. J. P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int. J. Quantum Chem.* **1996**, *58*, 133–146.

[29] Vasilyev, V.; Bliznyuk, A. Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor. Chem. Acc.* **2004**, *112*, 313–317.

[30] Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules. *J. Chem. Phys.* **2004**, *121*, 1266–1270.

[31] Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V. Validation of linear scaling semiempirical LocalSCF method. *J. Chem. Theory Comput.* **2006**, *2*, 1685–1692.

[32] Anisimov, V. M.; Bugaenko, V. L. QM/QM docking method based on the variational finite localized molecular orbital approximation. *J. Comput. Chem.* **2009**, *30*, 784–798.

[33] Gao, J. L.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.

[34] Gao, J. L.; Ma, S. H.; Major, D. T.; Nam, K.; Pu, J. Z.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3188–3209.

[35] Cheng, Y. H.; Cheng, X. L.; Radic, Z.; McCammon, J. A. Acetylcholinesterase: Mechanisms of covalent inhibition of H447I mutant determined by computational analyses. *Chem. Biol. Interact.* **2008**, *175*, 196–199.

[36] Hu, H.; Boone, A.; Yang, W. T. Mechanism of OMP decarboxylation in orotidine 5 '-monophosphate decarboxylase. *J. Am. Chem. Soc.* **2008**, *130*, 14493–14503.

[37] Lameira, J.; Alves, C. N.; Moliner, V.; Marti, S.; Kanaan, N.; Tunon, I. A quantum mechanics/molecular mechanics study of the protein–ligand interaction of two potent inhibitors of human O-GlcNAcase: PUGNAc and NAG-thiazoline. *J. Phys. Chem. B* **2008**, *112*, 14260–14266.

[38] Mladenovic, M.; Junold, K.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. Atomistic insights into the inhibition of cysteine proteases: First QM/MM calculations clarifying the regiospecificity and the inhibition potency of epoxide- and aziridine-based inhibitors. *J. Phys. Chem. B* **2008**, *112*, 5458–5469.

[39] Suresh, C. H.; Vargheese, A. M.; Vijayalakshmi, K. P.; Mohan, N.; Koga, N. Role of structural water molecule in HIV protease-inhibitor complexes: A QM/MM study. *J. Comput. Chem.* **2008**, *29*, 1840–1849.

[40] Wu, R. B.; Cao, Z. X. QM/MM study of catalytic methyl transfer by the N-5-Glutamine SAM-dependent methyltransferase and its inhibition by the nitrogen analogue of coenzyme. *J. Comput. Chem.* **2008**, *29*, 350–357.

[41] Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **2009**, *113*, 1253–1272.

[42] Lodola, A.; Mor, M.; Sirirak, J.; Mulholland, A. J. Insights into the mechanism and inhibition of fatty acid amide hydrolase from quantum mechanics/molecular mechanics (QM/MM) modelling. *Biochem. Soc. Trans.* **2009**, *37*, 363–367.

[43] Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

[44] Gentilucci, L.; Squassabia, F.; Demarco, R.; Artali, R.; Cardillo, G.; Tolomelli, A.; Spampinato, S.; Bedini, A. Investigation of the interaction between the atypical agonist c[YpwFG] and MOR. *FEBS J.* **2008**, *275*, 2315–2337.

[45] Mukherjee, P.; Desai, P. V.; Srivastava, A.; Tekwani, B. L.; Avery, M. A. Probing the structures of leishmanial farnesyl pyrophosphate synthases: homology modeling and docking studies. *J. Chem. Inf. Model.* **2008**, *48*, 1026–1040.

[46] Fanfrlik, J.; Brynda, J.; Rezac, J.; Hobza, P.; Lepsik, M. Interpretation of protein/ligand crystal structure using QM/MM calculations: case of HIV-1 protease/metallacarborane complex. *J. Phys. Chem. B* **2008**, *112*, 15094–15102.

[47] Mladenovic, M.; Arnone, M.; Fink, R. F.; Engels, B. Environmental effects on charge densities of biologically active molecules: Do molecule crystal environments indeed approximate protein surroundings? *J. Phys. Chem. B* **2009**, *113*, 5072–5082.

[48] Gleeson, M. P.; Gleeson, D. QM/MM calculations in drug discovery: A useful method for studying binding phenomena? *J. Chem. Inf. Model.* **2009**, *49*, 670–677.

[49] Gao, X.; Huang, X.; Sun, C. Role of each residue in catalysis in the active site of pyrimidine nucleoside phosphorylase from Bacillus subtilis: A hybrid QM/MM study. *J. Struct. Biol.* **2006**, *154*, 20–26.

[50] Blumberger, J.; Lamoureux, G.; Klein, M. L. Peptide hydrolysis in thermolysin: Ab initio QM/MM investigation of the Glu143-assisted water addition mechanism. *J. Chem. Theory Comput.* **2007**, *3*, 1837–1850.

[51] Zheng, F.; Yang, W.; Ko, M.; Liu, J.; Cho, H.; Gao, D.; Tong, M.; Tai, H.; Woods, J. H.; Zhan, C. Most efficient cocaine hydrolase designed by virtual screening of transition states. *J. Am. Chem. Soc.* **2008**, *130*, 12148–12155.

[52] Wallrapp, F.; Masone, D.; Guallar, V. Electron transfer in the P450cam/PDX complex. The QM/MM e-pathway. *J. Phys. Chem. A* **2008**, *112*, 12989–12994.

[53] Reddy, M. R.; Singh, U. C.; Erion, M. D. Ab initio quantum mechanics-based free energy perturbation method for calculating relative solvation free energies. *J. Comput. Chem.* **2007**, *28*, 491–494.

[54] Magistrato, A.; Ruggerone, P.; Spiegel, K.; Carloni, P.; Reedijk, J. Binding of novel azole-

bridged dinuclear Platinum(II) anticancer drugs to DNA: Insights from hybrid QM/MM Molecular Dynamics simulations. *J. Phys. Chem. B* **2006**, *110*, 3604–3613.

[55] Friesner, R. A. Modeling polarization in proteins and protein–ligand complexes: Methods and preliminary results. *Adv. Protein Chem.* **2006**, *72*, 79–104.

[56] Trout, B. L.; Parrinello, M. The dissociation mechanism of H_2O in water studied by first-principles molecular dynamics. *Chem. Phys. Lett.* **1998**, 343–347.

[57] Ridder, L.; Mulholland, A. J. Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity. *Curr. Top. Med. Chem.* **2003**, *3*, 1241–1256.

[58] Car,; Parrinello, Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

[59] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[60] Ho, M.; Vivo, M. D.; Peraro, M. D.; Klein, M. L. Unraveling the catalytic pathway of metalloenzyme farnesyltransferase through QM/MM computation. *J. Chem. Theory Comput.* **2009**, *5*, 1657–1666.

[61] Feenstra, K. A.; Starikov, E. B.; Urlacher, V. B.; Commandeur, J. N. M.; Vermeulen, N. P. E. Combining substrate dynamics, binding statistics, and energy barriers to rationalize regioselective hydroxylation of octane and lauric acid by CYP102A1 and mutants. *Protein Sci.* **2007**, *16*, 420–431.

[62] Alves, C. N.; Marti, S.; Castillo, R.; Andres, J.; Moliner, V.; Tunon, I.; Silla, E. A quantum

mechanic/molecular mechanic study of the wild-type and N155S mutant HIV-1 integrase complexed with diketo acid. *Biophys. J.* **2008**, *94*, 2443–2451.

[63] Reddy, M. R.; Erion, M. D. Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. *J. Am. Chem. Soc.* **2007**, *129*, 9296–9297.

[64] Sponer, J.; Leszczynski, J.; Hobza, P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **2001**, *61*, 3–31.

[65] Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **2005**, *48*, 5437–5447.

[66] Khandelwal, A.; Balaz, S. QM/MM linear response method distinguishes ligand affinities for closely related metalloproteins. *Proteins* **2007**, *69*, 326–339.

[67] Friedman, R.; Caflisch, A. The protonation state of the catalytic aspartates in plasmepsin II. *FEBS Lett.* **2007**, *581*, 4120–4124.

[68] Klein, C. D. P.; Schiffmann, R.; Folkers, G.; Piana, S.; Rothlisberger, U. Protonation states of methionine aminopeptidase and their relevance for inhibitor binding and catalytic activity. *J. Biol. Chem.* **2003**, *278*, 47862–47867.

[69] Fong, P.; McNamara, J. P.; Hillier, I. H.; Bryce, R. A. Assessment of QM/MM scoring functions for molecular docking to HIV-1 protease. *J. Chem. Inf. Model.* **2009**, *49*, 913–924.

[70] Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J. Comput. Chem.* **2008**, *29*, 673–685.

[71] Park, H.; Lee, S. Determination of the active site protonation state of β-secretase from

molecular dynamics simulation and docking experiment: Implications for structure-based inhibitor design. *J. Am. Chem. Soc.* **2003**, *125*, 16416–16422.

[72] Rajamani, R.; Reynolds, C. H. Modeling the protonation states of the catalytic aspartates in β-secretase. *J. Med. Chem.* **2004**, *47*, 5159–5166.

[73] Polgar, T.; Keseru, G. M. Virtual screening for β-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749–3755.

[74] Yu, N.; Hayik, S. A.; Wang, B.; Liao, N.; Reynolds, C. H.; Merz, K. M. Assigning the protonation states of the key aspartates in β-secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.* **2006**, *2*, 1057–1069.

[75] Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737.

[76] Gorfe, A. A.; Caflisch, A. Functional plasticity in the substrate binding site of β-secretase. *Structure* **2005**, *13*, 1487–1498.

[77] Dudev, T.; Lim, C. Metal binding affinity and selectivity in metalloproteins: Insights from computational studies. *Annu. Rev. Biophys.* **2008**, *37*, 97–116.

[78] Seebeck, B.; Reulecke, I.; Kamper, A.; Rarey, M. Modeling of metal interaction geometries for protein–ligand docking. *Proteins* **2008**, *71*, 1237–1254.

[79] Lin, Y. L.; Lim, C. Factors governing the protonation state of Zn-bound histidine in proteins: A DFT/CDM study. *J. Am. Chem. Soc.* **2004**, *126*, 2602–2612.

[80] Dudev, T.; Lim, C. Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.* **2003**, *103*, 773–787.

[81] Riccardi, D.; Schaefer, P.; Cui, Q. pKa calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

[82] Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1250.

[83] Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential energy surface for the benzene dimer. Results of ab initio CCSD(T) calculations show two nearly isoenergetic structures: T-shaped and parallel-displaced. *J. Phys. Chem.* **1996**, *100*, 18790–18794.

[84] Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Estimates of the ab initio limit for π–π interactions: The benzene dimer. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

[85] Villar, R.; Gil, M. J.; Garcia, J. I.; Martinez-Merino, V. Are AM1 ligand-protein binding enthalpies good enough for use in the rational design of new drugs? *J. Comput. Chem.* **2005**, *26*, 1347–1358.

[86] Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. Molecular clusters of π-systems: Theoretical studies of structures, spectra, and origin of interaction energies. *Chem. Rev.* **2000**, *100*, 4145–4185.

[87] Wu, R. H.; McMahon, T. B. Investigation of cation–π interactions in biological systems. *J. Am. Chem. Soc.* **2008**, *130*, 12554–12555.

[88] Sinnokrot, M. O.; Sherrill, C. D. Substituent effects in π–π interactions: Sandwich and T-shaped configurations. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697.

[89] Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: A quantum-chemical study. *Biophys. J.* **1997**, *73*, 76–87.

[90] Hobza, P.; Sponer, J. Significant structural deformation of nucleic acid bases in stacked base pairs: an ab initio study beyond Hartree-Fock. *Chem. Phys. Lett.* **1998**, *288*, 7–14.

[91] Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4 ',6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study. *J. Am. Chem. Soc.* **2002**, *124*, 3366–3376.

[92] Jurecka, P.; Hobza, P. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine center dot center dot center dot cytosine, adenine center dot center dot center dot thymine, and their 9-and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.

[93] Meijer, E. J.; Sprik, M. A density-functional study of the intermolecular interactions of benzene. *J. Chem. Phys.* **1996**, *105*, 8684–8689.

[94] Ye, X. Y.; Li, Z. H.; Wang, W. N.; Fan, K. N.; Xu, W.; Hua, Z. Y. The parallel π–π stacking: a model study with MP2 and DFT methods. *Chem. Phys. Lett.* **2004**, *397*, 56–61.

[95] Xu, X.; Goddard, W. A. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673–2677.

[96] Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. Application of 25 density functionals to dispersion-bound homomolecular dimers. *Chem. Phys. Lett.* **2004**, *394*, 334–338.

[97] Cerny, J.; Hobza, P. The X3LYP extended density functional accurately describes H-bonding but fails completely for stacking. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624–1626.

[98] Perez-Jorda, J. M.; San-Fabian, E.; Perez-Jimenez, A. J. Density-functional study of van der Waals forces on rare-gas diatomics: Hartree-Fock exchange. *J. Chem. Phys.* **1999**, *110*, 1916–1920.

[99] Walsh, T. R. Exact exchange and Wilson-Levy correlation: a pragmatic device for studying complex weakly-bonded systems. *Phys. Chem. Chem. Phys.* **2005**, *7*, 443–451.

[100] Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. Hybrid density functional theory for pi-stacking interactions: Application to benzenes, pyridines, and DNA bases. *J. Comput. Chem.* **2006**, *27*, 491–504.

[101] McCammon, J. A. Free energy calculations in rational drug design. *Abstracts of Papers of the American Chemical Society* **2004**, *227*, U896–U896.

[102] Curioni, A.; Mordasini, T.; Andreoni, W. Enhancing the accuracy of virtual screening: molecular dynamics with quantum-refined force fields. *J. Comput. Aided Mol. Des.* **2004**, *18*, 773–784.

[103] Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Protein Simulations* **2003**, *66*, 27–85.

[104] Ryjacek, F.; Kubar, T.; Hobza, P. New parameterization of the Cornell et al. empirical force field covering amino group nonplanarity in nucleic acid bases. *J. Comput. Chem.* **2003**, *24*, 1891–1901.

[105] Maple, J. R.; Dinur, U.; Hagler, A. T. Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 5350–5354.

[106] Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. Potential energy functions: From consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers* **2003**, *68*, 383–394.

[107] Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. Automated parametrization of biomolecular force fields from quantum mechanics/molecular mechanics (QM/MM) simulations through force matching. *J. Chem. Theory Comput.* **2007**, *3*, 628–639.

[108] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, in press.

[109] Spiegel, K.; Magistrato, A.; Maurer, P.; Ruggerone, P.; Rothlisberger, U.; Carloni, P.; Reedijk, J.; Klein, M. L. Parameterization of azole-bridged dinuclear platinum anticancer drugs via a QM/MM force matching procedure. *J. Comput. Chem.* **2007**, *29*, 38–49.

[110] Sugiyama, A.; Takamatsu, Y.; Nishikawa, K.; Nagao, H.; Nishikawa, K. Docking stability and electronic structure of azurin-cytochrome c551 complex system. *Int. J. Quantum Chem.* **2006**, *106*, 3071–3078.

[111] Dykstra, C. E. Electrostatic interaction potentials in molecular-force fields. *Chem. Rev.* **1993**, *93*, 2339–2353.

[112] Narayszabo, G.; Ferenczy, G. G. Molecular electrostatics. *Chem. Rev.* **1995**, *95*, 829–847.

[113] Engkvist, O.; Astrand, P. O.; Karlstrom, G. Accurate intermolecular potentials obtained from molecular wave functions: Bridging the gap between quantum chemistry and molecular simulations. *Chem. Rev.* **2000**, *100*, 4087–4108.

[114] Dardenne, L. E.; Werneck, A. S.; Neto, M. D.; Bisch, P. M. Electrostatic properties in the catalytic site of papain: A possible regulatory mechanism for the reactivity of the ion pair. *Proteins* **2003**, *52*, 236–253.

[115] Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. General methodology to optimize damping functions to account for charge penetration effects in electrostatic calculations using multicentered multipolar expansions. *J. Phys. Chem. A* **2008**, *112*, 268–280.

[116] Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular

docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915–931.

[117] Kumar, S.; Jaller, D.; Patel, B.; LaLonde, J. M.; DuHadaway, J. B.; Malachowski, W. P.; Prendergast, G. C.; Muller, A. J. Structure based development of phenylimidazole-derived inhibitors of indoleamine 2,3-dioxygenase. *J. Med. Chem.* **2008**, *51*, 4968–4977.

[118] Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

[119] Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

[120] Fischer, B.; Fukuzawa, K.; Wenzel, W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins* **2008**, *70*, 1264–1273.

[121] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

[122] Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.

[123] Illingworth, C. J. R.; Morris, G. M.; Parkes, K. E. B.; Snell, C. R.; Reynolds, C. A. Assessing the role of polarization in docking. *J. Phys. Chem. A* **2008**, *112*, 12157–12163.

[124] Wang, M. L.; Wong, C. F. Calculation of solvation free energy from quantum mechanical charge density and continuum dielectric theory. *J. Phys. Chem. A* **2006**, *110*, 4873–4879.

[125] Gacche, R.; Khsirsagar, M.; Kamble, S.; Bandgar, B.; Dhole, N.; Shisode, K.; Chaudhari, A. Antioxidant and anti-inflammatory related activities of selected synthetic chalcones: structure-

activity relationship studies using computational tools. *Chem. Pharm. Bull. (Tokyo)* **2008**, *56*, 897–901.

[126] Occhiato, E. G.; Ferrali, A.; Menchi, G.; Guarna, A.; Danza, G.; Comerci, A.; Mancina, R.; Serio, M.; Garotta, G.; Cavalli, A.; De Vivo, M.; Recanatini, M. Synthesis, biological activity, and three-dimensional quantitative structure-activity relationship model for a series of benzo[c]quinolizin-3-ones, nonsteroidal inhibitors of human steroid 5 α-reductase 1. *J. Med. Chem.* **2004**, *47*, 3546–3560.

[127] Lepp, Z.; Chuman, H. Connecting traditional QSAR and molecular simulations of papain hydrolysis - importance of charge transfer. *Bioorg. Med. Chem.* **2005**, *13*, 3093–3105.

[128] Wan, J.; Zhang, L.; Yang, G. F.; Zhan, C. G. Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2099–2105.

[129] Yamagami, C.; Motohashi, N.; Akamatsu, M. Quantum chemical- and 3-D-QSAR (CoMFA) studies of benzalacetones and 1,1,1-trifluoro-4-phenyl-3-buten-2-ones. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2281–2285.

[130] Singh, P. P.; Srivastava, H. K.; Pasha, F. A. DFT-based QSAR study of testosterone and its derivatives. *Bioorg. Med. Chem.* **2004**, *12*, 171–177.

[131] Pasha, F. A.; Muddassar, M.; Jung, H.; Yang, B.; Lee, C.; Oh, J. S.; Cho, S. J.; Cho, H. QM and pharmacophore based 3D-QSAR of MK886 analogues against mPGES-1. *Bull. Korean Chem. Soc.* **2008**, *29*, 647–655.

[132] Good, A. C.; So, S. S.; Richards, W. G. Structure-activity-relationships from molecular similarity-matrices. *J. Med. Chem.* **1993**, *36*, 433–438.

[133] Malde, A.; Khedkar, S.; Coutinho, E.; Saran, A. Geometry, transition states, and vibrational spectra of boron isostere of N-methylacetamide by ab initio calculations. *Int. J. Quantum Chem.* **2005**, *102*, 734–742.

[134] Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. Stationary points on the PES of N-methoxy peptides and their boron isosteres: An ab initio study. *J. Chem. Theory Comput.* **2006**, *2*, 1664–1674.

[135] Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. Isosteres of peptides: boron analogs as dipolar forms of alpha-amino acids - a theoretical study. *J. Phys. Org. Chem.* **2007**, *20*, 151–160.

[136] Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. The B(OH)–NH analog is a surrogate for the amide bond (CO–NH) in peptides: An ab initio study. *J. Chem. Theory Comput.* **2007**, *3*, 619–627.

[137] Carbó, R.; Besalu, E.; Amat, L.; Fradera, X. Quantum molecular similarity measures (QMSM) as a natural way leading towards a theoretical foundation of quantitative structure-properties relationships (QSPR). *J. Math. Chem.* **1995**, *18*, 237–246.

[138] Fradera, X.; Amat, L.; Besalu, E.; Carbó-Dorca, R. Application of molecular quantum similarity to QSAR. *Quantitative Structure-Activity Relationships* **1997**, *16*, 25–32.

[139] Lobato, M.; Amat, L.; Besalu, E.; Carbó-Dorca, R. Structure-activity relationships of a steroid family using quantum similarity measures and topological quantum similarity indices. *Quantitative Structure-Activity Relationships* **1997**, *16*, 465–472.

[140] Amat, L.; Robert, D.; Besalu, E.; Carbó-Dorca, R. Molecular quantum similarity measures tuned 3D QSAR: An antitumoral family validation study. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 624–631.

[141] Besalu, E.; Girones, X.; Amat, L.; Carbó-Dorca, R. Molecular quantum similarity and the fundamentals of QSAR. *Acc. Chem. Res.* **2002**, *35*, 289–295.

[142] Bultinck, P.; Carbó-Dorca, R.; Van Alsenoy, C. Quality of approximate electron densities and internal consistency of molecular alignment algorithms in molecular quantum similarity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1208–1217.

[143] Bultinck, P.; Kuppens, T.; Girone, X.; Carbó-Dorca, R. Quantum similarity superposition algorithm (QSSA): A consistent scheme for molecular alignment and molecular similarity based on quantum chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1143–1150.

[144] Girones, X.; Carbó-Dorca, R.; Ponec, R. Molecular basis of LFER. Modeling of the electronic substituent effect using fragment quantum self-similarity measures. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2033–2038.

[145] Constans, P.; Carbó, R. Atomic shell approximation - electron-density fitting algorithm restricting coefficients to positive values. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1046–1053.

[146] Amat, L.; Carbó-Dorca, R. Quantum similarity measures under atomic shell approximation: First order density fitting using elementary Jacobi rotations. *J. Comput. Chem.* **1997**, *18*, 2023–2039.

[147] Constans, P.; Amat, L.; Carbó-Dorca, R. Toward a global maximization of the molecular similarity function: Superposition of two molecules. *J. Comput. Chem.* **1997**, *18*, 826–846.

[148] Gallegos, A.; Robert, D.; Girones, X.; Carbó-Dorca, R. Structure-toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity. *J. Comput. Aided Mol. Des.* **2001**, *15*, 67–80.

[149] Girones, X.; Carbó-Dorca, R. Modelling toxicity using molecular quantum similarity measures. *QSAR Comb. Sci.* **2006**, *25*, 579–589.

[150] Girones, X.; Ponec, R. Molecular quantum similarity measures from Fermi hole densities: Modeling Hammett sigma constants. *J. Chem. Inf. Model.* **2006**, *46*, 1388–1393.

[151] Bultinck, P.; Carbó-Dorca, R. Molecular quantum similarity matrix based clustering of molecules using dendrograms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 170–177.

[152] O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. 3. QTMS descriptors. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764–775.

[153] Popelier, P. L. A. Quantum molecular similarity. 1. BCP space. *J. Phys. Chem. A* **1999**, *103*, 2883–2890.

[154] O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. Part 2: The relation between properties in BCP space and bond length. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **1999**, *77*, 28–36.

[155] Chaudry, U. A.; Popelier, P. L. A. Estimation of pK(a) using quantum topological molecular similarity descriptors: Application to carboxylic acids, anilines and phenols. *J. Org. Chem.* **2004**, *69*, 233–241.

[156] Esteki, M.; Hemmateenejad, B.; Khayamian, T.; Mohajeri, A. Multi-way analysis of quantum topological molecular similarity descriptors for modeling acidity constant of some phenolic compounds. *Chem. Biol. Drug Des.* **2007**, *70*, 413–423.

[157] Singh, N.; Loader, R. J.; O'Malley, P. J.; Popelier, P. L. A. Computation of relative bond dissociation enthalpies (Delta BDE) of phenolic antioxidants from quantum topological molecular similarity (QTMS). *J. Phys. Chem. A* **2006**, *110*, 6498–6503.

[158] Hemmateenejad, B.; Mohajeri, A. Application of quantum topological molecular similarity descriptors in QSPR study of the O-methylation of substituted phenols. *J. Comput. Chem.* **2008**, *29*, 266–274.

[159] von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U. Variational particle number approach for rational compound design. *Phys. Rev. Lett.* **2005**, *95*, Doi 10.1103.

[160] von Lilienfeld, O. A.; Tuckerman, M. E. Molecular grand-canonical ensemble density functional theory and exploration of chemical space. *J. Chem. Phys.* **2006**, *125*, 154104–154113.

[161] Keinan, S.; Hu, X. Q.; Beratan, D. N.; Yang, W. T. Designing molecules with optimal properties using the linear combination of atomic potentials approach in an AM1 semiempirical framework. *J. Phys. Chem. A* **2007**, *111*, 176–181.

[162] von Lilienfeld, O. A.; Tuckerman, M. E. Alchemical variations of intermolecular energies according to molecular grand-canonical ensemble density functional theory. *J. Chem. Theory Comput.* **2007**, *3*, 1083–1090.

[163] Balamurugan, D.; Yang, W. T.; Beratan, D. N. Exploring chemical space with discrete, gradient, and hybrid optimization methods. *J. Chem. Phys.* **2008**, *129*, 174105–174113.

[164] Hu, X. Q.; Beratan, D. N.; Yang, W. T. A gradient-directed Monte Carlo approach to molecular design. *J. Chem. Phys.* **2008**, *129*, 64102–64110.

[165] Keinan, S.; Paquette, W. D.; Skoko, J. J.; Beratan, D. N.; Yang, W. T.; Shinde, S.; Johnston, P. A.; Lazo, J. S.; Wipf, P. Computational design, synthesis and biological evaluation of para-quinone-based inhibitors for redox regulation of the dual-specificity phosphatase Cdc25B. *Org. Biomol. Chem.* **2008**, *6*, 3256–3263.

[166] Keinan, S.; Therien, M. J.; Beratan, D. N.; Yang, W. T. Molecular design of porphyrin-based nonlinear optical materials. *J. Phys. Chem. A* **2008**, *112*, 12203–12207.

[167] Wang, M. L.; Hu, X. Q.; Beratan, D. N.; Yang, W. T. Designing molecules by optimizing potentials. *J. Am. Chem. Soc.* **2006**, *128*, 3228–3232.

[168] Huang, D.; Caflisch, A. Library screening by fragment-based docking. *J. Molec. Recogn.* **2009**, in press.