When cited, please refer to:
R. A. Bryce and I. H. Hillier. *Current Pharmaceutical
Design* **2014**, 20, 3293-3302.

**Quantum Chemical Approaches: Semiempirical
molecular orbital and hybrid quantum
mechanical/molecular mechanical techniques.**

Richard A Bryce¹ and Ian H Hillier²

¹*School of Pharmacy and Pharmaceutical Sciences, University of Manchester,
Manchester M13 9PT,*

²*School of Chemistry, University of Manchester, Manchester M13 9PL.*

Keywords: quantum chemistry, QM/MM method, semiempirical MO method, carbohydrate, metalloenzyme, enzyme inhibitor, docking of substrate, dispersion correction.

ABSTRACT

The use of computational quantum chemical methods to aid drug discovery is surveyed. An overview of the various computational models spanning *ab initio*, density function theory, semiempirical molecular orbital (MO), and hybrid quantum mechanical (QM)/molecular mechanical(MM) methods is given and their strengths and weaknesses are highlighted, focussing on the challenge of obtaining the accuracy essential for them to make a meaningful contribution to drug discovery. Particular attention is given to hybrid QM/MM and semiempirical MO methods which have the potential to yield the necessary accurate predictions of macromolecular structure and reactivity. These methods are shown to have advanced the study of many aspects of substrate–ligand interactions relevant to drug discovery. Thus, the successful parametrization of semiempirical MO methods and QM/MM methods can be used to model non-covalent substrate-protein interactions, and to lead to improved scoring functions. QM/MM methods can be used in crystal structure refinement and are particularly valuable for modelling covalent protein-ligand interactions and can thus aid the design of transition state analogues. An extensive collection of examples from the areas of metalloenzyme structure, enzyme inhibition and ligand binding affinities and scoring functions are used to illustrate the power of these techniques.

INTRODUCTION.

Computational methods based on molecular mechanical (MM) models are traditionally used to model substrate-protein binding, and are thus the most widely used computational chemistry technique to aid rational drug discovery. Quantum mechanical (QM) methods, although more computationally intensive, can be used to model both covalent and non-covalent intermolecular interactions, and are one of the many techniques now being used to relate the structure of enzymes and proteins to their biological function, and thus can contribute to rational drug discovery. Such calculations can model enzyme mechanisms at a molecular level, often providing information difficult or impossible to obtain from experiment, such as transition structures which may be used to design transition state analogues as inhibitors. The free energies of binding of substrates can also be predicted, which can aid the screening of potential drug molecule candidates. These two areas involve covalent and non-covalent intermolecular interactions respectively, and in both cases their prediction at a level to be of value in drug discovery presents a continuing challenge to the computational chemist.

With ever-greater financial pressure to produce drugs faster, more cheaply and with fewer side-effects, pharmaceutical companies are keen to exploit the full potential of computational techniques to reduce the time from lead discovery through optimisation and into later stages of drug discovery. Advances in hardware as well as the algorithmic developments have meant that increasing attention is being focused on QM methods and their potential in drug discovery. In this review, we will present an overview of the basic QM methods that are now available to study interactions of biological importance and present some examples of their application to drug discovery.

AN OVERVIEW OF QUANTUM MECHANICAL METHODS.

The solution of the Schrodinger equation can in principle yield all of the microscopic properties of a molecular system, statistical mechanics providing the link to the corresponding macroscopic properties. Extremely accurate QM calculations of small molecules can be carried out which provide structural and energetic data to rival experiment. For example, a computational study of the argon-phenol intermolecular complex[1] which is bound by dispersive interactions predicts the binding

energy to be 5.9 kJ mol^{-1} , a value within 1.5 kJ mol^{-1} of experiment. Such interactions are often crucial for biomolecular recognition. This extremely close agreement with experiment was achieved using so-called *ab initio* QM methods which include an accurate treatment of the correlated motion of the electrons, which is usually important in both covalent chemical reactions and non-covalent complexes. These *ab initio* methods have no empirical parameters, but do depend on the quality of the basis sets used to describe the molecular orbitals (MO). There is a hierarchy of *ab initio* methods, ranging from Hartree Fock methods which consider each electron to move in the average field of the other electrons, to those which properly describe the correlated motion of the electrons (eg. CCSD(T), MP2). It would be desirable to use these latter highly correlated methods to describe both enzyme catalysis and substrate-protein binding. However, the basic problem is that these methods scale poorly with an increase in the size of the molecular system, and even with the use of newer linear scaling local correlation methods[2], they can only be applied to molecules having at most a few tens of atoms, particularly if a range of structures or a full potential energy surface needs to be studied.

Even with the continuing reduction of the real cost of computational resources, the use of such high level QM methods will not directly contribute to the drug discovery area in the foreseeable future, and there is a continuing search for more computationally economic, yet sufficiently accurate methods. Hartree Fock calculations do scale more favourably, but still depend on the fourth power of the molecular size, and can be used to study molecular systems having a few thousand atoms, but often lack the necessary accuracy.

For molecular systems containing less than a few hundred atoms, QM calculations are now usually carried out using density functional theory (DFT) methods[3], which can often produce results of chemical accuracy, since electron correlation is indeed included. However, care must be taken to use a functional properly validated for the type of chemical system being modelled.[4] There are often problems using DFT for molecules containing transition metal atoms, particularly those having unpaired electrons, and there is continuing effort to produce more reliable functionals. DFT calculations do scale more favourably with molecular size than high level *ab initio* methods, but are

still too computationally time consuming to be used routinely for systems containing more than a few hundred atoms. An alternative to traditional DFT schemes employing atom centred Gaussian basis sets is to employ a plane-wave basis and atomic pseudopotentials as in the ONETEP code[5], which is now used to model large biological systems.

There has recently been a resurgence in the use of semiempirical MO methods, which can be applied to systems comprising many thousands of atoms, and with suitable parameterization these can achieve an acceptable level of accuracy. The computational speed of semi-empirical methods, being 3-4 orders of magnitude faster than DFT, means that these methods are potentially valuable for the rapid modelling of biological systems. We will discuss the use of semiempirical methods more fully later in this review.

Molecular mechanical (MM) force fields do not treat the electrons explicitly, but use empirical potentials to describe the interactions between the various atoms in the molecule in terms of bonds, angles, dihedrals, and van der Waals and electrostatic energies. The computation of these interactions is both simple and rapid, and MM methods are routinely used to study proteins having many thousands of atoms. With suitable parameters, MM force fields can successfully model the structure and energies of molecules near their equilibrium geometry. However, they often fail when bond breaking or forming takes place, and thus are not generally suited to model chemical reactions, which need QM methods. MM methods are widely used to aid drug discovery, and an excellent review of this area can be found elsewhere in this special issue. In this review we focus on how methods based upon QM may also be used to study systems having many thousands of atoms.

HYBRID QM/MM METHODS

The problem of modelling large systems employing QM methods has been addressed by combining QM and MM methods in the QM/MM model[6,7], where QM is used to describe the region of the system where changes in the electron distribution occurs, such as the active site of an enzyme, or the

binding pocket of a protein, the rest of the enzyme being modelled using MM. An example of this partitioning is illustrated in Figure 1.

In the QM/MM method, the effective Hamiltonian for the whole molecular system is written as the sum of the Hamiltonians for the QM and MM regions and the interaction between them:

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \quad (1)$$

The corresponding total energy (E_{tot}) can be similarly written as

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \quad (2)$$

The interaction term is given by

$$\hat{H}_{\text{QM/MM}} = \sum_{i,s} \frac{q_s}{R_{is}} + \sum_{m,s} \frac{Z_m q_s}{R_{ms}} + \sum_{m,s} \left[\frac{A_{ms}}{R_{ms}^{12}} - \frac{B_{ms}}{R_{ms}^6} \right] \quad (3)$$

where m and s label the QM and MM atoms respectively, and i is the index of the electrons of the QM region. The first term of the right hand side of Eq. (3) accounts for the effect of the formal charges of the MM region (q_s) on the QM region, the final two terms giving the Coulombic and Lennard-Jones interaction between the QM and MM regions. It is possible to use any QM method in a QM/MM model, as long as the integrals describing the interaction of the QM electron density with the MM point charges can be evaluated. In the QM/MM model, the QM region feels both the electrostatic effect of the formal point charges on the MM atoms, and also the van der Waals forces from the MM atoms, so that both an enzymatic reaction and ligand protein binding are properly responsive to the structure of the bulk protein. The QM/MM method was developed by a number of groups, and was later generalized for more than two different regions in the ONIOM scheme of Morokuma and

coworkers[8], and has been the subject of a number of reviews[9,10,11]. We will outline a few central considerations in the use of this approach in drug discovery.

As in all QM studies, the choice of the QM method depends on the size of the QM region, and a balance between accuracy and computational speed. High level *ab initio* methods are rarely employed, except for calibration purposes using structures at fixed geometries[12], and most QM/MM studies employ either DFT or the very fast semiempirical methods which we will discuss in more detail later.

The QM-MM boundary.

A central problem in all QM/MM models is how to treat the boundary between the QM and MM regions. In some situations there is no covalent bond linking the two regions, for example, in solvation studies when the solute is the QM region, and the solvent the MM region. There is a similar situation for non-covalent binding of a QM ligand in an MM binding pocket. Here the interaction between the two regions is due to van der Waals and electrostatic forces, and the QM/MM method can be readily applied. However, in most examples of enzyme catalysis, the partitioning scheme will give rise to covalent bonds linking the QM and MM atoms. (The enzyme chorismate mutase is a much studied enzyme[13,14,15] where there is no such covalent bond, which may account for its popularity with computational chemists). In the presence of such a covalent link, the valency of the QM atom must be satisfied in some way for a realistic QM calculation to be carried out. This can be achieved by the use of a link atom, usually hydrogen (as in Figure 1), or by a localised orbital attached to the QM atom[16,17].

The link atom method is widely used, being simple in concept and in computational implementation. However, it introduces a QM atom, not present in the real system, which can experience unrealistic electrostatic interactions with the MM region. This problem can be alleviated by setting the charges on the junction atoms to zero, the charges being redistributed on the other MM atoms to conserve the total charge. The use of localized orbitals is formally more satisfactory, since they provide a QM

description of the electron distribution in the boundary region, and thus may be included in the QM calculation. However, the optimal localized orbitals will vary depending on the chemical environment, and thus unlike link atoms, they will not generally be transferable. It is usually found that either approach is satisfactory if used carefully.

The MM region.

The QM/MM model allows any force field to be coupled to the QM method. QM/MM simulations of enzyme catalysis and substrate protein binding have been carried out using the full range of commonly used MM force fields; AMBER, CHARMM, GROMOS, and OPLS-AA. There are many surveys of these and other force fields in the literature[18,19].

Although the QM/MM method does include the polarization of the QM region by the MM point charges, the polarization of the MM by the QM region is usually ignored. However, polarization of the individual MM atoms can be included *via* classical induced dipoles, fluctuating charges or Drude oscillators[20]. The charge distributions of both the MM and the QM regions must then be iterated to self-consistency during the calculation, and this procedure incurs a considerable computational overhead, which must be justified by the more realistic model, and hopefully more accurate biological predictions.

SEMIEMPIRICAL MO METHODS.

A central problem in the use of QM/MM methods to study biological problems is the level of QM to be employed, particularly as many calculations may be required, such as to screen potential drug candidates, or to evaluate the many configurations required to predict macroscopic thermodynamic properties such as free energy changes. The ideal strategy would be to use a high level QM method which includes electron correlation explicitly, such as MP2 or CCSD(T) methods, but the routine use of these methods is generally precluded, due to their computational expense. The use of semiempirical methods to model covalent and non-covalent biological interactions has recently been reviewed[21].

We here summarize the important aspects of these methods and give some recent applications of their use.

There was considerable progress in the development of semiempirical MO methods during the 1970s, when *ab initio* calculations could be applied to only the smallest of molecules. These methods are based upon a Hartree Fock model, but include implicit electron correlation effects by the use of empiric parameters. These parameters are optimized by fitting either to experimental data or to the results of high level *ab initio* calculations, which themselves include the effects of electron correlation. The huge computational speed up of these methods, compared to *ab initio* Hartree Fock calculations, is due to the enormous reduction in the number of two-electron repulsion integrals by means of the zero differential overlap (ZDO) approximation. This approximation was implemented in the early CNDO and INDO methods of Pople and coworkers[22,23], which were later refined in the MNDO method of Dewar and Thiel[24], retaining more of the integrals, and parameterizing against experimental data. Further modification of the Hamiltonian and new parameterization schemes led to the AM1 (Austin Model 1)[25] and PM3 (Parameterized Model 3)[26] methods which have been widely used. More recently, reparametrizations of PM3 and AM1 have been carried out resulting in the PM6[27], PM7[28], OMx[29] and RM1[30] schemes. The code MOZYME which employs localized orbitals within the PM6 method has been used to optimize the geometry of large proteins of up to 14000 atoms.[31].

The main drawback of these very fast methods is that the required chemical accuracy cannot be achieved by a universal set of parameters. This problem is particularly acute in the description of the varied chemistry of molecules containing transition metal atoms, due to the complexity of their electronic states, and the restrictions inherent in the use of a minimal basis set which semiempirical methods employ. In view of these problems, Rossi and Truhlar[32] have suggested the use of semiempirical schemes with specific reaction parameters (SRP) for different chemical situations. This philosophy is now generally followed, and there continues to be developments of SRP for use in biomolecular modelling. We refer the reader to [21] where details of the fitting procedures, the choice

of reference data and error function to be employed, are all discussed. We here note a few recent applications of semiempirical MOs methods using SRPs to model interactions of biological importance.

The accurate modelling of carbohydrate structure and binding to proteins are important in a wide range of biological areas such as bacterial adhesions and toxins, viral glycoproteins, and in various amyloid-forming proteins such as those associated with Alzheimer and Creutzfeldt diseases. The flexibility and polar nature of both mono- and poly-saccharides has prompted the development of a number of elaborate force fields in order to properly describe the subtle stereoelectronic anomeric effects in such molecules. An alternative to this MM approach is to use a semiempirical MO method which incorporates explicit electronic polarization. Standard AM1 and PM3 parameterizations lack sufficient accuracy to accomplish this, but a new set of parameters, denoted PM3CARB-1, developed by fitting to MP2 *ab initio* data for small molecule carbohydrate analogues, does have the necessary accuracy[33]. This new parameterization is available within the AMBER simulation package. The PM3CARB-1 method has been used to study the dynamics of the disaccharide 4-*O*- α -D-xylopyranosyl- α -D-xylopyranose in aqueous solution within a QM/MM model, with the QM solute being solvated by 492 water molecules described by the TIP3P MM potential.[33] Previous hybrid QM/MM dynamics simulations using a PM3/TIP3P potential required ring constraints to prevent unphysical ${}^1\text{C}_4$ conformations being adopted, which were not required using the PM3CARB-1 potential.

Hydrogen motion, H^+ , H^- or H , is often involved in the rate-limiting step of many enzyme catalysed reactions, and the hydrogen atom often tunnels through the reaction barrier, thus accelerating the reaction. An accurate calculation of the potential energy surface for the reaction, as well as the degree of tunnelling are both required to relate the protein structure to the rate of hydrogen atom motion. Calculations at many points on the potential energy surface are required, and a computationally feasible approach is to use a semiempirical MO method to do this. A much studied enzyme where proton tunnelling is exceptionally large is methylamine dehydrogenase (MADH), where we have used

the PM3 method with SRPs to successfully predict the degree of tunnelling which is reflected in the very large kinetic isotope effects (KIE) which are observed for this enzyme[34].

The modelling of the mechanism of metalloenzymes is particularly challenging, even with the use of high level *ab initio* and DFT methods. The use of semiempirical methods with appropriate SRPs does allow metalloenzymes to be studied. The iron-sulfur protein, rubredoxin (Rb), which is an important electron transfer protein, is one such example[35]. The active site of Rb involves high spin iron surrounded by four sulphur atoms from cysteine residues. We have used the PM3 method to successfully predict the electron detachment energies of clusters which mimic the active site of this protein, and also used the QM(PM3)/MM approach to predict how both these energies and the Fe-S bond lengths are modified by the environment of the protein.

Intermolecular interactions involving π systems have a major role in governing molecular recognition and biomolecular structure, π - π stacking, C-H/ π , N-H/ π , and cation- π interactions being identified as major stabilizing forces for host-guest interaction including drug binding. It has been estimated that around 60% of aromatic side chains in proteins participate in such π -interactions, which although often individually weak, can cumulatively lead to large effects, DNA structure being a well known example.[36,37] Thus, an accurate knowledge of these nonbonded interactions is essential in the design of new drugs towards specific targets. One such area which is now being actively pursued is that of carbohydrate-protein interactions. Amino acids having aromatic side chains (tryptophan, tyrosine and phenylalanine) are frequently found in protein active sites which recognize carbohydrates, and the importance of such arene-carbohydrate interactions is now recognized.[38] The contribution of dispersion forces to these interactions is frequently extremely important, but the use of *ab initio* methods is here far too computationally costly, and more cost effective approaches have been sought, such as the use of new density functionals. An alternative strategy developed by Grimme[39] is to add an atom-atom pair-wise potential of the form C_6/R^6 to the QM energy in order to account for dispersion effects. This has been widely used in DFT calculations since many functionals do not accurately describe dispersion forces. For example such a DFT-D scheme based

upon the commonly employed B3LYP functional[40] has been shown to describe the full range of noncovalent interactions found in large biomolecules. Such a strategy may also be used when the QM method is a semiempirical one.

Thus, in the case of the PM3 method the dispersion corrected semiempirical energy ($E_{\text{PM3-D}}$) is now given by:

$$E_{\text{PM3-D}} = E_{\text{PM3}} + E_{\text{disp}} \quad (4)$$

where E_{PM3} is the normal QM PM3 energy and E_{disp} is an empirical term containing the dispersion correction. The necessary adjustment of the PM3 parameters was carried out by fitting to the interaction energies of a wide number of prototypical biological interactions computed using high level *ab initio* methods, and resulted in a mean unsigned error in these energies of only a little more than 1 kcal mol⁻¹ [41]. In order to achieve such high accuracy for carbohydrate- π interactions, a modified scheme is required which we have denoted PM3-D*[42]. This scheme has been successful in predicting the interaction energies of model carbohydrate-arene complexes, and also the binding of benzene and fluorobenzene to cyclodextrin[43]. More recently the PM6[44] and OMx[45] schemes have been corrected in a similar way.

Thus, we have seen that semiempirical MO schemes can clearly be successfully parameterized, and then have potential for modelling many substrate-protein interactions of importance in drug discovery. We now discuss some recent applications of QM methods in the area of rational drug discovery.

CHARACTERISATION OF LIGAND AND PROTEIN STRUCTURES AND PROPERTIES

Historically quantum mechanics has played an important role in ligand-based drug design strategies[46]. Semiempirical QM methods have been used for many decades to characterise 3D ligand structures. From the molecule's wavefunction, various useful properties can be calculated, including molecular dipole moment and atom-centred partial charges for polar atoms. Such properties are calculated for a congeneric series of ligands, constituting descriptors that can then be used in

establishing a quantitative structure activity relationship (QSAR). This statistical approach seeks to correlate trends in biological activity for a set of molecules (eg. measured enzyme inhibition constants) to measured and/or calculated properties derived from the structures of the molecules. Once a meaningful QSAR is determined, it can then be used to guide further ligand optimisation. Other descriptors can be derived from QM calculations and include the energy of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), and other indices along similar lines, such as electrophilic and nucleophilic superdelocalizabilities, related to electron-acceptor and donor capacities respectively. Another property which can be derived from computation at the QM level is the molecular electrostatic potential (MEP). The MEP may be projected onto the isodensity molecular surface of the ligand[47] and then used in 3D QSAR analysis[48], to identify regions around a common pharmacophore for a set of small molecules where introducing substituents of a given polarity would be beneficial or detrimental to biological activity.

With the advent of linear scaling QM approaches, it is now possible to calculate the QM electrostatic potential of entire biomacromolecules. Thus these methods have been used, in combination with reaction field solvation models, to compute the electrostatic potential surfaces of A-, B- and Z-DNA[49]. Another example is characterisation of the HIV-1 nucleocapsid protein, where the QM electrostatic potential was shown to provide improved interpretation of experimental observations relative to equivalent MM approaches[50]. These approaches can equally be applied to the prediction of pKa values for titratable residues in proteins. Identification of the most likely protonation states for proteins and protein-ligand complexes are important in providing the right starting point for subsequent receptor-based virtual screening campaigns. Such QM-based approaches have been applied for example to the aspartyl dyads of β -secretase[51] and HIV-1 protease[52]. QM/MM methods have also been used in this context:[53,54] , for example, Ryde and Nilsson predicted the protonation state of a Zn-bound water molecule in alcohol dehydrogenase; best agreement with crystallographic geometry was found when the solvent molecule was modelled as water rather than hydroxide[55]. Similarly, obtaining the correct ligand protonation state is thought to be important in determining its final docked pose, although there is some evidence that questions this assertion[56].

Structure-based drug design of course depends on the generation of accurate structural information, including from experimental sources. Spectroscopic methods have a crucial role to play in determination of the 3D structure of a receptor or receptor-ligand complex. Crystal structures are widely used but vary in resolution and most commonly do not include information about location of the hydrogen atoms. It has been proposed that QM/MM approaches can assist in assignment where the electron density of the crystal structure is poorly resolved, providing an accurate description of the internal geometry and interactions of ligand chemical groups where the standard MM parameterization may be unsatisfactory or unavailable[57].

QM/MM methods have recently been used to reoptimise crystal structure atomic coordinates. In this approach, the MM component of an ONIOM-like QM/MM scheme includes a weighting function reflecting deviation from the crystallographic structure factor amplitudes.[58] Using this approach for a crystal structure cytochrome c553, improvement of the haem geometry was observed, with the QM/MM structure being in better agreement with a more recently determined high resolution cytochrome c553 crystal structure. Whilst it could be argued that high accuracy protein crystal structures are not an absolute requirement for structure-based drug design, such structures are needed for accurate computation of enzyme reaction mechanisms or interaction energies which can guide design efforts. We note however that determining the structure of unresolved regions of significant size (eg. involving flexible protein loops) remains a significant challenge to modelling methods.

In addition to comparison of structure, QM calculations and spectroscopy are complementary in other ways. For example, the crystal structure of the extremophile *T. ferrooxidans*, a blue copper protein with unusually high redox potential and acid stability, was solved at high resolution, both as wild-type and a M148Q mutant[59] (Figure 2). Using this structure, the relative redox potential was calculated using the BP86 density functional to model the copper-containing active site and the AMBER force field for the surrounding protein. The predicted value (-101 mV) agreed very well with experiment (-117 mV), indicating that the M148Q mutant had a lower redox potential. Furthermore, the DFT calculations

produced Cu-ligand geometries in good agreement with bond distances and angles obtained from EXAFS, both for the oxidised and reduced forms of the protein[59].

COVALENT PROTEIN-LIGAND INTERACTIONS

QM calculations provide electronic information unavailable from classical force field methods, enabling the prediction of properties such as redox potential as well as the study of covalent bond formation and scission. Consequently, QM/MM methods are widely employed to characterise enzyme reaction mechanisms,[60] through calculation of the substrate/enzyme energetic profile along suitably chosen reaction coordinates. In this way, it is possible to identify transition states, providing inspiration for design of transition state analogues as potential therapeutic inhibitors. Similarly, the formation of covalent intermediates of suicide inhibitors in complex with their target enzyme can be studied using computational methods; the energetic and structural information can subsequently be used to guide inhibitor design and optimisation.

The latter approach was used in a QM/MM study of adenosine deaminase inhibitors[61]. At the B3LYP/6-31G* level of theory, a quantitative correlation was predicted between the structure of a set of five adenosine deaminase inhibitors and their associated enzyme activity. Albeit a small data set, the QM/MM energy difference between the Michaelis complex of the inhibitor and the corresponding hydroxylated intermediate correlated to IC_{50} values (the half-maximal inhibitory concentration) with a coefficient of 0.88 (Figure 3). It was found that the binding energies of the substrate or the intermediate alone did not correlate well, but that the potency was determined by the stability of the bound intermediate relative to that of the substrate; this finding was largely explained as a result of changes in hydrogen bonding to the deaminase.

In a similar study, the binding and reactivity of seven human neutrophil elastase inhibitors was examined *via* QM/MM methods[62]. These peptidyl α -ketoheterocyclic inhibitors irreversibly form a tetrahedral intermediate as their mode of action. The study concluded that the activity of the least

potent inhibitors, with only weakly electron withdrawing heterocyclic rings and rather endothermic reaction energies, could be explained by their calculated binding energy. Conversely, for the more potent inhibitors, with highly electron withdrawing heterocyclic rings, their activity correlated with the stability of the intermediate. These studies indicate that hybrid QM/MM methods can be usefully employed in explaining enzyme inhibition mechanisms where a transition state analogue can form.

We also note that studies using QM levels of theory on model receptors can also provide valuable insights for drug design. Inhibitors of histone deacetylase can directly chelate to the Zn ion of the enzyme's catalytic core. A small library of potential Zn-binding inhibitor groups, containing mono- and bi-dentate chelators, was assessed for their ability to bind to a model of the Zn-containing catalytic core[63], at the B3LYP/6-31G* level of theory. A good correlation (R^2 of 0.74) was obtained between the QM interaction energy and the experimentally determined pKi (the -log of enzyme inhibition constant, K_i) for the Zn-binding molecular fragments.

NON-COVALENT PROTEIN-LIGAND INTERACTIONS.

Interest continues to grow in the application of QM and QM/MM approaches to analysis and prediction of non-covalent interactions[64]. Indeed calculation of the bound pose of a small molecule ligand to the active site of its target macromolecule forms the basis of receptor-based computational drug design. This molecular docking can be resolved into two steps: (a) geometric sampling of potential ligand/protein binding modes (sometimes itself referred to as “docking”) and (b) scoring, usually using an equation and specific parameters to estimate a ligand's binding affinity. Scoring gives the relative rank of the docked compounds according to their predicted binding affinity. Sometimes a second round of scoring of predicted protein-ligand poses is performed with a more accurate scoring function. Therefore docking algorithms are used to identify different binding modes of ligand in the protein active site and scoring functions are used to distinguish the correct binding mode from all alternative modes. Scoring functions also rank the results of virtual screening according to the binding affinities of different ligands.

Docking has been intensively researched for over twenty years[65] and most current docking programs can achieve an average accuracy of about 1.5 to 2 Å in RMSD with success rates of 70 - 80% in predicting known protein-ligand binding modes[66,67,68]. However, it is generally agreed that further improvements are mainly limited by the imperfections of scoring functions. Wang *et al.*[69] compared the binding affinities of 100 protein-ligand complexes calculated from 11 scoring functions with experimentally determined binding affinities. The result shows that only four of them were able to give correlation coefficients over 0.50. Similarly, a study by Brooks *et al.*[70] assessed nine scoring functions using a set of 189 protein-ligand complexes. The result shows that five of the scoring functions perform well in discriminating near-native from misdocked structures and their recognition rate was between 80 - 90%. However, only one of the scoring functions achieved even a moderate correlation (R^2 of 0.51) between the calculated binding score and the experimentally determined binding energies.

Therefore, improvements in scoring functions are required to avoid incorrect prediction of binding energies and a high number of false positives and negatives during virtual screening. Scoring functions suffer from a range of deficiencies. The treatment of binding entropies and solvation effects are major issues. Some problems arise from the way in which many scoring functions are parameterised (eg too narrow a training set or the presence of experimental errors in the data). Another deficiency of certain scoring functions is the neglect of cation- π interactions and aromatic rings interactions. Both interactions are found to be common in protein structure and are important at the surface of proteins. For example, a cationic sidechain such as Lys or Arg can interact with an aromatic sidechain with significant attractive energy.[71] Aromatic ring interactions have also been shown to occur between aromatic sidechains [72], but most current empirical scoring functions do not model these interactions directly.

The presence of polarization effects in protein-ligand interactions has been recognised.[73] However, it is generally not explicitly included in scoring functions. Hensen *et al.*[73,74] assessed the polarization effects between HIV-1 protease and three potent inhibitors. Their results demonstrated

that the polarization energy could contribute up to one-third of the total electrostatic interaction. It seems not unreasonable that implementation of explicit polarization effects in scoring functions has the potential to increase accuracy. Moreover, metal-containing systems increase the complexity of protein-ligand interactions, through partial covalent bonding and unusual hybridization states. We note that such systems include situations where the metal exists in the ligand rather than the receptor.[75] These complexities are difficult to predict using common scoring functions and can lead to inaccurate ranking.

QM approaches.

Consequently, the use of QM in scoring non-covalent protein-ligand interactions is being increasingly explored. A pragmatic advantage of a QM scoring function, besides potentially improved accuracy over MM, is that there is no requirement for parameterization of new ligands. However, the computational cost for high level QM methods has limited its application in docking and scoring. Therefore, QM-based scoring functions generally focus on semiempirical potentials only.

In a trial study, Villar *et al.*[76] compared AM1 and PM3 calculated geometries and binding enthalpies of a set of 16 small molecules, representative of typical protein-ligand interactions, with higher levels of QM theory, namely B3LYP/6-31G*//B3LYP/6-311+G(2d,p) and MP2/6-31G*. The set of small molecules contained ammonium, carboxylate ions, alkanes, aromatic systems and hydrogen bonding donor and acceptor groups. The results demonstrated that the interaction enthalpies calculated by AM1 had a reasonable correlation with those from *ab initio* MP2 calculations. However PM3 showed poor correlation.[76] In terms of geometry prediction, both showed poor performance, particularly in systems with aromatic interactions. The conclusion of the study was that the semi-empirical AM1 Hamiltonian may serve as a scoring function to rank relative binding affinities of different ligands. However, it is likely to be less valuable in a docking algorithm, i.e. in predicting binding geometries.

Using a linear scaling QM approach, Merz *et al.*[52] has compared the performance of their semi-empirical QM-based scoring functions, QMScore and TotalScore, in predicting binding affinity with eleven other scoring functions for a set of 56 protein-ligand complexes from the data set of Wang *et al.*[69]. The eleven scoring functions compared included four from the LigFit module in Cerius2 (LigScore, PLP, PMF and LUDI);[77] four from SYBYL version 6.8 (F-Score, G-Score, D-Score and ChemScore);[78] and DrugScore,[79,80] X-Score[81] and Autodock[82]. The functions fall into the three categories: (a) force field-based (AutoDock, G-Score and D-Score); (b) empirical (LigScore, PLP, LUDI, F-Score, ChemScore and X-Score); and (c) knowledge-based (PMF and DrugScore). The study demonstrated a slightly improved accuracy of semiempirical QM-based scoring functions in predicting binding affinities over the other methods. The study also highlighted the ability of QM-based scoring functions to discriminate native from decoy poses.

In addition to linear scaling semiempirical QM approaches, more recently a linear scaling density functional theory approach has been applied to the study of protein-inhibitor interactions. A PBE/TZP molecular dynamics study of the complexes of five inhibitors with the cyclin-dependent kinase CDK2[83] concluded that polarization effects resulting from the specific hydration patterns in the ATP pockets of CDK2 contributed to the relative potency of its inhibitors.

A particularly interesting approach adopted by Peters and Merz[84] was to apply QM calculations of protein-ligand interactions in developing receptor-based QSARs. The Comparative Binding Energy Analysis (COMBINE) approach[85] was integrated with the semi-empirical QM method, pairwise energy decomposition (PWD), to produce the SE-COMBINE approach. The method permits the investigation of the gain or loss of interaction energy upon fragment substitution. For a set of 88 benzamidine inhibitors and trypsin, this new method proved competitive when compared to other non-receptor-based QSARs.

QM/MM approaches.

Whilst QM-based scoring functions, such as QMScore discussed above, have been reported to show promise in predicting binding affinity and binding mode for protein-ligand interactions,[52,86] a QM-only approach is currently considered to be too slow to be practical in docking due to the large size of the receptor. Thus, a combined QM/MM approach is a potential alternative method for estimating protein-ligand interaction energies and can provide further insight for developing other scoring functions.[87,88] With sufficient computational resources, the accuracy of QM/MM approaches can be stepwise improved from semiempirical QM methods to *ab initio* QM or density functional theory levels. Additional corrections for accuracy could also include a term for describing dispersive interactions, important in π - π stacking and to some extent in hydrogen-bonding interactions.[42]

Indeed, there are an ever-growing number of QM/MM studies used to successfully gain insights into protein-ligand interactions, including the prediction of active site structure, ligand binding mode and interaction energies calculations[34,73,87,89,90,91]. We discuss some examples below.

Grater *et al.*[88] used a combined AM1/CHARMM potential with a Poisson-Boltzmann/surface area (PBSA) term to calculate the binding energies of 47 benzamidine ligands to trypsin. The results were then compared to experimental binding energies and binding energies derived from empirical scoring functions. The final correlation coefficient for the QM/MM-PB/SA method relative to experiment was 0.6, slightly higher than that of the empirical scoring functions (0.5).

Friesner[74] used a QM/MM model to calculate the atomic charges of the ligand in the protein environment. The ligand was treated at the B3LYP/6-31G* level and the protein by the MM force field, OPLS-AA. The use of these QM-derived ligand atomic charges in docking showed significant improvements over MM-based charges in detection of the correct binding mode of the inhibitor.

A straightforward approach to combining docking with QM/MM methods was adopted by Beierlein *et al.*[74,87]. In their method, ligands were docked to proteins using Autodock, via a Lamarkian genetic algorithm and AMBER-based empirical scoring function. Resulting structures were then optimised at

the AM1/MM level of theory. The method was tested for redocking of crystallographic complexes of β -trypsin, HIV-1 protease, thrombin, carboxypeptidase A and dihydrofolate reductase. Broadly similar results were observed with and without the QM/MM refinement. However, the optimisation step permitted protein as well as ligand flexibility and was able to reproduce induced fit of the flap regions of HIV-1 protease on ligand binding. This was attributed in part to the enhanced electrostatic receptor-inhibitor interaction afforded by the QM/MM potential relative to comparative MM receptor-inhibitor optimisations.

Fong et al.[92] examined the ability of QM/MM, MM and empirical scoring functions to discriminate native from non-native poses in six HIV-1 protease complexes. This was based on large numbers of geometric decoys for each ligand/protein complex, obtained from the Ligand Protein Data Bank (LPDB).[70] Geometric decoys, which are misdocked poses of the ligand in the active site, were generated using a replica-based Langevin molecular dynamics protocol, in conjunction with an NOE-like restraint in order to distance the ligand from its native pose.[93] The QM/MM scoring functions examined all displayed a relatively smooth funnel-shaped topology of the binding surface required for efficient docking; ie. misdocked solutions are ranked higher in energy than native-like solutions (Figure 4).

The most discriminatory QM/MM scoring function for the six HIV-1 protease inhibitors was found to be the solvent-corrected HF/6-31G*:AMBER model, exceeding the performance of the ChemScore and GoldScore empirical scoring functions. This QM/MM function, however, performed comparably to a solvent-corrected MM scoring function. Indeed, the polarization component of the QM/MM interaction energy, although comprising 12-18% of the total protein-ligand electrostatic interaction for the native pose, appeared relatively insensitive to changes in RMSD from the native pose. For one inhibitor, XK216, the polarization component actually served to impede accurate discrimination of native from non-native poses. This stabilization of certain decoy poses by electric polarization may be physically correct but distorts the funnel-shaped topology of the binding surface. Correspondingly, the MM approach, which omits explicit ligand polarization, performed similarly to the Hartree-Fock

QM/MM model. Semiempirical methods, which may underestimate the polarization contributions of decoys relative to native, exhibited lower discriminative power.

The accuracy and practicability of the combined QM/MM method in docking has shown some promising recent results and presents a good avenue for improving docking and scoring of protein-ligand complexes but further careful refinement and assessment of methodologies is clearly required.

CONCLUSIONS

QM approaches are being increasingly brought to bear on problems of pharmaceutical interest. In particular, recent developments in more accurate semiempirical QM approaches, and methods for their application to large molecules, appear promising for the future integration of accurate modelling of intermolecular potential energies with the computational expediency required to assess the binding of large ligand data sets, or to describe molecular flexibility and obtain thermodynamic properties via statistical mechanics.

QM approaches afford various advantages, ranging from calculation of spectroscopic properties, through to improved structure determination, and more accurate modelling of chemical systems where MM parametrization strategies would be expected to struggle, eg. metalloprotein complexes, or where bond making and breaking occurs (covalent inhibitors). Whilst algorithmic and hardware developments are beginning to enable such QM and QM/MM approaches, in many cases the outweighing of additional benefits versus the cost of employing a QM approach in drug discovery over classical or empirical methods remains to be clearly demonstrated.

References

1. Vincent MA, Hillier IH, Morgado CA, Burton NA, Shan X. The structure and binding energies of the van der Waals complexes of Ar and N₂ with phenol and its cation, studied by high level ab initio and density functional theory calculations. *J. Chem. Phys.* 2008;128:044313-044317.

2. Werner HJ, Manby FR, Knowles PJ. Fast linear scaling second-order Moller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* 2003;118:8149-8160.
3. Cramer, C. J., *Essentials of Computational Chemistry*, Wiley, Chichester, 2002
4. Zhao Y, Truhlar DG. Applications and validations of the Minnesota density functionals. *Chem.Phys.Lett.* 2011;502:1-13.
5. Mostofi AA, Haynes PD, Skylaris C-K, Payne MC. ONETEP: linear-scaling density-functional theory with plane-waves. *Mol.Sim.* 2007;33:551-555.
6. Field MJ, Bash PA, Karplus M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J.Comput.Chem.* 1990;11:700-
7. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J.Mol.Biol.* 1976;103:227-249.
8. Lundberg, M. and Morokuma, K. in *Multi-scale Quantum Models for Biocatalysis* (Ed. York, D. M. and Lee, T.-S.), Springer, New York, 2009, pp.21-55
9. Hillier IH. Chemical reactivity studied by hybrid QM/MM methods. *Journal of Molecular Structure: THEOCHEM* 1999;463:45-52.
10. Lin H, Truhlar D. QM/MM: What have we learned, where are we, and where do we go from here? *Theor Chem Acc* 2007;117:185-199.
11. Senn HM, Thiel W. QM/MM Methods for Biomolecular Systems. *Angew.Chem.Int.Ed.* 2009;48:1198-1229.
12. van der Kamp MW, Zurek J, Manby FR, Harvey JN, Mulholland AJ. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* 2010;114:11303-11314.
13. Lyne PD, Mulholland AJ, Richards WG. Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction. *J. Am. Chem. Soc.* 1995;117:11345-11350.
14. Davidson MM, Gould IR, Hillier IH. The mechanism of the catalysis of the Claisen rearrangement of chorismate to prephenate by the chorismate mutase from *Bacillus subtilis*. A molecular mechanics and hybrid quantum mechanical/molecular mechanical study. *J.Chem.Soc., Perkin Trans.2* 1996;525-532.
15. Wiest O, Montiel DC, Houk KN. Quantum Mechanical Methods and the Interpretation and Prediction of Pericyclic Reaction Mechanisms. *J.Phys.Chem.A* 1997;101:8378-8388.
16. Hall RJ, Hindle SA, Burton NA, Hillier IH. Aspects of hybrid QM/MM calculations: The treatment of the QM/MM interface region and geometry optimization

- with an application to chorismate mutase. *J. Comput. Chem.* 2000;21:1433-1441.
17. Thery V, Rinaldi D, Rivail JL, Maigret B, Ferenczy GG. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *J. Comput. Chem.* 1994;15:269-282.
 18. Ponder, Jay W. and Case, David A. in *Advances in Protein Chemistry Protein Simulations* (Ed. Valerie, Daggett), Academic Press 2003, pp.27-85
 19. Liljefors, T., Gundertofte, K., Norrby, P.-O., and Pettersson, I. in *Computational Medicinal Chemistry for Drug Discovery* (Ed. Bultnick, P., DeWinter, H., Wilfried, L., and Tollenaere, J. P.), Dekker, New York, 2004, pp.1-28
 20. Lu Z, Zhang Y. Interfacing ab Initio Quantum Mechanical Method with Classical Drude Oscillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *J. Chem. Theory Comput.* 2008;4:1237-1248.
 21. McNamara, J. P. and Hillier, I. H. in *Multi-scale Quantum Models for Biocatalysis* (Ed. York, D. M. and Lee, T.-S.), Springer, New York, 2009,
 22. Pople JA, Segal GA. Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *J. Chem. Phys.* 1965;43:S136-S151.
 23. Pople JA, Beveridge DL, Dobosh PA. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *J. Chem. Phys.* 1967;47:2026-2033.
 24. Dewar MJS, Thiel W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* 1977;99:4899-4907.
 25. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 1985;107:3902-3909.
 26. Stewart JJP. Optimisation of parameters for semi-empirical methods. 2. Applications. *J. Comput. Chem.* 1989;10:221-264.
 27. Stewart JJP. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Modell.* 2007;13:1173-1213.
 28. Stewart J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model* 2013;19:1-32.
 29. Korth M, Thiel W. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods Are Almost As Accurate and Robust As DFT-GGA Methods for Organic Molecules. *J. Chem. Theory Comput.* 2011;7:2929-2936.

30. Rocha GB, Freire RO, Simas AM, Stewart JJP. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* 2006;27:1101-1111.
31. Stewart J. Application of the PM6 method to modeling proteins. *J Mol Model* 2009;15:765-805.
32. Rossi I, Truhlar DG. Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem.Phys.Lett.* 1995;233:231-236.
33. McNamara JP, Muslim AM, Abdel-Aal H, Wang H, Mohr M, Hillier IH, Bryce RA. Towards a quantum mechanical force field for carbohydrates: a reparametrized semi-empirical MO approach. *Chem.Phys.Lett.* 2004;394:429-436.
34. Tresadern G, Faulder PF, Gleeson MP, Tai Z, MacKenzie G, Burton NA, Hillier IH. Recent advances in quantum mechanical/molecular mechanical calculations of enzyme catalysis: hydrogen tunnelling in liver alcohol dehydrogenase and inhibition of elastase by alpha-ketoheterocycles. *Theor. Chem. Acc.* 2003;109:108-117.
35. Sundararajan M, Hillier IH, Burton NA. Structure and Redox Properties of the Protein, Rubredoxin, and Its Ligand and Metal Mutants Studied by Electronic Structure Calculation. *J.Phys.Chem.A* 2005;110:785-790.
36. Hunter CA, Singh J, Thornton JM. Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* 1991;218:837-846.
37. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 1985;229:23-28.
38. Dwek RA. Glycobiology: Toward Understanding the Function of Sugars. *Chem.Rev.* 1996;96:683-720.
39. Grimme S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* 2004;25:1463-1473.
40. Morgado C, Vincent MA, Hillier IH, Shan X. Can the DFT-D method describe the full range of noncovalent interactions found in large biomolecules? *Phys. Chem. Chem. Phys.* 2007;9:448-451.
41. McNamara JP, Hillier IH. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. *Phys.Chem.Chem.Phys.* 2007;9:2362-2370.
42. Sharma R, McNamara JP, Raju RK, Vincent MA, Hillier IH, Morgado CA. The interaction of carbohydrates and amino acids with aromatic systems studied by density functional and semi-empirical molecular orbital calculations with dispersion corrections. *Phys.Chem.Chem.Phys.* 2008;10:2767-2774.

43. Raju RK, Hillier IH, Burton NA, Vincent MA, Doudou S, Bryce RA. The effects of perfluorination on carbohydrate- π interactions: computational studies of the interaction of benzene and hexafluorobenzene with fucose and cyclodextrin. *Phys.Chem.Chem.Phys.* 2010;12:7959-7967.
44. Rezac J, Fanfrik J, Salahub D, Hobza P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* 2009;5:1749-1760.
45. Tuttle T, Thiel W. OMx-D: Semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys.Chem.Chem.Phys.* 2008;10:2159-2166.
46. Richards, W. G., *Quantum Pharmacology*, Butterworths & Co. Ltd., 1977
47. Murray JS, Ranganathan S, Politzer P. Correlations between the solvent hydrogen bond acceptor parameter, β , and the calculated molecular electrostatic potential. *J.Org.Chem.* 1991;56:3734-3737.
48. Chalk AJ, Beck B, Clark T. A quantum mechanical/neural net model for boiling points with error estimation. *J.Chem.Inf.Comp.Sci.* 2001;41:1053-1059.
49. Khandogin J, Hu AG, York DM. Electronic structure properties of solvated biomolecules: A quantum approach for macromolecular characterization. *J.Comput.Chem.* 2000;21:1562-1571.
50. Khandogin J, York DM. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins: Struct.Funct.Bioinf.* 2004;56:724-737.
51. Rajamani R, Reynolds CH. Modeling the protonation states of the catalytic aspartates in beta-secretase. *J.Med.Chem.* 2004;47:5159-5166.
52. Raha K, Merz KM. 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.
53. Li H, Hains AW, Everts JE, Robertson AD, Jensen JH. The prediction of protein $pK(a)$'s using QM/MM: The $pK(a)$ of lysine 55 in Turkey ovomucoid third domain. *J.Phys.Chem.B* 2002;106:3486-3494.
54. Alonso H, Cummins PL, Gready JE. Methyltetrahydrofolate:corrinoid/iron-sulfur Protein Methyltransferase (MeTr): Protonation State of the Ligand and Active-Site Residues. *J.Phys.Chem.B* 2009;113:14787-14796.
55. Nilsson K, Ryde U. Protonation status of metal-bound ligands can be determined by quantum refinement. *J.Inorg.Chem.* 2004;98:1539-1546.
56. Todorov NP, Monthoux PH, Alberts IL. The influence of variations of ligand protonation state and tautomerism on protein-ligand recognition and binding energy landscape. *J.Chem.Inf.Model.* 2006;46:1134-1142.

57. Gleeson MP, Gleeson D. QM/MM Calculations in Drug Discovery: A Useful Method for Studying Binding Phenomena? *J.Chem.Inf.Modell.* 2009;49:670-677.
58. Ryde U, Nilsson K. Quantum chemistry can locally improve protein crystal structures. *J.Am.Chem.Soc.* 2003;47:14232-14233.
59. Barrett ML, Harvey I, Sundararajan M, Surendran R, Hall JF, Ellis MJ, Hough MA, Strange RW, Hillier IH, Hasnain SS. Atomic resolution crystal structures, EXAFS, and quantum chemical studies of rusticyanin and its two mutants provide insight into its unusual properties. *Biochemistry* 2006;45:2927-2939.
60. Burton NA, Harrison MJ, Hart JC, Hillier IH, Sheppard DW. Prediction of the mechanisms of enzyme-catalysed reactions using hybrid quantum mechanical molecular mechanical methods. *Faraday Discuss.* 1998;110:463-475.
61. Gleeson MP, Burton NA, Hillier IH. Prediction of the potency of inhibitors of adenosine deaminase by QM/MM calculations. *Chem.Commun.* 2003;2180-2181.
62. Gleeson MP, Hillier IH, Burton NA. Theoretical analysis of peptidyl alpha-ketoheterocyclic inhibitors of human neutrophil elastase: Insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. *Org.Biomol.Chem.* 2004;2:2275-2280.
63. Vanommeslaeghe K, Loverix S, Geerlings P, Tourwe D. DFT-based ranking of zinc-binding groups in histone deacetylase inhibitors. *Bioorg.Med.Chem.* 2005;13:6070-6082.
64. Raha K, Peters MB, Wang B, Yu N, WollaCott AM, Westerhoff LM, Merz KM. The role of quantum mechanics in structure-based drug design. *Drug Discov.Today* 2007;12:725-731.
65. Leach AR, Shoichet BK, Peishoff CE. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J.Med.Chem.* 2006;49:5851-5855.
66. Westhead DR, Clark DE, Murray CW. A comparison of heuristic search algorithms for molecular docking. *J.Comput.-Aided Mol.Des.* 1997;11:209-228.
67. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct.Funct.Bioinf.* 2004;57:225-242.
68. Bursulaya BD, Totrov M, Abagyan R, Brooks CL. Comparative study of several algorithms for flexible ligand docking. *J.Comput.-Aided Mol.Des.* 2003;17:755-763.
69. Wang RX, Lu YP, Wang SM. Comparative evaluation of 11 scoring functions for molecular docking. *J.Med.Chem.* 2003;46:2287-2303.
70. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing scoring functions for protein-ligand interactions. *J.Med.Chem.* 2004;47:3032-3047.

71. Gallivan JP, Dougherty DA. Cation- π interactions in structural biology. *Proc.Natl.Acad.Sci.USA* 1999;96:9459-9464.
72. McGaughey GB, Gagne M, Rappe AK. π -stacking interactions - Alive and well in proteins. *J.Biol.Chem.* 1998;273:15458-15463.
73. Hensen C, Hermann JC, Nam KH, Ma SH, Gao JL, Holtje HD. A combined QM/MM approach to protein-ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J.Med.Chem.* 2004;47:6673-6680.
74. Friesner RA. Modeling polarization in proteins and protein-ligand complexes: methods and preliminary results. *Adv.Prot.Chem.* 2006;72:79-104.
75. Ciancetta A, Genheden S, Ryde U. A QM/MM study of the binding of RAPTA ligands to cathepsin B. *J. Comput.-Aided Des.* 2011;25:729-742.
76. Villar R, Gil MJ, Garcia JJ, 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.
77. Accelrys, Inc., Cerius2 Modeling Environment, San Diego, CA, USA., 2006
78. Tripos Inc., SYBYL 6.8, St. Louis, MO, USA., 2003
79. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J.Mol.Biol.* 2000;295:337-356.
80. Velec HFG, Gohlke H, Klebe G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J.Med.Chem.* 2005;48:6296-6303.
81. Wang RX, Lai LH, Wang SM. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J.Comput.-Aided Mol.Des.* 2002;16:11-26.
82. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 1998;19:1639-1662.
83. Heady L, Fernandez-Serra M, Mancera RL, Joyce S, Venkitaraman AR, Artacho E, Skylaris CK, Ciacchi LC, Payne MC. Novel structural features of CDK inhibition revealed by an ab initio computational method combined with dynamic simulations. *J. Med. Chem.* 2006;49:5141-5153.
84. Peters MB, Merz KM. Semiempirical comparative binding energy analysis (SE-COMBINE) of a series of trypsin inhibitors. *J. Chem. Theory Comput.* 2006;2:383-399.
85. Ortiz AR, Pisabarro MT, Gago F, Wade RC. Prediction of Drug-Binding Affinities by Comparative Binding-Energy Analysis. *J. Med. Chem.* 1995;38:2681-2691.

86. Peters MB, Raha K, Merz KM. Quantum mechanics in structure-based drug design. *Curr.Opin.Drug Discovery Dev.* 2006;9:370-379.
87. Beierlein F, Lanig H, Schurer G, Horn AHC, Clark T. Quantum mechanical/molecular mechanical (QM/MM) docking: an evaluation for known test systems. *Mol.Phys.* 2003;101:2469-2480.
88. Grater F, Schwarzl SM, Dejaegere A, Fischer S, Smith JC. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J.Phys.Chem.B* 2005;109:10474-10483.
89. Cho AE, Guallar V, Berne BJ, Friesner R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* 2005;26:915-931.
90. Rungrotmongkol T, Mulholland AJ, Hannongbua S. Active site dynamics and combined quantum mechanic s/molecul ar mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. *J. Mol. Graph. Modell.* 2007;26:1-13.
91. Morgado CA, Hillier IH, Burton NA, McDouall JJW. A QM/MM study of fluoroaromatic interactions at the binding site of carbonic anhydrase II, using a DFT method corrected for dispersive interactions. *Phys.Chem.Chem.Phys.* 2008;10:2706-2714.
92. Fong P, McNamara JP, Hillier IH, Bryce RA. Assessment of QM/MM Scoring Functions for Molecular Docking to HIV-1 Protease. *J.Chem.Inf.Modell.* 2009;49:913-924.
93. Roche O, Kiyama R, Brooks CL. Ligand–Protein DataBase: Linking Protein–Ligand Complex Structures to Binding Data. *Journal of Medicinal Chemistry* 2001;44:3592-3598.

FIGURE CAPTIONS

Figure 1 QM/MM model to study adenosine deaminase inhibitors. Residues included in the QM region (white), Glu 217, His 238 and the catalytic water molecule. The rest of the enzyme is represented by an MM potential (shaded). Link atoms are represented by circled hydrogen atoms.

Figure 2 Active site electron density for H143M rusticyanin.

Figure 3 Plot of energy of intermediate relative to that of the substrate for five adenosine deaminase inhibitors (QM/MM B3LYP/6-31G(d)//HF/3-21G) *versus* $-\ln \text{IC}_{50}$. $R^2 = 0.87$. The same trend is observed at the QM/MM HF/3-21G//HF/3-21G level, with an R^2 of 0.88.

Figure 4 Interaction energies (ΔE_{inter}) of six HIV-1 proteases inhibitors native poses and their geometric decoys. From left to right columns, ΔE_{inter} calculated at HF/6-31G*:AMBER and AM1d:AMBER levels of theory. Outlying decoy score for **1** is circled.

Figure 1

Figure 2

Figure 3

Figure 4