EXPERT OPINION

1. Introduction
2. Development of quantum chemical approaches for biomolecular modeling
3. Application of quantum chemical approaches to computational medicinal chemistry
4. Expert opinion

informa
healthcare

The application of quantum mechanics in structure-based drug design

Daniel Mucs & Richard A Bryce[†]

University of Manchester, School of Pharmacy and Pharmaceutical Sciences, Manchester, UK

Introduction: Computational chemistry has become an established and valuable component in structure-based drug design. However the chemical complexity of many ligands and active sites challenges the accuracy of the empirical potentials commonly used to describe these systems. Consequently, there is a growing interest in utilizing electronic structure methods for addressing problems in protein–ligand recognition.

Areas covered: In this review, the authors discuss recent progress in the development and application of quantum chemical approaches to modeling protein–ligand interactions. The authors specifically consider the development of quantum mechanics (QM) approaches for studying large molecular systems pertinent to biology, focusing on protein–ligand docking, protein–ligand binding affinities and ligand strain on binding.

Expert opinion: Although computation of binding energies remains a challenging and evolving area, current QM methods can underpin improved docking approaches and offer detailed insights into ligand strain and into the nature and relative strengths of complex active site interactions. The authors envisage that QM will become an increasingly routine and valued tool of the computational medicinal chemist.

Keywords: binding free energies, ligand strain, linear scaling, QM/MM docking, quantum mechanics, semiempirical quantum mechanics, structure-based drug design

Expert Opin. Drug Discov. (2013) 8(3):263–276

1. Introduction

The design and development of new medicines is an expensive and complex enterprise, requiring the interplay of numerous scientific facets via integrated multidisciplinary teams. The role of the computational chemist within this team has become firmly established. Metrics point to the measurable impact of computer-aided drug design across the discovery time line [1], from target identification and hit discovery to hit-to-lead optimization and safety profiling. More than ever, computational medicinal chemists have a wide array of *in silico* tools, chemical databases and informatics at their disposal. The arsenal of tools available include advanced molecular graphics, pipelining tools, substructure searching, tools for deriving quantitative structure–activity relationships, pharmacophore-based analysis, statistical packages and receptor–ligand docking tools.

In a recent analysis of most valued computational tools in one major pharmaceutical company, among the lowest ranked was quantum mechanics (QM), alongside molecular dynamics (MD) [2]. This is interesting to note, as QM featured among the earliest tools within the field of computer-aided drug design, with its application to calculate descriptors of polarity for use in quantitative structure–activity relationships (QSAR). These descriptors included atom-centered partial charges derived from the wavefunction of the molecule and energies of its highest occupied and lowest unoccupied molecular orbitals. This work continues today with more

Article highlights.

- Recent advances in theory, methodology, software and hardware, present increased opportunity to apply quantum chemistry to systems relevant to medicinal chemistry.
- Ligand poses predicted from docking to receptors, such as metalloproteins, have been shown to more closely resemble experiments when partial atomic charges are derived from QM or QM/MM calculations.
- The use of QM methods in docking is anticipated to increase. Although not computationally tractable currently, ultimately this could evolve to the point where direct QM or QM/MM calculations at each ligand-receptor pose sampled become a possibility, as well as incorporation of polarizability of the entire protein-ligand system.
- The use of QM and QM/MM approaches in computation of protein-ligand binding affinities has met with mixed success, and often depends on details of method and receptor active site; improvements are only likely in concert with developments in modeling of the effects of solvent, entropy and target flexibility.
- QM methods have proved valuable in accounting for ligand strain on binding and in QM/MM refinement of protein-ligand crystal structures; the latter approach appears to be of most benefit for low resolution X-ray structures, where an incorrectly assigned ligand structure due to its MM force field is more likely.
- QM methods have considerable potential for accurate resolution of protein-ligand interaction energies into residue and functional group contributions, particularly for systems where MM parameterization is not straightforward.

This box summarizes key points contained in the article.

sophisticated QSAR approaches, based on atoms-in-molecules (AIM) [3] and other topological indices derived from the wavefunction [4].

QM historically has also been used within medicinal chemistry to understand and potentially mitigate the energetic strain in a ligand that results from its binding to a receptor. This played a role in the development of dorzolamide in the early 1990s, a carbonic anhydrase II inhibitor to treat glaucoma [5]. The lead compound during the development of dorzolamide had (R)- and (S)-enantiomers. The crystallographic poses of these two forms explained the greater potency of the (S) form, in terms of internal conformation of the N-S-C-S moiety and isobutyl side-chain conformations; these were supported by QM calculations at the HF/6-31G* level of theory. Further QM calculations suggested that a 6-methyl group would lock the ring conformation of the ligand; this was reflected by experiment and is a feature retained in the final marketed compound.

The potential of QM to furnish informative effective QSAR descriptors and to assess a ligand's preferred 3D conformation were features highlighted by Graham Richards in his 1977 book, *Quantum Pharmacology* [6]. The subsequent

view outlined in the book was where theoretical chemistry, and QM approaches in particular, would develop to the point that they would play a role in guiding the selection of compounds for the medicinal chemist to synthesize and assay. Indeed, from earliest days of QM, Schrodinger, Dirac and others had had a sense of the potential of QM to impact chemistry and biology. Exponents such as Pauling, Löwdin and the Pullmans subsequently drove the application of QM to problems in biology and biochemistry; for example, see a 1969 review by Löwdin which includes early work on nucleotide base pair hydrogen bonds [7]. The field has since evolved to a point which could perhaps barely have been conceived originally. Nevertheless, while computational chemistry has become an invaluable spoke within the wheel of drug discovery, in this the significance of the methods of QM may be less apparent. However, with increased compute power now available to exhaustively test empirical potentials, via long time-scale simulation of equilibrium properties and through very high-level quantum chemical calculations, deficiencies in the ability of the classical potentials underlying many biomolecular modeling approaches are being exposed [8-10]. This is in part responsible for prompting a greater focus on exploring QM-based approaches to provide, directly or indirectly, more accurate descriptions of protein-ligand interactions.

In this paper, we briefly review the development of QM approaches for studying large molecular systems pertinent to biology; we then focus on recent applications of QM-based methods, principally for the prediction of protein-ligand geometries (docking); protein-ligand affinities (scoring) and changes of ligand internal energetics on binding (ligand strain).

2. Development of quantum chemical approaches for biomolecular modeling

Over the last two decades, there has been a step change in the computational power and accessibility of quantum chemical software programs available to the computational chemist, enabling the application of QM schemes to ever larger and more complex systems. It remains true however that the highest levels of QM can only be applied to reasonably modest system sizes (10s of atoms). Among the most computationally intensive of these are the correlated wavefunction-based methods, CCSD(T) and MP2 levels of theory, which scale approximately as N^7 and N^5 , respectively, where N is the number of basis functions. CCSD(T) is considered chemically accurate [11], that is, able to provide relative energies accurate to within 1 kcal/mol. CCSD(T) has been fruitfully applied in providing biological relevant benchmarking for testing the accuracy of lower level methods, both QM-based methods and those based on molecular mechanics (MM). These sets typically contain hydrogen bonding and dispersive interactions of small bimolecular complexes and include various aromatic-aromatic interaction types [8,12-14].

The Hartree-Fock method offers a more computationally expedient approach but lacks the necessary accuracy to properly

treat the effect of correlated electrons motion. Approaches based on density functional theory (DFT) incorporate the effects of electron correlation, although the ability to accurately describe these effects varies widely between functional. A known failing of density functionals relates to their inadequate treatment of the dispersion interaction. Considerable recent effort has produced functionals which are more capable of modeling these biologically key effects, for example, the functionals which combine non-local correlation with a long-range corrected hybrid density functional [15]. An alternative approach popularized by Grimme [16] is to correct deficiencies in the treatment of dispersion by a functional with the addition of an empirically parameterized term (typically with a r^{-6} dependence). These DFT-D approaches have found widespread application in modeling biological interactions and recent implementations of the approach, such as DFT-D3 [17] and ω B97X-D [18], reproduce higher level calculations rather well.

Concomitant with the maturing of DFT methods there has been a renewed interest in semiempirical quantum chemical approaches. These semiempirical QM Hamiltonians are readily applicable to many thousands of atoms due to a judicious set of approximations. The majority of semiempirical methods are based on the neglect of differential diatomic overlap (NDDO) approximation [19], which ignores a large number of two-electron integrals and assumes experimentally derived parameters for others. In their original form, NDDO methods, such as AM1, PM3 and MNDO, did not handle noncovalent interactions well. More recently, attention has been focused on an alternative to these approaches which is based on a second-order expansion of the DFT total energy; this self-consistent charge density functional tight binding (SCC-DFTB) method [20], however, to some extent is subject to the same problems of DFT functionals, for example, an inability to model dispersion forces.

Correspondingly, there has been considerable interest in improving the accuracy of these computationally expedient semiempirical QM methods. Analogous to DFT-D techniques, these approaches can be empirically corrected to improve their ability to model dispersive contributions, leading to approaches, such as AM1-FS1 [21], OMx-D [22], PM3-D [23] and SCC-DFTB/D [24]. In a similar fashion, deficiencies in their treatment of hydrogen bonding can also be compensated in a number of ways [25,26]. These include additional or modified core terms, as found in the PM3-PIF [27-29] and PDDG/PM3 [30] approaches, or modified parameters for SCC-DFTB [31]. A further example of a semiempirical QM method corrected for improved hydrogen bonding is provided by the PM6-DH methods of Korth *et al.* [32], the most recent of which is the D3H4 implementation. These methods employ an r^{-6} term which is used for the dispersion correction. For the hydrogen bonding correction, donor and acceptor radial and angular terms are employed, as well as a proton transfer term dependent on the position of the hydrogen between the donor and acceptor. Further scaling terms are applied for charged

hydrogen bonding groups or those where water acts as a hydrogen donor. It has been found that PM6-D3H4 can reproduce the interaction energetics of noncovalent complexes in benchmark data sets with errors < 1 kcal/mol [33]. It can be difficult to obtain a generally applicable set of semiempirical QM parameters, a fact first recognized by Rossi and Truhlar, who developed specific reaction parameters for studying chemical reactivity [34]. Subsequently semiempirical approaches have been developed specific to model noncovalent interactions, eg, for carbohydrates: the PM3CARB-1 approach [35] successfully corrected for deficiencies of PM3 in the modeling of ring pucker and, although presenting a fairly stiff pyranose conformational hypersurface, provides reasonable relative free energies between 1C_4 and 4C_1 chair conformations [36].

A number of general strategies can be employed to further extend the size of systems to which quantum chemical calculations can be applied. One such approach is to describe a subregion of interest in electronic detail via QM and couple it to its larger environment modelled at the MM level. This QM/MM approach has been widely employed in the study of enzyme reactivity [37]. A range of QM/MM schemes exist, many of which hinge on the coupling of the QM region to the MM region. Couplings may be purely MM in nature or use link atoms or localized orbitals. For further details on this, the readers are directed to the review of Senn and Thiel [37]. QM/MM approaches have also been generalized to multiple layers via the ONIOM formalism [38]. Analogous to developments in semiempirical Hamiltonians, Wang and Bryce recently reformulated QM/MM interaction terms to improve hydrogen bonding at the QM/MM interface [39]. Using this approach, a corrected PM3/MM* potential showed improved prediction of geometry and interaction energy over unmodified PM3/MM for hydrogen-bonded small molecule complexes typical of biomolecular interactions, without significantly impacting the modeling of other interaction types. Using this potential, more quantitative prediction of interaction energies was also found for HIV-1 protease/inhibitor complexes (Figure 1).

A second approach to increase the reach of QM methods is in the application of fragment molecular orbital (FMO) approaches [40]. FMO schemes fragment the larger molecules to smaller moieties. This can be achieved as isolated monomers, using hybrid projection operators or adaptive frozen orbitals. Nonadditive terms can be recovered by considering combinations of dimers and trimers of the fragments. A similar approach is molecular fractionation of conjugated caps (MFCC) [41] which also has been applied to a range of biologically relevant systems. In this approach, fragments are capped by methyl groups to provide a more realistic context for the monomers. These approaches have utility in calculating the binding energies and their resolution into specific amino acid contributions, for example, via the pair interaction energy (PIE) approaches [42].

A third approach is through linear scaling strategies [43]. These seek to reduce the scaling of computational cost to

Figure 1. A. Comparison of PM3/MM (+) and PM3/MM* (Δ) models with reference HF/6-31G*/MM calculations in predicting total electrostatic interaction energies (kcal/mol) of native and decoy ligand poses with HIV-1 protease. B. Example poses of native ligand (magenta) and decoys (yellow, cyan).

Reproduced from [39] with permission of the American Chemical Society.

order N basis functions. A range of approaches exist within this category, which include the use of localized molecular orbitals, such that elements of the density matrix between distant parts of a large molecule can be neglected. A divide-and-conquer approach partitions the density matrix into blocks that depend on spatial proximity and solves each block independently. Alternatively, linear scaling plane wave approaches have been applied to protein–ligand interactions at the QM level [44] and even to protein–protein interactions [45].

A final approach that we note here arises from utilization of computational architectures to accelerate QM calculations. Distribution of the computational overhead of QM algorithms on parallel CPU-based architectures leads to tractable *ab initio* and density functional QM calculations on thousands of basis functions [46]. Recently, work has focused on the adaptation of QM algorithms to utilization of graphics processing unit (GPU) architectures [47]. GPU-based approaches to perform SCF [48–50], DFT [51] and MP2 [52] calculations have been developed. The GPU architectures have broad appeal from their ready availability at low cost. The challenges of GPU architectures include the relatively small memory available to each processor, which has currently limited QM basis sets to fairly modest sizes [53].

3. Application of quantum chemical approaches to computational medicinal chemistry

A number of exciting developments, as outlined above, are permitting quantum chemical calculations to increasingly relevant biological systems. We now consider the application

of quantum chemical approaches to specific problems in computational medicinal chemistry.

3.1 Protein–ligand docking

A common activity of the computational medicinal chemist is in the prediction of the bound geometry of a ligand – or millions of ligands – to a given receptor binding pocket. In many cases, the widely used knowledge-based, empirical or force field-based scoring functions are capable of providing docked poses with reasonable accuracy, typically with root-mean-square deviations (RMSDs) between a ligand's docked and crystallographic positions of < 2 Å. However, limitations in these scoring functions are being increasingly exposed, particularly as more challenging and electronically complex pockets are being probed (e.g., with a metal or highly polar site present) [54]. Improving the fidelity of modeling these interactions via direct inclusion of a quantum mechanical Hamiltonian is attractive as a more general approach; however, the large number of energy evaluations required in conformational analysis of ligand–protein poses is limiting. Alternatively re-parameterization of scoring functions to improve results has been explored [54,55]. One focus has been to employ more sophisticated MM approaches for scoring which are informed at some level by QM.

MM approaches approximate the electronic charge distribution in various ways, very commonly as atom-centered partial point charges. Typically these point charges are derived from fitting to the QM electrostatic potential (ESP) of the molecule. The established and widely used DOCK [56] and AutoDock [57] programs utilize a force field-based scoring function in their evaluation of receptor–ligand

conformations. Recently, in the derivation of the AutoDock4 scoring function by Wang *et al.* [58], two approaches to obtaining ligand partial charges were explored: the restrained ESP (RESP) approach [59], using the electrostatic potential obtained at the HF/6-31G* level and the AM1-BCC method [60], which uses a bond charge correction (BCC) to produce charges consistent with the HF/6-31G* ESP. For a training set of 147 protein–ligand complexes, the AM1-BCC model achieved a root-mean-square error of 1.64 kcal/mol. Subsequently, for a test set of 1,427 compounds based on the PDBbind database [61,62], an error of 2.2 kcal/mol in free energy was found.

Using this approach, for assessment of binding pose prediction using 100 ligands with associated decoy sets, 87% of ligands were predicted to be in their native pose to within 2 Å (Table 1). This compares well against other scoring functions, including PLP (76%) and PMF (52%). Interestingly, for neutral molecules with small dipole moments, the more sophisticated QM-based charge models produced results comparable to a scoring function based on empirically-derived Gasteiger charges. However, for molecules with larger dipole moments, improved scoring was observed with the QM-derived charge models. We do note, nevertheless, that the regression approach employed in fitting the AutoDock4 scoring function made several approximations, regarding the treatment of solvent, protein flexibility, protonation state and entropy.

In the above work, Wang *et al.* [58] used the QM-derived partial atomic charges of the *isolated* ligands in their docking calculations. It is possible to further incorporate the polarizing effect of the receptor on the ligand. This has been explored by Cho *et al.* [63]. In their work, ligand atomic charges are obtained via a QM/MM calculation of the QM ligand in the MM active site of the receptor, taken from the B3LYP/6-31G* ESP rather than resolved directly from the wavefunction as Mulliken charges. Calculations on X-ray structures indicated generally improved docked poses, with lower RMSD relative to the X-ray complex, and higher, more negative scores, compared to regular force field charges. This idea has recently been adopted and extended in a method for MD simulations of the condensed phase [65]. In this study, partial charges on the solute are periodically obtained from fitting to the QM/MM ESP of the MD configuration, where the ligand is modeling at the QM level, in the presence of the atomic partial point charges of the protein, solvent and ions. Thus, dynamic polarization of the system is permitted over the course of the MD trajectory [64].

Returning to the QM/MM-based docking study of Cho *et al.* [63], they then attempted to unbias their calculations: in a realistic medicinal chemistry environment, there would be no prior knowledge of the geometry of the ligand–receptor complex. This approach, called the QM-polarized ligand docking (QPLD) method, uses force field-based scoring to initially dock the ligand into the receptor; a survival of the fittest approach then selects

QM-derived ligand charge sets from these docked poses: top-ranked poses from the initial docking are used to compute new ligand charge sets via QM/MM. A second round of docking is then performed using the new ligand charge sets. The top-ranked pose is selected as the overall winner. This approach leads to lower RMSDs than via the force field-based approach alone and via QM-based charges derived from the isolated ligand [63]. The latter, however, also exhibits improvements over force field-based docking. The QPLD approach was applied more recently to a set of 445 cocrystal complexes from the PDBbind set (Table 1) [65]. QPLD outperformed Glide for predominantly hydrophobic binding pockets that contain a proportion of polar interaction sites (hydrogen bonding groups).

The ability of this QM/MM docking approach to predict poses of metalloprotein–ligand complexes was also evaluated (Table 1) [66]. Redocking was performed for a series of ligand–metalloprotein crystal structures, including human carbonic anhydrase II. This protein has Zn(II) coordinated by three His residues (Figure 2). For force field-based docking, the RMSD in bound pose of a sulfamoyl benzamide inhibitor was 4.5 Å using Glide SP. A QM/MM docking which incorporated the Zn(II) and His atoms in the QM region led to an improvement in RMSD to 0.9 Å, although computationally expensive. The charge on the Zn(II) center reduced to a value of +1.16 *e*, when sidechains of amino acids within 4 Å of the ligand were included (> 100 atoms). This reduction proved important in preventing undesirable folding of the ligand arising from interaction of the F atom of its fluorobenzyl moiety with the Zn(II) center. A similar docking calculation on an inhibitor complex with cytidine deaminase led to a reduced charge of +0.84 *e* on the Zn; this was accompanied by comparable improvements in RMSD of the docked pose via the QM-based protocol.

This study can be compared with the work of Illingworth *et al.* [67], in which protein–ligand docking was performed for a set of 12 complexes, including metalloproteins, using AutoDock. Their scoring function also took ligand charges based on QM/MM calculations to incorporate the influence of the receptor (Table 1). Receptor polarization by its environment was modeled by an induced charge scheme. The final polarized charge set, derived from the receptor–ligand crystal structure, was then employed in docking. Although the RMSD in pose did not improve compared to a force field approach, the population of the largest cluster increased via use of QM/MM-derived charges (Table 1).

In work by Fischer *et al.* [68], rather than employing QM/MM calculations, a full QM calculation was performed in order to obtain partial charges, for the ligand and the receptor. This was achieved by using a fragment molecular orbital approach, based on 50 amino acid clustered around the ligand. The FMO approach was used to obtain Mulliken charges which were subsequently used in ligand docking. For both human estrogen receptor α and retinoic acid receptor γ , QM-based docking performed better than the force

Table 1. Selected recent examples of protein–ligand docking using QM-based ligand partial atomic charges.

Source of ligand charges (QM region)	Context of ligand charge derivation	Update of ligand charges during docking	No. of complexes in test set	Software	RMSE < 2 Å	Year [Ref.]
B3LYP/6-31G* QM/MM ESP (ligand)	Receptor-bound	Initial update	445 complexes from PDBbind database	Glide, Qsite	87% (77%, Glide only)	2009 [65]
B3LYP/6-31G* QM/MM ESP (active site + metal + ligand)	Receptor-bound	Initial update	8 metalloproteinase complexes	Glide, Qsite	88%	2009 [66]
AM1-BCC or RESP HF/6-31G* (ligand)	<i>In vacuo</i>	No update	100 complexes from PDBbind database	AutoDock4, Gaussian 09	87%	2011 [58]
Distributed multipole analysis of B3LYP/6-31G* QM/MM using a polarizable MM region (ligand)	Receptor-bound	No update	12 diverse complexes	AutoDock4, Gaussian 03	66%, ranked according to most populated cluster (50% using Gasteiger charges)	2008 [67]

field-based approach (using FlexScreen). We, however, observe that potentially further improvement might result from advancing beyond the use of Mulliken charges for ligand and receptor in this study.

Therefore, there does appear to be evidence that QM-based charge models provide scoring functions that can improve the quality of predicted docking pose for challenging receptors. Nevertheless, electrostatic and non-electrostatic parameters are required for modelling the ligand and protein. Ideally, one could dock a ligand while describing the ligand and possibly some of the active site directly at the QM level. This obviates the need for parameters, which is attractive for modeling chemically diverse and complex systems accurately. For example, Fong *et al.* have demonstrated the ability of a series of QM/MM scoring functions to discriminate native from non-native pose in six HIV-1 protease complexes [69]. This was based on large numbers of geometric decoys for each ligand–protein complex obtained from the Ligand Protein Data Bank (LPDB) [70]. The examined QM/MM scoring functions displayed a relatively smooth funnel-shaped topology of the binding surface required for efficient docking and demonstrated good discrimination, exceeding the performance of the GoldScore [71] and ChemScore [72] scoring functions.

It is possible to use a QM rather than QM/MM potential for discriminating native from non-native protein–ligand poses. Raha *et al.* developed linear scaling semiempirical QM methods, based on AM1 and PM3, for treating large molecules [73]. Using divide-and-conquer linear scaling approach, they constructed semiempirical QM-based scoring functions, QMScore and TotalScore [74]. Analogous to the QM/MM study of Fong *et al.* [69], they found the QM-based scoring functions capable of discriminating native from decoy poses. However, due to the computational cost of evaluating large numbers of poses, protein–ligand docking is in large part limited to using empirically based scoring functions as in the indirect QM approaches outlined above.

3.2 Protein–ligand binding affinities

Consequently, the focus of direct QM applications has been to score protein interactions with distinct ligands, rather than binding pose prediction. The levels of complexity vary considerably in the range of existing approaches to computing binding affinity. In their study of QMScore and TotalScore, Raha and Merz [74] compared their performance in predicting binding affinity with 11 other scoring functions for a set of 56 protein–ligand complexes from the data set of Wang *et al.* [61]. The 11 scoring functions compared included four from the LigFit module in Cerius2 (LigScore, PLP, PMF and LUDI) [75], four from SYBYL version 6.8 (F-Score, G-Score, D-Score and ChemScore) [76] and DrugScore [77], X-Score [78] and AutoDock [57]. The functions fall into the three categories: i) force field-based (AutoDock, G-Score and D-Score); ii) empirical (LigScore, PLP, LUDI, F-Score,

Figure 2. Active site of human carbonic anhydrase II, with bound aminosulfonyl ligand (PDB code 1g52).

ChemScore and X-Score); and iii) 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. More recently, Hayik *et al.* [79] have applied the divide-and-conquer approach within a QM/MM scheme, to score the affinities of a set of 18 carbonic anhydrase and 5 carboxypeptidase protein–ligand complexes. Using AM1 and AMBER for the QM and MM regions, respectively, the study obtained a correlation r^2 for the binding affinity of the set of metalloproteins of 0.64 without fitting, and an r^2 of 0.71 with adjusting the weights of individual scoring terms (Table 2).

Subsequent to the initial, ground-breaking work of Merz *et al.* in this area, large-scale semiempirical QM approaches have emerged which utilize increasingly sophisticated underlying semiempirical QM Hamiltonians. In particular, we highlight the PM6-DH2 approach by Hobza *et al.* which possesses empirical corrections for dispersion and hydrogen bonding. The method achieved accuracy to within 1 kcal/mol of high level *ab initio* QM benchmark calculations on small noncovalent complexes. The method was applied to the docking of 22 HIV-1 protease ligands [80]. These were first docked using DOCK6, and then the two top poses rescored using the enthalpy of binding at the PM6-DH2 level, in conjunction with a force field-based entropy estimate, and strain and desolvation estimates (using the SMD solvation model [81]). For the 22 ligands, a correlation r^2 of 0.62 was obtained using the PM6-DH2-based scoring function, as opposed to 0.15 and 0.10 when using DOCK and AMBER scorings, respectively. Interestingly, omission of vibrational binding entropy contributions considerably worsened correlation of the PM6-DH2 function. The importance of the entropic term to the binding of the large flexible HIV-1 protease inhibitors was noted by the authors.

Other work applied the PM6-DH2-based function to score the affinities of 15 cyclin-dependent kinase-2 (CDK2) inhibitors [82]. A correlation r^2 of 0.77 was found (Table 2). In this case, inclusion of the entropy term led to a decrease in correlation r^2 to 0.52. The justification here was provided by the neutral, more rigid and less polar CDK2 inhibitors relative to HIV-1 protease ligands. We note that alternative

methods in ranking these inhibitors performed less well, including MM/generalized Born/surface area (GBSA) rescoring of GlideXP (r^2 of 0.69) but not as well as a pharmacophore-based method (r^2 of 0.83) [83].

Using a combined AM1/CHARMM potential with a Poisson-Boltzmann/surface area (PBSA) term, Grater *et al.* [84] evaluated the binding energies of 47 benzamidine ligands to trypsin. Correlation to experiment improved only slightly compared to a force field-based MM/PBSA approach. Khandelwal *et al.* [85], integrated QM/MM calculations with a linear interaction energy approach to study matrix metalloproteinase inhibitors. The study employed docking, MD simulations and QM/MM calculations of the MD structures to rank 28 hydroxamate inhibitors of a zinc-dependent matrix metalloproteinase (MMP-9). Their solvent-corrected QM/MM approach, using the B3LYP density functional, led to smaller unsigned errors in inhibition constants (of ~ 0.3 vs > 0.6 log units without QM/MM). More recently, Natesan *et al.* [86] developed this QM/MM-linear response approach to take account of multiple tautomer and protonation states. Thus, using a protocol based on short (1 ns) classical MD simulations and single-point electronic embedding at the B3LYP/6-31G*/AMBER *ff99SB* level of theory, they considered 233 microspecies for 66 inhibitors of MAPK-activated protein kinase. Incorporation of these multiple species improved the correlation r^2 with experimental affinities (from 0.66 to 0.91, Table 2).

A study by Mikulskis *et al.* [87], also rescored configurations from MD simulations, using an entirely semiempirical QM Hamiltonian rather than QM/MM methods. Solvation energies were calculated using the conductor-like screening model (COSMO) of implicit solvent. A total of 25 ligands were studied, across the three proteins, avidin, factor Xa and ferritin. Three semiempirical QM Hamiltonians were compared (AM1, RM1 and PM6), as well as the effect of hydrogen-bonding and dispersion corrections. The study highlighted the need to include a dispersion correction to the semiempirical QM calculation. In general, the hydrogen bond correction also improved agreement with experiment. Interestingly, the authors recommended dispersion- and hydrogen bonding-corrected AM1 as giving overall slightly better results for the complexes considered by the study, compared to PM6-DH2, although they acknowledged the superior performance of the latter in estimating the (*in vacuo*) interaction energetics of binary small molecule complexes representative of the noncovalent interactions found in biology [88]. Application of the AM1-DH2 Hamiltonian led to correlation r^2 of 0.86, 0.12 and 0.92 for ligand-binding affinities to avidin, factor Xa and ferritin, respectively (Table 2). The authors also concluded that the performance of the semiempirical QM scoring approach, relative to MM/PBSA and MM/GBSA classical approaches [89], varied depending on target, and further work is required (e.g., on entropic terms) in order to obtain absolute binding affinities more in accord with experiment.

A similar conclusion was obtained in a comparison of MM/PBSA with the MM/QM-COSMO approach [90]. For

Table 2. Selected recent examples of protein–ligand scoring using QM methods.

Level of theory (QM region)	Receptor (no. of ligands)	MD	Software	Binding affinity method	r^2 for best models	Year [Ref.]
AM1/ff99SB [optimisations]; (ligand + residues within 5 Å of Zn)	Carbonic anhydrase II (18) carboxypeptidase (5)	no	AMBER 10	Endpoint scheme using PBSA solvent	0.71 (with fitting weights of scoring terms) 0.64 (no fitting)	2010 [79]
PM6-DH2/COSMO [optimisations] SMD correction for ligand desolvation; (protein + ligand)	CDK2 (15)	no	MOPAC2009	Endpoint method using COSMO/SMD solvent models	0.87 (enthalpy) 0.77 (enthalpy+ solvation+ deformation)	2011 [82]
B3LYP/6-31G**/ff99SB QM/MM [ONIOM EE]; (binding site + ligand + one water)	MAPK-activated protein kinase 2 (66)	yes	AMBER 10, Gaussian 09	Linear interaction energy approach (based on $\Delta E_{QM/MM} >$)	0.91 (tautomeric and ionization substates included); 0.66 (without multiple states)	2012 [86]
comparison of AM1, RM1, PM6, with COSMO and various hydrogen bond and dispersion corrections; (protein + ligand)	Avidin (7), factor Xa (9) and ferritin (9)	yes	MOPAC2009	Endpoint scheme using COSMO solvent	0.86, 0.12, 0.92 (AM1-DH2 model, for avidin, factor Xa and ferritin, respectively)	2012 [87]

the latter method [91], the enthalpy of protein–ligand binding was estimated by the difference in PM3 energies of the species, solvated by the COSMO model. These endpoint approaches were applied to a rescoring of classical MD simulations series of seven phosphopeptides, interacting with the C-terminal domain of BRCA1. Albeit a small test set, the QM/PBSA approach did not appear to offer significant improvement in correlation over the MM/QM-COSMO approach.

In addition to semiempirical QM approaches, DFT methods have been applied to the study of protein–inhibitor interactions. A linear scaling PBE/TZP MD study of the complexes of five inhibitors with the CDK2 [44] obtained a reasonable agreement between calculated and measured relative potencies when dynamics were incorporated. The study concluded that polarization effects resulting from the specific hydration patterns in the ATP pockets of CDK2 made important contributions to the relative potency of the five inhibitors.

A recent study by Antony *et al.* [11] sought to compute a series of protein–ligand interaction energies at a range of very high levels of QM. The protein was truncated, with the protein–ligand systems ranged in size from 50 to 300 atoms. The benchmark level for their study used the local pair natural orbital coupled-electron pair approximation (LPNO-CEPA/1) method at the complete basis set (CBS) limit. Using a variety of dispersion-corrected density functionals, the error relative to these benchmarks was found to be in the order of 1 – 2 kcal/mol for weaker binding complexes, and for complexes bound by a potential energy of more than 10 kcal/mol, the error rose to 2 – 4 kcal/mol, amounting to about 10% of the binding energy. The dispersion-corrected density functional, DFT-D3, was found to overbind relative to LPNO-CEPA/1; in turn, LPNO-CEPA/1 was found to slightly underbind relative to CCSD(T)/CBS calculations for the S22 set of small bimolecular complexes. This underlines the errors inherent in computing protein–ligand interactions. Interestingly for 15 complexes, the mean absolute deviation was 8.1 kcal/mol for PM6 and 5.1 kcal/mol for PM6-DH2, as compared to typical errors of 1.7 kcal/mol for the dispersion-corrected M06/TZVP density functional level of theory. However, such methods even at a density functional level are computationally expensive to apply to large numbers of ligand–receptor structures.

We also note a recent QM/MM study of the binding of a series of six Ru(II)-based ligands to cathepsin B [92]. These ligands form a Ru–S coordination complex, with the S atom belonging to the catalytic cysteine of the protease. This study was novel in attempting to rank the binding affinities of a series of metal complexes to a protein using QM. In their study, solvation effects were addressed using the PBSA model. It was found that the absolute energies of interaction considerably exceeded the experimental values (IC_{50}). A level of correlation in energies r^2 of 0.6 was observed. The study concluded that current methods struggled to satisfactorily account for protein flexibility on binding and

balance interaction energies with desolvation energies. In particular, the solvation models lacked the accuracy to adequately predict the complexation of a charged metal complex to a charged protein active site.

This last approach leads to the consideration of covalent ligand–protein interactions, a more traditional demesne of QM and QM/MM methods. Briefly, we provide two examples of this type of approach. First, a density functional/MM study characterized the covalent intermediates of suicide inhibitors of five adenosine deaminase inhibitors [93]. Although a small data set, the study found a good correlation (with a r^2 value of 0.88) connecting observed IC_{50} and the energy difference between the Michaelis complex of the inhibitor and the corresponding hydroxylated intermediate. The dependence of the correlation on the stability of the bound intermediate relative to the substrate, rather than the energies of the independent species, was due to a change in hydrogen bonding to the deaminase.

A second example involved QM/MM analysis of the covalent interaction of seven peptidyl α -ketoheterocyclic inhibitors of human neutrophil elastase inhibitors [94], which irreversibly form a tetrahedral intermediate. A good correlation was found between the stability of their covalent complexes of the transition state analogs and their activity, which related to the electron withdrawing ability of the heterocyclic rings of the inhibitors.

3.3 Ligand strain on binding

Although an actively evolving area of research, accurate prediction of the binding free energy of a ligand to a protein remains a major challenge for QM-based and other approaches [95,96]. Work has also focused on the perhaps more readily attainable goal of reliable computation of the change in internal energy of a ligand on binding to its receptor. From a study of 150 crystal structures of medicinally relevant protein–ligand complexes using the OPLS-AA and MMFF force fields, Perola and Charifson [97] estimated that over 60% of ligands bind in a strained conformation. The MM study concluded that while $\sim 60\%$ of the ligands were calculated to bind with strain energies < 5 kcal/mol, strain energies > 9 kcal/mol were calculated in at least 10% of the cases, regardless of the method used. Subsequently, a study [98] of the bound versus unbound pose of 99 drug-like molecules at the B3LYP/6-31G* level with an implicit solvent model found that the level of strain does not appear as great as suggested by Perola and Charifson: for a set of these, it was found that two-thirds of the bioactive conformations considered were within 0.5 kcal/mol of a local energy minimum [98]. Internal energies predicted to be > 2 kcal/mol could generally be related to problems in structure determination. The discrepancy between the MM and QM studies, although also depending on several other factors, does highlight the care that must be taken in modeling small molecule conformation, and the potential benefits of directly utilizing QM calculations.

This very much applies to the methods of X-ray structure determination, which rely on a suitable force field for ligand

and receptor in the refinement process. While reasonable force field parameters exist for the protein, there is often a lack of validated parameters to describe the small molecule ligand. The ability of QM to accurately model ligand conformational preferences, bound and unbound, has been exploited for improved refinement of crystal structures by a QM/MM approach [99,100]. The method was applied to refine inhibitor complexes of the multidrug-binding repressor protein, QacR. These improved structures provide details that could be important in guiding ligand design, for example, observing the existence of an orthogonal ring conformation of a pentadiamide ligand in complex with QacR that was assigned as planar in the original MM-led refinement. A QM/MM refinement approach appears to be most valuable for low resolution X-ray structures, where the MM potential can dominate and potentially incorrectly assign ligand structure.

Furthermore, from the QM/MM refined crystal structure, ligand strain energies can be derived. Fu *et al.* used a QM/MM re-refinement of the complex of ibuprofen with human serum albumin to obtain its bound conformation [101]. In this work, the QM region, modeled at the HF/6-31G* level, was defined as the ligand plus amino acid residues and water molecules within 2 Å of the ligand. An 88% reduction in predicted ibuprofen strain energy (as predicted at the MP2/CBS level of theory) was found when QM/MM re-refinement was performed. Thus, the strain energy of the ligand reduced by ~ 25 kcal/mol. A further conclusion was that the MM force field did not adequately describe the strain energy of charged molecules such as ibuprofen. This suggests that strain energies are largely overestimated in PDB structures of protein–ligand complexes.

In conclusion, advances in theory, methodology, software and hardware mean that the methods of quantum chemistry have never been more available and more powerful than today. More sophisticated QM-based QSAR approaches are available; the 3D conformation of large number of ligands can be evaluated rapidly at a reasonable level of QM theory. A number of QM or QM/MM-based docking approaches exist; selected examples are summarized in Table 1. The prediction of ligand poses when interacting with complex receptors, for example, metalloproteins, has been shown to more closely resemble experimentally determined poses when partial charges are derived from QM or QM/MM calculations. A number of studies have sought to directly incorporate QM and QM/MM approaches into methodologies for calculating ligand–protein binding affinities; selected examples of these are given in Table 2. These approaches have shown promise but require further development to be broadly applicable. Finally, QM methods have proved valuable in quantitative analysis of the energetics of ligand deformation on binding. Related to this, QM/MM refinement of protein–ligand crystal structures have been shown to improve the quality of structure determination over an approach which depends on a MM description of the ligand.

4. Expert opinion

QM or QM/MM methods have been used to enhance the accuracy of protein–ligand docking via their use in deriving more accurate partial atomic charges on the ligand. As this area continues to evolve, we anticipate that the frequency with which the partial atomic charges of the ligand will update during docking, according to QM/MM ESP calculations, will increase. Ultimately this could develop to the point where direct QM or QM/MM calculations at each sampled ligand–receptor conformation becomes a possibility. Furthermore, it is possible to incorporate the electronic polarization of the receptor due to the ligand via a polarizable MM force field [102,103]. In a recent study of ion– π small molecule complexes [104], Wang and Bryce examined the performance of a semiempirical QM/MM approach and the corresponding QM/MMpol model, where inducible point dipoles are additionally employed in the MM region such that the QM and MM regions can mutually polarize. In these studies, the ion was modeled via QM and the π -molecule at the MM level. The predicted potential energy surfaces of cation–benzene complexes improved by inclusion of explicit MM polarization of the π -molecule. For cation–formamide complexes, inducible dipoles appreciably improved energetic estimates at geometries forming non-optimal interactions. Interestingly, for complexes at non-optimal orientations, as might be found in protein–ligand docking, the QM/MM model exhibits higher errors than the QM/MMpol approach, being unable to capture the orientation-dependent variation in polarization energy.

In contrast with docking, attempts have been made to more directly incorporate QM and QM/MM approaches into methodologies for ranking ligand–protein complexes according to binding affinities; selected examples of these are given in Table 2. Unfortunately, as stated above, the success of these approaches is rather mixed, and often depends on the details of the method and receptor active site. Furthermore, it is evident that factors, such as solvent model, entropy and target flexibility, are at least as important as accurate protein–ligand binding potential energies obtained from QM-based methods. Attempts to incorporate these other factors typically employ endpoint methods, such as LIE and MM/PBSA, with their accompanying assumptions. Although in its early stages, we note that some work has been done in attempting to incorporate QM effects into more rigorous approaches to compute binding free energies which are based on statistical perturbation theory [105,106]. In conclusion, further work is required on a number of fronts before such binding affinity predictions become routine. However, whether directly or indirectly, QM approaches will be part of the solution, as the shortcomings of MM potentials become increasingly apparent.

Given the evolving nature of methodology for computing binding affinities, perhaps of more current utility to the computational medicinal chemist are QM methods for

resolution of protein–ligand interaction energies into residue and functional group contributions, to allow a better understanding of molecular recognition at a QM level of accuracy. QM analysis is particularly useful for systems where MM parameterization is not straightforward or likely to be inaccurate. QM energetic analysis was used by Barroso-Neto *et al.* [107] in understanding the inactivation of ovine COX-1 by bromoaspirin and aspirin. Using a MFCC scheme based on DFT to compute ligand–residue QM interaction energies, they were able to identify the most important amino acid residues in binding salicylic acid after acetylation of the protein by these ligands. Similarly, Mazanetz *et al.* [108] used an FMO approach to resolve the QM interaction of 28 CDK2 inhibitors into residue-based components; their pair interaction energy analysis outperformed three MM force field-based scoring functions in producing predictive correlations with experiment. Such approaches are of practical use in guiding the development of new inhibitors to recruit potential active site interactions. In a somewhat similar vein, graphical representation of the information afforded by QM, such as visualization of molecular orbitals and group energetic contributions, could be a powerful and more widely used tool in understanding the subtleties of ligand–protein interactions.

To close, we highlight a study which employed QM calculations prospectively, guiding structure-based design of inhibitors targeting the ATP site of the tyrosine kinase erythropoietin-producing human hepatocellular carcinoma receptor B4 (EphB4). The approach, by Zhou and Caflisch [109], was a hierarchical filtering of a virtual database of compounds. Part of this filtering involved using QM probe molecules oriented in the geometry of polar amino acid groups found in the binding site of EphB4. The interaction energy between these probes and docked ligands were then computed at the semiempirical QM level (PM6). The main requirements for the probes were to represent the local electronic structures of the binding pocket well, to be as simple as possible (so as to be computationally inexpensive), and to be able to distinguish sensitively between preferable and adverse interactions. The probes varied in size from 2 to 10 atoms. Ligand–probe QM energies were calculated in the absence of solvent, protein or other probes.

Thus, flexible docking of a library of 2.7 million compounds into the homology modeled EphB4 structure was performed using AutoDock4. The docked structures were then minimized using CHARMM. In total, the QM probe filtering was then carried out on ~ 100 million different small molecule poses generated by AutoDock. After QM probe filtering using five QM probe molecules (one gatekeeper, one entrance pocket and three hinge amino acid sites), 1% of these poses were retained. Further screening using ligand strain, non-polar interactions and visual inspection left only 23 compounds, which were tested in an enzyme assay. Out of the 23, three compounds were found to be active at fairly low micromolar concentrations.

The application of quantum chemical approaches in guiding structure-based design is not new. Indeed, quantum chemistry has featured in some of the earliest medicinally relevant chemistry calculations, in providing informative descriptors for QSAR, in determining the 3D conformation of ligands, and in estimating the strain energies of their bound poses. Modern QM methods offer the ability to provide an accurate representation of the electrostatics of ligands and proteins and have ready application to improved structure determination and modeling of chemical systems where MM parameterization strategies can struggle, for example, for metalloprotein complexes and for covalently bound inhibitors. Recent developments such as linear scaling methodologies and more accurate semiempirical QM approaches hold promise in enabling quantum chemical approaches to address pharmacological

problems on the time scale demanded by drug discovery projects. Although computation of binding energetics remains a challenging and evolving area, current QM approaches can offer detailed information on the nature and relative strengths of complex active site interactions which can be valuable in guiding molecular design. Increasingly, it is likely that QM will become a more prominent tool in the repertoire of the computational medicinal chemist. Thus, modern QM approaches will play a more direct role in informing and streamlining the drug discovery process.

Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.

Bibliography

Papers of special note have been highlighted as either of interest (●) or of considerable interest (●●) to readers.

- Loughney D, Claus BL, Johnson SR. To measure is to know: an approach to CADD performance metrics. *Drug Discov Today* 2011;16:548-54
- Green D, Leach A, Head M. Computer-aided molecular design under the SWOTlight. *J Comput Aided Mol Des* 2012;26:51-6
- Popelier PLA, Smith PJ. QSAR models based on quantum topological molecular similarity. *Eur J Med Chem* 2006;41:862-73
- Chalk AJ, Beck B, Clark T. A quantum mechanical/neural net model for boiling points with error estimation. *J Chem Inf Comput Sci* 2001;41:1053-9
- Greer J, Erickson JW, Baldwin JJ, Varney MD. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J Med Chem* 1994;37:1035-54
- Richards WG. *Quantum pharmacology*. Butterworth & Co. Ltd., London; 1977
- Löwdin PO. Some aspects of the hydrogen bond in molecular biology. *Ann N Y Acad Sci* 1969;158:86-95
- Korth M, Hobza P. On extension of the current biomolecular empirical force field for the description of halogen bonds. *J Chem Theory Comput* 2012;8:1325-33
- Zgarbova M, Otyepka M, Sponer J, et al. Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 2011;7:2886-902
- Hornak V, Abel R, Okur A, et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712-25
- Antony J, Grimme S, Liakos DG, Neese F. Protein-ligand interaction energies with dispersion corrected density functional theory and high-level wave function based methods. *J Phys Chem A* 2011;115:11210-20
- **This density functional and high-level ab initio QM treatment of truncated protein-ligand complexes of 50 – 300 atoms provides useful insight into the errors associated with calculation of intermolecular interaction energies.**
- Hobza P. Calculations on noncovalent interactions and databases of benchmark interaction energies. *Acc Chem Res* 2012;45:663-72
- Jurecka P, Sponer J, Cerny J, Hobza P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys Chem Chem Phys* 2006;8:1985-93
- Ramraj A, Raju RK, Wang Q, et al. An evaluation of the GLYCAM06 and MM3 force fields, and the PM3-D* molecular orbital method for modelling prototype carbohydrate-aromatic interactions. *J Mol Graph Modell* 2010;29:321-5
- Sato T, Tsuneda T, Hirao K. Long-range corrected density functional study on weakly bound systems: balanced descriptions of various types of molecular interactions. *J Chem Phys* 2007;126:234114-12
- Grimme S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J Comput Chem* 2004;25:1463-73
- Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010;132:154104-19
- Burns LA, Mayagoitia AV, Sumpter BG, Sherrill CD. Density-functional approaches to noncovalent interactions: a comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *J Chem Phys* 2011;134:084107-25
- Pople JA, Santry DP, Segal GA. Approximate self-consistent molecular orbital theory. I. Invariant procedures. *J Chem Phys* 1965;43:S129
- Elstner M, Porezag D, Jungnickel G, et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev* 1998;B58:7260-8
- Foster ME, Sohlberg K. A new empirical correction to the AM1 method for macromolecular complexes. *J Chem Theory Comput* 2010;6:2153-66
- Tuttle T, Thiel W. OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and

- biochemical application. *Phys Chem Chem Phys* 2008;10:2159-66
23. 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;9:2362-70
 24. Elstner M, Frauenheim T, Kaxiras E, et al. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. In: *Computer simulation of materials at atomic level*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim; 2000. p. 357-76
 25. Clark T. Quo vadis semiempirical MO-theory? *J Mol Struct: Theochem* 2000;530:1-10
 26. Winget P, Seluki C, Horn AHC, et al. Towards a next generation neglect of diatomic differential overlap based semiempirical molecular orbital technique. *Theoret Chem Acc* 2003;110:254-66
 27. Monard G, Bernal-Uruchurtu MI, van der Vaart A, et al. Simulation of liquid water using semiempirical hamiltonians and the divide and conquer approach. *J Phys Chem A* 2005;109:3425-32
 28. Bernal-Uruchurtu MI, Ruiz-Lopez MF. Basic ideas for the correction of semiempirical methods describing H-bonded systems. *Chem Phys Lett* 2000;330:118-24
 29. Bernal-Uruchurtu MI, Martins-Costa MTC, Millot C, Ruiz-Lopez MF. Improving description of hydrogen bonds at the semiempirical level: water-water interactions as test case. *J Comput Chem* 2000;21:572-81
 30. Repasky MP, Chandrasekhar J, Jorgensen WL. PDDG/PM3 and PDDG/MNDO: improved semiempirical methods. *J Comput Chem* 2002;23:1601-22
 31. Yang Y, Yu H, York D, et al. Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J Phys Chem A* 2007;111:10861-73
 32. Korth M, Pitonak M, Rezac J, Hobza P. A transferable h-bonding correction for semiempirical quantum-chemical methods. *J Chem Theory Comput* 2010;6:344-52
 33. Rezac J, Hobza P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J Chem Theory Comput* 2011;8:141-51
 34. 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;233:231-6
 35. McNamara JP, Muslim AM, Abdel-Aal H, et al. Towards a quantum mechanical force field for carbohydrates: a reparameterized semi-empirical MO approach. *Chem Phys Lett* 2004;394:429-36
 36. Barnett CB, Naidoo KJ. Ring puckering: a metric for evaluating the accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM simulations. *J Phys Chem B* 2010;114:17142-54
 37. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 2009;48:1198-229
 38. Maseras F, Morokuma K. IMOMM: a new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J Comput Chem* 1995;16:1170-9
 39. Wang Q, Bryce RA. Improved hydrogen bonding at the NDDO-type semiempirical quantum mechanical/molecular mechanical interface. *J Chem Theory Comput* 2009;5:2206-12
 40. Fedorov DG, Nagata T, Kitaara K. Exploring chemistry with the fragment molecular orbital method. *Phys Chem Chem Phys* 2012;14:7562-77
 41. Zhang DW, Xiang Y, Zhang JZH. New advance in computational chemistry: full quantum mechanical ab Initio computation of streptavidin-biotin interaction energy. *J Phys Chem B* 2003;107:12039-41
 42. Fedorov DG, Kitaara K. Pair interaction energy decomposition analysis. *J Comput Chem* 2007;28:222-37
 43. Zhou T, Huang DZ, Caffisch A. Quantum mechanical methods for drug design. *Curr Top Med Chem* 2010;10:33-45
 44. Heady L, Fernandez-Serra M, Mancera RL, et al. Novel structural features of CDK inhibition revealed by an ab initio computational method combined with dynamic simulations. *J Med Chem* 2006;49:5141-53
 45. Cole DJ, Skylaris CK, Rajendra E, et al. Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations. *Epl* 2010;91(3):37004
 46. Valiev M, Bylaska EJ, Govind N, et al. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput Phys Commun* 2010;181:1477-89
 47. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K. GPU-accelerated molecular modeling coming of age. *J Molr Grap Model* 2010;29:116-25
 48. Ufimtsev IS, Martinez TJ. quantum chemistry on graphical processing units. 2. direct self-consistent-field implementation. *J Chem Theory Comput* 2009;5:1004-15
 49. Ufimtsev IS, Martinez TJ. Quantum chemistry on graphical processing units. 3. analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J Chem Theory Comput* 2009;5:2619-28
 50. Ufimtsev IS, Martinez TJ. Quantum chemistry on graphical processing units. 1. strategies for two-electron integral evaluation. *J Chem Theory Comput* 2008;4:222-31
 51. Yasuda K. Accelerating density functional calculations with graphics processing unit. *J Chem Theory Comput* 2008;4:1230-6
 52. Vogt L, Olivares-Amaya R, Kermes S, et al. Accelerating resolution-of-the-identity second-order moller-Plesset quantum chemistry calculations with graphical processing units. *J Phys Chem A* 2008;112:2049-57
 53. Luehr N, Ufimtsev IS, Martinez TJ. Dynamic Precision for electron repulsion integral evaluation on Graphical Processing Units (GPUs). *J Chem Theory Comput* 2011;7:949-54
 54. Seebeck B, Reulecke I, Kamper A, Rarey M. Modeling of metal interaction

- geometries for protein-ligand docking. *Proteins* 2008;71:1237-54
55. Kirton SB, Murray CW, Verdonk ML, Taylor RD. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. *Proteins* 2005;58:836-44
 56. Kuntz ID, Blaney JM, Oatley SJ, et al. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982;161:269-88
 57. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* 1996;9:1-5
 58. Wang JC, Lin JH, Chen CM, et al. Robust scoring functions for protein-ligand interactions with quantum chemical charge models. *J Chem Inf Model* 2011;51:2528-37
 - **Derivation of more accurate docking scoring functions based on improved approaches to determination of partial charges from ESPs.**
 59. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J Phys Chem* 1993;97:10269-80
 60. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 2000;21:132-46
 61. Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem* 2004;47:2977-80
 62. Wang R, Fang X, Lu Y, et al. The PDBbind database: methodologies and updates. *J Med Chem* 2005;48:4111-19
 63. 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;26:915-31
 64. Kimura SR, Rajamani R, Langley DR. Communication: quantum polarized fluctuating charge model: A practical method to include ligand polarizability in biomolecular simulations. *J Chem Phys* 2011;135:231101-4
 65. Chung JY, Hah JM, Cho AE. Correlation between performance of QM/MM docking and simple classification of binding sites. *J Chem Inf Model* 2009;49:2382-7
 66. Cho AE, Rinaldo D. Extension of QM/MM docking and its applications to metalloproteins. *J Comput Chem* 2009;30:2609-16
 67. Illingworth CJR, Morris GM, Parkes KEB, et al. Assessing the role of polarization in docking. *J Phys Chem A* 2008;112:12157-63
 68. 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;70:1264-73
 69. Fong P, McNamara JP, Hillier IH, Bryce RA. Assessment of QM/MM scoring functions for molecular docking to HIV-1 protease. *J Chem Inf Model* 2009;49:913-24
 70. Ferrara P, Gohlke H, Price DJ, et al. Assessing scoring functions for protein-ligand interactions. *J Med Chem* 2004;47:3032-47
 71. Eldridge MD, Murray CW, Auton TR, et al. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997;11:425-45
 72. Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using GOLD. *Proteins* 2003;52:609-23
 73. Raha K, Peters MB, Wang B, et al. The role of quantum mechanics in structure-based drug design. *Drug Discov Today* 2007;12:725-31
 74. 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;48:4558-75
 75. Accelrys, Inc., Cerius2 Modeling Environment. [Release 4.8] Accelrys Software Inc, San Diego; 2006
 76. Tripos Inc. SYBYL 6.8. 2003
 77. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000;295:337-56
 78. 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
 79. Hayik SA, Dunbrack R, Merz KM. Mixed quantum mechanics/molecular mechanics scoring function to predict protein-ligand binding affinity. *J Chem Theory Comput* 2010;6:3079-91
 80. Fanfrlik J, Bronowska AK, Rezac J, et al. A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and h-bonding: HIV-1 protease with 22 ligands. *J Phys Chem B* 2010;114:12666-78
 81. Marenich AV, Cramer CJ, Truhlar DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* 2009;113:6378-96
 82. Dobež P, Fanfrlik J, Rezac J, et al. Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J Comput Aided Mol Des* 2011;25:223-35
 83. Vadivelan S, Sinha BN, Irudayam SJ, Jagarlapudi SARP. Virtual screening studies to design potent CDK2-cyclin A inhibitors. *J Chem Inf Model* 2007;47:1526-35
 84. Grater F, Schwarzl SM, Dejaegere A, et al. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J Phys Chem B* 2005;109:10474-83
 85. Khandelwal A, Lukacova V, Comez D, et al. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J Med Chem* 2005;48:5437-47
 86. Natesan S, Subramaniam R, Bergeron C, Balaz S. Binding affinity prediction for ligands and receptors forming tautomers and ionization species: inhibition of mitogen-activated protein kinase-activated protein kinase 2 (MK2). *J Med Chem* 2012;55:2035-47
 87. Mikulskis P, Genheden S, Wichmann K, Ryde U. A semiempirical approach to ligand-binding affinities: dependence on

- the Hamiltonian and corrections. *J Comput Chem* 2012;33:1179-89
- **This work provides a systematic assessment of ligand-binding affinity as a function of semiempirical QM potential and protein receptor.**
88. Riley KE, Pitonak M, Jurecka P, Hobza P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem Rev* 2010;110:5023-63
 - **An excellent review of progress in the evaluation of noncovalent interactions using QM approaches.**
 89. Srinivasan J, Cheatham TE, Cieplak P, et al. Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J Am Chem Soc* 1998;120:9401-9
 90. Anisimov V, Ziemys A, Kizhake S, et al. Computational and experimental studies of the interaction between phospho-peptides and the C-terminal domain of BRCA1. *J Comput Aided Mol Des* 2011;25:1071-84
 91. Anisimov VM, Cavasotto CN. Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. *J Comput Chem* 2011;32:2254-63
 92. Ciancetta A, Genheden S, Ryde U. A QM/MM study of the binding of RAPTA ligands to cathepsin B. *J Comput Aided Mol Des* 2011;25:729-42
 93. Gleeson MP, Burton NA, Hillier IH. Prediction of the potency of inhibitors of adenosine deaminase by QM/MM calculations. *Chem Commun* 2003(17):2180-1
 94. 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;2:2275-80
 95. Michele J, Foloppe N, Essex JW. Rigorous free energy calculations in structure-based drug design. *Mol Inf* 2010;29:570-8
 96. Bryce RA. Physics-based scoring of protein-ligand interactions: explicit polarizability, quantum mechanics and free energies. *Future Med Chem* 2011;3:683-98
 97. Perola E, Charifson PS. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J Med Chem* 2004;47:2499-510
 98. Butler KT, Luque FJ, Barril X. Toward accurate relative energy predictions of the bioactive conformation of drugs. *J Comput Chem* 2009;30:601-10
 99. Ryde U, Nilsson K. Quantum refinement – a combination of quantum chemistry and protein crystallography. *J Mol Struct* 2003;632:259-75
 100. Yu N, Yennawar HP, Merz KM Jr. Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics. *Acta Crystallogr D Biol Crystallogr* 2005;61:322-32
 101. Fu Z, Li X, Merz KM. Accurate assessment of the strain energy in a protein-bound drug using QM/MM X-ray refinement and converged quantum chemistry. *J Comput Chem* 2011;32:2587-97
 102. Bryce RA, Buesnel R, Hillier IH, Burton NA. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem Phys Lett* 1997;279:367-71
 103. Bryce RA, Vincent MA, Malcolm NOJ, et al. Cooperative effects in the structuring of fluoride water clusters: ab initio hybrid quantum mechanical/molecular mechanical model incorporating polarizable fluctuating charge solvent. *J Chem Phys* 1998;109:3077-85
 104. Wang Q, Bryce RA. Accounting for non-optimal interactions in molecular recognition: a study of ion- π complexes using a QM/MM model with a dipole-polarisable MM region. *Phys Chem Chem Phys* 2011;13:19401-8
 105. Hori K, Yamaguchi T, Uezu K, Sumimoto M. A free-energy perturbation method based on Monte Carlo simulations using quantum mechanical calculations (QM/MC/FEP method): application to highly solvent-dependent reactions. *J Comput Chem* 2011;32:778-86
 106. Beierlein FR, Michel J, Essex JW. A simple QM/MM approach for capturing polarization effects in protein-ligand binding free energy calculations. *J Phys Chem B* 2011;115:4911-26
 107. Barroso-Neto IL, Marques JP, da Costa RF, et al. Inactivation of Ovine Cyclooxygenase-1 by Bromoaspirin and Aspirin: a Quantum Chemistry Description. *J Phys Chem B* 2012;116:3270-9
 108. Mazanetz M, Ichihara O, Law R, Whittaker M. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *J Cheminformatics* 2011;3:2
 109. Zhou T, Caffisch A. High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors. *Chem Med Chem* 2010;5:1007-14
 - **An example of using QM interaction energies as a filter in a virtual screen which contributes to identification and experimental validation of low micromolar inhibitors of a kinase.**

Affiliation

Daniel Mucs & Richard A Bryce[†]

[†]Author for correspondence

University of Manchester,
School of Pharmacy and Pharmaceutical
Sciences, Oxford Road,
Manchester, M13 9PT, UK
Tel: +0 161 275 8345;

E-mail: r.a.bryce@manchester.ac.uk