LUND UNIVERSITY

Quantum mechanics in structure-based ligand design

Söderhjelm, Pär; Genheden, Samuel; Ryde, Ulf

Published in:
Protein-ligand interactions

DOI:
10.1002/9783527645947.ch7

2012

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Söderhjelm, P., Genheden, S., & Ryde, U. (2012). Quantum mechanics in structure-based ligand design. In H. Gohlke (Ed.), *Protein-ligand interactions* (pp. 121-143). (Methods and principles in medicinal chemistry; Vol. 53). Wiley-VCH Verlag. https://doi.org/10.1002/9783527645947.ch7

Total number of authors:
3

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Quantum mechanics in structure-based ligand design

Pär Söderhjelm [1], Samuel Genheden [2], Ulf Ryde [2]*

[1] Department of Chemistry and Applied Biosciences – Computational Science, ETH Zürich,

Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

[2] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124,

SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,

Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

18/03/17

1. Introduction

The prediction of the free energy for the binding of a small ligand to a macromolecule is undoubtedly one of the most important challenges of computational chemistry. If the affinity of any ligand could be predicted with good accuracy, enormous amounts of money could be saved in pharmaceutical industry because the number of drug candidates synthesised could be strongly reduced. Consequently, many methods have been developed with this aim, ranging from methods based on statistical mechanics and a physical formulation of the problem, e.g. free-energy perturbations and thermodynamic integration to methods based on a statistical analysis of available experimental data.[1]

Owing to the size of the macromolecule, the physical methods have normally been based on a molecular-mechanics (MM) description of the molecules, i.e. by treating molecules as a collection of balls, interacting with each other by an empirical potential, a force field. Typically, such a force field for a macromolecule includes terms for bonded interactions (bonds, angles, and dihedral angles), as well as electrostatic and van der Waals interactions between non-bonded atoms. This provides a classical-mechanics description of the system and treats the electrons only implicitly. The advantage of MM methods is their speed – the energy of a biomacromolecule, including the ligand and several thousands of water molecules, can be calculated within seconds. The disadvantages are that the force field is an approximation with a limited accuracy[2] and that you need to determine the force field for all molecules of interest (typically at least two parameters for each bond and angle, three for each dihedral, two for each pair of non-bonded atoms, and a charge on every atom), a very significant task, especially as drug candidates often show a great variation in their chemistry. In addition, some molecules, e.g. metal complexes, are hard to treat with MM methods.

These disadvantages of MM can be cured by using quantum mechanics (QM) methods: They do not require any parametrisation and they are applicable to any type of ligand or macromolecule. On the other hand, they are computationally much more demanding. Unfortunately, the Schrödinger equation (which is the basis of QM) cannot be solved analytically for systems with more than one electron. Therefore, a large number of QM methods have been developed that solves this equation approximately, having varying computational demands and accuracies, ranging from semiempirical methods (which require parametrisation, but can treat a full macromolecule), via Hartree–Fock (HF) and density functional theory (DFT), to high-level correlated methods, e.g. Møller–Plesset perturbation theory and coupled-cluster methods (which may give an accuracy similar to experiments but are applicable only for small molecules).[3]

Lately, there has been a great interest in using QM methods in structure-based ligand design.[4,5,6,7,8,9] In this chapter, we review such efforts. The subject is wide and QM methods in some way are now used in most projects of computational drug design. Therefore, we will focus on methods to calculate binding affinities (scoring), whereas studies of structures[10] and reaction mechanisms,[5,11,12] as well as the use of QM methods to obtain QSAR descriptors[4,13,14,15] are excluded.

2. Three MM-based methods

Before turning to the QM-based methods, we will shortly describe three commonly used MM approaches for ligand binding, because they nicely introduce the problems encountered in binding-affinity calculations and they are the basis of many QM approaches.

The physically most strict method to obtain binding affinities is free-energy perturbation (FEP),[16] in which the binding affinity is calculated by converting one ligand to another or to nothing by several small steps involving non-physical intermediate states. For each state, extensive sampling is performed by either molecular dynamics (MD) or Monte Carlo simulations. Consequently, the method is computationally expensive and therefore little used in drug design. Instead, methods based on sampling only of the end-states (the complex, the

free macromolecule, and the free ligand) have been more popular.

For example, Kollman and coworkers have developed the MM/PBSA method,[17] in which the free energy of a system is estimated from

$$G = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdW}} + G_{\text{solv}} + G_{\text{np}} - TS \tag{1}$$

where the three first terms on the right-hand side are the internal (i.e. bonds, angles, and dihedrals), electrostatic, and van der Waals energies, calculated at the molecular-mechanics (MM) level, G_{solv} is the polar solvation energy, calculated either with the Poisson–Boltzmann (PB) or the generalised Born (GB; giving the MM/GBSA approach) continuum-solvation methods, G_{np} is the non-polar solvation energy, estimated from a linear relation to the solvent-accessible surface area (SASA), and the last term is the product of the absolute temperature and the entropy, estimated from vibrational frequencies calculated at the MM level. All these energy terms are typically averaged over a MD simulation and the binding free energy is estimated from the difference in free energy of the complex, the free macromolecule and the free ligand. Normally, only the complex is simulated, in which case the internal energy cancels.

Åqvist has developed another end-point approach, the linear interaction energy (LIE) method,[18] in which the binding affinity of a ligand is estimated from

$$\Delta G_{bind} = \beta \left(\langle E_{el}^{L-S} \rangle_{bound} - \langle E_{el}^{L-S} \rangle_{free} \right) + \alpha \left(\langle E_{vdW}^{L-S} \rangle_{bound} - \langle E_{vdW}^{L-S} \rangle_{free} \right) \tag{2}$$

where $\langle E_{el}^{L-S} \rangle$ and $\langle E_{vdW}^{L-S} \rangle$ are the electrostatic and van der Waals interaction energy between the ligand and the surroundings (macromolecule and solvent), averaged over MD simulations of the ligand free in solution or bound to the macromolecule. a and b are parameters; b should be 0.5 according to linear-response theory, but has later been shown to depend on the functional groups of the ligand.[18,19] a was originally found to be 0.18, but it seems to depend on the system and is often treated as a fitting parameter.[18] Later, this approach has been extended by additional terms (and fitting parameters), making it approach the quantitative structure–activity relationship (QSAR) methods.[20]

3. QM-based force fields

The first step towards the use of QM methods for drug design is to perform the calculations at the MM level, but employing QM calculations in the parametrisation of the MM force field. In fact, essentially all modern MM force fields are based at least partly on QM calculations and most studies of the binding of drug candidates employ charges of the ligand calculated with QM. However, three potentials stand out by being specifically developed to give close agreement with QM methods with a minimum of fitted parameters: SIBFA (sum of interactions between fragments ab initio computed),[21] EFP (effective fragment potential),[22] and NEMO.[23] SIBFA has been extensively used for protein–ligand interactions, especially for metalloproteins.[21]

Moreover, in a several studies, QM calculations have been used to obtain specific MM parameters for certain macromolecule–ligand complexes. For example, Hayes and coworkers have shown that predictions of ligand-binding energies can be improved by reparametrisation of the charges and torsion parameters, based on QM calculations.[24]

Curioni et al. have developed a method to improve the MM charges in MD simulations of protein–ligand complexes.[25] They recalculated the charges for the ligand and possibly the closest residues by fitting them to the electrostatic potential (ESP), including a harmonic restraint to the original MM charges. These QM-refined charges improved the correlation between averaged MM protein–ligand interaction energies and experimental binding affinities

for 36 inhibitors to HIV-1 protease. This method has also been used for the scoring of 38 inhibitors of cyclin-dependent kinase 2 using an extended LIE approach.[26] The QM charges gave slightly improved results over other scoring functions.

Likewise, Friesner and coworkers have developed a docking strategy in which the ligand charges are updated by a QM/MM calculation (see below).[27] They include only the ligand in the QM system and treat it at the B3LYP/6-31G* level. They demonstrate that the QM/MM charges in many cases give clear improvements in the docking accuracy. This approach, called QM-polarised ligand docking (QPLD), is now a standard tool in the Schrödinger software package and it has been employed in many studies.[28,29,30,31,32,33,34,35] For metal-binding ligands, the metal and all its ligating groups need to be included in the QM system.[36,37,38] This approach has also been used to obtain charges for MM/GBSA calculations of binding affinities of two DNA-binding ligands.[39] It was shown that the QM/MM charges perform better than standard MM charges for both the root-mean-squared deviation (RMSD) during MD simulations and the calculated binding affinities.

Reynolds and coworkers have made a systematic investigation of how the docking of twelve ligands change when the charges were recalculated by QM methods.[40] They compared the results of Gasteiger, Amber, and QM charges for the ligand, the latter obtained both in vacuum or polarised by the surrounding protein. Moreover, they also tested to change the charges of all amino acids within 5.5 Å of the ligand (polarised by the ligand and the protein), both from a single-point calculation and after an iterative procedure. They did not observe any consistent improvement in the docked poses when the charges were improved.

A few groups have tried to recalculate all charges in the protein and ligand with QM methods. For example, Fischer et al. used the FMO method (see below) at the HF/6-31G* level.[41] Using a simple scoring function with Coulomb, Lennard-Jones, and hydrogen-bond interaction energies for single docked structures, they obtained good correlations between calculated and experimental affinities for both the binding of eleven ligands to the human oestrogen a receptor and four ligands to the human retinoic acid g receptor (r^2 = 0.66 and 0.90, respectively).

A similar approach has been developed for the MFCC method (see below), in which charges are calculated iteratively for the whole protein in a PB continuum solvent. This approach has been applied to the binding of biotin and an analogue to avidin using the MM/PBSA approach, indicating that the calculated difference in the binding affinity is improved with the polarised charges.[42]

Menikarachchi and Gascón have evaluated the effect of using polarised charges obtained for the whole protein by a moving-domain QM/MM approach, in which QM charges are iteratively calculated for each amino acid in turn with an updated point-charge model of the remaining protein.[8] They studied the docking of a ligand to the catalytic zinc ion of carbonic anhydrase and obtained improved results compared to a fixed-charge model.

We have recalculated charges of all atoms in 20 snapshots from MD simulations of seven biotin analogues to avidin by QM methods (HF/6-31G*).[43] The charges were used in a MM/PBSA calculation, but no improvement in the correlation between calculated and experimental affinities were found and they gave rise to unstable MD trajectories. Further analyses showed that the electrostatic interaction energy between the protein and the ligands differ by 43 and 8 kJ/mol on average between QM charges calculated for the correct conformation and standard MM charges for charged and neutral ligands, respectively, illustrating the conformational dependence of the charges.[44] Fortunately, this difference is reduced by 7 and 3 kJ/mol when solvation effects are included. For accurate results, QM charges need to be recalculated for all residues within 7 Å of the ligand.

Moreover, the instability of MD simulations can be avoided by averaging the QM charges over the snapshots or over all residues of the same type in the protein.[44] This represents an alternative approach to the restrained ESP method used to obtain MM charges in the AMBER force fields,[45] avoiding the arbitrary restraint used in that approach. Such charges are

transferable between different proteins and gave slightly improved correlation for MM/GBSA binding affinities of two different proteins (avidin and factor Xa).[46]

A full NEMO potential, i.e. multipoles up to quadrupoles and anisotropic polarisabilities in all atoms and bond midpoints, has also been calculated at the B3LYP/6-31G* level for ten MD snapshots.[47] It was combined with standard MM/PBSA van der Waals, SASA, and entropy terms, as well as a polarised continuum model (PCM) solvation energies to give complete binding affinities for the same seven biotin analogues, but still no significant improvement was seen compared to a standard MM force field.

4. QM calculations of ligand-binding sites

Another possible use of QM methods is to study only the ligand and the closest protein residues. The advantages with such an approach is that no force-field parameters are needed and that electronic polarisation is included in the calculation. The disadvantage is of course that parts of the macromolecule are ignored. Moreover, entropic and dynamic effects are typically ignored.

Several investigations have had this aim. The simplest approach is to compare the raw QM energies directly to experimental binding affinities.[48,49,50,51] However, it is more common to include solvation, either in the way of a number of explicit water molecules[52,53,54] or by a continuum-solvation method.[55,56] Unfortunately, including solvation does not always lead to a consistent improvement in the predicted affinities.[55] Peräkylä & Pekkanen went one step further by including also MM electrostatic interaction energy between the ligand and the rest of the protein,[57,58] thereby approaching a QM/MM approach. Rogachev and coworkers used the QM stacking energies (including a continuum-solvent correction) to enhance a normal docking score.[59]

Most studies have employed the ligand and a rather small number of nearby residues (typically within 4–5 Å, 100–200 atoms). DeChancie and Houk even truncated the ligand.[56] However, Nikita et al. use up to 450 atoms[52] and Thirot & Monard up to over 700 atoms.[51] In one case, the calculations were performed for the ligand and one amino-acid model at turn, i.e. a fractionation approach.[48] Most of the studies have been performed at the semiempirical level (AM1 or PM3), especially with the larger QM systems.[51,52,53,54,55] A few studies were performed at the Hartree–Fock (HF) level with small basis sets (3-21G or 6-31G),[57,58] whereas most recent studies are performed with DFT and basis sets of DZP quality.[48,49,50,56] Only a few studies have been performed at the MP2 level,[56,59] at which dispersion effects are included. However, the basis sets are still too small (6-31G* or 6-31+G**) to give any quantitative results. For example, it has been shown that the interaction energy between biotin and avidin changes by 160 kJ/mol if the basis set is increased from 6-31G* to aug-cc-pVTZ.[60]

Another important effect that needs to be considered in QM binding calculations is the basis-set superposition error. For the binding of drug-like ligands with medium-sized basis sets, this effect is very significant, being 105 and 214 kJ/mol for the binding of biotin to avidin calculated with the 6-31G* basis set at the HF and MP2 levels, respectively.[2] Unfortunately, only a few studies correct for this error.[50,56,57,58]

This approach is often used for metal ligands, because it is likely that the binding affinity is dominated by the metal–ligand interaction, whereas interactions with the surrounding protein may be less important. Then, it is enough to include in the calculations the ligand, the metal, and the other ligating groups. For example, many theoretical studies have been performed on the structure, binding, reactivity, and toxicity of cisplatin and similar anti-cancer drugs.[61,62,63] Likewise this approach has been tested for the binding of M-arene-PTA ligands to cathepsin B (M = Ru^{II}, Os^{II}, Rh^{III}, or Ir^{III}).[63,64,65] It was shown that a correlation of $r^2 = 0.52$ could be obtained between calculated and experimental affinities, including only the first-sphere ligands of the metal. The best results were obtained without any continuum solvation,

but with zero-point, thermal, and entropic corrections to the Gibbs free energy included (from a frequency calculation). The results were also improved by a conformational search of the orientation of the arene ligand.

QM calculations have mainly been used to estimate ligand-binding affinities, but Thirot and Monard have developed a genetic algorithm for protein–ligand docking using semiempirical calculations.[51] Moreover, Zhou & Caflisch have devised an approach to perform extensive virtual screening with semiempirical QM calculations.[66] They calculate interaction energies between small models of polar groups in the binding site and the ligands of interest. Using only five such groups as a model of the human hepatocellular carcinoma receptor B4, they screen ~100 million poses of 2.7 million commercially available compounds, obtained from high-throughput docking. Supplementing the QM energy with hydrophobic matching and ligand-strain calculations, a micromolar inhibitor was obtained after experimental tests of only 23 molecules.

5. QM/MM calculations

At the next level of approximation, the whole macromolecule is included in the calculations by treating the most interesting parts by QM methods, whereas the surroundings are treated by MM methods, the QM/MM approach, which has received much interest in the study of biochemical reactions.[67,68] Often only the ligand is included in the QM system, [35,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83] but for metalloproteins, it is necessary to include all first-sphere ligands of the metal in the QM system.[65,84,85,86] At the highest level of QM/MM approximation, several groups close to the ligand are also included in the QM system. [49,87,88,89,90,91,92,93,94,95,96]

The focus of the QM/MM investigations have varied between polarisation,[71] structures, [78,91] docking (rescoring),[70,79,80,89,90] and affinity estimation.[35,65,72,73,74,75,76,77,81,87,95,96] Only two studies[87,88] used QM methods (MP2/6-31G* or BLYP-D) that includes dispersion, whereas the other studies were performed at the semiempirical (AM1 or sometimes PM3), [69,70,71,72,73,74,75,77,81,91,92,95] DFT,[35,49,65,76,78,79,80,82,84,85,86,90,91,93,94,96] or HF levels.[89,90,93] However, if only the ligand is included in the QM system, the interaction between the ligand and the surroundings is treated by MM, in which dispersion is properly included.

Alzate-Morales et al. used a non-standard approach, in which most of the ligand and four amino-acid side chains were treated by B3LYP/6-31G*, the rest of the ligand was treated with HF/3-21G, whereas ~10 additional residues were treated at the PM3 level (i.e. a three-layer QM/QM/QM approach)[97]

Many studies have considered only the pure QM/MM energies. However, single minimised QM/MM structures will miss effects from dynamics and entropy, and the solvation and long-range electrostatics may be misleading because the systems are truncated after a few layers of water molecules and the outer atoms typically have fixed positions.

Several studies have addressed these problems. Ideally, they should be solved by performing QM/MM FEP, as was done in a study of the relative free energy of binding for five AMP analogues to fructose-1,6-bisphosphatase at the AM1/MM level.[75] They reproduce experimental observations within 1 kJ/mol, but FEPs at the MM level give the same results (within the statistical uncertainty). Essex and coworkers have used QM/MM calculations to improve the results of FEPs at the MM level. They do this by also performing one-step FEPs from the MM to the QM/MM description.[83]

Balaz and coworkers[84,85,86] have instead developed a QM/MM variant of the extended LIE approach with two or three terms, using single-point calculations on averaged MD structures.

Several groups[65,72,76,81,96,98] have used QM/MM calculations in the MM/PBSA approach, by replacing the first three terms in Eqn. 1 with the QM/MM interaction energy. The advantage with this approach is that is does not contain any adjustable parameters and

includes a proper solvation, involving both polar and non-polar terms, as well as entropy and some dynamic effects. The results of the QM/MM-PBSA approaches have been varying. For example, no consistent improvement has been observed compared to a pure DFT study for the binding of seven Ru^{II}-arene-PTA ligands to cathepsin B.[65]

6. QM calculations of entire proteins

Finally, the whole protein can be treated by QM methods. The advantage of this approach is that polarisation of both the ligand and the macromolecule is treated at equal footing and that the calculations allow charge transfer between the macromolecule and the ligand. However, despite the recent progress in computer technology and parallelisation techniques, a standard QM calculation of a full protein-ligand system is technically impossible, but also unnecessary, because electron correlation is a local effect.

Many approximate methods have been developed that take advantage of the locality of QM. Normally, these are referred to as either *linear-scaling methods*, if they compute the electron density for the full system, or *fragmentation methods,* if they combine independent calculations of smaller subsystems to directly compute the quantities of interest. Linear-scaling methods have been recently reviewed,[99] so we will only mention some applications to ligand binding, whereas the fragmentation methods will be somewhat more thoroughly described. It should be noted that some of the linear-scaling methods, such as the divide and conquer (D&C) approach,[100] also use fragmentation as part of their solution.

6.1. Linear-scaling methods

Although it is possible to study at least small proteins with DFT methods,[101] all linear-scaling whole-protein studies of ligand binding have been performed at the semiempirical level, AM1, PM3, PM5, or PM6-DH2.[102,103,104,105,106,107,108,109,110,111,112] This approach was pioneered by Merz and coworkers. They studied 18 carbonic anhydrase and 5 carboxypeptidase inhibitors with the AM1 method, using their D&C approach.[102] They supplemented the QM energy by the dispersive part of the Amber MM potential, a PB-based continuum-solvation energy, a non-polar solvation term based on the surface area burial for heavy atoms, as well as an entropy term, calculated from the number of degrees of freedom that was lost in the protein and the ligand upon binding. Thus, this approach already included most important terms for ligand binding, except dynamic sampling. The authors obtained correlation coefficients (r^2) of 0.69 or 0.80 without or after fitting a single weight factor before the SASA term. In a later study, the same approach (QMScore) was applied to 165 protein–ligand complexes and 49 metalloenzyme complexes, giving r^2 = 0.48 and 0.55 without and with fitting of weights of each of the five terms in the energy function.[105] They have also devised a method for pairwise decomposition of the observed interaction energies to understand the observed differences.[113]

The same group has also developed a more MM/PBSA-like version of this method, in which the QM energies (calculated at the AM1 and PM3 levels) are supplemented by a dispersive term, PB polar solvation, SASA non-polar solvation, and an entropy term from the MM vibrational frequencies, calculated for a truncated model.[104] They compared the binding of a penicillin and a celphalosporin to the TEM-1 b-lactamase. Unfortunately, the standard deviations of the QM/PBSA estimates (20–36 kJ/mol) were too large to give any significant difference in the binding energy between the two substrates.

QMScore has also been used to study the binding of 45 inhibitors to protein kinase B with reasonable results (r^2 = 0.68).[111] However, the main aim of that study was to improve this model by the QSAR-like comparative binding energy approach.

Pichierri has studied the binding of a phosphotyrosyl peptide to the SH2 domain of Lck kinase.[103] He only studied a single complex and included continuum-solvation, but no dispersion, non-polar solvation, entropy, or dynamic effects. Consequently, he reports an

unrealistic binding energy of ~800 kJ/mol.

Sakurai et al. have studied how the binding of an antigen to the 48G7 is improved during maturation.[106] They enhance the QM energies with COSMO continuum-solvation, a non-polar SASA term, and an MM entropy term from an earlier study.

Anikin et al. have developed a semiempirical QM approach to study the special case when many ligands are docked to a fixed protein structure, using a fixed density of the protein.[107] Using only the pure QM energy, they dock 1783 ligands with 30–144 atoms to the FKBP-12 protein with an average time consumption of only 5 minutes per ligand. With a similar approach, 200 000 poses of 20 000 ligands were docked to the p56 LCK SH2 domain.[109]

Zhou and coworkers have tested semiempirical QM calculations in a LIE-like approach.[108] They supplement the QM energy with a PB continuum-solvation term and a van der Waals energy term, all calculated for minimised structures. They fit three multiplicative parameters to the experimental data and compare the QM results to standard LIE calculations (also with a continuum-solvation term). For the binding of 44 peptidic inhibitors to West Nile virus NS3 serine protease, the QM approach gave slightly better results than conventional MM calculations (the RMS error decreased from 4 to 3 kJ/mol), whereas for the binding of 24 peptidic inhibitors to HIV-1 protease and 73 neutral inhibitors to human cyclin-dependent kinase 2, both approaches showed a similar accuracy.

Li & Reynolds have used the PM5 method to estimate the affinities of six stromelysin-1 inhibitors with two different zinc-binding groups (carboxylate or hydroxamate).[110] They supplemented the QM energies with a COSMO continuum-solvation energy and a SASA term, but ignored dispersion, entropy, and dynamics. They obtain an excellent correlation between calculated and experimental affinities ($r^2 = 0.95$), although the calculated relative energies are almost ten times larger than the experimental ones.

Finally, Hobza and coworkers have employed the recent PM6 method, combined with corrections for dispersion and hydrogen bonding.[112] They include entropies calculated from vibrational frequencies calculated at the MM level, as well as deformation and continuum-solvation energies of the ligand. They study the binding of 22 inhibitors of HIV-1 protease and are able to distinguish between binders and non-binders in a docking rescoring, in contrast to conventional DOCK calculation.

6.2. Fragmentation methods

Two types of fragmentation methods have been used for ligand binding. The first aims at calculating the total energy for any system, whereas the second type only calculates binding energies, assuming that the geometry of the macromolecule and the ligand do not change upon binding (as in MM/PBSA). Of course, the former methods are more general and versatile, whereas the latter typically are more effective and faster.

The fragment molecular orbital (FMO) method is probably the most developed and used approach of the first type.[114] For a cluster of n molecules, FMO first optimises the wavefunction of each monomer in separate calculations, including the electrostatic potential from the $n - 1$ other fragments as a fixed external potential. This has to be done iteratively, approximating the potential from distant fragments, e.g. by a point-charge model. Next, the energy of each pair of fragments (dimers) is computed, using the electrostatic potential from the $n - 2$ other fragments as external potential (taken from the converged monomer stage, so no iterations are needed). If greater accuracy is needed, the procedure can be continued by also computing trimers (the FMO3 method), which has been shown to give essentially exact results (cf. Table 1), but to a much higher cost.

If the fragments are connected with covalent bonds, each nucleus and electron pair are assigned to one fragment using chemistry-based rules. The monomer calculation of a given fragment is then performed in the basis set spanned by its assigned nuclei and if the fragment

has a dangling-bond, the valence orbital of the missing atom is included in the basis set by a projection operator.[114]

FMO formulations for many QM methods have been developed, out of which MP2 is most relevant for ligand binding. However, it should be noted that the electron correlation is not treated self-consistently (it is only a dimer energy correction) and thus normally gives a larger error than at the HF level.[115] Recently, the FMO method has been interfaced with the EFP force field[116] and with two implicit solvent models: PCM[117] and PB.[118] However, FMO has two problems relevant to ligand-binding energies. Due to the neglect of Pauli effects,[119] the method does not work well with large and diffuse basis sets, which are needed for a quantitative account of e.g. dispersion. Moreover, no rigorous correction for the basis set superposition error has been presented.

Nevertheless, there have been several applications of FMO to ligand binding. For human oestrogen receptor with eleven ligands, it was found that raw FMO-HF/STO-3G binding energies were much better correlated to experiment ($r^2 = 0.70$) than binding energies from a standard force field.[120] Similarly, a good correlation ($r^2 = 0.83$) was obtained for progesterone receptor with eight ligands.[121] On the other hand, for the FK506 binding protein, studied with a higher level of theory (MP2/6-31G*) and including solvent effects, no correlation with experimental values was found.[122] This demonstrates the significant system-dependence of performances and highlights that it is impossible to assess a theoretical method using only one target. A more pragmatic way is to use FMO interaction energies as a QSAR descriptor, as was done for HIV-1 protease with twelve ligands.[123,124] A simple model using the interaction energy and SASA as descriptors gave $r^2 = 0.85$.

Other complexes studied with FMO include the catabolite activator protein with cAMP,[125] vitamin D receptor with 1α,25-dihydroxyvitamin D3,[126, 127] avian influenza A virus hemagglutinin with human and avian receptors,[128,129,130,131] peroxisome proliferator-activated receptor-gamma with rosiglitazone and farglitazar,[132] cyclic AMP receptor protein with cyclic AMP bound to DNA[133] (for which the interactions were quite different between FMO and MM), and retinoid X receptor with steroid receptor coactivating factor-1 coactivator.[134] Several of these studies were performed at the MP2 level.[127,132,133,134,122,130,131] In most studies, it was assumed that differences in vacuum interaction energy correlates with the binding free energy in water. However, two of the studies included solvation through the PCM method[130,131] and only one of them[130] shows such a correlation. FMO has also been used as an integrated tool for drug development in the visualized cluster analysis of the protein-ligand interaction.[135]

The simplest fragmentation methods of the other type, aiming at calculating binding energies, rely on the approximate atom-wise additivity of interaction energies. For a polymer $P_1P_2..P_n$ (of monomers P_i) interacting with a ligand L, the total interaction energy is related to the sum of P_i–L interactions, but the problem is that the subsystems P_i are not closed-shell systems and thus the P_i–L interaction energies have no meaning. The basic idea of the fragmentation reconstruction method is to cap each P_i system, compute the interaction energies, sum them, and then correct for the artificial introduction of caps.[136] In the original formulation, the correction was done atom-wise, but Zhang and Zhang introduced the notion of conjugate caps,[137] which are simply two neighbouring caps joined together, as is illustrated in Figure 1. By subtracting all interaction energies involving conjugate caps, the effect of the caps is eliminated to a first order. This approach is called the molecular fractionation with conjugate caps (MFCC).

Applications of MFCC to ligand binding include streptavidin with biotin,[137] adipocyte lipid-binding protein with propanoic acid,[138] b-trypsin with benzamidine,[139] HIV-1 protease with various ligands,[140,141] HIV-1 reverse transcriptase with efivarenz and nevirapine,[142,143] a-thrombin with various ligands,[144,145] and p53 with MDM2.[146] Most of these studies employed the very approximate HF/3-21G level of theory, but some have used B3LYP/6-31G* to assess correlation effects.[140,144,145] Several studies have used MP2/6-31G*[142,143] or MP2/6-

31+G*[141,144,145] to get a more quantitative account of the most important interactions, and one study adds bond-centred basis functions to test the limitations of a small basis set.[146] Only one investigation (addressing the interaction energy with a specific water molecule) includes solvation effects through a PCM model,[141] and none of the studies attempts to do any quantitative comparison to experimental binding affinities.

Similar add-and-subtract schemes can also be used to compute total energies. For example, the systematic fragmentation method (SFM)[147] uses classical expressions for well-separated dimers, most recently the EFP expressions.[148] Bettens and Lee adapted the SFM method for protein–ligand interaction energies,[149] by ignoring intra-protein fragment dimers and improved the level of theory to MP2/6-311(+)G(2d,p). They studied the binding of an inhibitor to neuraminidase.

The polarisable multipole interaction with supermolecular pairs (PMISP) method[2,60] combines the MFCC approach for short-range interactions with the NEMO polarisable multipole description for long-range electrostatics and many-body polarisation of the whole system. The method can also be seen as a limiting case of a polarisable force field, in which the short-range terms are computed directly by fragmented QM on the fly. The same idea but implemented using another force field (EFP) is used in the EFMO method.[150]

The accuracy of this approach was tested for a set of ligands binding to avidin[2] and compared to the standard MFCC approach, the electrostatically embedded pairwise additive (EE-PA) model,[151] and the FMO method. As can be seen from Table 1, the inclusion of many-body effects significantly reduces the errors compared to the MFCC model. In fact, the accuracy of the PMISP method is better than that of the computationally more expensive EE-PA and FMO methods, unless trimers are explicitly included. The remaining error for charged ligands (11 kJ/mol) comes from the difficulty of the polarisable multipole model to treat strong interactions and can be reduced by combining the model with an embedding approach.[2]

The PMISP method has been used to compute protein–ligand interaction energies at a high level of theory, MP2/aug-cc-pVTZ.[60] We also investigated the distance-dependence of various approximations to the energy, showing that the explicit QM dimer calculations can be replaced by a sum of a polarisable multipole model and a standard Lennard-Jones term for distances greater than 6 Å, whereas the use of standard MM also for the classical terms has a much more long-range effect, 15–20 Å.

The PMISP method has been interfaced with PCM to provide self-consistent solvation energies and used in an MM/PBSA-like approach to compute binding free energies for seven ligands to avidin at the MP2/cc-pVTZ level using ten MD snapshots per ligand.[47] Unfortunately, no significant improvement was obtained compared to a standard force field, probably because the remaining terms, in particular the non-polar solvation energy, were not simultaneously improved.[152]

7. Concluding remarks

In this paper, we have reviewed various approaches to improve calculated ligand-binding estimates by the use of QM methods. QM methods have many attractive characteristics, e.g. that the accuracy can be systematically improved, that there is no need of any parameterisation, that any systems can be modelled, including metal complexes or systems where covalent bonds are formed or broken, and that polarisation and charge-transfer effects are explicitly accounted for. Moreover, many interesting properties are only available from QM methods. On the other hand, QM methods are time-consuming.

The most important lesson from the applications of QM methods to ligand binding is that even if a method works for one target, it might fail for another. Of course, this type of varying performance is what one would like to avoid by using QM methods. In fact, it has not yet been clearly demonstrated that QM gives better binding free energies than MM, despite the

greater computational cost.

For the simplest QM methods (e.g. semiempirical methods and HF or DFT with small basis sets), this is understandable, because it is not evident that the QM calculations give more accurate interaction energies than well-calibrated MM force fields. Moreover, we have pointed out that dispersion is missing in all QM methods currently used for ligand binding, except MP2 and that large basis sets (at least TZP) are needed to get near-quantitative results. Very few studies have been performed at this level.[47,60,149] It is also important that the basis-set superposition error is properly treated.

However, as has been discussed above, many other terms contribute to the ligand binding, besides the macromolecule–ligand interaction energy, e.g. polar and non-polar solvation, entropy, dynamic effects, and geometric reorganisation. It is most likely that ligand-binding predictions in many cases are limited by these terms rather than the interaction energy. This may explain the varying performance of QM-based ligand-affinity estimates. Although many QM methods include continuum-solvation effects, it is mainly the QM extensions of the LIE and MM/PBSA approaches that include all relevant terms for ligand binding.[47,65,72,76,81,84,85,86,96,102,104,105] Unfortunately, it seems that even methods like LIE and MM/PBSA have problems to provide accurate and consistent binding affinities for all types of macromolecules, probably because of the limitations in the end-point or continuum approximations.[1,16,17,18,99,152,153] The only method that is guaranteed to give correct results (with perfect interaction potentials and sampling) is FEP, which has so far been used only once with a QM potential.[75]

The main problem with the QM methods is the cost of the calculations in terms of computer time, which makes proper sampling problematic. It is interesting to note that the only QM-based method that has reached a wide use for ligand binding also outside the developers is the QPLD method,[27] i.e. the recalculation of the ligand charges by single-point QM/MM calculations during a docking procedure. This is probably because these calculations are rather fast and automatic. However, it also indicates that there are some gain of using QM charges that are obtained for the correct conformation of the ligand, polarised by the surrounding protein.

Thus, even if it yet has not been shown that the expense of QM approaches is offset by improved accuracy in drug design applications,[4,6,9] we can most likely expect an increasing use of QM methods in structure-based drug discovery the coming years. In the nearest future, the largest gain may come from improving the MM description by QM calculations, but at the end we will probably see more and more pure QM calculations, based on combinations of various methods (like QM/MM and PMISP) and perhaps within the FEP approach.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2010-5025) and from the Research school in pharmaceutical science.

References

1 Gohlke, H., Klebe, G. (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed*. **41**, 2644-2676.

2 Söderhjelm, P., Ryde, U. (2009) How accurate can a force field become? - A polarizable multipole model combined with fragment-wise quantum-mechanical calculations *J. Phys. Chem. A*, **113**, 617-627

3 Jensen, F. (2007) *Introduction to computational chemistry*, John Wiley & Sons, Ltd, Chichester.

4 Peters, M. B., Raha, K., Merz, K. M. (2006) Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discov. Develop.* **9**, 370-379.

5 Cavalli, A., Carloni, P., Recanatini, M. (2006) Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* **106**, 3497-3519.

6 Raha, K., Peters, M. B., Wang, B., Yu, N., Wollacott, A. M., Weterhoff, L. M., Merz, K. M. (2007) The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* **12**, 725-731.

7 Zhou, T., Huang, D., Caflisch, A. Quantum mechanical methods for drug design. *Curr. Top. Med. Chem.* **10**, 33-45.

8 Menikarachchi, L. C., Gascón, J. A. (2010) QM/MM approaches in medicinal chemistry research. *Curr. Top. Med. Chem*. **10**, 46-54.

9 Söderhjelm, P., Kongsted, J., Genheden, S., Ryde, U. (2010) Estimates of ligand-binding affinities supported by quantum mechanical methods. *Interdiscip. Sci. Comput. Life Sci.*, **2**, 21-37

10 Ryde, U. (2007) Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry, *Dalton Trans.* 607-625.

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

12 Mulholland, A. J. (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug. Discov. Today* **10**, 1393-1402.

13 Karelson, M., Lobanov, V. S., Katritzky, A. R. (1996) Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **96**, 1027-1044.

14 Khandogin J., York, D. M. (2004) Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins*. **56**, 724-737.

15 Dixon, S., Merz, K. M., Lauri, G., Ianni, J. C. (2005) QMQSAR: Utilization of a semiempirical probe potential in a field-based QSAR method. *J. Comput. Chem.* **26**, 23-34.

16 Christ, C. D., Mark, A. E., van Gunsteren, W. E. (2010) Basic ingredients of free energy calculations: A review. *J. Comput. Chem.* **31**, 1569-1582.

17 Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham III, T. E. (2000) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* 33, 889-897.

18 Brandsdal, B. O., Österberg, F., Almlöf, M., Feierberg, I., Luzhkov, V. B., Åqvist, J. (2003) Free energy calculations and ligand binding. *Adv. Prot. Chem.* 66, 123-158.

19 Almlöf, M, Carlsson, J, Åqvist, J (2007) Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J. Chem. Theory. Comput.* **3**, 2162–2175.

20 Tominaga, Y., Jorgensen, W. L. (2004) General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* **47**, 2534-2549.

21 Gresh, N., Cisneros, G. A., Darden, T. A., Piquemal, J.-P. (2007) Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and

Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **3**, 1960-1986

22 Gordon, M. S., Slipchenko, L. V., Li, H., Jensen, J. H. (2007) The Effective Fragment Potential: a general method for predicting intermolecular interactions. Annual Reports in Computational Chemistry, Volume 3, pp 177-193

23 Engkvist, O., Åstrand, P.-O., Karlström, G. (2000) Accurate Intermolecular Potentials Obtained from Molecular Wave Functions: Bridging the Gap between Quantum Chemistry and Molecular Simulations. *Chem. Rev.* **100**, 4087-4108

24 Hayes, J. M., Stein, M., Weiser, J. (2004) Accurate calculations of ligand binding free energies: Chiral separation with enatioselective receptors. *J. Phys. Chem. A* **108**, 3572-3580.

25 Curioni, A., Mordasini, T., Andreoni, W. (2004) Enhancing the accuracy of virtual screening: molecular dynamics with quantum-refined force fields. J. *Comp.-Aided Mol. Design* **18**, 773-784.

26 Ferrara, P., Curioni, A., Vangrevelinghe, E., Meyer, T., Mordasini, T., Andreoni, W., Ackling , Jacoby, E. (2006) New scoring functions for virtual screening from molecular dynamics simulations with a quantum-refined force-field (QRFF-MD). Application to cyclin-dependent kinase 2. *J. Chem. Inf. Model*. **46**, 254-263.

27 Cho, A. T., Guallar, V., Berne, B. J., Friesner, R. (2005) Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) aproach. *J. Comput. Chem.* **26**, 915-931.

28 Sander, T., Liljefors, T., Balle, T. (2008) Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis. *J. Mol. Graph. Model.* **26**, 1259-1268.

29 Somsak, L.; Czifrak, K.; Toth, M.; Bokor, E.; Chrysina, E. D.; Alexacou, K. M.; Hayes, J. M.; Tiraidis, C.; Lazoura, E.; Leonidas, D. D.; Zographos, S. E.; Oikonomakos, N. G. (2008) New Inhibitors of Glycogen Phosphorylase as Potential Antidiabetic Agents. *Curr. Med. Chem*. **15**, 2933-2983.

30 Zhong, H., Kirschner, K. N., Lee, M. Bowen, J. P. (2008) Binding free energy calculation for duocarmycin/DNA complex based on the QPLD-derived partial charge model. *Bioorg. Med. Chem. Lett.* **18**, 542-545.

31 Sgrignai, J., Bonaccini, C., Grazioso, G., Chioccioli, M., Cavalli, A., Gratteri, P. (2009) Insights into docking and scoring neuronal a4b2 nicotinic receptor agonists using molecular dynamics simulations and QM/MM calculations. *J. Comp. Chem.* **30**, 2443-2454.

32 Benltifa, M.; Hayes, J. M.; Vidal, S.; Gueyrard, D.; Goekjian, P. G.; Praly, J. P.; Kizilis, G.; Tiraidis, C.; Alexacou, K. M.; Chrysina, E. D.; Zographos, S. E.; Leonidas, D. D.; Archontis, G.; Oikonomakos, N. G. (2009) Glucose-based spiro-isoxazolines: A new family of potent glycogen phosphorylase inhibitors. *Bioorg. Med. Chem*. 17, 7368-7380.

33 Das, D., Koh, Y., Tojo, Y., Ghosh, A. K., Mitsuya, H. (2009) Prediction of potency of protease inhibitors using free energy simulations with polarizable quantum mechanics-based ligand charges and a hybrid water model. *J. Chem. Inf. Model.* **49**, 2851-2862.

34 Koldso, H., Severinsen, K., Tran, T. T., Celik, L., Jensen, H. H., Wiborg, O. Schiott, B., Sinning, S. (2010) The two enantiomers of citalopram bind to the human serotonin transporter in reversed orientations. *J. Am. Chem. Soc.* **132**, 1311-1322.

35 Tsirkone, V. G., Tsoukala, E. Lamprakis, C., Manta, S., Hayes, J. M., Skamnaki, V. T., Drakou, C., Zographos, S. E., Komiotis, D., Leonidas, D. D. (2010) 1-(3-Deoxy-3-fluoro-β-d-glucopyranosyl) pyrimidine derivatives as inhibitors of glycogen phosphorylase b: Kinetic, crystallographic and modelling studies. *Bioorg. Med. Chem.* **18**, 3413-3425.

36 Cho, A. E. (2007) Effect of quantum mechanical charges in binding sites of metalloproteins. *Biochip J.* **1**, 70-75.

37 Cho, A. E. (2008) Quantum mechanical calculations for binding sites of metalloproteins. *Biochip J.* **2,** 148-153.
38 Cho, A. E., Rinaldo, D. (2009) Extension of QM/MM docking and its applications to metalloproteins. *J. Comput. Chem.* **30**, 2609-2616.
39 Zhong, H, Kirschner, K. N., Lee, M., Bowen, J. P. (2008) Binding free energy calculation for duocarmycin/DNA complex based on the QPLD-derived partial charge model. *Bioorg. Med. Chem. Lett.* **18**, 542-545.
40 Illingworth, C. J. R., Morris, G. M., Parkes, K. E. B., Snell, C. R., Reynolds, C. A. (2008) Assessing the role of polarization in docking. *J. Phys. Chem. A* **112**, 12157-12163.
41 Fischer, B., Fukuzawa, K., Wenzel, W. (2008) Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins* **70**, 1264-1273.
42 Tong, Y., Mei, Y., Li, Y. L., Ji, C. G., Zhang, J. Z. H. (2010) Electrostatic polarisation makes a substantial contribution to the free energy of avidin-biotin binding. *J. Am. Chem. Soc.* **132**, 5137-5142.
43 Weis, A., Katebzadeh, K., Söderhjelm, P., Nilsson, I., Ryde, U. (2006) Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field, *J. Med. Chem.* **49**, 6596-6606
44 Söderhjelm, P., Ryde, U. (2009) Conformational dependence of charges in protein simulations. *J. Comput. Chem.*, **30**, 750-760.
45 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269-10280.
46 Genheden, S., Söderhjelm, P.,Ryde, U. (2011) Transferability of conformational dependent charges from protein simulations. *Int. J. Quant. Chem.*, in press; DOI 10.1002/qua.22967.
47 Söderhjelm, P., Kongsted, J., Ryde U. (2010) Ligand affinities estimated by quantum chemical calculations. *J. Chem. Theory Comput.*, **6**, 1726-1737.
48 Morreale, A., Maseras, F., Iriepa, I, Gálvez, E. (2002) Ligand-receptor interaction at the neural nicotinic acetylcholine binding site: a theoretical model. *J. Mol. Graph. Model.* **21**, 111-118.
49 Saen-oon, S., Aruksakunwong, O., Wittayanarakul, K., Somornpisut, P., Hannongbua, S. (2007) Insight into analysis of interactions of saquinavir with HIV-1 protease in comparison between the wild-type and G48V and G48V/L90M mutants based on QM and QM/MM calculations. *J. Mol. Graph. Model.* **26**, 720-727.
50 Rosso, L., Gee, A. D., Gould, I. R. (2008) Ab initio computational study of positron emission tomography ligands interacting with lipid molecule for the prediction of nonspecific binding. *J. Comput. Chem.* **29**, 2397-2405.
51 Thirot, E., Monard, G. (2009) Combining a genetic algorithm with a linear scaling semiempirical method for protein-ligand docking. *J. Mol. Struct. (THEOCHEM)* **898**, 31-41.
52 Nikitina, E., Sulimov, V., Zayets, V., Zaitseva, N. (2004) Semiempirical calculations of binding enthalpy for protein-ligand complexes. *Int. J. Quant. Chem.* **97**, 747-763.
53 Villar, R., Gil, M. J., Garcia, J. I., Martínez-Merino, V. (2005) Are AM1 ligand-protein binding enthalpies good enough for use in the rational design of new drugs? *J. Comput. Chem*. **26**, 1347-1358.
54 Nikitina, E., Sulimov, V., Grigoriev, F., Kondakova, O., Luschekina, S. (2006) Mixed implicit/explicit solvation models in quantum mechanical calculations of binding enthalpy for protein-ligand complexes. *Int. J. Quant. Chem.* **106**, 1943-1963.
55 Vasilyev, V., Bliznyuk, A. (2004) Application of semiempirical quantum chemical methods as a scoring function in docking. *Theo. Chem. Acc.* **112**, 313-317.

56 DeChancie, J., Houk, K. N. (2007) The origins of femtomolar protein-ligand binding: Hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *J. Am. Chem. Soc.* **129**, 5419-5429.

57 Peräkylä, M., Pakkanen, T. A. (1994) Quantum mechanical model assembly study on the energetics of binding of arabinose, fucose, and galactose to L-arabinose-binding protein. *Proteins*, **20**, 367-372.

58 Peräkylä, M., Pakkanen, T. A. (1995) Model assembly study of the ligand binding by *p*-hydroxybenzoate hydroxylase: Correlation between the calculated binding energies and the experimental dissociation constants. *Proteins*, **21**, 22-29.

59 Rogacheva, O. N., Popov, A. V., Savvateeva-Popova, E. V., Stefanov, V. E., Shchegolev, B. F. (2010) Thermodynamic analysis of protein kinase A Iα activation. *Biochem. (Moscow)* **75**, 233-241.

60 Söderhjelm, P., Aquilante, F., Ryde, U. (2009) Calculation of Protein-Ligand Interaction Energies by a Fragmentation Approach Combining High-Level Quantum Chemistry with Classical Many-Body Effects. *J. Phys. Chem. B*, **113**, 11085-11094.

61 McCarthy, S. L., Hidne, R. J., Miller, K. J., Anderson, J. S., Basch, H., Krauss, M. (1990) Theoretical studies of cis-Pt(II)-diammine binding to duplex DNA. *Biopol.* **29**, 823-836.

62 Dans, P. D., Coitiño E. L. (2009) Density functional theory characterization and descriptive analysis of cisplatin and related compounds. *J. Chem. Inf. Model.* **49**, 1407-1419.

63 Ciancetta, A. (2010) Density functional theory and combined QM/MM studies on selected transition metal based anticancer complexes. *Ph.D. Thesis* Universitá degli studi "G. D'Annunzio", Chieti.

64 Casini, A., Edafe, F., Erlandsson, M., Gonsalvi, Ciancetta, A., Re, N., Ienco, A., Messori, L., Peruzzini, M., Dyson, P. J. (2010) Rationalization of the inhibition activity of structurally related organometallic compounds against the drug target cathepsin B by DFT. *Dalton Trans.* **39**, 5556-5563.

65 Ciancetta, A., Genheden, S., Ryde, U. (2011) A. QM/MM study of the binding of RAPTA ligands to cathepsin B. *J. Comput.-Aided Mol. Design*, **25**, 729-742.

66 Zhou, T., Caflisch, A. (2010) Virtual screening using quantum mechanical probes: Discovery of selective kinase inhibitors. *Chem. Med. Chem.* **5**, 1007-1014.

67 Senn H. M.; Thiel W. (2009) QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* 48, 1198-1229.

68 Ryde, U. (2003) Combined quantum and molecular mechanics calculations on metalloproteins, *Curr. Opin. Chem. Biol.,* **7**, 136-142.

69 Alex, A., Finn, P. (1997) Fast and accurate predictions of relative binding energies. *J. Mol. Struct. THEOCHEM* 398-399, 551-554.

70 Beierlein, F., Lanig, H., Schürer, G., Horn, A. H. C., Clark, T. (2003) Quantum mechanical/molecular mechanical (QM/MM) docking: an evaluation for known test systems. *Mol. Phys.* **15**, 2469-2480.

71 Hensen, C., Hermann, J. C., Nam, K., Ma, S., Gao, J., Höltje, H.-D. (2004) A combined QM/MM approach to protein-ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors *J. Med. Chem.* **47**, 6673-6680.

72 Gräter, F., Schwarzl, S. M., Dejaegere, A., Fischer, S., Smith, J. C. (2005) Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J. Phys. Chem. B* **109**, 10474-10483.

73 Alves C. N., Marti, S., Castillo, R., Andrés, J., Moliner, V., Tuñón, I., Silla, E. (2007) A quantum mechanics/molecular mechanics study of the protein-ligand interaction for inhibitors of HIV-1 integerase. *Chem. Eur. J.* **13**, 7715-7724.

74 Alzate-Morales, J. H., Contreras, A., Soriano, A., Tuñon, I., Silla, E. (2007) A computational study of the protein-ligand interactions in CDK2 inhibitors: Using

quantum mechanics/molecular mechanics interaction energy as a predictor of the biological activity. *Biophys. J.* **92**, 430-439.

75 Reddy, M. R., Erion, M. D. (2007) Relative binding affinities of fructose-1,6-bisophsphatase inhibitors calculated using a quantum mechanics-base free energy perturbation method. *J. Am. Chem. Soc.* **129**, 9296-9297.

76 Wang, M., Wong, C. F. (2007) Rank-ordering protein-ligand binding affinity by a quantum mechanics/molecular mechanics/Poisson-Boltzmann-surface area model. *J. Chem. Phys.* **126**, 026101.

77 Alves, C. N., Marti, S., Castillo, R., Andrés, J., Moliner V., Tuñón, I, Silla, E. (2008) A quantum mechanic/molecular mechanic study of the wild-type and N155S mutant HIV-1 integrase complexed with diketo acid. *Biophys. J.* **94**, 2443-2451.

78 Fanfrlík, J., Brynda, J., Rezác, J, Hobza, P. Liepsík, M. (2008) Interpretation of protein/ligand crystal structure using QM/MM calculations: Case of HIV-1 protease/metallacarborane complex. *J. Phys. Chem. B* **112**, 15094-05102.

79 Parks, J. M., Kondru, R. K., Hu, H., Beratan D. N., Yang, W. (2008) Hepatitis C virus NS5B polymerase: QM/MM calculations show the important role of the internal energy in ligand binding. *J. Phys. Chem. B* **112**, 3168-3176.

80 Cho, A. E., Chung, J. Y., Kim, M., Park, K. (2009) Quantum mechanical scoring for protein docking. *J. Chem. Phys.* **131**, 134108.

81 Retegan, M., Milet, A., Jamet, H. (2009) Exploring the binding of inhibitors derived from tetrabrombenzimidazole to the CK2 protein using a QM/MM-PB/SA approach *J. Chem. Inf. Model.* **49**, 963-971.

82 Burger, S. K., Thompson, D. C., Ayers, P. W. (2011) Quantum mechanics/molecular mechanics strategies for docking pose refinement: Distinguishing between binders and decoys in cytochrome c peroxidas. *J. Chem. Inf. Model.* **51**, 93-101.

83 Beierlein, F. R.; Michel, J.; Essex, J. W. (2011) *J. Phys. Chem. B* **115**, 4911-4926.

84 Khandelwal A., Lukacova, V., Comez, D., Kroll, D. M., Raha, S., Balaz, S. (2007) A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **48**, 5437-5447.

85 Khandelwal A., Balaz, S. (2007) Improved estimation of ligand-macromolecule binding affinities by linear response approach using a combination of multi-mode MD simulation and QM/MM methods. *J. Comput.-Aided Mol. Des.* **21**, 131-137.

86 Khandelwal A., Balaz, S. (2007) QM/MM linear response method distinguishes ligand affinities for closely related metalloproteins. *Proteins* **69**, 326-339.

87 Saen-oon, S., Kuno, M., Hannongbua, S. (2005) Binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-Cl TIBO complex structures: Quantum chemical calculations based on the OINOM method. *Proteins* **61**, 859-869.

88 Morgado, C. A., Hillier, I. H., Burton, N. A., McDouall, J. J. W. (2008) 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*. **10**, 2706-2714

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

90 Gleeson, M. P., Gleeson, D. (2009) QM/MM as a tool in fragment based drug discovery: A cross-docking, rescoring study of kinase inhibitors. *J. Chem. Inf. Model* **49**, 1437-1448.

91 Gleeson, M. P., Gleeson, D. (2009) QM/MM calculations in drug discovery: A useful method for studying binding phenomena? *J. Chem. Inf. Model* 49, 670-677.

92 Li, Y., Yang, Y., He, P., Yang, Q. (2009) QM/MM study of epitope peptides binding to HLA-A*0201: The roles of anchor residues and water. *Chem. Biol. Drug Des.* **74**, 611-618.

93 Li, Q., Gusarov, S., Evoy, S., Kovalenko, A. (2009) Electronic structure, binding energy, and solvation structure of the strepavidin-biotin supramolecular complex: ONIOM and 3D-RISM study. *J. Phys. Chem. B* **113**, 9958-9967.
94 Ruiz, R., García, B., Ruisi, G., Silvestri, A., Barone, G. (2009) Computational study of the interaction of proflavine with d(ATATATATAT)$_2$ and d(GCGCGCGCGC)$_2$. *J. Mol Struct. THEOCHEM* **915**, 86-92.
95 Hayik, S. A., Dunbrack, R., Merz, K. M. (2010) Mixed quantum mechanics/molecular mechanics scoring function to predict protein-ligand binding affinity. *J. Chem. Theory Comput.* **6**, 3079-3091.
96 Shi, J., Lu, Z., Zhang, Q, Wang, M., Wong, C. F., Liu, J. (2010) Supplementing the PBSA approach with quantum mechanics to study the binding between CDK2 and N2-substituted O6-cyclohexylmethoxyguanine inhibitors. *J. Theor. Comput. Chem.* **9**, 543-559.
97 Alzate-Morales, J. H., Caballero, J., Vergara Jague A., Gonzáles Nilo, F. D. (2009) Insights into the structural basis of N2 and O6 substituted guanine derivatives as cyclin-depndent kinase 2 (CDK2) inhibitors: Prediction of the binding modes and potency of the inhibitors by docking and ONIOM calculations. *J. Chem. Inf. Model*. **49**, 886-899.
98 Kaukonen, M., Söderhjelm, P., Heimdal, J., Ryde, U. (2008) A QM/MM-PBSA method to estimate free energies for reactions in proteins. *J. Phys. Chem. B* **112**, 12537-12548.
99 Ochsenfeld, C., Kussmann, J., Lambrecht, D. S. (2007) Linear-Scaling Methods in Quantum Chemistry, in Reviews in Computational Chemistry, edited by K. B. Lipkowitz and T. L. Cundari, volume 23, pages 1-82, VCH Publishers, New York
100 Yang, W. (1991) Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **66**, 1438-1441.
101 Sulpizi, M. Raugei, S. VandeVondele, J. Carloni, P. Sprik, M. (2007) Calculation of Redox Properties: Understanding Short- and Long-Range Effects in Rubredoxin *J. Phys. Chem. B* **111**, 3969-3976.
102 Raha, K., Merz, K. M. (2004) A quantum mechanics-based scoring function: Study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.* **126**, 1020-1021.
103 Pichierri, F. (2004) A quantum mechanical study on phosphotyrosyl peptide binding to the SH2 domain of p56^{lck} tyrosine kinase with insights into the biochemistry of intracellular signal transduction events. *Biophys. Chem.* **109**, 295-304.
104 Díaz, N., Suárez, D., Merz, K. M., Sordo, T. L. (2005) Molecular dynamics simulations of the TEMP-1 beta-lactamase complexed with cephalothin. *J. Med. Chem.* **48**, 780-791.
105 Raha, K., Merz, K. M. (2005) Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **48**, 4558-4575.
106 Ohno, K., Wada, M., Saito, S., Inoue, Y., Sakurai, M. (2005) Quantum chemical study on the affinity maturation of 48G7 antibody. *J. Mol. Struct. THEOCHEM* **772**, 203-211.
107 Anikin, N. A., Andreev, A. M., Kuzminskii, M. B., Mendkovich, A. S. (2008) A fast method of large-scale serial semiempirical calculations of docking complexes. *Russ. Chem. Bull., Int. Ed.* **57**, 1793-1798.
108 Zhou, T., Huang, D., Caflisch, A. (2008) Is quantum mechanics necessary for predicting binding free energy? *J. Med. Chem.* **51**, 4280-4288.
109 Anisimov, V. M., Bugaenko, V. L.(2009) QM/QM docking method based on the variational finite localized molecular orbital approximation. *J. Comput. Chem.* **30**, 784-798.
110 Li, J., Reynolds, C. H. (2009) A quantum mechanical approach to ligand binding – Calculation of ligand-protein binding affinities for stromelysin-1 (MMP-3) inhibitors. *Can. J. Chem.* **87**, 1480-1484.
111 Zhang, X., Gibbs, A. C., Reynolds, C. H., Peters, M. B., Westerhoff, L. M. (2010)

Quantum mechanical pairwise decomposition analysis of protein kinase B inhibitors. Validating a new tool for guiding drug design. *J. Chem. Inf. Model*, **50**, 651-661.

112 Fanfrlík, J.,Bronowska, A. K.; Rezác, J., Prenosil, O, Konvalinka, J., Hobza, P. (2010) 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* **114**, 12666-12678.

113 Raha, K. van der Vaart, A. J., Riley, K. E., Peters, M. B., Westerhoff, L. M., Kim, H., Merz, K. M. (2005) Pairwise decomposition of residue interaction energies using semiempirical quantum mechanical methods in studies of protein-ligand interactions. *J. Am. Chem. Soc.* **127**, 6583-6594.

114 Fedorov, D.G., Kitaura, K. (2007) Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **111**, 6904-6914.

115 Fedorov, D. G., Ishimura, K., Ishida, T., Kitaura, K., Pulay, P., Nagase, S. (2007) Accuracy of the three-body fragment molecular orbital method applied to Møller–Plesset perturbation theory. *J. Comp. Chem.* **28**, 1476-1484.

116 Nagata, T., Fedorov, D. G., Kitaura, K., Gordon, M. S. (2009) A combined effective fragment potential-fragment molecular orbital method. I. The energy expression and initial applications. *J. Chem. Phys.* **131**, 024101.

117 Fedorov, D. G., Kitaura, K., Li, H., Jensen, J.H., Gordon, M.S. (2006) The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (FMO). *J. Comp. Chem.* **27**, 976-985.

118 Watanabe, H., Okiyama, Y., Nakano, T., Tanaka, S. (2010) Incorporation of solvation effects into the fragment molecular orbital calculations with the Poisson-Boltzmann equation. *Chem. Phys. Lett.* **500**, 116-119.

119 Söderhjelm, P., Öhrn, A., Ryde, U., Karlström, G. (2008) Accuracy of typical approximations in classical models of intermolecular polarization, *J. Chem. Phys.* **128**, 014102

120 Fukuzawa, K., Kitaura, K., Uebayasi, M., Nakata, K., Kaminuma, T., Nakano, T. (2005) Ab initio quantum mechanical study of the binding energies of human estrogen receptor with its ligands: An application of fragment molecular orbital method. *J. Comp. Chem.* **26**, 1-10.

121 Harada, T., Yamagishi, K., Nakano, T., Kitaura, K., Tokiwa, H. (2008) Ab initio fragment molecular orbital study of ligand binding to human progesterone receptor ligand-binding domain. *Naunyn-Schmiedebergs Archives of Pharmacology* **377**, 607-615.

122 Nakanishi, I., Fedorov, D. G., Kitaura, K. (2007) Molecular recognition mechanism of FK506 binding protein: An all-electron fragment molecular orbital study. *Proteins: Structure, Function, and Bioinformatics* **68**, 145-158.

123 Yoshida, T., Yamagishi, K., Chuman, H. (2008) QSAR study of cyclic urea type HIV-1PR inhibitors using ab initio MO calculation of their complex structures with HIV-1PR. *QSAR & Combinatorial Science* **27**, 694-703.

124 Yoshida, T., Fujita, T., Chuman, H. (2009) Novel Quantitative Structure-Activity Studies of HIV-1 Protease Inhibitors of the Cyclic Urea Type Using Descriptors Derived from Molecular Dynamics and Molecular Orbital Calculations. *Current Computer-aided Drug Design* **5**, 38-55.

125 Sugiki, S., Matsuoka, M., Usuki, R., Sengoku, Y., Kurita, N., Sekino, H., Tanaka, S. J. (2005) Density functional calculations on the interaction between catabolite activator protein and cyclic AMP using the fragment molecular orbital method. *J. Theor. Comp. Chem.* **4**, 183.

126 Yamagishi, K., Yamamoto, K., Yamada, S., Tokiwa, H. (2006) Functions of key residues in the ligand-binding pocket of vitamin D receptor: Fragment molecular orbital-interfragment interaction energy analysis. *Chem. Phys. Lett.* **420**, 465-468.

127 Yamagishi, K., Tokiwa, H., Makishima, M., Yamada, S. (2010) Interactions between 1 alpha,25(OH)(2)D-3 and residues in the ligand-binding pocket of the vitamin D receptor: A correlated fragment molecular orbital study. *Journal of Steroid Biochemistry and Molecular Biology* **121**, 63-67.

128 Sawada, T., Hashimoto, T., Nakano, H., Suzuki, T., Ishida, H., Kiso, M. (2006) Why does avian influenza A virus hemagglutinin bind to avian receptor stronger than to human receptor? Ab initio fragment molecular orbital studies. *Biochemical and biophysical research communications* **351**, 40-43.

129 Sawada, T., Hashimoto, T., Nakano, H., Suzuki, T., Suzuki, Y., Kawaoka, Y., Ishida, H., Kiso, M. (2007) Influenza viral hemagglutinin complicated shape is advantageous to its binding affinity for sialosaccharide receptor. *Biochemical and biophysical research communications* **355**, 6-9.

130 Sawada, T., Fedorov, D. G., Kitaura, K. (2010) Role of the Key Mutation in the Selective Binding of Avian and Human Influenza Hemagglutinin to Sialosides Revealed by Quantum-Mechanical Calculations. *J. Am. Chem. Soc.* **132**, 16862-16872.

131 Sawada, T., Fedorov, D. G., Kitaura, K. (2010) Binding of Influenza A Virus Hemagglutinin to the Sialoside Receptor Is Not Controlled by the Homotropic Allosteric Effect. *J. Phys. Chem. B* **114**, 15700-15705.

132 Yamagishi, K., Yamamoto, K., Mochizuki, Y., Nakano, T., Yamada, S., Tokiwa, H. (2010) Flexible ligand recognition of peroxisome proliferator-activated receptor-gamma (PPAR gamma). *Bioorg. Med. Chem. Lett.* **20**, 3344-3347.

133 Fukuzawa, K., Komeiji, Y., Mochizuki, Y., Kato, A., Nakano, T., Tanaka, S. (2006) Intra-and intermolecular interactions between cyclic-AMP receptor protein and DNA: Ab initio fragment molecular orbital study. *J. Comp. Chem.* **27**, 948-960.

134 Ito, M., Fukuzawa, K., Mochizuki, Y., Nakano, T., Tanaka, S. (2007) Ab initio fragment molecular orbital study of molecular interactions between liganded retinoid X receptor and its coactivator: roles of helix 12 in the coactivator binding mechanism. *J. Phys. Chem. B* **111**, 3525-3533.

135 Amari, S., Aizawa, M., Zhang, J., Fukuzawa, K., Mochizuki, Y., Iwasawa, Y., Nakata, K., Chuman, H., Nakano, T. (2006) VISCANA: Visualized Cluster Analysis of Protein-Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening. *J. Chem. Inf. Model* **46**, 221-230.

136 Amovilli, C., Cacelli, I., Campanile, S., Prampolini, G. (2002) Calculation of the intermolecular energy of large molecules by a fragmentation scheme: Application to the 4-n-pentyl-4′- cyanobiphenyl (5CB) dimer. *J. Chem. Phys.* **117**, 3003-3012.

137 Zhang, D. W., Xiang, Y., Zhang, J. Z. H. (2003) New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin- Biotin Interaction Energy. *J. Phys. Chem. B* **107**, 12039-12041.

138 Xiang, Y., Zhang, D. W., Zhang, J. Z. H. (2004) Fully quantum mechanical energy optimization for protein-ligand structure. *J. Comp. Chem.* **25**, 1431-1437.

139 Zhang, D. W., Xiang, Y., Gao, A. M., Zhang, J. Z. H. (2004) Quantum mechanical map for protein-ligand binding with application to beta-trypsin/benzamidine complex. *J. Chem. Phys.* **120**, 1145-1148.

140 Zhang, D. W., Zhang, J. Z. H. (2005) Full quantum mechanical study of binding of HIV-1 protease drugs. *Int. J. Quantum Chem.* **103**, 246-257.

141 Tong, Y., Mei, Y., Zhang, J. Z. H., Duan, L. L., Zhang, Q. G. (2009) Quantum calculation of protein solvation and protein-ligand binding free energy for hiv-1 protease/water complex. *J. Theor. Comp. Chem.* **8**, 1265-1279.

142 He, X., Mei, Y., Xiang, Y., Zhang, D. W., Zhang, J.Z.H. (2005) Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations. *Proteins: Structure, Function, and Bioinformatics* **61**, 423-432.

143 Mei, Y., He, X., Xiang, Y., Zhang, D.W., Zhang, J.Z.H. (2005) Quantum study of

mutational effect in binding of efavirenz to HIV-1 RT. *Proteins: Structure, Function, and Bioinformatics* **59**, 489-495.
144 Wu, E. L., Mei, Y., Han, K. L., Zhang, J. Z. H. (2007) Quantum and Molecular Dynamics Study for Binding of Macrocyclic Inhibitors to Human a-Thrombin. *Biophysical Journal* **92**, 4244-4253.
145 Wu, E. L., Han, K. L., Zhang, J. Z. H. (2009) Computational study for binding of oscillarin to human alpha-thrombin. *J. Theor. Comp. Chem.* **8**, 551-560.
146 Ding, Y., Mei, Y., Zhang, J. Z. H. (2008) Quantum mechanical studies of residue-specific hydrophobic interactions in p53-MDM2 binding. *J. Phys. Chem. B* **112**, 11396-11401.
147 Collins, M. A., Deev, V. A. (2006) Accuracy and efficiency of electronic energies from systematic molecular fragmentation. *J. Chem. Phys.* **125**, 104104.
148 Mullin, J. M., Roskop, L. B., Pruitt, S. R., Collins, M. A., Gordon, M. S. (2009) Systematic Fragmentation Method and the Effective Fragment Potential: An Efficient Method for Capturing Molecular Energies. *J. Phys. Chem. A* **113**, 10040-10049.
149 Bettens, R. P. A., Lee, A.M. (2007) On the accurate reproduction of ab initio interaction energies between an enzyme and substrate. *Chem. Phys. Lett.* **449**, 341-346.
150 Steinmann, C., Fedorov, D. G., Jensen, J. H. (2010) Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **114**, 8705-8712.
151 Dahlke, E. E., Truhlar, D. G. (2007) Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput* **3**, 46-53.
152 Genheden, S., Kongsted, J., Söderhjelm, S., Ryde, U. (2010) Nonpolar solvation free energies of protein–ligand complexes, *J. Chem. Theory Comput.* **6**, 3558-3568.
153 Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., Ryde, U. (2010) An MM/3D-RISM approach for ligand-binding affinities. *J. Phys. Chem. B* **114**, 8505-8516

Table 1. Mean absolute errors in kJ/mol for interaction energies between a 216-atom model of avidin and three charged ligands (12 structures) or four neutral ligands (4 structures), respectively, relative to the exact HF/6-31G* results.[2]

Method	Charged	Neutral
MFCC	29.4	6.2
EE-PA	14.2	4.2
FMO	12.1	3.8
FMO3	0.9	0.3
PMISP	11.0	1.1

Figure 1. The cutting scheme used for a polypeptide in e.g. the MFCC and PMISP methods, giving two capped fragments (middle row) and a conjugated caps fragment (lower row), the interactions of which are subtracted instead of added.