Modeling this phenomenon is more of a challenge when classical potentials are used.

Linear-scaling QM and binding affinity calculation

The QM/MM approaches described above clearly show promise for calculating the binding affinity in protein/ligand interaction. However, it is obvious from the above discussion that, first, these approaches still require extensive sampling of ligand/receptor conformations through molecular simulation and are very time-consuming and, second, in all RBDD QM/MM studies reported to date, only the ligand is treated quantum mechanically, because including even small portions of the protein is computationally too expensive. Third, if bonded regions of the protein/ligand complex are to be divided into QM and MM regions, then there are well-documented pitfalls associated with the boundary region in the QM/MM approach.[81]

These problems have to some extent been surmounted by the development of linear-scaling QM technology in the past decade. Semiempirical Hamiltonians such as AM1 and PM3 can now be employed to calculate the molecular wave function for proteins with thousands of atoms. One of the first applications of linear-scaling methods to RBDD was reported by Raha and Merz, where they calculated the binding affinity of ligands bound to the metalloenzyme human carbonic anhydrase with reasonable accuracy.[82] As described by the authors, the free energy of binding in solution was calculated using the following set of equations:

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

Here, the free energy of binding in solution was calculated as the sum of the gas phase interaction energy and a solvation correction. The gas phase interaction energy consisted of enthalpic and entropic components. The electrostatic part of the enthalpic component was calculated with the program DivCon, using semiempirical Hamiltonians. The solvation correction was calculated as a difference between the solvation free energies of the protein/ligand complex (PL) with the protein (P) and the ligand (L) free in solution. The solvation free energy was calculated using a Poisson-Boltzmann based self-consistent reaction field (PB/SCRF) method in which the polarization of the solute electron density due to the presence of the solvent reaction field is calculated self-consistently using a QM Hamiltonian.[83] This is a major advantage of using the QM-based solvation method wherein the dielectric relaxation (or the internal dielectric) of the protein in response to a solvent reaction field is not preset.

In subsequent studies, the authors carried out a very large-scale and detailed validation of this quantum-mechanics-based scoring function, named QMScore, for predicting binding affinity. They calculated interaction energies for a diverse range of protein/ligand complexes comprising of 165 noncovalent complexes and 49 metalloenzyme complexes.[63] For the 165 noncovalent complexes the interaction energies, without any fitting, agreed with experimental binding affinity within 2.5 kcal/mol. When different parts of the scoring function were fit to the experimental free energy of binding using regression methods, the agreement was within 2.0 kcal/mol. For metalloenzymes, the agreement with experiments without fitting was within 1.7 kcal/mol and with fitting was within 1.4 kcal/mol. The authors thus demonstrated the inherent predictive ability of this first-generation full QM-based scoring function that takes into account all aspects of binding.

In another study, Nikitina et al., using linear-scaling QM methods, calculated the binding enthalpy of eight ligands bound to protein conformations from the PDB.[84] The authors chose enthalpy to examine the ability of the semiempirical Hamiltonian PM3 to calculate the enthalpy of binding. The choice of the enthalpy of binding instead of the free energy of binding was a prudent choice because the computation of entropy is far more challenging and generally introduces further simplifying approximations. Another important aspect of the study was inclusion of water molecules in the calculation of enthalpy. The structural water molecules were included in the computation of reference state enthalpies of the protein and ligand. They tried two different schemes where water molecules that were hydrogen bonded to both the protein and the ligand in the complex were considered in both reference state calculations of the protein and the ligand. One drawback of the study is the exclusion of solvation effects or the solvation correction to the enthalpy of binding. However, the authors argue that solvation effects are modeled enthalpically by including explicit water molecules. The calculated enthalpies agreed with the experimental enthalpies within 2 kcal/mol.

Other recent examples of using of linear-scaling QM in RBDD include a study by Vasilyev and Bliznyuk where the computer program MOZYME was used to rescore the top 100 predicted ligands from another docking program. The authors evaluated the feasibility of using a linear-scaling QM program for such a task.[85] In another application of MOZYME, Ohno et al. studied the affinity maturation of an antibody by calculating the binding free energy of the hapten bound to a germline antibody and the mature form.[86] The authors emphasize the importance of polarization and charge transfer in the maturation process.

Recent development of linear-scaling technology has focused on higher levels of theory, such as Hatree-Fock or DFT to calculate the wave functions of macromolecules. Gao et al. have described the development and application of a density matrix (DM) scheme based on molecular fractionation with conjugate caps (MFCC).[87] Using this method the density matrix is calculated for capped fragments of a macromolecule at high levels of theory. The total energy is then calculated from the full DM that is assembled from the fragment DMs. In an application of this method, Chen and