classical polarizability terms has seldom been done, though we will briefly mention attempts to include these. Computing free energies using mixed QM/MM simulations can be done but its use has been even more restricted and so will not be discussed here.[19]

Basic equations

The binding affinity K_d of a small molecule ligand L to a protein P can be expressed simply by

$$K_d = \frac{[L][P]}{[PL]}, \tag{5.1}$$

where the brackets denote an equilibrium concentration, L is the ligand, P is the protein, and PL is the protein/ligand complex. This definition makes the assumption that the difference between bound and unbound states can be well defined, an assumption that is essentially always valid for tight, specific binders but becomes more complicated for very weak and nonspecific binders.

This binding affinity can then be related to the free energy of binding by

$$\Delta G_{\text{Bind}} = -kT \ln \frac{K_d}{C^o}, \tag{5.2}$$

where C° indicates the standard state concentration (by convention, 1 M for solutions). We use the Gibbs free energy G in our equations, because situations of pharmaceutical interest are usually under constant pressure.

The free energy of binding can also be expressed as

$$\Delta G_{\text{Bind}} = -kT \ln \frac{Z_P Z_L}{C^o Z_{PL}}, \tag{5.3}$$

where Z represents the partition function of the system. It is this quantity that we wish to calculate via simulation.

MM-PBSA

As a compromise between speed and accuracy for physics-based estimates of protein/ligand binding affinities, we first discuss the end-point free-energy method molecular mechanics with Poisson–Boltzmann and surface area (MM-PBSA).[20] As an end-point method, MM-PBSA requires direct simulation of only the bound and unbound states. This simplification comes with the expectation of significantly larger intrinsic errors with MM-PBSA than with other more rigorous methods we will address later in the chapter.

The free energy of binding can be written as a difference in the solvation free energies of each of the components:

$$\Delta G_{\text{Bind}}\, \Delta G_{\text{PL-solv}} - \Delta G_{\text{L-solv}} + \Delta G_{\text{P-solv}}. \tag{5.4}$$

Each of these solvation energies can be written as

$$\Delta G_{\text{solv}} = \Delta\, H_{\text{solv}} - T\Delta S_{\text{solv}}. \tag{5.5}$$

If we average out the coordinates of the solvent over all the configurations, then we can approximate each of these free energies as

$$\Delta G_{\text{X-solv}} = \langle E_{\text{X-MM}} \rangle + \Delta G_{\text{X-solvent}} - T\Delta S_{\text{X-MM}}, \tag{5.6}$$

where $\langle E_X \rangle$ is the average molecular mechanics energy of X alone (without water), ΔS_X is the internal entropy of X (without water), and $\Delta G_{\text{X-solvent}}$ is the energy and entropy due to the solvation of X waters. These solvation energies for P, L, and PL can then be combined to compute a full binding energy.

In practice, a variety of implementations of the MM-PBSA protocol have been reported, and particular care needs to paid to a number of details in setting up the calculations. In general, protocols can be separated into three steps. First, coordinate sampling [such as molecular dynamics (MD)] is performed on the protein/ligand complex to sample configurations for energy analysis. In the next step, calculation of gas-phase potential energies and solvation free energies is performed on each structure collected from the previous step to produce ensemble averages. Finally, some measure of estimated change in solute entropy is computed for the set of structures. The final binding free energy is then obtained by combining these various components.

To generate the structures in the first step, one can perform separate MD simulations for the isolated ligand, apo protein, and bound protein/ligand complex. Alternatively, one can use a single trajectory of the bound complex as the source of conformations for the unbound (and bound) states.[21] This second case is equivalent to assuming that the conformations explored in the protein/ligand complex in solution are sufficiently similar to those conformations explored by the apo protein and isolated ligand. This assumption is not necessarily reasonable and in fact is guaranteed to be grossly incorrect in some contexts; however, the amount of noise added when taking differences between averages produced from independent bound and unbound trajectories substantially increases the sampling required for convergence, so by simulating one structure, lower variance is traded for some bias.[22,23] In theory, one could then perform a single MD run of the apo protein, and all additional runs would involve only isolated ligands. In any case, determining arrival at a stable average can be challenging.[24] A possible alternative formulation for the case of running the three separate trajectories is to disregard all energies but the interaction energies in an attempt to dampen the contributions to noise due to noncanceling internal-energy differences.

The potential energy $E_{\text{X-MM}}$ is that of only the protein and ligand and consists of

$$E_{\text{X-MM}} = E^{\text{elec}} + E^{\text{vdW}} + E^{\text{int}}, \tag{5.7}$$

where E^{elec} is the electrostatic energy, E^{vdW} is the van der Waals dispersion and repulsion, and E^{int} is composed of internal-energy terms for the ligand and protein, such as bond, angle, and torsion terms.

The solvation energy term $\Delta G_{\text{X-sovlent}}$ is subdivided into a sum of two components, one due to electrostatic interaction and the other due to nonpolar interactions:

$$\Delta G_{\text{X-solvent}} = \Delta G_{\text{PBSA}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}}, \tag{5.8}$$