differ by 3–5 kcal/mol depending on the starting configuration of a valine side chain.[112,113,203] In another study of HIV-1 NNRTIs, computed relative binding free energies had the wrong sign (an error of roughly 4 kcal/mol) when beginning from a crystal structure with an alternate rotamer conformation near the active site.[63] As noted previously, these are not simply issues of having a wrong protein structure: when protein conformational changes occur on binding, multiple metastable configurations are relevant.[112] Thus, some authors have suggested starting simulations from different regions of phase space as a test for these sorts of issues[111,155] or have in some cases performed this test.[112,155]

Another potential pitfall is the possibility of multiple potentially relevant bound ligand orientations that can be separated by kinetic barriers.[179] This has been observed not only in absolute free-energy calculations on the lysozyme binding site, where interpretation and analysis is simplest,[113,179] but also in relative free-energy calculations for ligands binding to neutrophil elastase[169] and the estrogen receptor.[129] Ligand symmetries can also play a complicating role.[167,179]

In many cases, failures are more difficult to interpret. In a recent study on squalene-hopene cyclase inhibitors, relative free energies computed with a single-step perturbation method had large errors and in some cases had the wrong sign;[131] this was also the case in a study by some of the same authors on phosphodiesterase inhibitors.[216] Another study with single-step methods found results of varied quality depending on the (in principle arbitrary) choice of reference state, indicating poor convergence,[87] and in some cases resulting in very large errors. These errors may mostly be due to poor phase-space overlap with the single-step approach.

Using multiple routes around the same thermodynamic cycle can be a helpful way to check for errors when doing relative free-energy calculations. This approach was used by de Graaf et al.[218] and found very different results depending on the choice of pathway (indicating convergence problems along at least some pathways); cycle closure errors were up to 4.9 kcal/mol and for some paths and mutations, the sign of the relative free energies was even incorrect. Cycle closure errors were also large in the work of van den Bosch et al.[219] Of course, cycle closures only provide a lower bound on the error, and in some cases true errors are much larger than the cycle closure error, as is the case with the large cycle closure error and even larger true error in the work of Dolenc et al.[220]

In many cases, studies may have simply been somewhat too ambitious. Donnini and Juffer attempted to use absolute free-energy calculations to examine binding free energies between peptides and proteins and concluded that "it generally proved rather difficult to predict the absolute free energies correctly, for some protein families the experimental rank order was reproduced. . . ."[221]

Other types of binding free-energy studies

There are several other types of rigorous binding free-energy calculations that have occasionally been applied to interesting biomolecular problems. Grand canonical Monte Carlo (GCMC) techniques have been used in several applications to compute insertion free energies; in one study these techniques were used to compute favorable sites for displacing waters around a ligand in a binding site,[192] and in another case GCMC techniques were used to estimate binding free energies of ligands to the lysozyme model binding site, though in the absence of protein flexibility.[204] More recently, grand canonical techniques were used to insert water molecules into a protein/ligand binding site while the ligand was being alchemically removed, thereby speeding convergence.[184] Potential of mean force (PMF) methods have also been applied to several protein/ligand systems, including the binding affinities of FKBP inhibitors[107] and the binding affinity of a phosphotyrosine peptide to the SH2 domain of Lck.[108] Nonequilibrium free-energy methods have also been applied to FKBP inhibitors and peptides binding to the SH3 domain.[109,110]

As mentioned previously, the mining minima approach of Gilson and collaborators is also particularly interesting and has given promising results in calculations of binding free energies for host/guest systems[66] and ligands to artificial receptors.[164] However, because of computational limitations, it is difficult to apply it to the protein/ligand systems that are of interest in drug discovery. One recent study used this approach, however, to assess changes in HIV protease inhibitor configurational entropy on binding.[68]

DIRECTIONS FOR LIKELY IMPROVEMENT

There are a number of different aspects in which free-energy calculations will need to improve to become more accurate and reliable. One of the most important is in the realistic treatment of the environment of the protein/ligand systems. Typical ligand binding simulations include only the protein, the ligand, water, and perhaps a few ions to neutralize the simulated system. But many ligand binding affinities have significant dependence on pH, salt concentration, and metal-ion concentration. None of these additional aspects are typically modeled in ligand binding free-energy calculations and will need to be treated better in the future.

It is also likely that there will need to be further advances in atomistic force field parameters. A number of tests of solvation free energies have demonstrated that the current generation of force fields have fundamental problems that may restrict the ability of these force fields to obtain binding free energies that are accurate to within 1 kcal/mol.[81,104,205,222] Most common force field protein parameters are more than ten to fifteen years old, and only the torsions have generally been improved.[223,224] The