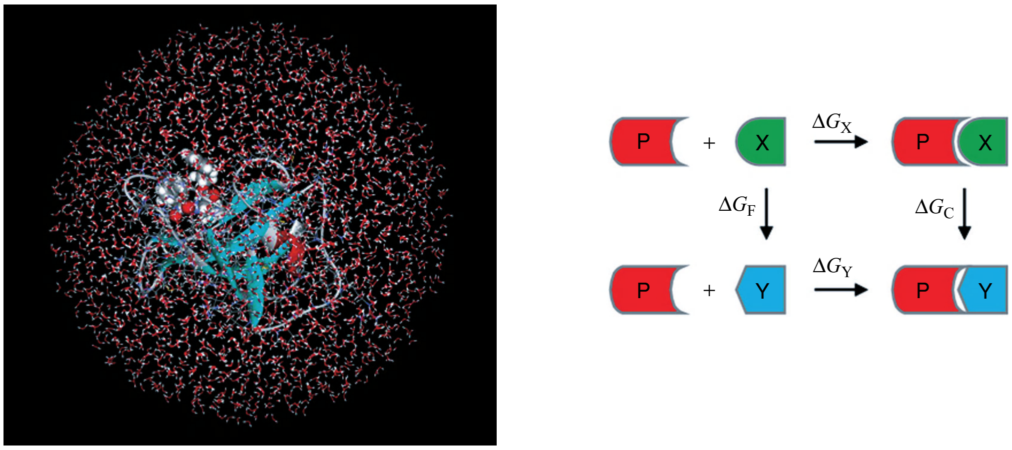

Figure 1.10. (Left) A protein/ligand complex surrounded by approximately 1,000 water molecules in a spherical shell or "cap." (Right) Thermodynamic cycle used to compute relative free energies of binding, $\Delta\Delta G_b$. P is a host and X and Y are two ligands.

the primary sulfonamide group can be associated with sulfa allergies and indiscriminant metal chelation.[36]

Overall, for the 1,712 oral drugs, 278 violate one or more of the four Lipinski rules ($MW < 500$, $\log P_{o/w} < 5$, H-bond donors ≤ 5, H-bond acceptors ≤ 10) with QPlog$P$ used for log $P_{o/w}$. There are 178, 82, and 18 oral drugs with one, two, and three violations, respectively. The group with two violations includes macrolides such as erythromycin and azithromycin and some other well-known drugs like atorvastatin, amiodarone, chloramphenicol, ketoconazole, and telmisartan. These examples all fail one member of the rule-of-three, either the solubility limit or number of primary metabolites, for example, respectively, atorvastatin and the macrolides. The group with three rule-of-five violations includes the HIV-protease inhibitors ritonavir and saquinavir, which are known to have low bioavailability. There are exceptions to such rules because they are based on 90th-percentile limits. Nevertheless, in all stages of lead generation, it would be imprudent to ignore property distributions for known drugs such as those in Figures 1.8 and 1.9.

LEAD OPTIMIZATION

It is assumed that inhibitory potency increases with increasing biomolecule-inhibitor binding. So, on the computational side, the key for lead optimization is accurate prediction of biomolecule-ligand binding affinities. There are many approaches, but the potentially most accurate ones are the most rigorous.[17] At this time, the best that is done is to model the complexes in the presence of hundreds or thousands of explicit water molecules using Monte Carlo (MC) statistical mechanics or molecular dynamics methods (Figure 1.10).[17] Classical force fields[16] are used, and extensive sampling is performed for key external (translation and rotation) and internal degrees of freedom for the complexes, solvent, and any counterions. FEP and thermodynamic integration (TI) calculations then provide formally rigorous means to compute free-energy changes.[37] For biomolecule/ligand affinities, perturbations are made to convert one ligand to another using the thermodynamic cycle in Figure 1.10. The conversions involve a coupling parameter, λ, that causes one molecule to be smoothly mutated to the other by changing the force field parameters and geometry.[38] The difference in free energies of binding for the ligands X and Y then comes from $\Delta\Delta G_b = \Delta G_X - \Delta G_Y = \Delta G_F - \Delta G_C$. Two series of mutations are performed to convert X to Y unbound in water and complexed to the biomolecule, which yield ΔG_F and ΔG_C.

Absolute free energies of binding are not obtained, but for lead optimization it is sufficient to assess the effects of making changes or additions to a core structure in the same spirit as synthetic modifications. Though the MC or MD plus FEP or TI calculations are rigorous, the accuracy of the results is affected by many issues, including the use and quality of force fields; missing energy terms, such as instantaneous polarization effects; and possible inadequate configurational sampling, which may be associated with, for example, infrequent conformational changes that are beyond the duration of the simulations. In the author's experience, more approximate methods are not accurate enough to provide satisfactory guidance in lead optimization.

The idea of using such calculations for molecular design goes back more than twenty years, at least to the report of the first FEP calculation for conversion of a molecule X to molecule Y in 1985[38] and to the earliest application of FEP calculations for protein-ligand binding by Wong and McCammon.[39] A final comment from McCammon's review on computer-aided molecular design in *Science* in 1987 was perspicacious: "The attentive reader will have noticed that no molecules were actually designed in the work described here."[40] The situation has remained basically unchanged since the late 1980s. As the convergence of FEP calculations was investigated, it was apparent that they were too computationally intensive for routine use in molecular