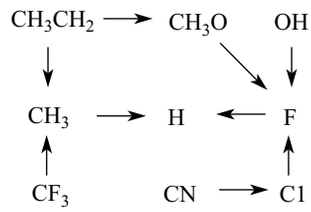

Figure 1.16. Possible series of FEP calculations to be run in parallel for optimization of a small substituent.

the logical next step; the order depends on the specific case. Optimization of peripheral rings and their substituents is likely to be easier synthetically than change of a central heterocycle. For the small groups, replacement of chlorine by fluorine (smaller, less lipophilic) and cyano (larger, more polar) can often be constructive, while replacements of methyl by OCH_3, CF_3, and CH_2OCH_3 can provide informative variety. It is straightforward to run a series of FEP modifications in parallel for optimization of a small substituent. A typical series is shown in Figure 1.16 for optimization of a substituent on an aromatic ring; the indicated conversions are intended to minimize steric and hydrogen-bonding changes. Avoidance of bromine, iodine, and nitro groups can be justified because of potential reactivity and metabolism drawbacks.

As a standard protocol, the necessary structure files are built using BOMB, and the nine indicated FEP calculations for the complexes are run simultaneously on nine processors; such calculations for an approximately 200-residue protein, 1,000 water molecules, and normal run lengths require six to seven days on a 3-GHz Pentium IV using MCPRO. A script is also used to extract the ligand from the complexes and to initiate the corresponding nine FEP calculations for the perturbations between the unbound ligands in water; these require one day each. So, with the commitment of eleven processors, the nine $\Delta\Delta G_b$ results are available in one week. For heterocycle optimization, it is convenient to use a reference that has the maximal number of hydrogens, for example, pyrrole, and perturb to other heterocycles with the same ring size and with the same or a smaller number of hydrogen atoms. Such isosteric FEP calculations converge well, and running about ten heterocycle perturbations in parallel is straightforward.[14,24] The default FEP procedure is to use eleven windows of overlap sampling (11-SOS), which is described in detail elsewhere.[57] If rapid turn-around is needed, it is easy to have a script distribute the eleven windows on eleven processors. An FEP calculation for a complex can then be completed in one day, and a twelfth processor can be used for the unbound leg of the cycle in Figure 1.10; that is, one $\Delta\Delta G_b$ result can be obtained in one day using twelve processors.

As noted in Figure 1.1, throughout the lead generation and optimization process it is also advisable to stay aware of the predicted ADME characteristics of the compounds to avoid potential bioavailability problems. It is often more difficult to change properties than potency because potency is so locally sensitive. For example, the predicted solubilities, QPlog*S*, and octanol/water partition coefficients, QPlog*P*, for the four compounds in Figure 1.14 are within approximately 0.5 log unit, whereas the activity range is nearly 3 log units. Molecular design for some drug classes can be particularly challenging, for example, for CNS-active compounds in view of the simultaneous needs for good potency, solubility, cell permeability, and blood/brain barrier penetrability, and for Gram-negative antibacterial agents because of the outer membrane structure. In general, a common problem that needs to be avoided is being lured by the Siren of in vitro potency into the Charybdis of insolubility.

CONCLUSION

Great progress has been made in the development and application of methodology to facilitate both drug lead generation and lead optimization. Computational chemistry has contributed significantly through advances in de novo design, virtual screening, prediction of pharmacologically important properties, and the estimation of protein-ligand binding affinities. Docking of large commercial and in-house libraries has evolved into being an essential approach for structure-based lead generation. All pharmaceutical companies also routinely use software for predictive ADME profiling. Furthermore, as summarized here, the long-standing promise of the utility of free-energy calculations for molecular design including thorough lead optimization has been fulfilled. The methodology allows broad exploration of the effects of potential modifications to a compound without the need for synthesis and without conceptual constraints associated with ease of synthesis. Depending on the outcome of the computational explorations, synthetic and biological resources can be focused on the most promising directions. In view of the ever-pressing needs for efficiency, free-energy-guided molecular design can be expected to become a mainstream activity in many contexts.

ACKNOWLEDGMENTS

The work described here has been supported by grants from the National Institutes of Health (GM32136, AI44616) and the Alliance for Lupus Research. The contributions from the coworkers listed in the references were also essential.

REFERENCES

1. Hughes, B. 2007 FDA drug approvals: a year of flux. *Nat. Rev. Drug Discov.* **2008**, *7*, 107–109.
2. Lahana, R. How many leads from HTS? *Drug Discov. Today* **1999**, *4*, 447–448.
3. Posner, B. A. High-throughput screening-driven lead discovery : Meeting the challenges of finding new therapeutics. *Curr. Opin. Drug Disc. Dev.* **2005**, *8*, 487–494.