The implicit-solvent predominant-states methods[29–31] mentioned by Shirts et al. replace the full configuration integrals of Equation (7.2) with a summation over configuration integrals for the most favorable minima on the potential energy surface:

$$Z_X = \sum_{i=1}^{M} z_i, \tag{7.6}$$

where the M individual configuration integrals z_i are computed using techniques such as estimation by harmonic approximation, harmonically biased sampling, or unbiased Monte Carlo integration.

Similar connections can be made between docking and Equation (7.1). As with the relative free-energy methods, the contribution due to the protein in solution is constant across a set of ligands binding to the same protein. For relative free-energy methods the protein configuration integrals cancel in the expression for $\Delta\Delta G$, Equation (7.4); in comparison, for docking Z_P is assumed constant:

$$\Delta G^{\circ}_{\mathrm{sol,PL}} \approx -RT \ln\left(\frac{Z_{\mathrm{PL}}}{Z_{\mathrm{L}}}\right) + K, \tag{7.7}$$

where K has been explicitly written to emphasize that the protein configuration integral Z_P has been subsumed into K with the symmetry-number and standard-concentration portions of Equation (7.1). For most but not all docking methods, the contribution due to the ligand free in solution is also treated as constant. Finally, in analogy to an extreme case of the predominant states methods, docking methods generally replace the full configuration integrals of Equation (7.2) with the single energetically most favorable state:

$$\Delta G^{\circ}_{\mathrm{sol,PL}} \approx -RT \ln\left(\frac{z_{\mathrm{PL},0}}{z_{\mathrm{L},0}}\right) + K \tag{7.8}$$

and the entropic component of the integral is either ignored:

$$z_{X,0} \equiv \int \exp^{-\beta E_X(\mathbf{r_0})}\, d\mathbf{r_0} \approx \exp^{-\beta E_{X,0}} \tag{7.9}$$

or is approximated by inclusion of terms to account for the entropic penalty of confining the ligand within the protein binding site (Chang, Chen, and Gilson and references therein).[32] On application of all of the approximations that are typically inherent in docking calculations, Equation (7.1) reduces to

$$\Delta G^{\circ}_{\mathrm{sol,PL}} \approx E_{\mathrm{PL}}, \tag{7.10}$$

where a docking "score" E_{PL} is computed for a single pose of a ligand docked into a protein binding site. In specific docking implementations, that score might be as simple as a counting of favorable interactions between protein and ligand or as complicated as a force field energy calculation supplemented by estimates of the free energy of the ligand in solution, the solvation differences between uncomplexed and complexed species, and the entropic cost for localizing the ligand to a specific location.

Given the approximations made to the underlying theory, it is no surprise that most studies have shown no correlation between docking score and affinity for closely related analogs.[24,33] There have of course been reports of specific examples where a correlation is seen between affinity and docking score[34,35] or between affinity and interaction energy.[36] As a general practice, therefore, the experienced computational chemist will explore all possible correlations in the hope that one will prove reliable enough to guide design and synthesis. The more typical case, however, is that there will be no reliable signal for decision-making, and the computational chemist must fall back on more computationally expensive methods such as those described by Shirts et al. or on more heuristic strategies – for example, docking large numbers of analogs and assessing emerging patterns of interactions.

FINDING NEW LEADS WITH DOCKING

The availability of large numbers of protein crystal structures (almost 54,000 public structures in the RCSB as of October 2008[37]) strengthens the impetus to make use of that structural information, particularly in those pharmaceutical companies that have made substantial internal investments in structural biology personnel and infrastructure. The intuitive hope has been that protein structures would prove particularly useful for finding novel starting points – leads – for drug discovery efforts, and docking and scoring technologies have seemed particularly relevant tools for virtual screens of large databases of compounds against these protein structures because docking-based virtual screens are in principle not limited by a need for or similarity to known ligands. Indeed, in many recent publications, authors have asserted that docking technologies have improved and that virtual screening successes have become more prevalent. In this section, a survey of the docking landscape will examine whether such assertions are well-founded pragmatism or unwarranted optimism.

Taking census of docking screens, 2000–2008

A census has been carried out for all docking-based virtual screens reported during the period January 2000 through October 2008; the census covered all peer-reviewed scientific journals up through the time of the submission of this chapter to the book editors. Literature searches using terms such as "docking," "virtual screening," and so on, were carried out in SciFinder,[38] PubMed,[39] and Google Scholar.[40] Non-English-language journals were included in the census only if abstracts provided sufficient detail concerning the virtual screening process and results. The initial manuscripts from these literature searches were supplemented to include examples from docking reviews that were not found using simple docking-related search terms. This set of literature searches returned substantially more than 1,000 publications for review and compilation.

From the collection of virtual-screen reports compiled, screens against DNA or RNA targets were removed; the