Docking: a domesday report

Martha S. Head

In 1085, most likely from a desire to audit his tax revenues, William the Conqueror commissioned a survey of the land and resources of the country over which he reigned.[1] The results of that survey come down to us in two tomes, the Little Domesday and the Great Domesday, in which were recorded voluminous amounts of data concerning the land, people, buildings, and chattel throughout England. By no means was this a complete record; large swathes of urban England – London, for example – were not included, nor was there any census of church personnel or property. The Little and Great Domesday books accordingly are an odd mix of completeness and incompleteness, leaving out such large parts of English society yet cataloguing to an excruciating level of detail within the areas surveyed.

Similarly, this chapter is a complete yet incomplete survey of the docking and scoring landscape. We do not review the general principles of docking technologies; a sufficient number of such reviews have been published in peer-reviewed journals alone.[2–20] Nor do we evaluate the state of the art for docking programs and scoring functions; a number of well-regarded and careful evaluations describe the current capabilities and limitations of the technology.[21–26] Instead, under "Comments on the Theory of Docking" we will make explicit the connections between docking and a theory of noncovalent association. Under "Finding New Leads with Docking" we take census of docking virtual screens carried out in this decade, and under "Predicting Bound Poses with Docking" we examine the role of expertise in predicting docked poses of small molecules bound to protein targets. From this mix of general overview and detailed analysis, it is hoped that we will develop a realistic snapshot of how docking is used in the pharmaceutical industry as a tool for drug discovery and design.

COMMENTS ON THE THEORY OF DOCKING

The standard free energy of noncovalent association of a protein and ligand in solution at constant pressure can be written as[27]

$$\Delta G^{\circ}_{\mathrm{sol,PL}} = -RT\ln\left(\frac{C^{\circ}}{8\pi^2}\frac{\sigma_{\mathrm{P}}\sigma_{\mathrm{L}}}{\sigma_{\mathrm{PL}}}\right) - RT\ln\left(\frac{Z_{\mathrm{PL}}}{Z_{\mathrm{P}}Z_{\mathrm{L}}}\right) + P^{\circ}\Delta\bar{V}, \tag{7.1}$$

where C° is the standard concentration (generally one molar), σ_X are symmetry numbers for each species, P° is the standard pressure (generally one atmosphere), $\Delta\bar{V}$ is the change in equilibrium volume, and Z_X are configuration integrals for each species:

$$Z_X \equiv \int e^{-\beta E(\mathbf{r})} d\mathbf{r}. \tag{7.2}$$

The work $P^{\circ}\Delta\bar{V}$ associated with changes in equilibrium volume due to complex formation is generally negligible, and the first term of Equation (7.1) is fully specified for a given protein/ligand pair at a particular standard concentration; the task of computing the free energy of association accordingly reduces to computation of the configuration integrals Z_{P}, Z_{L}, and Z_{PL}. For the rest of the discussion, therefore, equations will be written in a simpler and more compact format, for example:

$$\Delta G^{\circ}_{\mathrm{sol,PL}} = -RT\ln\left(\frac{Z_{\mathrm{PL}}}{Z_{\mathrm{P}}Z_{\mathrm{L}}}\right), \tag{7.3}$$

but it should be kept in mind that those missing terms are still implied.

In principle, Equation (7.1) provides an exact expression for the free energy of noncovalent association; putting the theory into practice, however, presents a number of computational challenges. In the previous chapter, Shirts, Mobley, and Brown described several strategies for computing both relative and absolute free energies,[28] many of which tie directly to the theory underlying Equation (7.1). For example, alchemical techniques for computing relative free-energy differences between two related ligands L_1 and L_2 binding to the same protein can be expressed as ratios of configuration integrals[27]:

$$\Delta\Delta G = \Delta G_{\mathrm{PL}_2} - \Delta G_{\mathrm{PL}_1} = -RT\ln\left(\frac{Z_{\mathrm{PL}_2}}{Z_{\mathrm{PL}_1}}\right) + RT\ln\left(\frac{Z_{\mathrm{L}_2}}{Z_{\mathrm{L}_1}}\right), \tag{7.4}$$

while the Zwanzig relationship discussed by Shirts et al. corresponds to replacing those ratios with the average free-energy difference extracted from simulations run under the Hamiltonian of ligand L_1:

$$\begin{aligned}\frac{Z_{\mathrm{PL}_2}}{Z_{\mathrm{PL}_1}} &= \left\langle \exp^{-\beta[H_{\mathrm{PL}_2}(\boldsymbol{x}) - H_{\mathrm{PL}_1}(\boldsymbol{x})]} \right\rangle_{PL_1} \\ \frac{Z_{\mathrm{L}_2}}{Z_{\mathrm{L}_1}} &= \left\langle \exp^{-\beta[H_{\mathrm{L}_2}(\boldsymbol{x}) - H_{\mathrm{L}_1}(\boldsymbol{x})]} \right\rangle_{L_1}.\end{aligned} \tag{7.5}$$