Empirical scoring functions for docking and virtual screening: Fundamentals, challenges and trends

Christoph Sotriffer

Institute of Pharmacy and Food Chemistry, University of Würzburg, Am Hubland, D - 97074 Würzburg

Computational structure-based drug design and virtual screening require methods for assessing the quality of protein-ligand interactions, both in structural (binding mode) and energetical (affinity) terms. For many applications, rigourous free energy calculations based on molecular simulations are computationally still too expensive to be practically useful. Accordingly, faster methods are applied in the context of docking and virtual screening, which are commonly referred to as scoring functions. A prominent example of this approach is given by the class of empirical scoring functions. Using a set of descriptors which encode structural features of protein-ligand complexes, these functions are derived empirically from experimental structures and affinity data by means of statistical regression techniques.

Recent tests on large data sets and community-wide benchmark exercises have revealed the strengths and shortcomings associated with various types of current scoring functions. Good results can often be obtained in docking applications with respect to the prediction of near-native binding modes. Also the ability to distinguish active ligands from unlikely binders appears at least to be sufficient to make virtual screening a practically useful endeavour. However, the correlation with the experimental binding free energy and the possibility to quantify the effects of small structural changes on the ligand affinity are frequently still disappointing.

For empirical scoring functions, improvements are conceivable along three lines of development: (1) compilation of larger high-quality training sets of experimental structures and affinity data; (2) development of new descriptors covering hitherto unconsidered interaction features and contributions; and (3) application of alternative regression techniques and machine-learning methods. Although advanced scoring functions with improved performance are indeed being obtained by these strategies, overcoming the rather fundamental limitations and the remaining deficiencies in accuracy and reliability of scoring functions appears much more challenging.