lected, but drug-like, commercially available compounds using DOCK (327) and the Vertex in-house tool Gambler. ChemScore (80, 188), the DOCK AMBER force field score, and PLP (185) performed consistently well in enriching active compounds. This result was partially attributed to the fact that a rigid-body optimization could be carried out with these functions because they include repulsive terms in contrast to many other tested functions. Stahl and Rarey compared DrugScore (226), PMF (317), PLP (185), and the original FlexX score using FlexX for docking (110,130,138). Interestingly, the two knowledge-based scoring functions performed differently. DrugScore achieved better ranking for the tight-binding ligands in narrow lipophilic cavities of COX-2 and the thrombin S1 pocket. In contrast, PMF obtained better enrichment for the case of the very polar binding site of neuraminidase. Obviously, a general strength of PMF is the description of complexes showing multiple hydrogen bonds. This has also been noted in the study by Bissantz et al., in which PMF was found to perform well for the polar target thymidine kinase and less well for the estrogen receptor (230).

5.3 Hydrogen Bonding versus Hydrophobic Interactions

A balanced description of the contribution of hydrogen bonding and hydrophobic interactions to the total score is of general importance, to avoid a bias toward either highly polar or completely hydrophobic molecules. The actual parameterization of a scoring function depends on the compilation of the data set used to develop the function. Empirical scoring functions are more likely affected by the data set composition used for parameterization, but can be quickly reparameterized. In the case of knowledge-based functions such a readjustment is more difficult to perform; however, because of the much larger databases used for their development, they are supposed to be less dependent on special data set compilations.

The PLP function, for example, addresses general steric complementarity and hydrophobic interactions based on rather long-range pair potentials, whereas the FlexX score concentrates on hydrogen-bond complementarity. This is clearly reflected in results of database-ranking experiments. To combine the virtues of both scoring functions and to construct a more robust general function, a combination of PLP and FlexX called ScreenScore has recently been published (287). It was derived by a systematic optimization of library ranking results over seven targets and covers a wide range of active sites with respect to form, size, and polarity. ScreenScore obtains good enrichments for COX-2 (highly lipophilic binding site) and neurarninidase (highly polar site), whereas the individual functions fail in one of the two cases. The authors of PLP have recently enhanced their scoring function by including directed hydrogen bonding terms (367). Similar to ScreenScore, this could also lead to a more robust scoring function.

5.4 Finding Weak Inhibitors

Seeding experiments are often carried out with a small number of active compounds that are already optimized for binding to the studied target. Enrichment factors based on the retrieval of these compounds are not very conclusive because the recovery of potent inhibitors from a large set of candidate molecules is significantly easier than the discovery of new, but usually rather weak inhibitors from a large majority of nonbinders. In general, as in HTS, one can only expect hits from virtual screening that bind in the low micromolar range.

Nevertheless, a recent study showed that library screening can also successfully detect very weak ligands. Approximately 4000 commercially available compounds had been screened for FKBP-binding by means of the SAR-by-NMR technique (368) and 31 compounds with activity in the low millimolar range were detected. This set of compounds was flexibly docked into the FKBP binding site using DOCK 4.0 with the PMF scoring function (369). Interestingly, significant enrichment factors of 2 to 3 were achieved, whereas scoring with the standard AMBER score of DOCK did not really provide an enrichment.

5.5 Consensus Scoring

Different scoring schemes focus on different aspects as most important contributions to