marginally improve the chances of finding more potent active compounds from docking-based virtual screens.

PREDICTING BOUND POSES WITH DOCKING

In the preceding sections, we have presented a theoretical framework underpinning docking algorithms and have examined the real-world performance of docking as a prospective virtual screening tool. In the pharmaceutical industry, however, it is generally the case that the primary use of docking is predicting the bound pose of a specific molecule to a suitable degree of reliability. One may have a known active compound from a high-throughput screen or ongoing drug discovery effort and want to understand how that compound interacts with the protein target. Or one may have a collection of known actives and want to examine protein/ligand complementarity to rationalize differences in activity or selectivity. Or one might have multiple active series and want to design novel hybrid molecules using the best possible three-dimensional overlay. Or one might have synthesis proposals from medicinal or computational chemists and want to assess the likelihood for maintaining potency or to propose modifications to increase the likelihood of success. In all of these examples, the first steps would be the prediction of a bound pose for one or more molecules and the estimation of the reliability of that prediction.

It is particularly important for successful lead optimization that the computational and medicinal chemists have an understanding of the level of confidence in their docking mode predictions. Is a prediction expected to be accurate enough that I can make design and synthesis decisions at an atomic level of detail? Is the prediction accuracy such that I can draw only general conclusions about types and locations of substitutions on the core scaffold? Or is the confidence level low enough that the prediction can at best provide multiple testable hypotheses, and I should therefore design molecules to probe those hypotheses? Numerous evaluations of the ability to predict docked poses have been published,[24,147] and these evaluations support one general conclusion: Many docking programs can generate poses near the crystal conformation, but no scoring function can consistently score the correct pose at the top of the list. Therefore, in everyday practice, a computational chemist uses a docking program to generate poses for visual inspection and then selects the pose thought to be "best" based on chemical intuition and compatibility with any available SAR data. And in practice we can all point to examples of successful predictions, so it is our instinct that human experience and expertise in combination with computational tools is sufficiently predictive. This assertion, however, has not been validated through analysis of success rates for blind predictions. In this section preliminary results for both manual and automated docking mode predictions from SAMPL-1, a blind prediction challenge, are described.[148] The full results and crystallographic data from SAMPL-1 are not yet public, so the results for automated procedures are not discussed in any detail but instead results for the single manual predictor to the aggregate performance of automated docking procedures.

The SAMPL challenge

The SAMPL challenge was made possible by the generous contribution of two protein/ligand data sets, urokinase plasminogen activator provided by Abbott Laboratories and JNK3 protein kinase provided by Vertex Pharmaceuticals. The challenge proceeded in three phases corresponding to three typical drug discovery activities: (1) identification of active compounds from a background pool of inactive compounds, (2) prediction of the bound poses of active compounds, and (3) rank-ordering of active compounds by affinity. Data for the challenge were released to individual predictors in a correspondingly phased manner: if a predictor requested and received data for phase 2 or 3, that predictor could not subsequently obtain data or make predictions for phase 1. Only the results for phase 2, docking mode prediction, will be discussed here.

Details of the docking mode challenge. The docking mode prediction portion of the challenge also proceeded in phases, an initial "cross-docking" phase followed by a second "self-docking" phase. The cross-docking exercise more closely mimics the real-life situation in drug discovery while the second self-docking exercise allows for assessment of the importance of protein flexibility for successful prediction. Manual docking was applied only to the cross-docking exercise so details for the self-docking exercise are not discussed. For the cross-docking exercise, the SAMPL-1 organizers provided one urokinase and two JNK3 protein structures; the primary difference between the two JNK3 structures was in the side-chain conformation of the "gatekeeper" methionine. None of the three docking structures contained a bound ligand. In addition, during the previous virtual screening portion of the challenge, the organizers had provided two urokinase and two JNK3 structures, each with a ligand bound. SAMPL-1 organizers further provided SD files containing thirty-four compounds for docking to urokinase and sixty-two compounds for docking to JNK3. The small molecule conformations contained in the provided SD files had been generated from 2D representations and so differed from the conformation in any crystal structure. Some inactive compounds were included in those two lists, but the organizers did not reveal which compounds were inactive until after the completion of the full challenge. After the cross-docking challenge closed, SAMPL-1 organizers provided the raw results to me; none of the automated predictors – neither the names of the predictors nor the programs used – were identified in this output. Results were reported as follows:

$$\begin{aligned} &\text{rmsd} - \text{DPI}; \quad \text{rmsd} > \text{DPI} \\ &0; \quad \text{rmsd} \leq \text{DPI}, \end{aligned}$$