

Specificity scoring

S. Tietze¹ J. Apostolakis¹

¹*Ludwig Maximilian University, Amalienstr 17. D-80333 Munich , Germany*

Structure based virtual screening (SVS) is now a common tool in the study of molecular recognition, specificity, and the development of novel molecules for pharmaceutical and technological purposes. For a given target molecule the structure of its complex with different candidate ligands (or receptors) is predicted, and from the predicted structure the stability of the complex (binding affinity) is estimated. For the estimation, simple scoring functions are used, which are parametrized either by fitting to binding affinities or using simple statistical approaches. It has recently been shown that the most common of these scoring functions do not predict binding affinity significantly better than simple features of the ligands, such as molecular weight. It is therefore not particularly surprising that score based rankings in SVS often yield poor enrichments in screening benchmarks and applications.

Here we suggest two independent approaches for obtaining specific empirical scoring functions. In the first approach the parameters of the scoring function are optimized solely with respect to performance in screening, without using binding affinity. The potential is trained on a so called cross screening benchmark set, where 100 different ligands are screened against 100 different proteins. The accuracy of the trained potential is evaluated on a complementary cross screening set, which has been selected as to have no overlap with the training set proteins. Significant improvement over our previously validated regression based parameterization of the same functional form (ChillScore¹) is observed, with the average Area Under ROC (AUC) over all 100 targets improving from 0.65 to 0.77.

The second approach is based on a simple two-step affinity regression method, where however, unspecific (based on ligand properties alone) contributions to the affinity are removed from the potential. A detailed comparison between the two obtained score parameterizations, among each other, and with a number of the generally used scoring functions was performed to highlight the peculiarities of both approaches. One of the interesting results of this study is that the two methods lead to similar parameterizations, even though the first is purely qualitative (no use of binding affinities and X-ray structures is ever made), while the second is based only on crystal structures and experimentally measured binding affinities (no use of predicted structures is made). To further validate the method we show results on standard screening benchmarks (taken from Jain et al. 2005²), where a significant improvement over the standard regression potential can be demonstrated: average AUC increases from 0.65 for our previous standard regression based potential to 0.87 for the potential trained on independent cross-screening data, and 0.9 for the two-step specificity fit. We finally discuss one of the most interesting results of this study, namely that scoring functions that show practically no correlation with binding affinity ($R=0.32$ and 0.26), are significantly better at screening than empirical scoring functions parametrized according to affinity ($R=0.53$).

1. Tietze S, Apostolakis J, GlamDock: development and validation of a new docking tool on several thousand protein-ligand complexes. *J. Chem. Inf. Model.* 2007, 47(4):1657-72.
2. Pham TA, Jain AN, Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J. Med. Chem.* 2005, 49(20): 5856-5868