

Abstract Submitted
for the MAR13 Meeting of
The American Physical Society

A Bayesian Statistical Approach for Improving Scoring Functions for Protein-Ligand Interactions SAM Z GRINTER, XIAOQIN ZOU, University of Missouri — Even with large training sets, knowledge-based scoring functions face the inevitable problem of sparse data. In this work, we present a novel approach for handling the sparse data problem, which is based on estimating the inaccuracy caused by sparse count data in a potential of mean force (PMF). Our new scoring function, STScore, uses a consensus approach to combine a PMF with a simple force-field-based potential (FFP), where the relative weight given to the PMF and FFP is a function of their estimated inaccuracies. This weighting scheme implies that less weight will be given to the PMF for any pairs or distances that occur rarely in the training data, thus providing a natural way to deal with the sparse data problem. Simultaneously, by providing the FFP as a substitute, the method provides an improved approximation of the interactions between rare chemical groups, which tend to be excluded or reduced in influence by purely PMF-based approaches. Using several common test sets for protein-ligand interaction studies, we demonstrate that this sparse data method effectively combines the PMF and FFP, exceeding the performance of either potential alone, and is competitive with other commonly-used sparse data methods.

Sam Z Grinter
University of Missouri

Date submitted: 11 Nov 2012

Electronic form version 1.4