

# Datamining Results of Docking-Based Virtual Screens

Richard M. Fine and Boris Klebansky, BioPredict, Inc. Oradell, NJ, USA 07649.

Despite progress docking-based virtual screens continue to suffer from high false-positive rates, high false-negative rates, and incorrect assignment of docked poses. The primary source of error is imperfect ranking by current scoring functions: the highest scoring compound is not always active while the highest scoring pose is not always correct. Selecting active compounds and correct docking poses therefore requires the use of additional information on the target and related targets to help select compounds. Our approach collects multiple poses of all docked compounds into a database and then uses datamining techniques to select compounds to test. For each protein atom in the active site an 'interaction footprint' is generated that describes the interaction of the ligand pose with that atom. Since 'Interaction footprints' are of a fixed size they are easily compared and used as input to datamining operations. Current datamining operations include *filtering*, to select compounds with similar footprints to co-crystallized or docked actives from the literature; *clustering*, to identify common binding motifs for high-ranking docked compounds; and *learning methods*, to differentiate footprint components of actives from inactives. Footprint filters and binding motifs comprise structure-based *hypotheses* whose imposition enhances detection using imperfect energetic methods. Hypotheses can be imposed post-docking as described here, or incorporated as constraints applied directly during docking. We acknowledge NIH grant support of NIH-GM061465.