Cheminformatics Analysis of Protein-Ligand Interface.

Alexander Tropsha.

Professor and Chair, Director, The Laboratory for Molecular Modeling, Division of Medicinal Chemistry & Natural Products, School of Pharmacy, CB# 7360 Beard Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, U.S.A.

I shall discuss the application of chemiformatics approaches to structure based virtual screening, i.e., ligand docking and scoring. The Complementary Ligands Based on Receptor Information (CoLiBRI) method is based on the representation of both receptor binding sites and their respective ligands in a space of universal chemical descriptors. We have established a protocol to map patterns of nearest neighbor active site vectors in a multidimensional TAE/RECON descriptors space onto those of their complementary ligands. This protocol affords the prediction of a virtual complementary ligand vector in the ligand chemical space from the position of a known active site vector. This prediction is followed by chemical similarity calculations between this virtual ligand vector and those calculated for molecules in a chemical database to identify real compounds most similar to the virtual ligand. The current results indicate that CoLiBRI is capable of identifying all known ligands of 260 test binding sites within the top 1% of the database of ca. 60,000 compounds in 95% of all cases.

To address scoring, novel geometrical chemical descriptors have been derived based on Pauling atomic electronegativities (EN) and the computational geometry of protein-ligand interfaces using Delaunay Tesselation. The ENTess descriptors were employed in the variable selection k-Nearest Neighbor quantitative structure-binding affinity relationship (QSBR) studies of 264 diverse protein-ligand complexes with known binding constants. 24 complexes with chemically dissimilar ligands were set aside as an independent validation set, and the remaining dataset of 240 complexes was divided into multiple training and test sets. The best models were characterized by the leave-one-out cross-validated correlation coefficient q^2 as high as 0.66 for the training set and the correlation coefficient R^2 as high as 0.83 for the external test set.