SiteKey: A Novel Binding Site Predictor for Ordered Proteins Interacting with Intrinsically Disordered Proteins

Maheshwari, Sagar

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are segments of proteins that lack a defined tertiary structure, giving them the ability to interconvert among a range of conformations and folds. This structural plasticity adds to the complexity of predicting protein behavior and treating numerous diseases. For example, a key tumor suppressor protein p53 has significant regions of disorder and has been associated with lung, breast, and brain cancer. While we now have effective algorithms for predicting ordered binding partners of IDRs, an algorithm for identifying and characterizing binding sites on such proteins has remained elusive, despite some effort. Here we present a novel machine learning algorithm SiteKey — a random forest classifier, trained on features derived from both protein sequence and structure, capable of identifying these binding sites with 88.4% accuracy and an area under the ROC curve of 0.9441. These results should provide a new approach to rational drug design in which binding regions can specifically be targeted to prevent major diseases, including cancer.

Awards Won:

Third Award of $1,000