STRUCTURE-BASED IDENTIFICATION OF AROMATIC SURFACE MOTIFS INVOLVED IN PROTEIN-CARBOHYDRATE RECOGNITION

Andrew C. Doxey, Zhenyu Cheng, Brendan J. McConkey

Department of Biology, University of Waterloo, Canada

Carbohydrate-binding proteins (CBPs) play important roles in the natural and industrial processing of carbohydrate-rich biomass. Due to the structural diversity of CBPs and the lack of a common sequence motif defining CBP binding sites, it is difficult to computationally identify novel members of these families that may be important for biotechnological purposes. We developed a computational method to detect aromatic structural motifs involved in protein-carbohydrate binding given only 3D structure. The algorithm, which searches for clusters of solvent-exposed coplanar aromatic residues, correctly predicted carbohydrate-binding sites in 18 different folds. Using linear-discriminant analysis in a post-processing step, CBPs could be effectively discriminated from other structures in a non-redundant dataset based on features of the detected binding sites. The computational approach was further validated experimentally by identifying a tobacco CBP not present within the training set whose binding site was correctly predicted and received a significant score. This work highlights the potential of our pattern-detection / machine-learning approach for structure-based prediction and annotation of protein function.