Binding affinity of protein-carbohydrate complexes: database development, analysis and prediction

M. Michael Gromiha

Department of Biotechnology, Indian Institute of Technology Madras
Chennai 600036, INDIA
Email: gromiha@iitm.ac.in

Abstract

Protein-carbohydrate interactions are involved in several biological functions such as host-pathogen recognition, inflammation, signal transduction, and cell adhesion. These functions are mainly dictated by their binding affinity. Elucidating the factors influencing the binding affinity and predicting their binding free energy provide deep insights for understanding the recognition mechanism (1). We have investigated the dual role of amino acid residues, which are involved in binding with carbohydrates and stability of the complex using structure based parameters (2). Further, we have developed a database on binding affinity of protein-carbohydrate complexes, ProCaff (Protein-Carbohydrate complex binding Affinity Database), which contains about 4000 entries on dissociation constant (K_d), Gibbs free energy change (ΔG), experimental conditions, sequence, structure and literature information (3). The relationship between binding affinity and structure-based features revealed that binding site residues, accessible surface area, interactions between various atoms and energy contributions are important to understand the binding affinity. We developed a multiple regression method for predicting the binding affinity, which showed an average correlation and mean absolute error (MAE) of 0.73 and 1.15 kcal/mol, respectively, between experimental and predicted binding affinities on a jackknife test (4). Further, we have constructed a method for predicting binding free energy change upon mutation ($\Delta\Delta G$) using the sequence-based features, which showed an average correlation of 0.74 and a mean absolute error of 0.70 kcal/mol between experimental and predicted $\Delta\Delta G$ (5). These resources help to relate the binding affinity with disease-causing mutations, and designing therapeutic strategies for diseases.