A Quantitative Pan-Receptor Predictive Method for Peptide Recognized by the SH2 Domain Family

Hao Zhang, Claus Lundegaard, Morten Nielsen, Ole Lund

Center for Biological Sequence Analysis, Department of Systems Biology,
Technical University of Denmark, Lyngby, Denmark

Abstract

As the largest class of known phosphotyrosine recognition modular structures, Src-homology 2 (SH2) domain controls specificity in signal transduction by recognizing short peptide motifs and subsequently inducing reversible protein phosphorylation by kinases. SH2-mediated protein phosphorylation is a key part of protein post-translational modification, responsible for fundamental intracellular regulation of a variety of essential cellular functions, including early signaling events in T and B lymphocytes and fine-tuning of immune responses. It is estimated that about one third of mammalian proteins contain covalently bound phosphate at least once during their life cycle. Recent clinical findings are indicative of an important role of Src homology 2 domain-containing proteins. A number of human diseases are known to be causally associated with dysfunction of phosphorylation-dependent signaling and regulation.

The SH2 family is very large with each receptor potentially having unique ligand specificity. The experimental effort required to characterize each receptor-specificity is labor-intensive, time-consuming, and costly. There are more than 120 SH2 domains known to recognize specific amino acid motifs, making it infeasible to determine the binding specificity for each variant experimentally. The challenge to computational approaches also comes from the vast genetic variability within SH2 domains. There are handful receptor-specific methods, but to our knowledge, yet few pan-specific methods published. There is a demand on generic methods that provide rational characterization of the highly polymorphic receptors.

Here we present a novel method to characterize SH2 receptors given receptor sequences aligned by binding pockets. The method can extrapolate from variants of SH2 receptors where the binding specificity is known to those where no experimental data is available. We demonstrate, through leave-one-receptor-out, that this pan-receptor method can accurately predict the binding specificity for a broad range of receptors. Our comparative study concludes that the presented method is an efficient tool for predicting SH2-mediated protein-protein interactions quantitatively, and it can be used to study molecular basis of Src kinase specificity, and to identify binding motifs for hitherto uncharacterized SH2 domains.