Predicting SH3 domain binding sequences using artificial neural networks and other methods.

BENJAMIN J. STAPLEY & TREVOR P. CREAMER

Kentucky Center for Structural Biology and Center for Computational Science, University of Kentucky, Lexington, KY, USA, 40536-0298.

Email: ben@sophocles.gws.uky.edu

SH3 domains have been implicated in a large number of cellular signaling processes in both morbid and healthy phenomena. They bind a consensus PXXP motif that adopts a type II Polyproline helical conformation. SH3 domains are unusual in their variety, ubiquity, subtle sequence specificity and in that they can bind ligands in two pseudo-symmetrical orientations.

We have gathered a collection of 120 sequences from the literature for which binding to the c-src SH3 domain has been proven or disproven. The experimental evidence for this binding comes from phage display and yeast two-hybrid assays as well as direct biophysical measurement. Armed with this data we have trained back-propagated, feed-forward artificial neural networks with hidden layers containing up to five units. We have also used position-specific scoring matrices and hidden Markov models for prediction. With ANN's using a jack-knifed test set and early stopping, prediction accuracy is 85% and exceeds that of other methods. Averaging over different networks improves prediction accuracy and class I ligands are more easily predicted than class II. A method for simultaneously predicting ligand specificity for a number of SH3 domains is discussed.

Bishop, CM. *Neural Networks for Pattern Recognition* Clarendon Press. 1995.
Pawson, T. Protein modules and signalling networks. *Nature* 373;573-579. 1995.