Structure-based detection of distant functional relations

Ke Chen[1], Lukasz Kurgan [1,*]

[1] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
* presenting author

Distant functional relations concern proteins with substantially different folds that perform the same function, e.g., they bind the same ligand. We investigate prediction of such relations from protein structure using a model problem involving protein-nucleotide interactions. The protein-nucleotides interactions are essential for numerous cellular processes and are highly ubiquitous, where the protein targets are very different structurally. The knowledge of the nucleotide-binding sites is crucial for function annotation and elucidation of cellular activities, and yet many of these interactions remain unknown [1]. This could be alleviated by employing prediction models derived from the known interactions. Our recent survey [2] shows that the threading-based Findsite [3], energy-based Q-SiteFinder [4], and a consensus geometry/energy-based MetaPocket [5] outperform other predictors, but their predictions leave a wide margin for improvements and the most accurate threading approach cannot accurately predict distant relations. We propose a novel method, NSiteMatch, which accurately predicts binding sites for 3 common nucleotides: ADP, ATP, and AMP. NSiteMatch combines geometrical, energy-, and threading-based approaches and draws from the observation that use of sufficiently similar templates leads to high quality predictions [1]. However, unlike Findsite that relies on the similarity of the fold, we introduce a new approach based on local similarity in the binding region to find the most suitable templates, which may share low homology with the predicted protein. We empirically evaluate NSiteMatch when predicting proteins that share low homology (from different SCOP folds or families) or low sequence identity (< 40%) with the templates to investigate whether our predictions can be used to annotate novel/uncharacterized folds and families. Tests show that NSiteMatch significantly outperforms other methods when predicting binding sites, see Figure 1. The evaluation of the prediction of binding residues shows that the average (over 3 nucleotides) MCC values obtained by NsiteMatch, Findsite, Q-SiteFinder, and MetaPocket are 0.56, 0.50, 0.37, and 0.24 at 40% similarity; 0.46, 0.37, 0.37, and 0.23 at the family level; and 0.39, 0.30, 0.37, and 0.23 at the fold level. We will illustrate how the distant functional relations are utilized in our predictions with a case study. Although we limit our study to nucleotides, we introduce a generic platform that can be adopted to derive predictors for other small ligands.

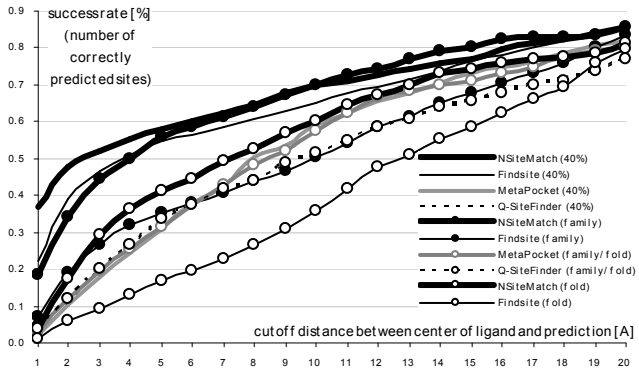

Figure 1 Average, over 3 nucleotides, success rates

References:

1. Saito M, Go M, Shirai T (2006) *Protein Eng Des Sel.* 19:67-75.
2. Chen K, Mizianty M, Gao J, Kurgan L (2011) *Structure*, 19(5):613-621.
3. Skolnick J, Brylinski M (2008) *Proc Natl Acad Sci U S A.* 105: 129-34.
4. Laurie AT, Jackson RM (2005) *Bioinformatics* 21: 1908-16.
5. Huang B (2009) *OMICS* 13: 325-30.