Poster I-52

Insights into protein-protein interfaces using a Bayesian network prediction method

Authors:
James Bradford (*University of Leeds*)
Chris Needham (*University of Leeds*)
Andrew Bulpitt (*University of Leeds*)
David Westhead (*University of Leeds*)

Short Abstract: We combine surface patch analysis with a Bayesian network to predict protein-protein binding sites with high accuracy even when no sequence homologues to the query protein are available. We show that our method can be applied to un-annotated structures solved by structural genomics projects, and can identify potential drug targets.

Long Abstract:
Identifying the interface between two interacting proteins provides important clues to the function of a protein, and is becoming increasing relevant to drug discovery. Here, surface patch analysis was combined with a Bayesian network to predict protein-protein binding sites with a success rate of 82% on a benchmark dataset of 180 proteins, improving by 6% on previous work and well above the 36% that would be achieved by a random method. A comparable success rate was achieved even when evolutionary information was missing, a further improvement on our previous method which was unable to handle incomplete data automatically. In a case study of the Mog1p family, we showed that our Bayesian network method can aid the prediction of previously uncharacterised binding sites and provide important clues to protein function. On Mog1p itself a putative binding site involved in the SLN1-SKN7 signal transduction pathway was detected, as was a Ran binding site, previously characterized solely by conservation studies, even though our automated method operated without using homologous proteins. On the remaining members of the family (two structural genomics targets, and a protein involved in the photosystem II complex in higher plants) we identified novel binding sites with little correspondence to those on Mog1p. These results suggest that members of the Mog1p family bind to different proteins and probably have different functions despite sharing the same overall fold. We also demonstrated the applicability of our method to drug discovery efforts by successfully locating a number of binding sites involved in the protein-protein interaction network of papilloma virus infection. In a separate study, we attempted to distinguish between the two types of binding site, obligate and non-obligate, within our dataset using a second Bayesian network. This proved difficult although some separation was achieved on the basis of patch size, electrostatic potential and conservation. Such was the similarity between the two interacting patch types, we were able to use obligate binding site properties to predict the location of non-obligate binding sites and vice versa.