BMC Bioinformatics

ORAL PRESENTATION

Open Access

Structure-based kernels for the prediction of catalytic residues and their involvement in human inherited disease

Fuxiao Xin[1], Steven Myers[1], Yong Fuga Li[1], David N Cooper[2], Sean D Mooney[3], Predrag Radivojac[1*]

Background
Enzyme catalysis is involved in numerous biological processes and the disruption of enzymatic activity has been implicated in human disease. Despite the functional importance, various aspects of catalytic reactions are not completely understood, such as the mechanics of reaction chemistry and the geometry of catalytic residues within active sites. As a result, the computational prediction of catalytic residues has the potential to identify novel catalytic pockets, aid in the design of more efficient enzymes and also predict the molecular basis of disease.

Results
We proposed a new kernel-based algorithm for the prediction of catalytic residues and functional sites in general in protein structures [1]. The method relies upon explicit modelling of similarity between residue-centred neighbourhoods in protein structures. Specifically, we start with a construction of oriented structural neighbourhoods followed by separating the neighbourhood volume into small cells. The similarities between two structural neighbourhoods are accumulation of their similarity in each cell. The kernel function is a product of three kernels, each addressing a separate aspect of protein function: (i) the geometric kernel addresses the shape similarity, (ii) the chemical kernel addresses the similarity in physicochemical properties, and (iii) the evolutionary kernel addresses the evolutionary similarity of conservation patterns for the residues in two structural neighbourhoods. Our approach was favourably evaluated against two of the leading alternative approaches, FEATURE [2] and GBT [3], as shown in Table 1. The new algorithm was used to identify known mutations associated with inherited disease whose molecular mechanism might be predicted to operate specifically though the loss or gain of catalytic residues. It should therefore provide a viable approach in identifying the molecular basis of disease in which the loss or gain of function is not caused solely by the disruption of protein stability. Our analysis suggests that both loss and gain of catalytic residues are actively involved in human inherited disease.

Conclusions
Our kernel method for functional sites prediction based on protein structures evaluates favourably against established methods on the same data set using the same evaluation procedure. The results from applying our catalytic residue predictor to disease mutations indicated that both loss and gain of catalytic residues are actively involved in human inherited disease.

* Correspondence: predrag@indiana.edu
[1]School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408, USA
Full list of author information is available at the end of the article

Table 1 Performance comparison between the three methods of catalytic residue prediction when evaluation was carried out by chain, family, superfamily and fold. Methods were evaluated on the same data set using 10-fold cross-validation. sn means the sensitivity when specificity is 0.95

	FEATURE		GBT		Structure Kernel	
	AUC	sn	AUC	sn	AUC	sn
Fold	76.4	0.20	80.4	0.34	86.1	0.40
Superfamily	76.5	0.20	80.8	0.34	86.1	0.40
Family	76.7	0.21	80.7	0.34	86.8	0.42
Chain	76.7	0.21	81.1	0.34	87.3	0.45

BioMed Central

Author details
¹School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408, USA. ²Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK. ³Buck Institute for Age Research, Novato, California 94945, USA.

References
1. Xin F, Myers S, Li Y, Cooper D, Mooney S, Radivojac P: **Structure-based kernels for the prediction of catalytic residues and their involvement in human inherited disease.** *Bioinformatics* 2010, **26**(16):1975-1982.
2. Wu S, Liang MP, Altman RB: **The SeqFEATURE library of 3D functional site models: comparison to existing methods and applications to protein function annotation.** *Genome Biol* 2008, **9**:R8.
3. Gutteridge A, Bartlett GJ, Thornton JM: **Using a neural network and spatial clustering to predict the location of active sites in enzymes.** *J Mol Biol* 2003, **330**(4):719-34.