BMC Systems Biology

BioMed Central

Poster presentation

Open Access

Discrimination of proteins using graph theoretic properties

Alper Küçükural* and O Ugur Sezerman

Address: Biological Sciences and Bioengineering, Sabanci University, Tuzla, Istanbul, Turkey

Email: Alper Küçükural* - kucukural@su.sabanciuniv.edu

* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology
Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, **1**(Suppl 1):P49 doi:10.1186/1752-0509-1-S1-P49

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

Background

Graph theoretic properties of proteins can be used to perceive the differences between correctly folded proteins and well designed decoy sets. Graphs are used to represent 3D protein structures. We used two different graph representations of protein structures which are Delaunay tessellations of proteins and contact map graphs. Graph theoretic properties for both graph types showed high classification accuracy to discrimination of proteins. Different type of linear classifiers and support vector classifier were used to classification of the protein structures. The best classifier accuracy was over 95% as shown in Table 1. The results showed that characteristic features of graph theoretic properties can be used many fields such as prediction of fold recognition, structure alignment and comparison, detection of similar domains and definition of structural motifs in high accuracy.

Conclusion

In this work we successfully showed that structural properties as well as potential scores can be used to discriminate native folds from the decoy sets. As far as graph types are concerned, the classification accuracy rates of the results obtained from contact map graphs are higher than the results obtained from Delaunay tessellated graphs for the same classification methods. Therefore contact map matrices are better representation method for protein structures. Support vector classifier and quadratic classifiers results are quite promising for the dataset which formed after outlier analysis. The accuracy rates are over 95%.

References

1. Strogatz SH: **Exploring Complex Networks.** 2001, **410:**268-276.
2. Albert R, Barabasi A-L: **Statistical mechanics of complex networks.** *Rev Mod Phys* 2002, **74:**47-97.

Table 1:

	Contact Map		Delaunay Tes.		
	After OA*	**Before OA***	**After OA***	**Before OA***	**Classifier Description**
Svc	95,54%	94,12%	92,49%	90,67%	Support vector classifier
Klldc	95,22%	91,25%	82,97%	84,60%	Linear classifier by KL exp. of common cov matrix
Ldc	95,22%	91,25%	82,97%	84,60%	Normal densities based linear classifier
Pcldc	95,22%	91,25%	82,97%	84,60%	Linear classifier by PCA expansion on the joint data
Loglc	95,16%	93,68%	90,42%	87,07%	Logistic linear classifier

*OA: Outlier Analysis.

3. McConkey BJ, Sobolev V, Eldman M: **Discrimination of native protein structures using atom-atom contact scoring.** *Proc Natl Acad Sci* 2003, **100:**3215-3220.
4. Wang K, Fain B, Levitt M, Samudrala R: **Improved protein structure selection using decoy-dependent discriminatory functions.** *BMC Struct Biol* 2004, **4:**8.
5. Taylor T, Vaisman II: **Graph theoretic properties of networks formed by the Delaunay tessellation of protein structures.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2006, **73:**041925.
6. Alper Küçükural, Uğur Sezerman: **Finding Common Domains of Proteins Using Parallelized Attributed Inexact Sub-graph Matching Algorithm.** *TAM* 2006.
7. Miyazawa S, Jernigan RL: *J Mol Biol* 1996, **256:**623-644.
8. Liang J, Dill KA: **Are proteins Well-Packed?** *Biophys J* 2001, **81:**751-766.
9. Atilgan AR, Akan P, Baysal C: **"Small-World Communication of Residues and Significance for Protein Dynamics,".** *Biophys J* 2004, **86:**85-91.
10. Vendruscolo M, Kussel E, Domany E: **Recovery of Protein Structure from Contact Maps.** *Fold Des* 1997, **2:**295-306.