Active site comparison: Assigning function to protein structures

Alexander Stark* & Robert B. Russell

EMBL, Heidelberg, Germany

The number of known protein three-dimensional structures is exponentially growing and structural genomics projects now target proteins of unknown function. As a result, there is a growing need to annotate protein structures for function via comparison to others. The most standard approaches involve structural alignment: given a new protein one can search for others that share a similar fold often in the absence of any sequence similarity. These methods are highly successful, but can fail if a protein adopts a new fold, or one that performs many functions.

We have developed an alternative approach that searches only for similarity between constellations of side-chains likely to be functionally important, thus allowing the detection of convergent functional similarities such as the trypsin catalytic triad. PINTS (Patterns in Non-Homologous Tertiary Structures) can compare structures to the database in seconds, and readily detects known and novel similarities.

An important feature is a statistical E-value (expectation value, like that used in Blast) that allows one to judge whether a database match is likely to be meaningful, and to compare results across searches involving functional sites of different sizes, something that is not possible with RMSD. The E-value is based on a statistical model that uses the geometrical meaning of increasing values of RMSD to predict the number of randomly matching patterns for different searches. In this talk, I will provide an overview of the method and the statistical model, discuss previously known and new examples that the method detects, and provide a guide to its use as a tool for structural genomics functional annotation.

References

Lorentzen, E., et al. (2003). Crystal structure of an archaeal class I aldolase and the evolution of $(\beta\alpha)_8$ barrel proteins. *J Biol Chem*, **278**, 47253-60.

PINTS (Patterns in Non-homologous Tertiary Structures) is available online:
    `http://pints.embl.de`

Stark, A. and Russell, R.B. (2003). Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. *Nucleic Acids Res*, **31**, 3341-4.

Stark, A., Shkumatov, A., and Russell, R.B. (2004). Finding functional sites in structural genomics proteins. *Structure*, accepted.

Stark, A., Sunyaev, S., and Russell, R.B. (2003) A model for statistical significance of local similarities in structure. *J Mol Biol*, **326**, 1307-16.