

Poster H-60
Improvements in ELM Resource
Filtering and Web Service
Technology Implementation



Authors:

Claudia Chica (*EMBL*)
Kate Gould (*EMBL*)
Allegra Via (*University of Rome Tor Vergata*)
Jan Christian Bryne (*Bergen Centre for Computational Science*)
Francesca Diella (*EMBL*)
Pal Puntervoll (*Bergen Centre for Computational Science*)
Rein Aasland (*University of Bergen*)
Manuela Helmer-Citterich (*University of Rome Tor Vergata*)
Toby Gibson (*EMBL*)

Short Abstract: ELM predicts functional sites in eukaryotic proteins. Overprediction is limited by biological context information filtering. By refinement of the structural filter and the addition of conservation filtering, the predictive power of ELM can be greatly improved. Grid technology implementation is in progress to aid extensibility and interoperability of the tool.

Long Abstract:

ELM (<http://elm.eu.org>) is a widely used resource dedicated to the study of eukaryotic functional sites. ELMs (Eukaryotic Linear Motifs) are typically 4-10 residues long and are associated with a particular molecular function (e.g. targeting peptides, glycosylation sites, phosphorylation sites etc).

The key parts of the ELM resource are (1) a database of experimentally verified functional site instances and (2) a prediction server which combines regular expression motif matches and context filtering to score the reliability of motifs found in sequences submitted by the user.

Due to their shortness in length, ELMs are prone to overprediction. For this reason different filters have been implemented in order to reduce the amount of false positives:

- taxonomy filter: functional sites have a specific taxonomic range
- cellular compartment filter: functional sites are only functional in certain cell compartments
- globular domain filter: functional sites predicted in globular domains are less likely to be true.

We have recently reevaluated and extended the problematic NES motif (nuclear export sequence) and this example will be used to illustrate the difficulties in motif assignment.

The predictive power of ELM is to be strengthened by the refinement of the structural filter and the addition of a conservation filter.

The structural filter initially subtracted the ELM matches inside globular domains identified with the SMART/Pfam domain databases. This may involve the removal of false negative candidate ELMs, as it is known that some ELMs reside within globular domains on surface exposed loops. It has been tuned by using the known three-dimensional information, such as residue solvent accessibility and secondary structure features, and benchmark data to

determine a score to classify ELMS inside globular domains as being possible candidates or true negatives. A desirable future addition to this filter would be a link to an exploratory structure display tool where the linear motif matches are highlighted.

The conservation filter uses the information retrieved from alignments of homologous proteins to score each predicted motif. The score is calculated by giving a weight to the motif's presence/absence in the homologous sequences depending on their similarity to the query sequence.

As a participant in the EMBRACE Network (<http://www.embracegrid.org>), the ELM resource is currently in the process of being re-wired as a set of web services using grid and web services technology. The aim of EMBRACE is to provide a better way to integrate bioinformatic resources in Europe.

As the ELM resource architecture is modular in character i.e. consisting of components that integrate ELM specific information and data from external resources, it is a typical scenario exemplifying and promoting the use and advantages of grid-enabled web services. A prototype ELMdb web service already exists which provides access to key information of interest in the ELM resource database. Once all the dependant modules are available as web services, the ELM resource web service can incorporate all their outputs and make ELM prediction results available to clients. This framework architecture will accommodate future extensibility of ELM and will provide a standard programmatic interface to the ELM resource.