

Eukaryotic Linear Motifs in the ELM Web Tool

Pål Puntervoll⁽²⁾, Rune Linding⁽¹⁾, Christine Gemünd⁽¹⁾, Sophie Chabanis-Davidson⁽¹⁾, Morten Mattingsdal⁽²⁾, Scott Cameron⁽³⁾, David M.A. Martin⁽³⁾, Gabriele Ausiello⁽⁴⁾, Barbara Brannetti⁽⁴⁾, Anna Costantini⁽⁴⁾, Andreas Zanzoni⁽⁴⁾, Vincenza Maselli⁽⁴⁾, Allegra Via⁽⁴⁾, Gianni Cesareni⁽⁴⁾, Francesca Diella⁽⁵⁾, Giulio Superti-Furga⁽⁵⁾, Lucjan Wyrwicz⁽⁶⁾, Chenna Ramu⁽¹⁾, Caroline McGuigan⁽¹⁾, Rambabu Gudavalli⁽¹⁾, Ivica Letunic⁽¹⁾, Peer Bork⁽¹⁾, Leszek Rychlewski⁽⁶⁾, Bernhard Kuster⁽⁵⁾, Manuela Helmer-Citterich⁽⁴⁾, William N. Hunter⁽³⁾, Rein Aasland⁽²⁾ and Toby J. Gibson⁽¹⁾

⁽¹⁾European Molecular Biology Laboratory

⁽²⁾Department of Molecular Biology, University of Bergen, Norway

⁽³⁾Division of Biological Chemistry and Molecular Microbiology, University of Dundee, UK

⁽⁴⁾Centre for Molecular Bioinformatics, Department of Biology, University of Rome Tor Vergata, Rome, Italy

⁽⁵⁾Cellzome AG, Heidelberg, Germany

⁽⁶⁾BioInfoBank Institute, Poznan, Poland

Reflecting the modular nature of eukaryotic proteins, several WWW servers (e.g. PFAM, SMART, PROSITE) are dedicated to revealing domains in protein sequences. However, there is no resource, which specifically focuses on short functional motifs (targeting peptides, docking modules, glycosylation sites, phosphorylation sites, etc), yet these modules are just as important for function as the larger protein domains. Domains are identified by conventional methods, such as patterns (regular expressions) profiles or HMM models. But statistically robust methods cannot usually be applied to small motifs, while pattern-based methods over-predict enormously so that the few true motifs are lost amongst the many false positives.

ELM (Eucariotic Linear Motifs - <http://elm.eu.org>) [1] is a new web based tool for the prediction of these small motifs on eukaryotic protein sequences. At the moment, the ELM database contains manually curated information about 114 known linear motifs in the form of regular expressions, profiles or hidden markov models that identify the motifs on the sequence. ELM addresses the over prediction deficiency of other methods by the use of context-based rules and logical filters that exclude false positives. The current version of the ELM server provides core functionality including filtering by cell compartment, phylogeny, globular domain clash (using the SMART/Pfam databases), secondary structure, and solvent accessibility. The current set of motifs is not at all exhaustive. Filters work by comparing the information on the motifs stored in the db (taxonomic, structural and cellular context) with the information submitted by the user together with his sequence. The structural filter works by automatically modeling the submitted protein sequences, whenever a good template is found in the SCOP database, and comparing predicted solvent accessibility values and secondary structure features with the corresponding values associated to ELM matches on true positive structures. The ELM server was launched on November 2002 and regularly enhanced since then. The server activity has been running for several months at > 45,000 hits from > 1700 unique internet sites.

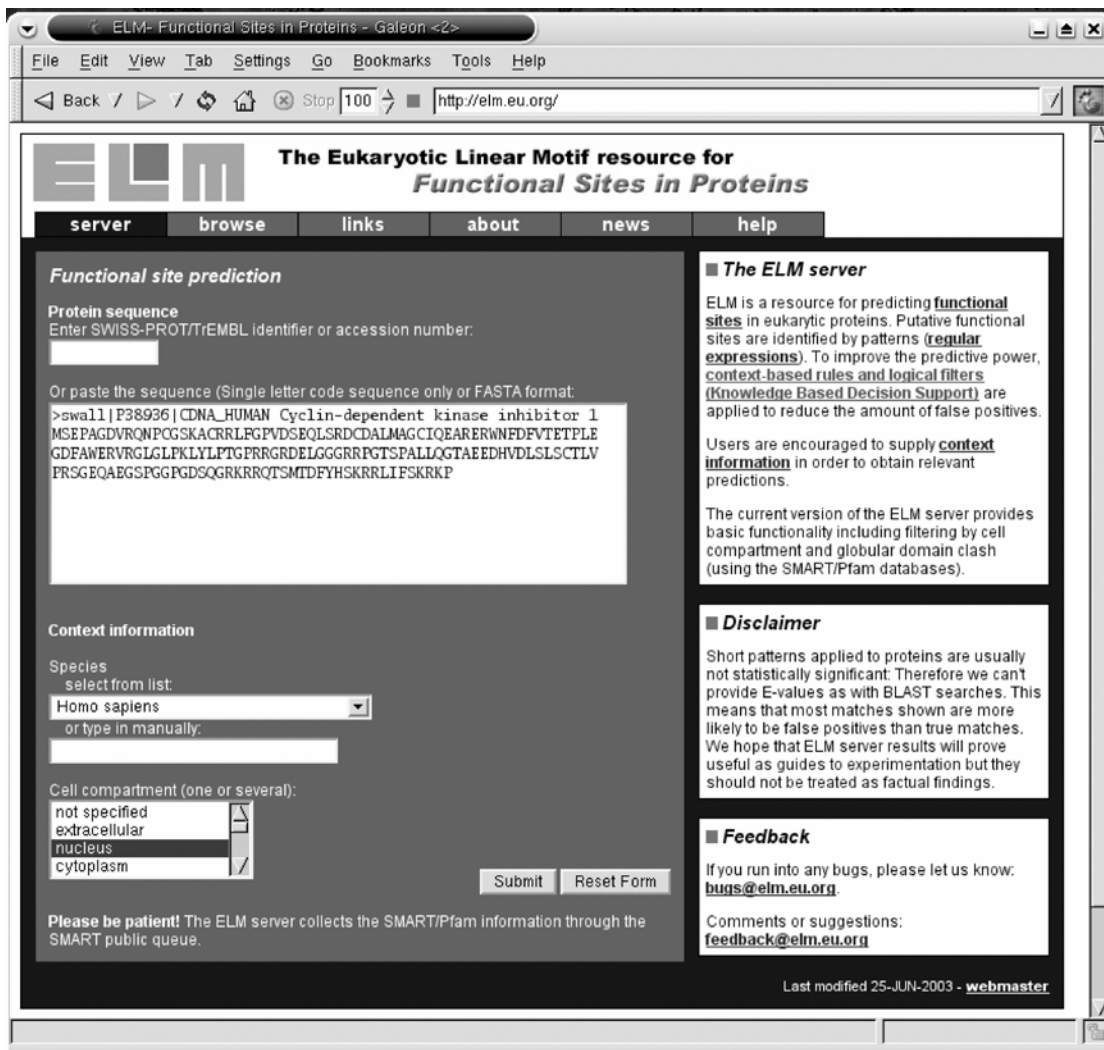


Fig. 1 The ELM resource for functional sites in proteins home page

References

- [1] P. Puntervoll, R. Linding, C. Gemünd, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D.M.A. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferrè, V. Maselli, A. Via, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Küster, M. Helmer-Citterich, W.N. Hunter, R. Aasland and T.J. Gibson, ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Research*, 31: 3625-3630, 2003