Short Linear Motifs and the Eukaryotic Linear motif resource

Toby Gibson, Manjeet Kumar & Holger Dinkel

[**Resources**]

UniProt http://www.uniprot.org/

ELM http://elm.eu.org

SlimSearch http://slim.ucd.ie/slimsearch/

[ELM exercises (contd)]

E1A adenoviral Protein

Objective: Apply the ELM (Eukaryotic Linear Motif) prediction tool to a viral protein.

Background Information: Adenoviruses are non-enveloped DNAds virus. Human adenoviruses are responsible for respiratory diseases, croup, and bronchitis outbreaks and gastroenteritis in children. The adenovirus E1A protein is unique to the Mastadenovirus genus. All members of the Mastadenovirus genus infects mammals. E1A plays a role in viral genome replication by driving entry of quiescent cells into the cell cycle. Stimulation of progression from G1 to S phase allows the virus to efficiently use the cellular DNA replicating machinery to achieve viral genome replication.

1. Search in ELM E1A_ADE05. Remember to define cellular compartments and taxonomic context.

a) What can you say about the structure of the protein?

b) How many annotated instances are?

c) How many annotated instances belong to cellular targets? How many are related?

d) How many phosphorylation sites are annotated in Phospho.ELM?

e) How many linear motifs for kinases are annotated and how many are predicted?

2. Search in ELM E1A_ADE02. Remember to define cellular compartments and taxonomic context.

a) What can you say about the structure of the protein?

Is this different from E1A_ADE05?

b) How many annotated instances are? Are those different from E1A_ADE05?

c) How many annotated instances belong to cellular targets?

How many are related?

d) How many instances are assigned by homology?

e) How many phosphorylation sites are annotated in Phospho.ELM?

f) How many linear motifs for kinases are annotated and how many are predicted?

3. If you have to test which kinase phosphorylates E1A, which of all the predictions would you test?

4. Search in ELM E1A_ADECR.

a) Which is the taxonomic context?

b) How many instances are annotated? Why do you think is that?

c) What can you say about the structure of the protein? What can you say in general about E1A proteins?

Helicobacter pylori CagA

Objective: Use ELM to predict Eukaryotic Linear Motifs in bacterial proteins.

Background Information: *H. pylori* infection causes gastritis, peptide ulcer or gastric cancer. There is a stronger probability to develop gastric cancer if an East Asian strain (like F32) is responsible for the infection compared to a Western strain (like NCTC 11637). East Asian and Western strains differ in the number and sequence context of the EPIYA motifs. (Higashi, H., et al., 2002; Jones, K.R., et al., 2009)

1. Paste in ELM prediction server the following sequences of CagA from a Western and an East Asian strain. Specify 'Cytosol' cell compartment, '*Homo sapiens*' and a Motif probability cutoff of 0.001.

> NCTC11637_CagA

MTNETIDQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKN
QYFSDFINKSNDLINKDNLIDIGSSIKSFQKFGTQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFL
KSAKQSFAGIIIGNQIRTDQKFMGVFDEFLKERQEAEKNGEPTGGDWLDIFLSFVFNKEQSSDVKEAINQEPVPHVQPDI
ATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYA
GNGGFGAKHDWNATVGYKNQQGDNVATLINVHMKNGSGLVIAGGEKGINNPSFCLYKEDQLTGSQRALSQEEIRNKIDFM
EFLAQNNAKLDNLSEKEKEKFQNEIEDFQKDSKAYLDALGNDRIAFVSKKDPKHSALITEFGKGDLSYTLKDYGKKADRA
LDREKNVTLQGNLKHDSVMFVNYSNFKYTNASKSPDKGVGVTNGVSHLDAGFSKVAVFNLPDLNNLAITSFVRRNLENKL
VTEGLSLQEANKLIKDFLSSNKELVGKALNFNKAVADAKNTGNYDEVKKAQKDLEKSLRKREHLEKEVEKKLESKSGNKN
KMEAKAQANSQKDKIFALINKEANRDARAIAYSQNLKGIKRELSDKLEKINKDLKDFSKSFDEFKNGKNKDFSKAEETLK
ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIVNQKITDKVDNLNQAVSMAKATGDF
SRVEQALADLKNFSKEQLAQQTQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSGIEATALAKNFSDIKKELNEKFKNFN
NNNNNGLENEPIYAKVNKKKTGQVASPEEPIYAQVAKKVNAKIDRLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVS
PEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
GPFPLKRHDKVDDLSKVGLSRNQELAQKIDNLSQAVSEAKAGFFSNLEQTIDKLKDSTKYNSVNLWVESAKKVPASLSAK
LDNYATNSHTRINSNIQNGAINEKATGMLTQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKL
NNAVKDVKSSFTQFLANAFSTGYYSLARENAEHGIKNVNTKGGFQKS

> F32_CagA

MTNETIDQTTTPDQTGFVPQRFINNLQVAFIKVDNAVASFDPDQKPIVDKNDKDNRQAYEKISQLREEYANKAIKNPAKK
NQYFSDFINKSNDLINKDNLIAVDSSVESFRKFGDQRYQIFTSWVSLQKDPSKINTQQIRNFMENVIKPPISDDKEKAEF
LRSAKQSFAGIIIGNQIRSDEKFMGVFDESLKARQEAEKNAEPAGGDWLDIFLSFVFNKKQSSDLKETLNQEPRPDFEQN
LATTTTDIQGLPPEARDLLDERGNFFKFTLGDVEMLDVEGVADKDPNYKFNQLLIHNNALSSMLMGSHSNIEPEKVSLLY
GDNGGPEARHDWNATVGYKNQQGNNVATLINAHLNNGSGLIIAGNEDGIKNPSFYLYKEDQLTGLKQALSQEEIQNKVDF
MEFLAQNNAKLDNLSEKEKEKFQTEIENFQKDRKAYLDALGNDHIAFVSKKDPKHLALVTEFGNGELSYTLKDYGKKQDK
ALDGETKTTLQGSLKYDGVMFVNYSNFKYTNASKSPNKGLGTTNGVSHLEANFSKVAVFNLPNLNNLAITNYIRRDLEDK
LWAKGLSPQEANKLIKDFLNSNKEMVGKVSNFNKAVAEAKNTGNYDEVKKAQKDLEKSLRKREHLEKEVAKKLESRNDNK
NRMEAKAQANSQKDKIFALISQEASKEARVATFDPYLKGVRSELSDKLENINKNLKDFGKSFDELKSGKNNDFSKAEETL
KALKDSVKDLGINPEWISKIENLNAALNDFKNGKNKDFSKVTQAKSDLENSIKDVIINQKITDKVDNLNQAVSEIKLTGD
FSKVEQALAELKNLSLDLGKNSDLQKSVKNGVNGTLVSNGLSKTEATTLTKNFSDIRKELNEKLFGNSNNNNNGLKNNTE
PIYAQVNKKKTGQATSPEEPIYAQVAKKVSAKIDQLNEATSAINRKIDRINKIASAGKGVGGFSGAGRSASPEPIYATID
FDEANQAGFPLRRSAAVNDLSKVGLSREQELTRRIGDLSQAVSEAKTGHFGNLEQKIDELKDSTKKNALKLWVESAKQVP
TSLQAKLDNYATNSHTRINSNVQSGTINEKATGMLTQKNPEWLKLVNDKIVAHNVGSAPLSAYDKIGFNQKNMKDYSDSF
KFSTKLNNAVKDIKSSFVQFLTNTFSTGSYSLMKANVEHGVKNTNTKGGFQKS

1. What are the differences in EPIYA motif predictions? Is the 'Assigned by homology' indicator showing any difference?

[References:]

Alexander et al. Sci. Sig 2011 “Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling” [URL]

Davey NE, Travé G and Gibson TJ (2011), *"How viruses hijack cell regulation"*, Trends Biochem Sci., Mar, 2011. Vol. 36, pp. 159-169. [DOI] [URL]

Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H and Gibson TJ (2012), *"Attributes of short linear motifs"*, Mol Biosyst., Jan, 2012. Vol. 8, pp. 268-281. [DOI] [URL]

Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kuban M, Strumillo M, Uyar B, Budd A, Altenberg B, Seiler M, Chemes LB, Glavina J, Sánchez IE, Diella F, Gibson TJ. (2015), "*The eukaryotic linear motif resource ELM: 10 years and counting.*" Nucleic Acids Res., Nov, 2013. [DOI] [URL]

Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ and Diella F (2011), *"Phospho.ELM: a database of phosphorylation sites--update 2011."*, Nucleic Acids Res., Jan, 2011. Vol. 39 (Database issue), pp. D261-D267. [DOI] [URL]

Dyson HJ and Wright PE (2005), *"Intrinsically unstructured proteins and their functions"*, Nat Rev Mol Cell Biol., Mar, 2005. Vol. 6, pp. 197-208. [DOI] [URL]

Van Roey K, Orchard S, Kerrien S, Dumousseau M, Ricard-Blum S, Hermjakob H and Gibson TJ (2013), *"Capturing cooperative interactions with the PSI-MI format"*, Database (Oxford). Vol. 2013, pp. bat066. [DOI] [URL]

Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. (2013) “*The switches.ELM resource: a compendium of conditional regulatory interaction interfaces.*” Sci Signal. 2013 Apr 2;6(269):rs7. [DOI] [URL]

Gibson TJ, Dinkel H, Van Roey K, Diella F (2015) “Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad”, Cell Communication & Signalling [URL]

Mészáros B, Kumar M, Gibson TJ, Uyar B, Dosztányi Z. (2017) "Degrons in cancer." Sci Signal. 2017 Mar 14 [URL]