

Short Linear Motifs and the Eukaryotic Linear motif resource

Toby Gibson, Manjeet Kumar & Holger Dinkel

Resources:

- ELM <http://elm.eu.org>
- UniProt <http://www.uniprot.org/>
- PeCan <https://pecan.stjude.cloud/proteinpaint/>
- ProViz <http://proviz.ucd.ie>

1 ELM exercises

Objective: Get familiar with the ELM (Eukaryotic Linear Motif) prediction tool.

1.1 Search in ELM by copy/pasting the following sequence and using the following parameters:

>P12931

```
MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
PKLFGGFNSSDTVTSPQ RAGPLAGGVTTTFVALYDYESRTETDLSFKKGERLQIVNNTEGD
WWLAHSLSTGQTGYIPSNYVAPSDSIQAEWYFGKITRRESERLLLNAENPRGTFLVRES
ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYSKHADGL
CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGC FGEVWMGTWNGTTRVAIKTL
KPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEPYIVTEYMSKGSLLDFLKGETGKY
LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADDFGLARLIEDNEYT
ARQGAKFPKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVL DQVER
GYRMPCPPECPESLHDL MCQCWRKEPEERPTFEYLQAFLEDYFTSTEPQYQPGENL
```

- Cell Compartment: **Not specified**
 - Motif Probability Cutoff: **100**
 - Context information: **(leave blank)**
1. Pay attention to many instances you find
 2. What can you say about the structure of the protein?
 - a. Do you find any domains?
 - b. Do you find any disordered regions?

1.2 Repeat the previous search (again accession P12931) using these parameters:

- Cell Compartment: **cytosol**
 - Motif Probability Cutoff: **0.01**
 - Context information: **Homo sapiens**
1. How many instances (roughly) do you find now?

2. How many of the instances are 'annotated'?
3. Do the structural predictors/filters (SMART, GlobPlot, IUPRED, Secondary Structure) agree in terms of which regions are structured/disordered?
4. Compare the location of the annotated instances with structural information at hand (IUPRED, Secondary Structure).

1.3 Submit the sequence of Paxillin (P49023) to ELM, using default parameters.

1. Compare the results with a search for the same sequence when using the cellular compartment 'plasma membrane'

1.4 Search protein SRC_MOUSE (P05480) for ELMs.

1. Do you find "annotated instances"?
2. If not, what's the closest to an 'annotated instance' that you can find? Investigate where this information might come from.

1.5 Submit the entry name 'P53_HUMAN'

1. Do the cell compartments make sense?
2. How many degrons are there in p53?
3. Is there a CDK site in p53? Is there a Cyclin Box in p53?

1.6 (Optional) Search ELM using the protein name 'MDM4_HUMAN' and look for the 'USP binding motif' DOC_USP7_MATH_1.

1. How many such motif instances are found in this protein sequence?

1.7 (Optional) Repeat this exercise with protein 'AMPH_HUMAN' and ELM class 'LIG_Clathr_ClatBox_1'

1. Try to assess the biological relevance of each of these instances.
2. Is the annotation for the biological relevance in accordance with the globular structure?

1.8 You're studying the cell surface expression of a receptor and find out that some isoforms are expressed at the surface (Q05586-2) while another isoform is retained in the endoplasmic reticulum (Q05586-5). You want to investigate a possible role of linear motifs in this phenomenon.

1. First, align these sequences to see which parts are similar/identical and which are different (go to <http://www.uniprot.org/uniprot/Q05586>, scroll down to 'sequences', select isoforms 2 and 5 and click 'align').
2. Then use <http://elm.eu.org> to scan these sequences for linear motifs, using cell compartment filter 'cytosol'. You're looking for targeting motifs (TRG_*).
3. By looking through the annotations of these targeting ELM classes, can you find motif instances that might be responsible for the different behaviour of the isoforms?
4. If there are multiple instances of that motif found per protein, can you use differential information (comparing the motifs found in the different isoforms) to narrow down the number of candidate instances?
5. Next, you sequence another isoform (Q05586-4) which also features this motif at a homologous position, but strangely does not get expressed at the cell surface. You discuss this with your colleague and he tells you that he recently found out that this protein also binds to the PDZ domain of DLG4. Can you come up with a hypothesis how this all fits together?

2 PeCan exercise

2.1 Go to [ELM](#) and enter ETV1_HUMAN as search term in the search box on top right; you should find a single annotated true positive instance of a degron.

1. Where/What ELM class is it? Note down it's amino acid position.

Next go to [PeCan](#) and enter ETV1 as search term.

1. Do you find any pediatric mutations?
2. How about Cosmic? Select only “Fusion transcript” mutations; can you find any near the degron motif?
3. Which tissue types are these mutations found in?
4. Are there differences in fusions and frameshifts in tissue types?

2.2 Go to [ELM](#) and enter NOTC1_HUMAN as search term.

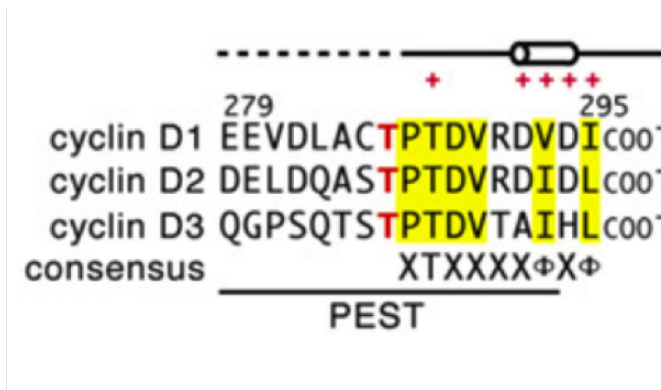
1. Find the annotated FBW7 degron (TP) and note down its position.

2.3 Now go to [PeCan](#) and enter notch1 as search term.

Select only “Frameshift” and “nonsense” mutations; if possible load COSMIC data as well; analyse the area around the degron; additionally, by adding “missense” is there another hotspot?

2.4 CCND2 possesses a degron in the C-Terminus, however the entry has not yet been annotated in elm.eu.org.

Fortunately, a new structure came out recently: PMID [29279382](#); open CCND2 in PeCan and look at the differences in pediatric & Cosmic data



(graphics from: " [Structural basis of the phosphorylation-independent recognition of cyclin D1 by the SCF^{FBXO31} ubiquitin ligase](#)" *Li* PNAS 2018)

3 ProViz exercise

ProViz aggregates and displays useful information from many resources where relevant to linear motif discovery.

3.1 Go to the ProViz server <http://proviz.ucd.ie> and put p53 into the “Search for a Protein” field

1. Explore the results, then check out these operations and questions:

- a. Are there any methyllysine modifications?
- b. What is a cumulative switch?
- c. Are there any motifs in p53 that can antagonise MDM2 ubiquitination?
- d. Are there any splice variants in the DNA-binding domain? Do any motifs get removed in splice variants?
Do you think p53 researchers study all the splice variants?
- e. What do the long structure modules correspond to?
- f. Can tetrameric p53 be exported from the nucleus?
- g. Click on the structure element of a short segment of p53. It goes to the PDB website (RCSB). Display the structure in the browser and adjust the viewing controls.
- h. What is the difference between mutagenesis sites and sequence variants?
- i. Which residue has the most sequence variants listed?
- j. Are there any sequence variants in the MDM2 degron (towards the N-terminus)?