Tools & Databases of Short Linear Motifs

Holger Dinkel
EMBO Practical Course:
“Computational Analysis of Protein-Protein Interactions: Sequences, Networks and Diseases”
Budapest, 03. 06. 2016
Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling; Alexander et al.; (Sci. Sig 2011)
**Protein Phosphorylation Sites**

<table>
<thead>
<tr>
<th>Kinase</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cdk1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>p[ST]</td>
<td>P</td>
<td>.</td>
<td>[KR]</td>
</tr>
<tr>
<td>Nek2</td>
<td>[FML]</td>
<td>[iP]</td>
<td>[iP]</td>
<td>p[ST]</td>
<td>[ILMV]</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

"Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling"; 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"; 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; Alexander et al.; (Sci. Sig 2011)
Kinase localization in Metaphase:
- **Cdk1**: whole cell
- **Plk1**: kinetochores
- **Aurora A**: centrosomes & microtubules
- **Aurora B**: centromeres & spindle
- **Nek2**: centrosomes

"Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling"; Alexander et al.; (Sci. Sig 2011)
Phospho.ELM

Database of experimentally verified phosphorylation sites in eukaryotic proteins. Current release contains 8,718 protein entries covering more than 42,500 instances. (Instances are fully linked to literature references.)
SEARCH

- for phosphorylation sites in proteins using protein name or gene name (e.g. Paxillin, Src, MAPK)

- by UniPROT accession or Ensembl identifier: (e.g. P12931 or P55211)

- by selected kinase (List):
  - None

- by selected phospho-peptide binding domain (List):
  - None

- Choose which organisms to include
  - All
  - Caenorhabditis
  - Drosophila
  - Vertebrates

- Do not show high throughput data

- Output as Comma-Separated-Values (.csv)
### Phospho.ELM

**a database of S/T/Y phosphorylation sites**

**Substrate:** p53 (Cellular tumor antigen p53)

**Seq-ID:** P04637 [Homo sapiens]

**Interaction Network(s):**
- STRING
- NetworKIN

**External Source(s):** PHOSIDA

**MINT Interaction(s):** [show]

**GO-Terms:** [show]

**Conservation:**

<table>
<thead>
<tr>
<th>Res</th>
<th>Pos</th>
<th>Sequence</th>
<th>Kinase</th>
<th>PMID</th>
<th>Src</th>
<th>Cons.</th>
<th>ELM</th>
<th>Binding Domain</th>
<th>SMART/Pfam</th>
<th>IUPRED score</th>
<th>P3D Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>9</td>
<td>KEEPQDSVPEPPLQETF</td>
<td></td>
<td></td>
<td>LTP</td>
<td>0.75</td>
<td></td>
<td>-</td>
<td>P53_TAD</td>
<td>0.94</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>15</td>
<td>QSDSVEPPLQGSTSFGLKL</td>
<td>DNA-PK</td>
<td>10446957</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_PIKK_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>15</td>
<td>QSDSVEPPLQGSTSFGLKL</td>
<td>ATM</td>
<td>11875057</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_PIKK_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>PSVEPPLQETFSDNWKLPE</td>
<td>CK1_group</td>
<td>10606744</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_CK1_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>PSVEPPLQETFSDNWKLPE</td>
<td>TTK</td>
<td>19332559</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_CK1_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>PSVEPPLQETFSDNWKLPE</td>
<td>VRK1</td>
<td>10951572</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_CK1_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>PSVEPPLQETFSDNWKLPE</td>
<td>VRK1</td>
<td>15542844</td>
<td>LTP</td>
<td>1.00</td>
<td>MOD_CK1_1</td>
<td>-</td>
<td>P53_TAD</td>
<td>0.66</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
<td>VEPPLQETFSDNWKLPE</td>
<td>-</td>
<td>15254178</td>
<td>LTP</td>
<td>0.95</td>
<td></td>
<td>-</td>
<td>P53_TAD</td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
<td>VEPPLQETFSDNWKLPE</td>
<td>-</td>
<td>15489221</td>
<td>LTP</td>
<td>0.95</td>
<td></td>
<td>-</td>
<td>P53_TAD</td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
<td>VEPPLQETFSDNWKLPE</td>
<td>-</td>
<td>10801407</td>
<td>LTP</td>
<td>0.95</td>
<td></td>
<td>-</td>
<td>P53_TAD</td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
<td>VEPPLQETFSDNWKLPE</td>
<td>-</td>
<td>12111733</td>
<td>LTP</td>
<td>0.95</td>
<td></td>
<td>-</td>
<td>P53_TAD</td>
<td>0.58</td>
<td>-</td>
</tr>
</tbody>
</table>
Tools & Databases of Short Linear Motifs

**Links to:**
- STRING
- NetworKin
- Phosida
- Phospho3D

**Display:**
- MINT interactions
- GO-Terms
Precalculated conservation scores for the phosphorylation sites are presented using Jalview.
PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, ubiquitination, acetylation and methylation. See About PhosphoSitePlus above for more information.


A PROTEIN MODIFICATION RESOURCE

WHAT'S NEW

Aug 2014  Download PTM-VarMut dataset: Overlap of disease missense mutations & genetic variants, with their corresponding PTMs and flanking sequences.

Jul 2012  Download Datasets of Regulatory or Disease-Associated Sites.


Jul 2011  Multiple Sequence Alignment (MSA) added to the Protein Page.


Phosphorylation Site Statistics

Non-redundant sites: 239,738
Non-redundant proteins: 19,680
Sites curated from literature: 136,109
All sites using site-specific (SS) methods: 12,528
All sites using discovery-mode MS (MS) methods: 127,064
Sites using both SS and MS methods: 6,010
MS sites observed at CST: 151,472
Number of curated papers: 16,428

Other Modification Site Statistics

Acetylation: 27,657  Caspase cleavage: 481
Di-methylation: 2,555  Methylation: 163
Mono-methylation: 4,992  O-GalNAc: 2,118
O-GlcNAc: 1,390  Succinylation: 4,657
Sumoylation: 816  Tri-methylation: 321
Ubiquitination: 51,255

PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at http://www.phosphosite.org/static/contact.do.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene/Symbol</th>
<th>ACC#</th>
<th>Organism</th>
<th>MW (Da)</th>
<th>Modifications (show legend)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td>TP53</td>
<td></td>
<td>human</td>
<td>43,683</td>
<td>H-m-1, K-ac, K-m-1, K-m-2, K- Sm, K-ub, R-m-1, S-yl, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>S3BP1</td>
<td>TP3BP1</td>
<td></td>
<td>human</td>
<td>213,374</td>
<td>D-ca, K-ac, K-m-1, K-ub, R-m-1, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>S3BP2</td>
<td>TP3BP2</td>
<td></td>
<td>human</td>
<td>125,610</td>
<td>K-ub, S-yl, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>AIFM2</td>
<td>AIFM1</td>
<td></td>
<td>human</td>
<td>40,527</td>
<td>K-ac, K-ub, S-p, Y-p</td>
</tr>
<tr>
<td>AK0S</td>
<td>AK0S</td>
<td></td>
<td>human</td>
<td>90,367</td>
<td>S-p, T-p, Y-p</td>
</tr>
<tr>
<td>CD1</td>
<td>CD1</td>
<td></td>
<td>human</td>
<td>21,892</td>
<td>K-ub, T-p</td>
</tr>
<tr>
<td>CYFIP2</td>
<td>CYFIP2</td>
<td></td>
<td>human</td>
<td>148,398</td>
<td>K-ac, K-ub, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>EFEMP2</td>
<td>EFEMP2</td>
<td></td>
<td>human</td>
<td>49,405</td>
<td>Y-p</td>
</tr>
<tr>
<td>ENC1</td>
<td>ENC1</td>
<td></td>
<td>human</td>
<td>66,130</td>
<td>K-ub, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>GADD45G1</td>
<td>GADD45G1</td>
<td></td>
<td>human</td>
<td>25,487</td>
<td>K-ub, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>IQCB1</td>
<td>IQCB1</td>
<td></td>
<td>mouse</td>
<td>68,734</td>
<td>K-m-2, K-ub, S-p, T-p</td>
</tr>
<tr>
<td>ERP52</td>
<td>ERP52</td>
<td></td>
<td>human</td>
<td>60,868</td>
<td>K-ac, K-ub, S-yl, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>JNK</td>
<td>JNK</td>
<td></td>
<td>mouse</td>
<td>111,345</td>
<td></td>
</tr>
<tr>
<td>LGALS7B</td>
<td>LGALS7B</td>
<td></td>
<td>human</td>
<td>15,075</td>
<td>Y-p</td>
</tr>
<tr>
<td>LITAF</td>
<td>LITAF</td>
<td></td>
<td>human</td>
<td>12,073</td>
<td>K-ub, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>MADIII1</td>
<td>MADIII1</td>
<td></td>
<td>human</td>
<td>83,647</td>
<td>K-ac, K-ub, S-p, T-p, Y-p</td>
</tr>
<tr>
<td>PhosphoSitePlus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.phosphosite.org/proteinaction.do?id=466&amp;showAffSites=true">www.phosphosite.org/proteinaction.do?id=466&amp;showAffSites=true</a></td>
</tr>
</tbody>
</table>
p53, a transcription factor and major tumor suppressor that plays a major role in regulating cellular responses to DNA damage and other genomic aberrations. Activation of p53 can lead to either cell cycle arrest and DNA repair or apoptosis. More than 50 percent of human tumors contain a mutation or deletion of the TP53 gene. p53 is modified post-translationally at multiple sites: DNA damage induces phosphorylation of p53 at S15, S20 and S37, reducing its interaction with the oncoprotein MDM2. MDM2 inhibits p53 accumulation by targeting it for ubiquitination and professional degradation. Phosphorylation by many kinases including Chk2 and Chk1 at S20, enhancing its tetramerization, stability and activity. The phosphorylation by Chk at S392 is increased in human tumors and has been reported to influence the growth suppressor function, DNA binding and transcriptional activation of p53. Phosphorylation of p53 at S46 regulates the ability of p53 to induce apoptosis. The acetylation of p53 appears to play a positive role in the accumulation of p53 during the stress response. Following DNA damage, p53 becomes acetylated at K382, enhancing its binding to DNA. Deacetylation of p53 can occur through interaction with SIRT1, a deacetylase that may be involved in cellular aging and the DNA damage response. p53 regulates the transcription of a set of genes encoding endosomal proteins that regulate endosomal functions. These include STEAP3 and CHMP4C, which enhance exosome production, and CAV1 and CHMP4C, which produce a more rapid endosomal clearance of the EGF receptor from the plasma membrane. DNA damage regulates p53-mediated secretory pathway, increasing the secretion of some proteins such as Hepp3, SERPIN1, SERPIN5, NEK1, and CypA, and inhibiting the secretion of others including CTSI and IGF-BP-2. Two alternatively spliced human isoforms have been reported. Isoform 2 is expressed in quiescent lymphocytes. Seems to be non-functional. May be produced at very low levels due to premature stop codon in the mRNA, leading to nonsense-mediated mRNA decay. Note: this description may include information from UniProtKB.

**Protein type:** DNA binding protein; Nuclear receptor co-regulator; Motility/polarity/chemotaxis; Transcription factor; Activator protein; Tumor suppressor

**Cellular Component:** PML body; transcription factor TFIIH complex; protein complex; nuclear matrix; mitochondrial; endoplasmic reticulum; replication fork; cytosol; nucleus; nuclear body; mitochondrial matrix; cytoplasm; nuclear chromatin; nucleolus; chromatin; nucleus

**Molecular Function:** Identical protein binding; protease binding; zinc ion binding; protein phosphatase 2A binding; p53 binding; protein N-terminus binding; receptor tyrosine kinase binding; protein phosphaestase binding; protein kinases binding; histone acetyltransferase binding; protein binding; copper ion binding; histone deacetylase regulator activity; enzyme binding; DNA binding; protein heterodimerization activity; chaperone binding; ubiquitin protein ligase binding; damaged DNA binding; chromatin binding; transcription factor activity; ATP binding

**Biological Process:** centrosomal process; development; viral reproduction; positive regulation of apoptosis; multiple cellular organism development; positive regulation of transcription, DNA-dependent; T cell differentiation in the thymus; gastrulation; determination of adult life span; DNA damage response; signal transduction by p53 class mediator resulting in cell cycle arrest; response to antibiotic; regulation of apoptosis; cellular response to glucose starvation; protein localization; positive regulation of neuronal proliferation; basic excision repair; transforming growth factor beta receptor signaling pathway; protein complex assembly; cell cycle arrest; ER overload response; response to X-radiation; somatogenesis; release of cytokine from mitochondrion; chromatin assembly; cell aging; RNA transcription; positive regulation of peptidyl-tRNA phosphorylation; negative regulation of DNA replication; negative regulation of ribosome biogenesis; positive regulation of transcription from RNA polymerase II promoter; regulation of mitochondrial membrane permeability; negative regulation of transcription, DNA-dependent; regulation of tissue remodeling; negative regulation of apoptosis; DNA damage checkpoint; DNA damage response; signal transduction by p53 class mediator; apoptosis; negative regulation of transcription from RNA polymerase II promoter; response to salt stress; negative regulation of cell proliferation; positive regulation of protein alkylation; negative regulation of histone deacetylation; DNA damage response; signal transduction by p53 class mediator resulting in transcription of p21 class mediator; regulation of transcription, DNA-dependent; T cell proliferation during immune response; double-strand break repair; positive regulation of neuron apoptosis; response to gamma radiation; cell differentiation; DNA damage response; signal transduction by p53 class mediator resulting in induction of apoptosis; cell growth pathway; in utero embryonic development; multicellular organismal growth; B cell lineage commitment; cell proliferation; neuron apoptosis; T cell lineage commitment; negative regulation of helicase activity; nucleotide-excision repair; protein import into nucleus; nucleocytoplasmic; Ras protein signal transduction; negative regulation of cell growth; negative regulation of transforming growth factor beta receptor signaling pathway; blood coagulation; response to DNA damage stimulus

**Reference:** P04637 (UniProtKB)
## Modification Sites in Parent Protein, Orthologs, and Isoforms

**Tools & Databases of Short Linear Motifs**

<table>
<thead>
<tr>
<th>SS</th>
<th>NS</th>
<th>Human</th>
<th>Mouse</th>
<th>Rat</th>
<th>Rabbit</th>
<th>Monkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>P4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>P5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>P6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>P7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>P8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>P9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>P10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>P11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>P12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>P13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>P14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questions?

CURIOSITY

"For every answer, there are but two more questions."

motifake.com
The ELM resource

is a collection of more than 240 thoroughly annotated motif classes with over 2700 annotated instances. It is also a prediction tool to detect these motifs in protein sequences employing different filters to distinguish between functional and non-functional motif instances.
The ELM resource

is a collection of more than 240 thoroughly annotated motif classes with over 2700 annotated instances.

It is also a prediction tool to detect these motifs in protein sequences employing different filters to distinguish between functional and non-functional motif instances.

<table>
<thead>
<tr>
<th>Functional sites</th>
<th>ELM classes</th>
<th>ELM instances</th>
<th>PDB structures</th>
<th>GO terms</th>
<th>PubMed links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>159</td>
<td>2702</td>
<td>348</td>
<td>549</td>
<td>2439</td>
</tr>
<tr>
<td>By category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIG</td>
<td>137</td>
<td>Human 1594</td>
<td></td>
<td>Biological Process 283</td>
<td>From class 1174</td>
</tr>
<tr>
<td>MOD</td>
<td>31</td>
<td>Mouse 253</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEG</td>
<td>25</td>
<td>Rat 130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOC</td>
<td>22</td>
<td>Yeast 94</td>
<td></td>
<td>Cellular Compartment 119</td>
<td>From instance 1746</td>
</tr>
<tr>
<td>TRG</td>
<td>20</td>
<td>Fly 90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLV</td>
<td>11</td>
<td>Other 541</td>
<td></td>
<td>Molecular Function 147</td>
<td></td>
</tr>
</tbody>
</table>
ELM Class

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]\text{L}\{0,1\}[FYLVMP] for Cyclin motif)
ELM Class

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]\times L\times \{0,1\}[FYLIVMP]" for Cyclin motif)

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Gene Name</th>
<th>Start</th>
<th>End</th>
<th>Subsequence</th>
<th>Logic</th>
<th>ORGanism</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB_HUMAN</td>
<td>RB1</td>
<td>873</td>
<td>877</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>CKB_HUMAN</td>
<td>CKB_HUMAN</td>
<td>394</td>
<td>398</td>
<td></td>
<td></td>
<td>Gallus gallus (Chicken)</td>
<td></td>
</tr>
<tr>
<td>PMY1_HUMAN</td>
<td>PMY1</td>
<td>466</td>
<td>469</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>E2F1_HUMAN</td>
<td>E2F1</td>
<td>90</td>
<td>94</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>CDN1C_HUMAN</td>
<td>CDN1C</td>
<td>31</td>
<td>34</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>RNR1_MOUSE</td>
<td>RNR1</td>
<td>248</td>
<td>251</td>
<td></td>
<td></td>
<td>Drosophila melanogaster (Fly)</td>
<td></td>
</tr>
<tr>
<td>E2F2_HUMAN</td>
<td>E2F2</td>
<td>87</td>
<td>91</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>E2F3_HUMAN</td>
<td>E2F3</td>
<td>134</td>
<td>138</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>AKAP12_MOUSE</td>
<td>AKAP12</td>
<td>501</td>
<td>504</td>
<td></td>
<td></td>
<td>Mus musculus (House mouse)</td>
<td></td>
</tr>
<tr>
<td>CDC5_HUMAN</td>
<td>CDC5</td>
<td>94</td>
<td>98</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>CDC19_HUMAN</td>
<td>CDC19</td>
<td>19</td>
<td>22</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>CDC24A_HUMAN</td>
<td>CDC24A</td>
<td>155</td>
<td>159</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>ORC8_YEAST</td>
<td>ORC8</td>
<td>378</td>
<td>382</td>
<td></td>
<td></td>
<td>Saccharomyces cerevisiae (Baker’s yeast)</td>
<td></td>
</tr>
<tr>
<td>P53_HUMAN</td>
<td>P53</td>
<td>381</td>
<td>385</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>RBL1_HUMAN</td>
<td>RBL1</td>
<td>658</td>
<td>661</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>RBL2_HUMAN</td>
<td>RBL2</td>
<td>680</td>
<td>684</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
<tr>
<td>HIRA_HUMAN</td>
<td>HIRA</td>
<td>629</td>
<td>633</td>
<td></td>
<td></td>
<td>Homo sapiens (Human)</td>
<td></td>
</tr>
</tbody>
</table>

24 Instances for DOC_CYCLIN_1
(cick table headers for sorting: Notes column: #Number of Switches, %Number of Interactions)

ELM Instance

An experimentally verified instance of an ELM class in a particular sequence.
**The ELM Database**

**ELM Class**

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]\times[Lx\{0,1\}][FYLIVMP]" for Cyclin motif)

| Instance |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Sequence | Start | End | Subsequence | Logic | PDB | Organism | Length |
| (Q99741) CDC5_HUMAN | 94 | 98 | SNTTXXX[LM]XXBRSTER | TP | Homo sapiens (Human) | 500 |

**ELM Instance**

An experimentally verified instance of an ELM class in a particular sequence.

- Experimental Evidences
- Methods
- References
- Interactions
**ELM Class**

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. "[KR]xLx{0, 1}[FYLIVMP]" for Cyclin motif)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Start</th>
<th>End</th>
<th>Subsequence</th>
<th>Logic</th>
<th>PDB</th>
<th>Organism</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDG5_HUMAN</td>
<td>94</td>
<td>98</td>
<td>SYNT18000335</td>
<td>TP</td>
<td>Homo sapiens (Human)</td>
<td>560</td>
<td></td>
</tr>
</tbody>
</table>

**Instance evidence**

- **Evidence class**: experimental
- **PSMI**: MDD11
- **Method**: X-ray crystallography
- **BioSource**: Homo sapiens (Human)
- **PubMed**: [Cheng, 2006]
- **Logic**: support
- **Reliability**: certain
- **Notes**: Interaction/Detection, Feature/Detection

This ELM Instance is part of the following switching mechanism(s) annotated at the Switches.ELM resource:

- SWNT800335: CDG5
  - **Cyclin**: DOC_CYCLIN
  - **CDG5**: DOC_CYCLIN

**ELM Instance**

An experimentally verified instance of an ELM class in a particular sequence.

- Experimental Evidences
- Methods
- References
- Interactions
ELM Class

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]xLx{0,1}[FYLIVMP] for Cyclin motif)

<table>
<thead>
<tr>
<th>Instance</th>
<th>Sequence</th>
<th>Start</th>
<th>End</th>
<th>Subsequence</th>
<th>Logic</th>
<th>PDB</th>
<th>Organism</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOC597441_CD45_HUMAN</td>
<td>94</td>
<td>98</td>
<td>KYTXIKDAW3KGPDLR5T3</td>
<td>TP</td>
<td>Homo sapiens (Human)</td>
<td>560</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ELM Instance**

An experimentally verified instance of an ELM class in a particular sequence.

- Experimental Evidences
- Methods
- References
- Interactions
**ELM Class**

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]xLx{0,1}[FYLIVMP] for Cyclin motif")

<table>
<thead>
<tr>
<th>Instance</th>
<th>Sequence</th>
<th>Start</th>
<th>End</th>
<th>Subsequence</th>
<th>Logic</th>
<th>PDB</th>
<th>Organism</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>▼0298741 CDCS_HUMAN</td>
<td>94</td>
<td>98</td>
<td>SNTIRI03530</td>
<td>TP</td>
<td>Homo sapiens (Human)</td>
<td>56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ELM Instance**

An experimentally verified instance of an ELM class in a particular sequence.

- Experimental Evidences
- Methods
- References
- Interactions
The ELM Database

ELM Class

Condensed information about a motif. Regular Expressions used to annotate the motif (eg. 
"[KR]xLx\{0,1\}[FYLIVMP] for Cyclin motif)
TRG_AP2beta_CARGO_1

Accession: ELME000247

Functional site class: AP-2 beta2 appendage CCV component motifs

Functional site description: Several motifs are responsible for the binding of accessory endocytic proteins to the beta2-subunit appendage of the adaptor protein complex AP-2 as part of their recruitment to the site of clathrin coated vesicle (CCV) formation. Proteins binding the platform subdomain have been found to be cargo family specific (for example can load all GPCRs, or all LDL receptor family members) clathrin adaptors. Accessory proteins which help in CCV formation bind the sandwich subdomain site or the alpha car domain.

ELM Description: Motif binding as a helix in a depression on the top surface of the AP-2 beta appendage platform subdomain. The pattern [ED]xF(1,2)Fxx[FL]xxxR is conserved in beta Arrestins, ARH and Epsin-1, -2 of vertebrates. It is also found in homologues of other metazoans, but the pattern is sometimes not matched exactly, meaning that the ELM regular expression will not provide a match. In other lineages, if there is an equivalent motif, the pattern is likely to have diverged.

Pattern: [DE]xF(1,2)Fxx[FL]xxxR

Pattern Probability: 0.0000182

Present in taxon: Metazoa

Interaction Domain: B2-adapt-app_C (PF09066) Beta2-adaptin appendage, C-terminal sub-domain (Stoichiometry: 1:1)

PDB Structure: 2G30
### Tools & Databases of Short Linear Motifs

ELM DATABASE

**The Eukaryotic Linear Motif resource for Functional Sites in Proteins**

#### Search ELM Instances

- **Full-Text Search:** Enter 'all' or '/*' ap2
- **Instance Logic:** true positive
- **Organism:** Homo sapiens

**58 Instances for search term 'ap2':**

<table>
<thead>
<tr>
<th>ELM identifier</th>
<th>Sequence</th>
<th>Start</th>
<th>End</th>
<th>Subsequence</th>
<th>Instance Logic</th>
<th>#Evidence</th>
<th>PDB</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>OPRD_HUMAN</td>
<td>241</td>
<td>246</td>
<td>GLMLRLPKYKLLSGSKEKD</td>
<td>true positive</td>
<td>8</td>
<td>---</td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_AP2beta_CARGO_1</td>
<td>ARRB1_HUMAN</td>
<td>385</td>
<td>395</td>
<td>TNDNFDVFDERFGUALGKMK</td>
<td>true positive</td>
<td>5</td>
<td>2IV8</td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>HG2A_HUMAN</td>
<td>19</td>
<td>24</td>
<td>DKPQWDMQRLBNSEQLP</td>
<td>true positive</td>
<td>5</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_2</td>
<td>EPS15_HUMAN</td>
<td>672</td>
<td>674</td>
<td>DPATSSSTDPRESSANSSSI</td>
<td>true positive</td>
<td>4</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_2</td>
<td>EPS15_HUMAN</td>
<td>692</td>
<td>694</td>
<td>SWTFLHDKFFAPGTUVA</td>
<td>true positive</td>
<td>4</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_2</td>
<td>EPS15_HUMAN</td>
<td>709</td>
<td>711</td>
<td>VAAATLDSATDPFNASVFGESF</td>
<td>true positive</td>
<td>4</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_AP2beta_CARGO_1</td>
<td>EPN1_HUMAN</td>
<td>377</td>
<td>386</td>
<td>FDTEFDSDFDRLRILP</td>
<td>true positive</td>
<td>4</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>ATP7A_HUMAN</td>
<td>1483</td>
<td>1488</td>
<td>SVTSEPKHSLVGRFDRED</td>
<td>true positive</td>
<td>4</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_SxIP_EBH_1</td>
<td>CLAP2_HUMAN</td>
<td>429</td>
<td>502</td>
<td>ASAQKRKIPFSGCSCRES</td>
<td>true positive</td>
<td>3</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_SxIP_EBH_1</td>
<td>CLAP2_HUMAN</td>
<td>493</td>
<td>525</td>
<td>LSVAKRSSRPRPSVSQGCSR</td>
<td>true positive</td>
<td>3</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>BCAM_HUMAN</td>
<td>604</td>
<td>609</td>
<td>MGVSGQPQETLHMGAGSGG</td>
<td>true positive</td>
<td>3</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>NPC1_HUMAN</td>
<td>1271</td>
<td>1276</td>
<td>KSCATERYKGYFRRHLNF</td>
<td>true positive</td>
<td>3</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_APGC_KENbox_2</td>
<td>CKAP2_HUMAN</td>
<td>700</td>
<td>84</td>
<td>KLKTMAQDKENKNKPAESKN</td>
<td>true positive</td>
<td>2</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_MAPK_1</td>
<td>MP2K1_HUMAN</td>
<td>3</td>
<td>11</td>
<td>MPKKPTIQMLNPAPDGSAV</td>
<td>true positive</td>
<td>2</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_MAPK_1</td>
<td>MP2K4_HUMAN</td>
<td>40</td>
<td>48</td>
<td>SSRQGRKKALKLFNAPPPFK</td>
<td>true positive</td>
<td>2</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_AP2beta_CARGO_1</td>
<td>ARH_HUMAN</td>
<td>256</td>
<td>266</td>
<td>DDDLEAEPSRALQRTNPQV</td>
<td>true positive</td>
<td>2</td>
<td>2G30</td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>TRG_LysEnd_APsAcLL_1</td>
<td>CD44_HUMAN</td>
<td>708</td>
<td>713</td>
<td>GEASKQMIVNYLKVNESET</td>
<td>true positive</td>
<td>2</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_1</td>
<td>AMPH_HUMAN</td>
<td>324</td>
<td>328</td>
<td>QENIISSFEDPFPEISVT</td>
<td>true positive</td>
<td>1</td>
<td>1KY7</td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_2</td>
<td>EP15R_HUMAN</td>
<td>599</td>
<td>601</td>
<td>RGSFGMDOPFKNKALLFSN</td>
<td>true positive</td>
<td>1</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
<tr>
<td>LIG_AP2alpha_2</td>
<td>EP15R_HUMAN</td>
<td>618</td>
<td>620</td>
<td>NWTQELHPDFDPDEPDPSD</td>
<td>true positive</td>
<td>1</td>
<td></td>
<td>Homo sapiens (Human)</td>
</tr>
</tbody>
</table>

**Tools & Databases of Short Linear Motifs**
Diseases mediated by short linear motifs

Several diseases are known which are caused by one or more mutations in linear motifs mediating important interactions. Below you find a selection of such diseases, for linear motifs abused by viruses, see the dedicated Viruses page. For a large-scale analysis on disease-causing mutations see [Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? Uyar B, et al., 2014]

Noonan Syndrome

The developmental disorder "Noonan Syndrome" can be caused by mutations in Raf-1 which abrogate the interaction with 14-3-3 proteins mediated by corresponding motifs and thereby deregulate the Raf-1 kinase activity [ Pandit et al., 2007]. The Raf-1 sequence features two LIG_14-3-3_1 binding sites, which are annotated at 256-261 and 618-623.

Noonan-like Syndrome

A S->G mutation at position 2 creates a novel MOD_NMyristoyl site (irreversible modification) resulting in aberrant targeting of SHOC2 to the plasma membrane and impaired translocation to the nucleus upon growth factor stimulation [ Cordeda et al., 2007].

Usher's Syndrome

"Usher's Syndrome" is the most frequent cause of hereditary deaf-blindness in humans [ Eudy and Samegi, 1996], affecting one child in 25 000. This disease can be caused by mutations in either PDZ domains in Harmonin or the corresponding PDZ interaction motifs in the SANS protein (annotated at 456-461) [ Weil et al., 2003, Kaly et al., 2005].

Another example implicating PDZ domains is "familial hypomagnesemia with hypercalciuria and nephrocalcinosis" (FHWN), an autosomal recessive wasting disorder of renal Mg^{2+} and Ca^{2+} that leads to progressive kidney failure. Here, motifs mediating interaction to PDZ domains are mutated in Claudin 16, abolishing important interactions to the scaffolding protein ZO-1 resulting in lysosomal mislocalization of the protein [ Müller et al., 2005, Müller et al., 2006].
Summary of features reported by the ELM resource.

**KEY**

**DOMAINS:**
- Smart/Pfam domain
- Signal peptide (pred.)
- Low-complexity region
- Coiled-coil (pred.)
- TM helix (pred.)

**GLOBPLOT:**
- Glob Dom
- Disorder

**2D STRUCT.:**
- Strand
- Helix
- Loop
- 3/10 Helix

**MOTIFS:**
- Favourable Context
- Sparse/Smart filtered
- Neutral
- Annotated: TP, FP, TN, FN
- Assigned by homology

**CONSCORE:**
- Low Conservation
- Medium Conservation
- High Conservation

(Mouseover the matches for more details)

**Feature**

PhosphoELM
SMART/Pfam domain
GlobPlot
IUPRED
Secondary Structure
CLY_PCSK_FUR_1
LIG_AP2alpha_2
LIG_APCC_KENbox_2
LIG_BIR_II_1
LIG_BRCT_BRCA1_1
LIG_Clathr_ClatBox_1
LIG_EH_1
LIG_JQ
LIG_NRBOX
LIG_PDZ_Class_2
LIG_RGD
LIG_SCF_FBW7_1
LIG_SH2_PTP2
LIG_SH2_STAT3
LIG_SH3_1
LIG_SH3_2
LIG_TRAF6

Sequence: EPN1_HUMAN|Q9Y6i3

**Conservation-Score**

LIG_EH_1
[546,550]
TNPL
(Cons.-Score: 0.988)

**Target Domain Structure:** d1h0aa
- Accessibility Score: 0.007
- Secondary Struct. Score: 0.55
- Combined Total Score: 0.557
- Total Score P-value: 0.7017

Tools & Databases of Short Linear Motifs
VIEW CONSERVATION IN JALVIEW

Tools & Databases of Short Linear Motifs
Questions?

CURIOSITY

Do you really want to know?

fakeposters.com
Short Linear Motifs

- are compact, degenerate protein interaction interfaces (in IDR)
- are ubiquitous in eukaryotic proteomes and mediate many regulatory functions:
  - directing ligand binding
  - providing docking sites for modifying enzymes
  - controlling protein stability
  - acting as signals to target proteins to specific subcellular locations

"The switches.ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces"; VAN ROEY, DINKEL, WEATHERITT, GIBSON AND DAVEY; (SCIENCE SIGNALING. 2013)
**Linear Motifs as Molecular Switches**

**Short Linear Motifs**
- are compact, degenerate protein interaction interfaces (in IDRs)
- are ubiquitous in eukaryotic proteomes and mediate many regulatory functions:
  - directing ligand binding
  - providing docking sites for modifying enzymes
  - controlling protein stability
  - acting as signals to target proteins to specific subcellular locations

**Motif-mediated interactions**
- occur with low affinity,
- are transient & reversible
- can be easily modulated.

"The switches.ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces"; VAN ROEY, DINKEL, WEATHERITT, GIBSON AND DAVEY; (SCIENCE SIGNALING. 2013)
LINEAR MOTIFS AS MOLECULAR SWITCHES

Short Linear Motifs
- are compact, degenerate protein interaction interfaces (in IDR)
- are ubiquitous in eukaryotic proteomes and mediate many regulatory functions:
  - directing ligand binding
  - providing docking sites for modifying enzymes
  - controlling protein stability
  - acting as signals to target proteins to specific subcellular locations

Motif-mediated interactions
- occur with low affinity,
- are transient & reversible
- can be easily modulated.

Motifs mediate switches
This makes SLiMs ideal regulatory modules and enable them to conditionally switch between “on” and “off” states or between multiple, functionally distinct on states.

"The switches.ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces"; VAN ROEY, DINKEL, WEATHERITT, GIBSON AND DAVEY; (SCIENCE SIGNALING. 2013)
**Linear Motifs as Molecular Switches**

PTM-induced binding

LAT → YxxL → PLCγ1 → YxxL → P

Tools & Databases of Short Linear Motifs
LINEAR MOTIFS AS MOLECULAR SWITCHES

PTM-induced incompatibility

NFATc1 → NLS  →  NLS

CYTOPLASM

NUCLEUS
LINEAR MOTIFS AS MOLECULAR SWITCHES

(a) Binary switch
PTM-induced binding

LAT
YxxL

PLCy1

(b) Specificity switch
Intrinsic affinity switch

talin
NPxY
Integrin [I3]

Dok 1

PP1

(c) Motif hiding
PTM-independent

MRTF
RPE1 NLS

PTM-dependent

G-actin

BRAP2

Figure legend

Protein

Protein

Small molecule

Post-translational modification

Motif (Regular expression)

Motif (Name / Abbreviation)
LINEAR MOTIFS AS MOLECULAR SWITCHES

(a) Cumulative switch
Positive rheostat

\[
p53 \xrightarrow{\text{TAD region}} P \xrightarrow{\text{TAD region}} P \xrightarrow{\text{TAD region}} P \xrightarrow{\text{TAD region}} P
\]

Affinity of p53 for CBP/p300

(b) Avidity-sensing switch
PTM-dependent

\[
\text{FcR}\gamma \xrightarrow{\text{ITAM}} P \xrightarrow{\text{ITAM}} P \xrightarrow{\text{ITAM}} P
\]

Cargofree

PTM-independent

\[
\text{Eps15} \xrightarrow{\text{DPF}} \text{Eps15} \xrightarrow{\text{DPF}} \text{Eps15} \xrightarrow{\text{DPF}} \text{Eps15} \xrightarrow{\text{DPF}} \text{Eps15}
\]

Abortive interactions

(c) Sequential switch
Priming PTM

\[
\text{HSF1} \xrightarrow{\text{KxExS}} P \xrightarrow{\text{KxExS}} P \xrightarrow{\text{KxExS}} P
\]

SUMO

Sequential specificity switch

\[
\text{Smad3} \xrightarrow{\text{PY box}} \text{Pin1} \xrightarrow{\text{SPxLSR}} \text{GSK3} \xrightarrow{\text{PY box}} \text{Nedd4L}
\]

Figure legend

- Protein
- Protein
- Small molecule
- Post-translational modification
- Motif (Regular expression)
- Motif (Name / Abbreviation)

Tools & Databases of Short Linear Motifs
The switches.ELM **database** curates experimentally validated motif-based molecular switches. In addition, based on these validated instances, the switches.ELM **prediction** tool was developed to identify possible switching mechanisms that might regulate a motif-containing protein of interest.
The switches.ELM **database** curates experimentally validated motif-based molecular switches. In addition, based on these validated instances, the switches.ELM **prediction** tool was developed to identify possible switching mechanisms that might regulate a motif-containing protein of interest.
ProViz [http://proviz.ucd.ie/] is a tool to visualize biological data allowing the investigation of functional and evolutionary protein features. The tool is designed to be an intuitive and accessible resource to allow users with limited bioinformatic skills to rapidly access and visualise data pertinent to their research.
"ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences."; JEHl P, MANGUY J, SHIELDS DC, HIGGINS DG, DAVEY NE.; (NUCLEIC ACIDS RES. 2016 APR 16)
"ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences."; JEHL P, MANGUY J, SHIELDS DC, HIGGINS DG, DAVEY NE.; (NUCLEIC ACIDS RES. 2016 APR 16)
Questions?

CURIOSITY KILLED THE CAT

Good boy curiosity.......  
Good boy!!!