

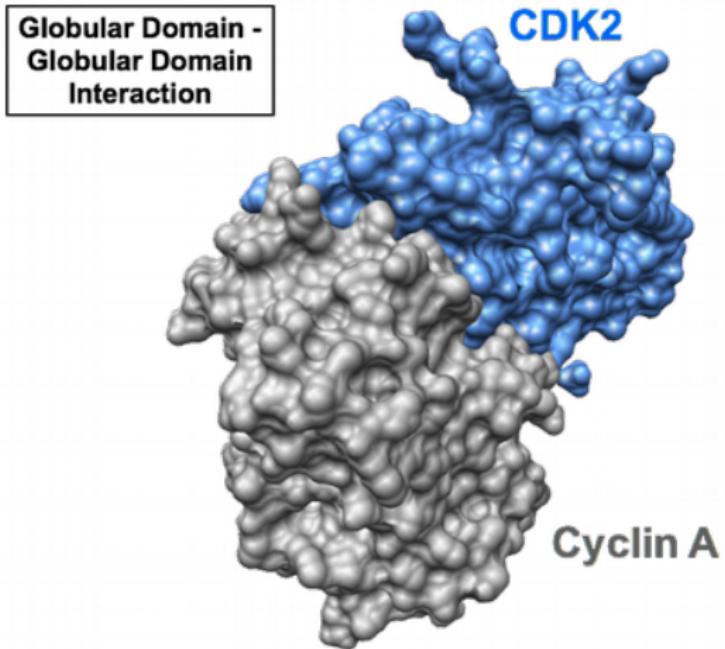


Short Linear Motifs and the Eukaryotic Linear Motif resource

Juliana Glavina & Hugo Samano

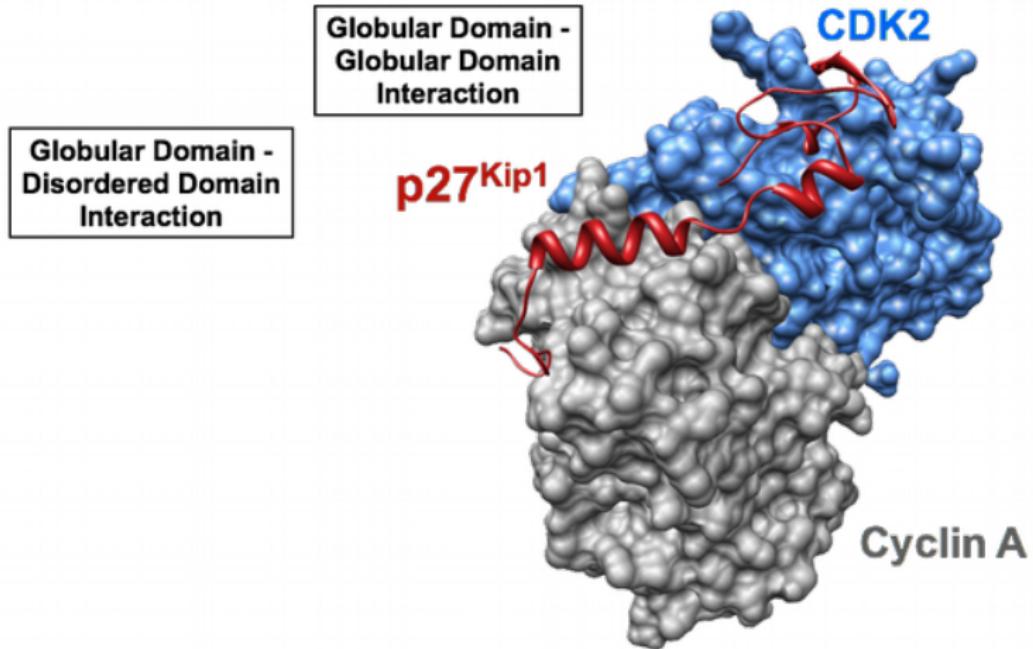
based on Holger Dinkel & Aidan Budd

WHAT ARE SLIMs?



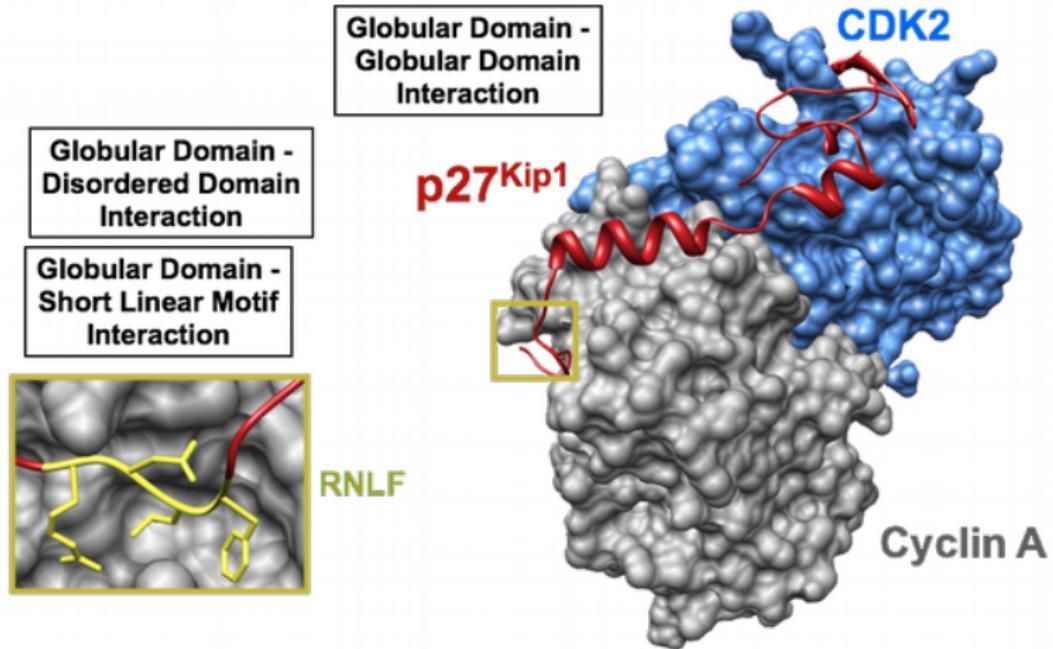
Russo et al., 1996: "Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex."

WHAT ARE SLIMs?



Russo et al., 1996: "Crystal structure of the p27^{Kip1} cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex."

WHAT ARE SLiMs?



Russo et al., 1996: "Crystal structure of the p27^{Kip1} cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex."

WHAT ARE SLiMs?

Globular Domain - Disordered Domain Interaction

PDB 1JSU
Russo *et al.*, Nature. 1996;
382: 325-331.

RNLF

Globular Domain - Globular Domain Interaction

p27^{Kip1}

CDK2

Regular Expression:

DOC_CYCLIN_1 [RK]xLx{0,1}[FYLVMP]

Defined positions

Fixed positions

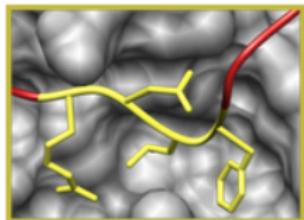
Degenerate positions

Undefined positions

Fixed-length wildcard

Flexible-length wildcard {min,max}

Cyclin A



Russo et al., 1996: "Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex."

WHAT ARE SLIMs?

Linear Motifs can be described by “Regular Expressions”

[KR]: different amino acids allowed at this position

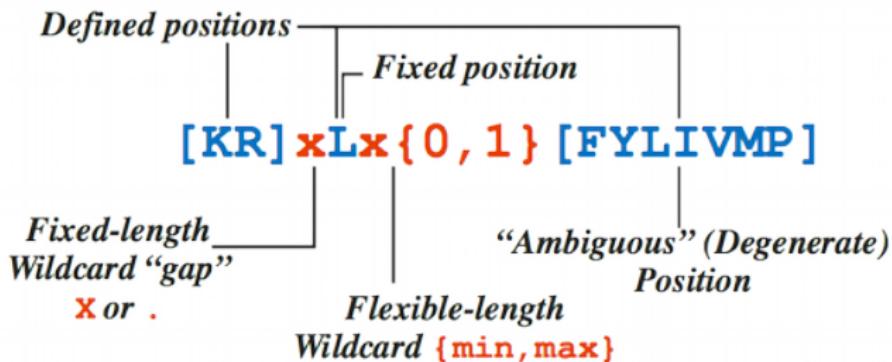
x or . : Wildcard

L: single amino acid “L”=Leucine

x{0,1} : variable length

[FYLIVMP]: different hydrophobic amino acids allowed

Example: DOC_CYCLIN_1:



WHAT ARE SLIMs?

Now that We are Experts in “Regular Expressions”

Example: CLV_PCSK_KEX2_1

RegEx: [KR] R .

How many possible combinations are there?

WHAT ARE SLIMs?

More Information in “Regular Expressions”

Character	Meaning
.	Any amino acid allowed
[xy]	Amino acids listed are allowed
[^xy]	Amino acids listed are not allowed
{ min, max }	Min required, max allowed
^	Matches the amino terminal
\$	Matches the carboxy terminal
ab cd	Matches either expression it separates
(xy)	Used to mark positions of specific interest (amino acid being covalently modified) or to group parts of the expression

WHAT ARE SLIMs?

Are We Really Experts ???

Example: DOC_MAPK_HePTP_8

RegEx:

```
( [LIV] [^P] [^P] [RK] . . . . [LIVMP] . [LIV] .  
[LIVMF] ) | ([LIV] [^P] [^P] [RK] [RK] G. {4,7}   
[LIVMP] . [LIV] . [LIVMF] )
```

TYPES

CLV Proteolytic **Cleavage** Sites

DEG Destruction Sites (**Degrons**)

DOC **Docking** Sites

LIG **Ligand** Binding Sites

MOD Post-Translational **Modification** Sites

TRG Subcellular **targeting** sites

MODIFICATION SITES

Description:

Modification Motifs mediate **specific binding** to the **active site** of a modifying **enzyme** to allow subsequent catalytic **post-translational** modification of the target site.

Example:

Name MOD_CDK_SPxK_1

RegEx ... ([ST])P.[KR]

Kinase domain

CDK site

MODIFICATION SITES

Description:

Modification Motifs mediate **specific binding** to the **active site** of a modifying **enzyme** to allow subsequent catalytic **post-translational** modification of the target site.

Example:

Name MOD_CDK_SPxK_1

RegEx ... ([ST])P.[KR]



MODIFICATION SITES

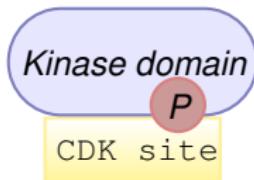
Description:

Modification Motifs mediate **specific binding** to the **active site** of a modifying **enzyme** to allow subsequent catalytic **post-translational** modification of the target site.

Example:

Name MOD_CDK_SPxK_1

RegEx ... ([ST])P.[KR]



MODIFICATION SITES

Description:

Modification Motifs mediate **specific binding** to the **active site** of a modifying **enzyme** to allow subsequent catalytic **post-translational** modification of the target site.

Example:

Name MOD_CDK_SPxK_1

RegEx ... ([ST])P.[KR]

Kinase domain



DEGRADATION MOTIFS

Description:

Degradation motifs (**Degrons**) recognized by E3 Ubiquitin Ligase complexes priming proteins for **degradation**, regulating protein half-life.

Example:

Name DEG_SCF_TRCP1_1

RegEx D (S) G . . ([ST])



DEGRADATION MOTIFS

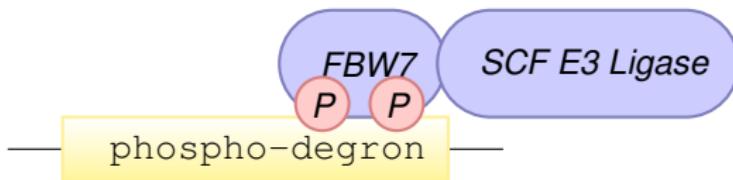
Description:

Degradation motifs (**Degrons**) recognized by E3 Ubiquitin Ligase complexes priming proteins for **degradation**, regulating protein half-life.

Example:

Name DEG_SCF_TRCP1_1

RegEx D (S) G . . ([ST])



DEGRADATION MOTIFS

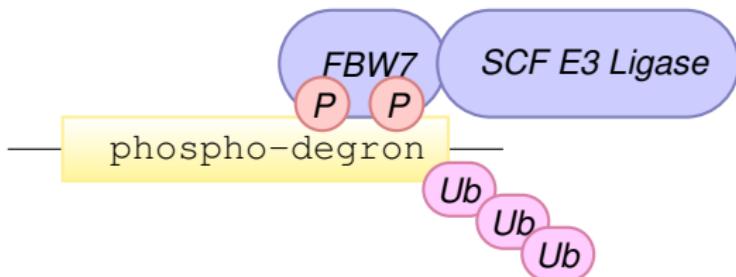
Description:

Degradation motifs (**Degrons**) recognized by E3 Ubiquitin Ligase complexes priming proteins for **degradation**, regulating protein half-life.

Example:

Name DEG_SCF_TRCP1_1

RegEx D (S) G . . ([ST])



DEGRADATION MOTIFS

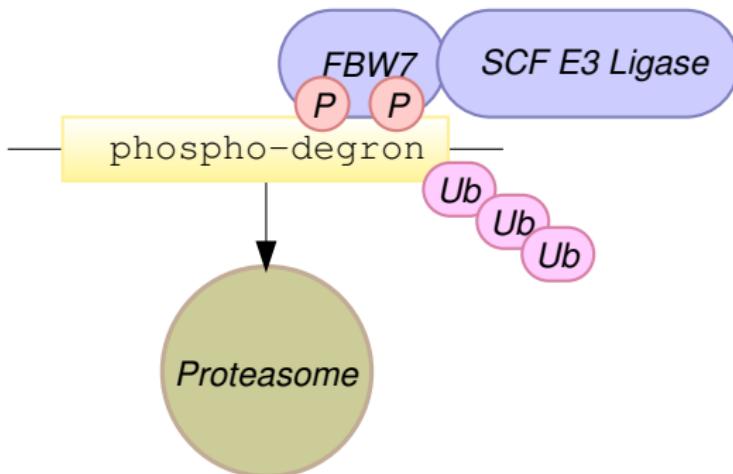
Description:

Degradation motifs (**Degrons**) recognized by E3 Ubiquitin Ligase complexes priming proteins for **degradation**, regulating protein half-life.

Example:

Name DEG_SCF_TRCP1_1

RegEx D (S) G . . ([ST])



TARGETING/ANCHORING MOTIFS

Description:

Targeting motifs allow a protein to bind to the **transport machinery** that **relocalizes** it to a particular sub-cellular location.

Anchoring motifs are recognized by biomolecules specific to a **sub-cellular** location and thereby **retain** the **motif**-containing protein at that location.

Example:

Name TRG_NLS_MonoCore_2

RegEx

[^DE] (K[RK] | RK) [KRP] [KR]
[^DE]

Importin α

NLS

TARGETING/ANCHORING MOTIFS

Description:

Targeting motifs allow a protein to bind to the **transport machinery** that **relocalizes** it to a particular sub-cellular location.

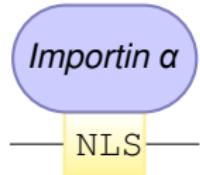
Anchoring motifs are recognized by biomolecules specific to a **sub-cellular** location and thereby **retain** the **motif**-containing protein at that location.

Example:

Name TRG_NLS_MonoCore_2

RegEx

[^DE] (K[RK] | RK) [KRP] [KR]
[^DE]



TARGETING/ANCHORING MOTIFS

Description:

Targeting motifs allow a protein to bind to the **transport machinery** that **relocalizes** it to a particular sub-cellular location.

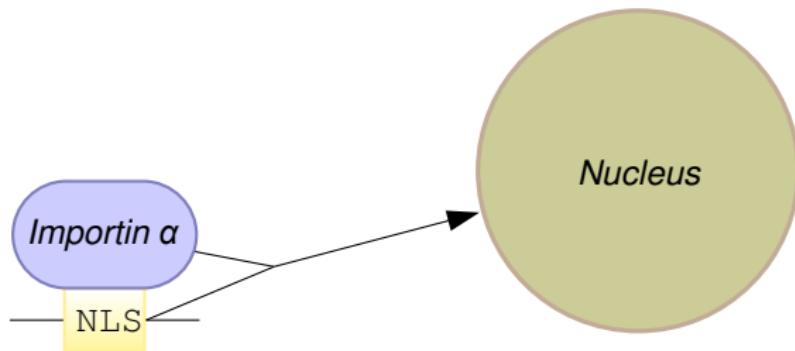
Anchoring motifs are recognized by biomolecules specific to a **sub-cellular** location and thereby **retain** the **motif**-containing protein at that location.

Example:

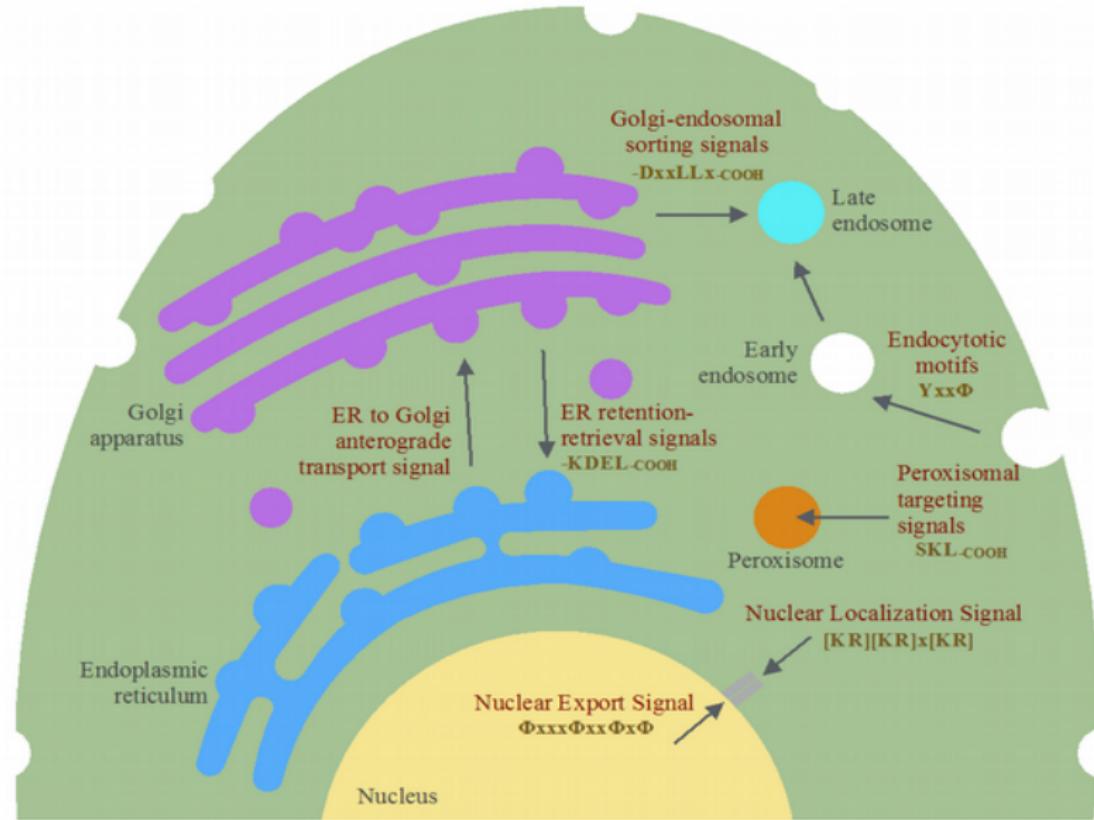
Name TRG_NLS_MonoCore_2

RegEx

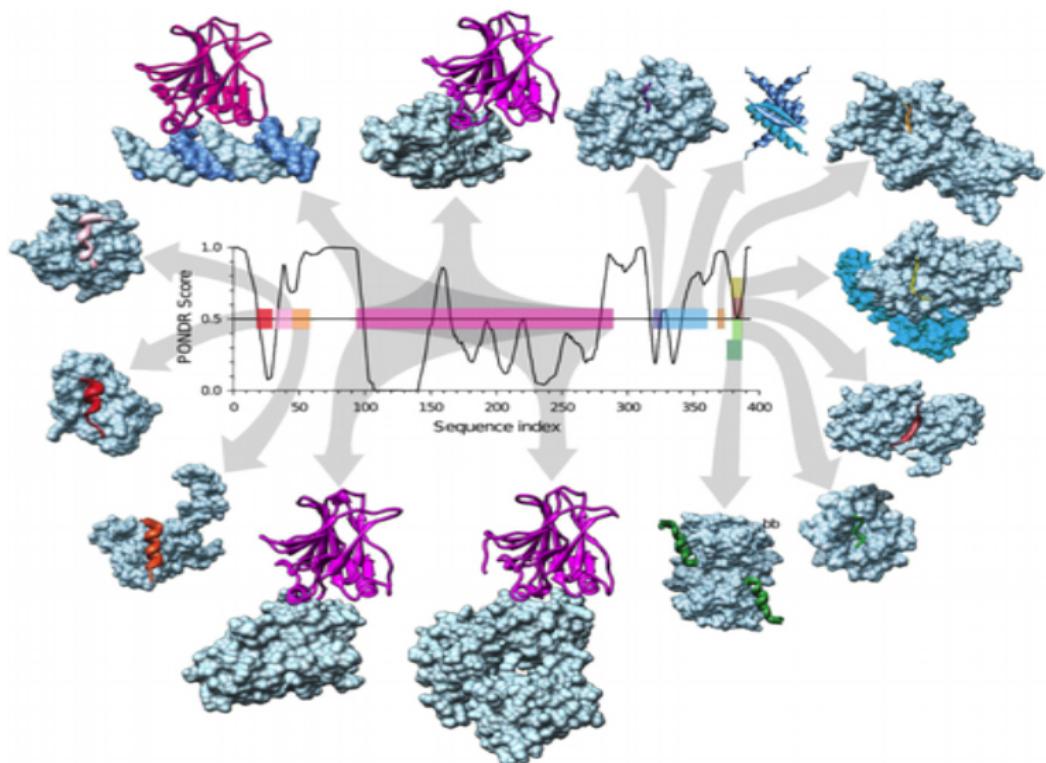
[^DE] (K[RK] | RK) [KRP] [KR]
[^DE]



IMPORTANCE OF MOTIFS



IMPORTANCE OF MOTIFS



The many interaction partners of p53
Uversky, Dunker, 2010: "Understanding protein non-folding"

IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

are **short protein modules** (\approx 3–10 AA length)

mediate **transient interactions**

often reside in **disordered** or low-complexity sequence regions

are involved in **protein-protein interactions**

IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

are **short protein modules** (\approx 3–10 AA length)

mediate **transient interactions**

often reside in **disordered** or low-complexity sequence regions

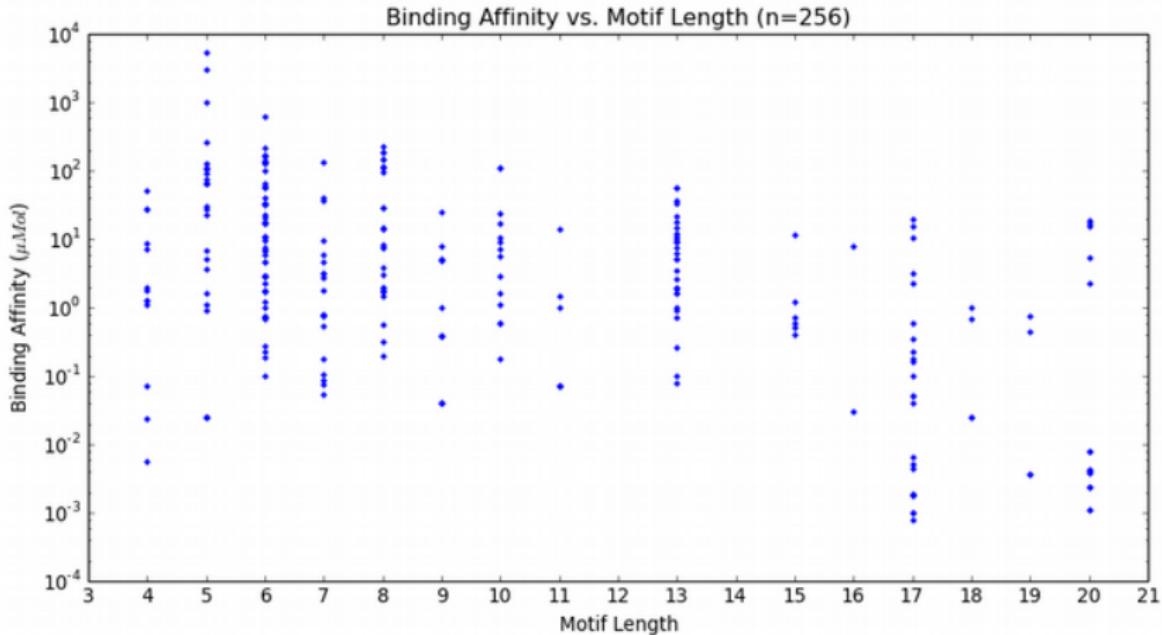
are involved in **protein-protein interactions**

IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

are short protein modules (\approx 3–10 AA length)

mediate transient interactions



IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

are **short protein modules** (\approx 3–10 AA length)

mediate **transient interactions**

often reside in **disordered** or low-complexity sequence regions

are involved in **protein-protein interactions**

IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

- are **short protein modules** (\approx 3–10 AA length)
- mediate **transient interactions**
- often reside in **disordered** or low-complexity sequence regions
- are involved in **protein-protein interactions**

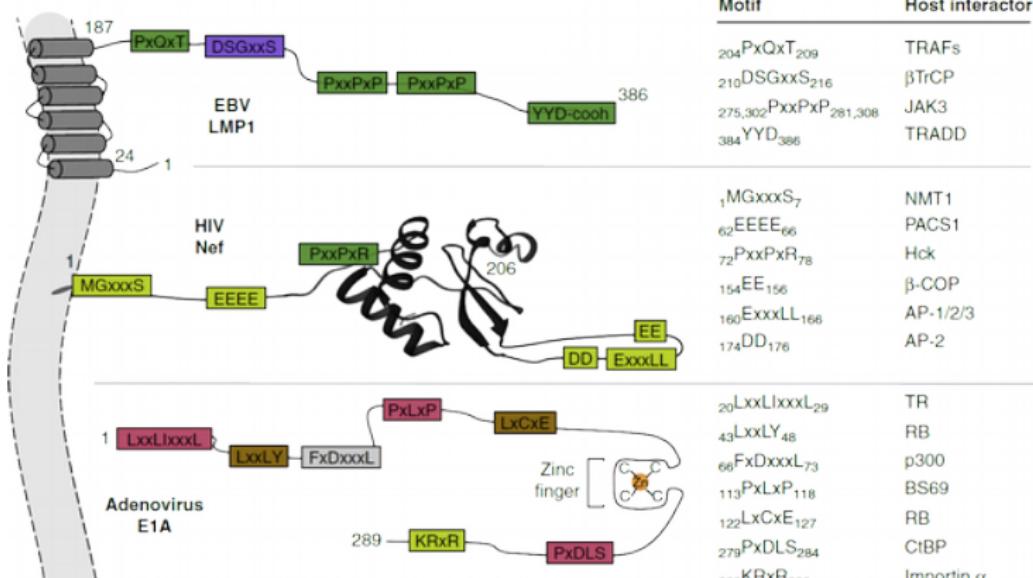
IMPORTANCE OF MOTIFS

SLiMs = “Short Linear Motif”:

- are **short protein modules** (\approx 3–10 AA length)
- mediate **transient interactions**
- often reside in **disordered** or low-complexity sequence regions
- are involved in **protein-protein interactions**

SLiMs are intrinsically disordered interaction modules that function independently of their tertiary structure

IMPORTANCE OF MOTIFS



TIBS-817; No. of Pages 11

ARTICLE IN PRESS

Review

Cell
PRESS

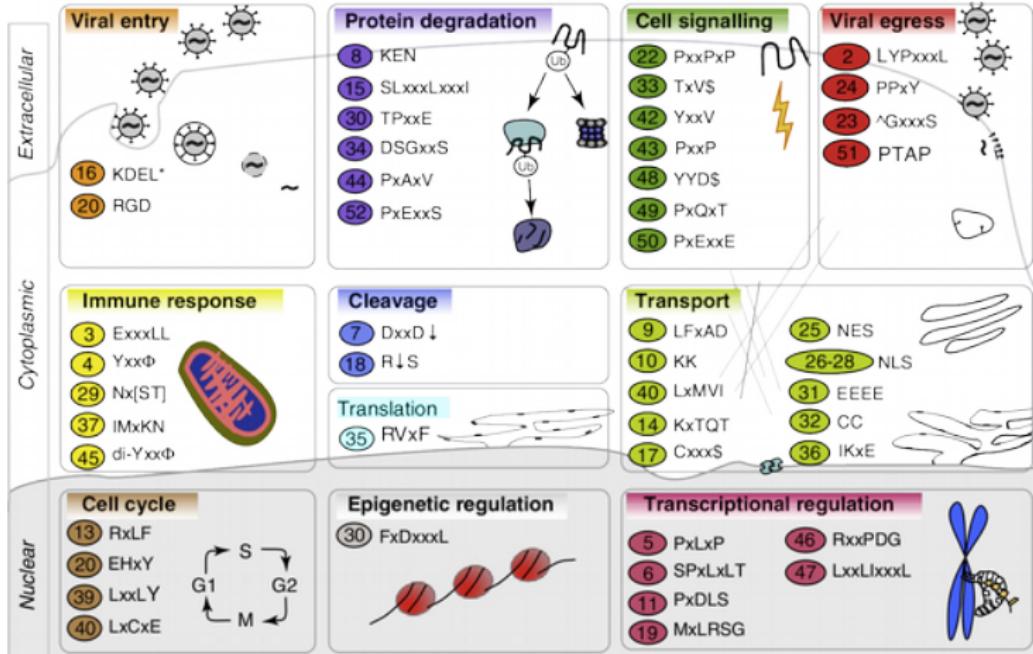
How viruses hijack cell regulation

Norman E. Davey¹, Gilles Travé² and Toby J. Gibson¹

¹Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

²Equipe Oncoproteines, FRE CNRS 3211, ESBS, 1, Bld Sébastien Brandt, BP10413, 67412 Illkirch, France

IMPORTANCE OF MOTIFS



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

The  resource

is a **collection** of over 250 thoroughly **annotated** motif classes with over 3000 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.

TALKING ABOUT SLIMs

TYPES	CLASSES	INSTANCES
CLV Proteolytic Cleavage Sites	CLV_C14_Caspase3-7	Q7KZF4 (812-817)
DEG Destruction Sites (Degrons)	DEG_SCF_TRCP1_1	O97143 (292-297)
DOC Docking Sites	DOC_CYCLIN_1	P04637 (381-385)
LIG Ligand Binding Sites	LIG_RB_LxCxE	Q13547 (411-428)
MOD Post-Translational Modification Sites	MOD_CDK_SPxK_1	Q99741 (71-77)
TRG Subcellular targeting sites	TRG_NLS_MonoCore_2	Q969H0 (10-15)

Nucleic Acids Research, 2003, Vol. 31, No. 13 3625–3630
DOI: 10.1093/nar/gkg545

ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins

Published online 17 November 2009

Nucleic Acids Research, 2010, Vol. 38, Database issue D167–D180
doi:10.1093/nar/gkp1016

ELM: the status of the 2010 eukaryotic linear motif resource

Published online 7 November 2013

Nucleic Acids Research, 2014, Vol. 42, Database issue D259–D266
doi:10.1093/nar/gkt1047

The eukaryotic linear motif resource ELM: 10 years and counting

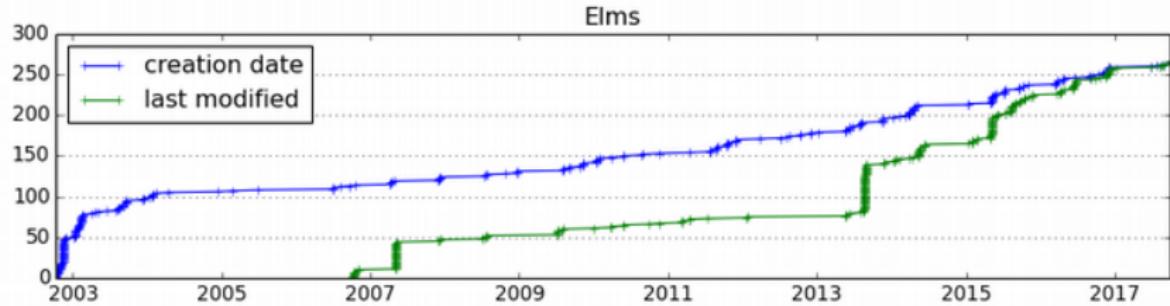
D294–D300 *Nucleic Acids Research*, 2016, Vol. 44, Database issue
doi: 10.1093/nar/gkv1291

Published online 28 November 2015

ELM 2016—data update and new functionality of the eukaryotic linear motif resource

Nucleic Acids Research, 2017 1
doi: 10.1093/nar/gkx1077

The eukaryotic linear motif resource – 2018 update



Please, go to the ELM site

elm.eu.org

THE ELM DATABASE

ELM



The Eukaryotic Linear Motif resource for
Functional Sites in Proteins

search ELM Database

ELM Home

ELM Prediction

ELM DB

ELM Candidates

ELM Information

ELM downloads

Help

Search ELM classes

submit

Reset

export 267 classes as: [tsv](#)

elm class	Description	RegEx	Instances	Instances in PDB
CLV_C14_Caspase3-7	Caspase-3 and Caspase-7 cleavage site.	[DSTE] [^P] [^DEHDFYC] D[GSAN]	39	0
CLV_MEL_PAP_1	Prophenoloxidase-activating proteinase (PAP) cleavage site ([ILV]-X-X-R- -[FV]-[GS]-X).	[ILV]..R [FV][GS]..	12	0
CLV_NRD_NRD_1	N-Arg dibasic convertase (NRD/Nardilysin) cleavage site (X- R-K or R- R-X).	(.RK . RK)[RK]	2	0
CLV_PCSK_FUR_1	Furin (PACE) cleavage site (R-X-[RK]R- X).	R.[RK]R..	13	0
CLV_PCSK_KEX2_1	Yeast kexin 2 cleavage site (K-R- X or R-R- X).	[KR]..	1	0
CLV_PCSK_PC1ET2_1	NEC1/NEC2 cleavage site (K-R- X).	KR..	6	0
CLV_PCSK_PCT_1	Proprotein convertase 7 (PC7, PCSK7) cleavage site (R-X-X-X-[RK]R- X).	R...[RK]R..	1	0
CLV_PCSK_SKI1_1	Subtilisin/kexin isozyme-1 (SKI1) cleavage site ([RK]X-[hydrophobic]-[LTKF]- X).	[RK]..[ATLHFV]..[LTKF]..	2	0
CLV_Separin_Fungi	Separase cleavage site, best known in sister chromatid separation. Also involved in stabilizing the anaphase spindle and centriole disengagement.	S[IVLNH]E[IVPPNLYQAQ]GR..	4	0
CLV_Separin_Metazoa	Separase cleavage site, best known in sister chromatid separation.	E[IVPVL][MLVFR..	5	0
CLV_TASPASE1	Taspase1 is a threonine aspartase which was first identified as the protease responsible for processing the trithorax (MLL) type of histone methyltransferases.	Q[MLV]D..[DE]	2	0
DEG_APCC_DBOX_1	An RxL-based motif that binds to the Cdh1 and Cdc20 components of APC/C thereby targeting the protein for destruction in a cell cycle dependent manner.	.R..L..[LIVH]..	11	0
DEG_APCC_KENBOX_2	Motif conserving the exact sequence KEN that binds to the APC/C subunit Cdh1 causing the protein to be targeted for 26S proteasome mediated degradation.	.KEN..	16	1
DEG_APCC_TPR_1	This short C-terminal motif is present in co-activators, the Doct1/APC10 subunit and some substrates of the APC/C and mediates direct binding to TPR-containing APC/C core subunits.	.+[ILR]RS	22	0
DEG_COP1_1	A destruction motif interacts with the COP1 WD 40 domain for target ubiquitination and degradation.	[STKE](1,3), (0,2)[TIDE], (2,3)VP[STKE](0,1)*PLIMVYPA	12	1
DEG_CRL4_CDT2_1	This degron overlaps a PCNA interaction protein (PIP) box and is recognised by the CRL4 ^{CDT2} ubiquitin ligase in a PCNA- and chromatin-dependent manner.	[HQ]{0,1}..[LAMV][ST][DGN][FY][Y], (2,3)[KR]{2,3}[DE]	6	0
DEG_CRL4_CDT2_2	This degron, occurring in non-Vertebrates, overlaps a PCNA interaction protein (PIP) box and is recognised by the CRL4 ^{CDT2} ubiquitin ligase in a PCNA- and chromatin-dependent manner.	[HQ]{0,1}..[LAMV][ST][DGN][MDRY][FY], (2,3)[KR]{2,3}[DE]	1	0
DEG_Kelch_actininfilin_1	A hydrophobic degron motif present in some kainate receptors necessary to interact with kelch domain of actininfilin protein for efficient ubiquitination and degradation.	[AP]P[KV][IM]V	1	0
DEG_Kelch_Keap1_1	Motif that binds to the Kelch domain of KEAP1 with high affinity. This high affinity motif is required for the efficient recruitment of target proteins to the Cul3-based E3 ligase.	[DNE]..[DES][TNS]GE	13	4
DEG_Kelch_Keap1_2	Motif that binds to the Kelch domain of KEAP1 with low affinity. This low affinity motif is important for ubiquitination and degradation of target proteins.	QD..DLOV	1	1
DEG_Kelch_KLHL3_1	An Acidic degron motif present in wrk kinases necessary to interact with kelch domain of KLHL2 and KLHL3 proteins for efficient ubiquitination degradation.	E..EE..E[A]DQE	4	0
DEG_MDM2_SWIB_1	An amphipathic α -helix found in p53 family members that binds in the hydrophobic cleft of MDM2's SWIB domain.	F[^P]{3}W[^P]{2,3}[VIL]	5	2



search ELM Database

ELM Home ELM Prediction ELM DB ELM Candidates ELM Information ELM downloads

Help

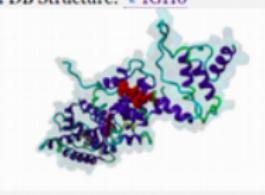
«LIG_PTB_Phospho_1»

»LIG_Rb_pABgroove_1»

LIG_Rb_LxCxE_1

Accession: **ELME000007****Functional site class:** Rb pocket B binding ligand**Functional site description:** The LxCxE motif is found in multiple host and viral interactors of the retinoblastoma protein family (Rb, p107 and p130).**ELM Description:** The LxCxE motif mediates binding to a highly conserved shallow groove in the B domain of Rb. The central Cysteine is highly conserved in all instances, however the Leucine and Glutamic Acid positions tolerate substitution of physicochemically similar residues allowing a less stringent definition of [LI]xCx[DE]. The staggered arrangement, evenly spaced and one residue apart, of the conserved residues cover one side of an extended, beta-strand-like conformation and bind the groove orthogonally, not by beta augmentation like many similar staggered motifs. The Leucine and Cysteine positions bind a hydrophobic region of the groove with tight complementarity. The Glutamic Acid forms hydrogen bonds with two backbone amide groups of an alpha helix forming one side of the binding groove. The interaction is further stabilized by additional hydrogen bonds to the peptide backbone adding rigidity. Phosphorylation of Rb at Thr821 and Thr826 inhibits LxCxE binding.**Pattern:** ([DEST] | ^) . {0,4} [LI] . C.E. {1,4} [FLMIVAWPHY] . {0,8} ([DEST] | \$)**Pattern Probability:** 0.0005417**Present in taxons:** Metazoa Viridiplantae**Interaction Domain:** RB_B (PF01857) Retinoblastoma-associated protein B domain
(Stoichiometry: 1 : 1)

PDB Structure: 1GH6



THE ELM DATABASE



■ 45 Instances for LIG_Rb_LxCxE_1

(click table headers for sorting; Notes column: =Number of Switches, =Number of Interactions)

Acc., Gene-, Name	Start	End	Subsequence	Logic	#Ev.	Organism	Notes
Q13547 HDAC1 HDAC1_HUMAN	411	428	SDKRTACEEEFS <u>DSEEEGEG</u>	TP	15	Homo sapiens (Human)	1
B8ZX42 LT-Ag B8ZX42_9POLY	211	229	DLFCDE <u>SLSPEPPSSSEEP</u>	TP	9	Merkel cell polyomavirus	
P03070 Large T antigen LT_SV40	101	119	ENILFCSEEMPSS <u>DDEATAB5</u>	TP	7	Simian virus 40	1GH6
Q5T8A7 PPP1R26 PRDM2_HUMAN	478	493	DTSAEL <u>MCAEAAILDISKTIL</u>	TP	5	Homo sapiens (Human)	
Q13029 PRDM2 PRDM2_HUMAN	309	328	ETR <u>CDEKPEDLLEEPKTTSE</u>	TP	9	Homo sapiens (Human)	1
P42753 CYCD3-1 CCD31_ARATH	20	38	DALY <u>CEEKKWDDGEEVEEN</u>	TP	4	Arabidopsis thaliana (Thale cress)	1
P42752 CYCD2-1 CCD21_ARATH	3	22	EN <u>LACGETSESWIIDNDODDO</u>	TP	4	Arabidopsis thaliana (Thale cress)	1
P42751 CYCD1-1 CCD11_ARATH	21	40	DLFC <u>GEDSGVFSGESTVDF5</u>	TP	3	Arabidopsis thaliana (Thale cress)	1
P30281 CCND3 CCND3_HUMAN	2	18	MELL <u>CCGTTRHAPRAGPDPR</u>	TP	3	Homo sapiens (Human)	1
P30279 CCND2 CCND2_HUMAN	2	17	MELL <u>CHEDVPVRRAVRDRLN</u>	TP	4	Homo sapiens (Human)	1
P17480 UBF1 UBF1_HUMAN	306	323	YSLYCAEL <u>MANMKDVPSTER</u>	FP	3	Homo sapiens (Human)	1
O14777 NDC80 NDC80_HUMAN	229	245	Y <u>TIKCYESPHMSGADSFDEN</u>	TP	5	Homo sapiens (Human)	2VE7
Q92769 HDAC2 HDAC2_HUMAN	412	429	SDKRT <u>IACDEEFSDSEDEGE</u> G	TP	5	Homo sapiens (Human)	1
Q62661 Hbp1 HBP1_RAT	31	51	S <u>LELLOCHENVPSSPQYN5D</u>	TP	9	Rattus norvegicus (Norway rat)	1
P51532 SMARCA4 SMC4A_HUMAN	1355	1371	VER <u>LTCEEEEEEKMFGRGSRH</u>	TP	4	Homo sapiens (Human)	1
Q63928 brg1 Q63928_9MURI	730	746	VER <u>LTCEEEEEEKMFGRGSRH</u>	TP	5	Mus sp.	1
Q77J94 Wsv056 Q77J94_WSSVS	214	226	GHGMGH <u>DLSCOEISEFLVQ</u>	TP	3	Shrimp white spot syndrome virus (isolate Tongan)	1
Q77JB9 Wsv069	215	224	GHGMGH <u>DLSCOEISEFLVQ</u>	TP	2		1

LIG Rb LxCxE_1

■ Instance

Accession	Acc. Gene-, Name	Start	End	Subsequence	Logic	PDB	Organism	Length
ELMI003141	Q13547 HDAC1 HDAC1_HUMAN	411	428	SDKRIACEEEFSOSEEAGE	TP	---	Homo sapiens (Human)	482

■ Instance evidence

Evidence class	PSI-MI	Method	BioSource	PubMed	Logic	Reliability	Notes
experimental	MI:0096	pull down	in vivo/in vitro		support	certain	InteractionDetection
experimental	MI:0519	glutathione s tranferase tag	in vitro	Vortmeyer,1999	support	certain	
experimental	MI:0065	isothermal titration calorimetry	in vitro	Singh,2005	support	certain	InteractionDetection
experimental	MI:0405	competition binding	in vitro	Vortmeyer,1999	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Vortmeyer,1999	support	certain	InteractionDetection
experimental	MI:0074	mutation analysis	in vivo/in vitro	Ferreira,1998	support	certain	FeatureDetection
experimental	MI:0405	competition binding	in vitro	Ferreira,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Ferreira,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Luo,1998	support	certain	InteractionDetection
experimental	MI:0074	mutation analysis	in vivo/in vitro	Magnaghi-Jaulin,1998	support	certain	FeatureDetection
experimental	MI:0405	competition binding	in vitro	Magnaghi-Jaulin,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Magnaghi-Jaulin,1998	support	certain	InteractionDetection
experimental	MI:0405	competition binding	in vitro	Brehm,1998	support	certain	InteractionDetection
experimental	MI:0018	two hybrid	in vivo	Brehm,1998	contradict	likely	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Brehm,1998	support	certain	InteractionDetection



The Eukaryotic Linear Motif resource for Functional Sites in Proteins

search ELM Database

ELM Home ELM Prediction

ELM DB

ELM Candidates

ELM Information

ELM downloads

Help

Search ELM Instance

Full-Text Search (use "*" to get all)

Filter by instance Logic

Filter by organism

submit

Reset

ELM classes

ELM instances

ELM pdb structures

ELM binding domains

ELM switches

ELM pathways

ELM references

ELM related diseases

ELM viral instances

ELM pathogenic abuse

latest 100 Instances

(click table headers for sorting; Nc:

export 100 instances as:

gff pir fasta tsv

Switches, =Number of Interactions)

CLV	DEG	DOC	LIG	MOD	TRG	ELM identifier	ELM experiments	Start	End	Subsequence	Logic	#Ev.	Organism	Notes
						LIG_Actin_WH2_2	ELM GO Terms B4Y9V9_BURPE	57	74	AFNAVIDQTIKKGDFKLKPVG	TP	3	Burkholderia pseu...	1
						LIG_Actin_WH2_2	C5IZN1_VopF C5IZN1_VIBCL	210	226	HSKLMKEELLNGAKLKKVST	TP	2	Vibrio cholerae	1
						LIG_Actin_WH2_2	Q87GE5_VPA1370 Q87GE5_VIBPA	166	182	HSKLMEEELLTGGRKLKKVAT	TP	4	Vibrio parahaemol...	1
						LIG_Actin_WH2_2	Q87GE5_VPA1370 Q87GE5_VIBPA	135	151	DHSKLMEQTRQGVKLKSATK	TP	5	Vibrio parahaemol...	3M1F 2
							C5IZN1_VopF	249	267	RSALLSETAGFSKDRLRKAG	TP	2	Vibrio cholerae	1

elm.eu.org/searchdb.html



[ELM Home](#) [ELM Prediction](#) [ELM DB](#) [ELM Candidates](#) [ELM Information](#) [ELM downloads](#)
[Help](#)

Search ELM Instances

Full-Text Search (use *** to get all instances)

Filter by instance Logic

Filter by organism

export 100 instances as: [xml](#) [pdf](#) [fasta](#) [tex](#)**CLV****DEG****DOC****LIG****MOD****TRG****latest 100 Instances**

(click table headers for sorting; Notes column: =Number of Switches, =Number of Interactions)

ELM Identifier	Acc., Gene-, Name	Start	End	Subsequence	Logic	#Ev.	Organism	Notes
LIG_Actin_WH2_2	B4Y9V9 bimA B4Y9V9_BURPE	57	74	A <u>FAAVIDG</u> IKK <u>GFLKL</u> KPG	TP	3	Burkholderia pseu...	1
LIG_Actin_WH2_2	C5IZN1 VopF C5IZN1_VIBCL	210	226	<u>RKLMKEELLINHGAKLKKVST</u>	TP	2	Vibrio cholerae	1
LIG_Actin_WH2_2	Q87GE5 VPA1370 Q87GE5_VIBPA	166	182	<u>RKLMKEELLTGGRKLLKKVAT</u>	TP	4	Vibrio parahaemol...	1
LIG_Actin_WH2_2	Q87GE5 VPA1370 Q87GE5_VIBPA	135	151	<u>RKLMKEELLTGGRKLLKKVATK</u>	TP	5	Vibrio parahaemol...	3M1F 2
LIG_Actin_WH2_1	C5IZN1 VopF C5IZN1_VIBCL	249	267	<u>R<u>S</u>ALLSETAGFSK<u>DRLRKAG</u></u>	TP	2	Vibrio cholerae	1
LIG_Actin_WH2_1	C5IZN1 VopF C5IZN1_VIBCL	179	195	<u>R<u>S</u>SKLMEEETR<u>QKVLRATPK</u></u>	TP	2	Vibrio cholerae	1
LIG_Actin_WH2_1	Q87GE5 VPA1370 Q87GE5_VIBPA	205	223	<u>R<u>N</u>ALLSETAGFSK<u>DRLRK</u>TG</u>	TP	4	Vibrio parahaemol...	1
DOC_MAPK_MEF2A_6	P0A2M9 spvC SPVC_SALTY	5	12	M <u>PIN<u>SPHILNNT</u>PPLNIVAA</u>	TP	4	Salmonella enteri...	2
DOC_MAPK_MEF2A_6	QBVSP9 ospF OSPF_SHIFL	5	12	M <u>PIN<u>KPLKLNLDSLNVR</u>S</u>	TP	1	Shigella flexneri	1
DOC_MAPK_MEF2A_6	P36507 MAP2K2 MP2K2_HUMAN	6	14	M <u>LARRK<u>KPLPALT</u>INPTIAE</u>	TP	2	Homo sapiens (Human)	4H3Q 1

Functional site prediction

Protein sequence

Enter Uniprot identifier or accession number: (auto-completion)
e.g. [EPN1_HUMAN, P04637, TAU_HUMAN, \[RANDOM\]](#)

EPN1_HUMAN

EPN1_HUMAN Q9Y6I3 (Homo sapiens):

```
>EPN1_HUMAN
MSTSSLLRQMKNVHNYYSEAIKVREATSDPWPQPSQLMSEIAIDLTYNNVVAFSEIMSMWIKRLNDGKNNMRHVYKANTILMEY
LIKTSKSERVSQCCKENNYAVQTLKDFQVYDGDKGDKDGVNVREKAKQLVALLRDEDRLEERAHALKTKELAQTAQATASSAAVG
SGPPPEAEQANPQSSGEELLQLALAMSKEEADQPPSCGPEDDAQQLQALSLISREEHDEERIRRGDDRLRQMAIEESKRET
GGKEESSLMDLADVFATAPAPAPTDPWGAPAPMAAAAVPTAAPTSDPNWGPFVPPAADPWGGPAFTPASGDPWRAAPAGPSVD
PNGGTAPAAAGEGGTPDPNGSSDGGVPVSGPSASDPMTPAPAFSDPWNWGSPAKPSTNGTTAAGGFDTPEDEFSDFDRLRLTALP
TGSAGAELELLAGEVPARSPGAFDMSGVRGSLAEAVGSPPPFAATFTTPPTRKTFESFLGPNAAVLDSLVSRPGPTPPGA
KASNPFLPGGGPATGPSVTNPFPQAPPATLTNQLRLSPVPVGAPPTYISPLGGGPGLPMMMPGPPAPNTNPFLL
```

- Cell compartment (one or several):

not specified
extracellular
nucleus
cytosol
peroxisome
glycosome
glycosome
golgi apparatus
endoplasmic reticulum
lysosome
endosome
plasma membrane
mitochondrion

- Taxonomic Context

Type in species name (auto-completion):
Homo sapiens

- Motif Probability Cutoff:

100

Submit

Reset Form

- ELM database update
A new ELM class has been annotated: [■ LIG_14-3-3_ChREBP_3](#)
- ELM database update
A new ELM class has been annotated: [■ LIG_IFRF3_LxS_1](#)
- ELM database update
New instances added for [■ LIG_Vh1_VBS_1](#)
- ELM database update
New instances have been added for [■ DEG_SCF_TRCP1_1](#) and [■ LIG_LRP6_Inhibitor_1](#)
- ELM database update
We have added new instances for: [■ LIG_APCC_ABBA_1](#), [■ LIG_APCC_ABBAyCdc20_2](#) as well as [■ DOC_MAPK_HePTP_8](#), [■ DOC_MAPK_MEF2A_6](#) and [■ DOC_MAPK_DCC_7](#)
- ELM Database Update
We have updated several MOD_CDK motifs and added new instances:
MOD_CDK_1 is now: [■ MOD_CDK_SPxK_1](#), and [■ MOD_CDK_SPK_2](#) [■ MOD_CDK_SPxxK_3](#) have been added.
- ELM database update
Several new ELM classes and instances have been added:

THE ELM PREDICTION TOOL

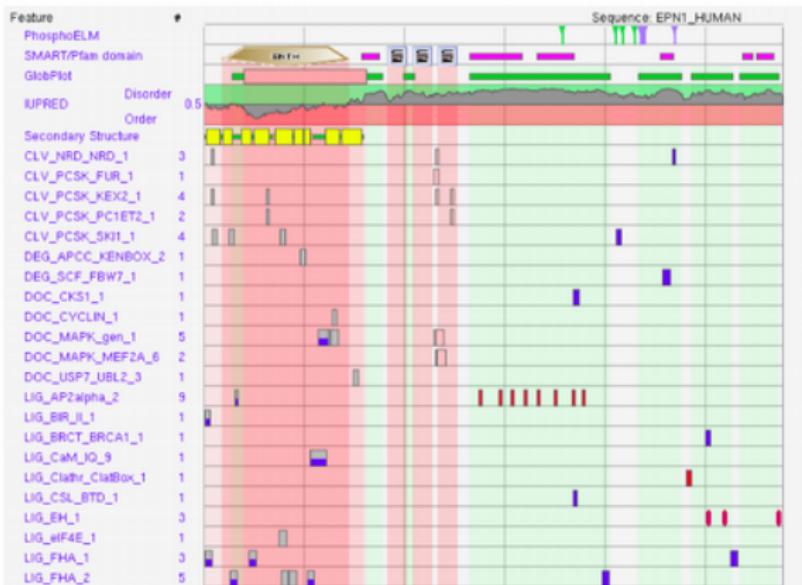
ELM

■ Summary for sequence 'EPN1_HUMAN'.

KEY

DOMAINS:	Smart/Pfam domain	Signal peptide (pred.)	Low-complexity region	Coiled-coil (pred.)	TM helix (pred.)
GLOBPLOT:	GlobDom	Disorder			
2D STRUCT:	Strand	Helix	Loop		3/10 Helix
MOTIFS:	Favourable Context	Sparse/Smart filtered	Neutral	Annotated:	TP FP TN U <= Assigned by homology
CONSCORE:	low Conservation	medium Conservation	high Conservation		
Phospho-ELM:	Phosphorylated Serine	Phosphorylated Threonine	Phosphorylated Tyrosine		

(Mouseover the matches for more details)



KEY

DOMAINS:	Smart/Pfam domain	Signal peptide (pred.)	Low-complexity region	Coiled-coil (pred.)	TM helix (pred.)
GLOBPLOT:	GlobDom	Disorder			Help
2D STRUCT:	Strand	Helix	Loop	3/10 Helix	
MOTIFS:	Favourable Context	Sparse/Smart filtered	Neutral	Annotated:	TP FP TN U < Assigned by homology
CONSCORE:	low Conservation	medium Conservation	high Conservation		
Phospho.ELM:	Phosphorylated Serine	Phosphorylated Threonine	Phosphorylated Tyrosine		

TP (True Positive): **Experimental evidence** supporting **functionality**.

FP (False Positive): **Experimental evidence hinting** at a function, but after **careful inspection** ELM annotators believe this instance to be **non-functional**.

TN (True Negative): **Experimental evidence** supporting **non-functionality**.

U (Unknown): **Not enough evidence** to determine whether this instance is functional or not.

THE ELM PREDICTION TOOL

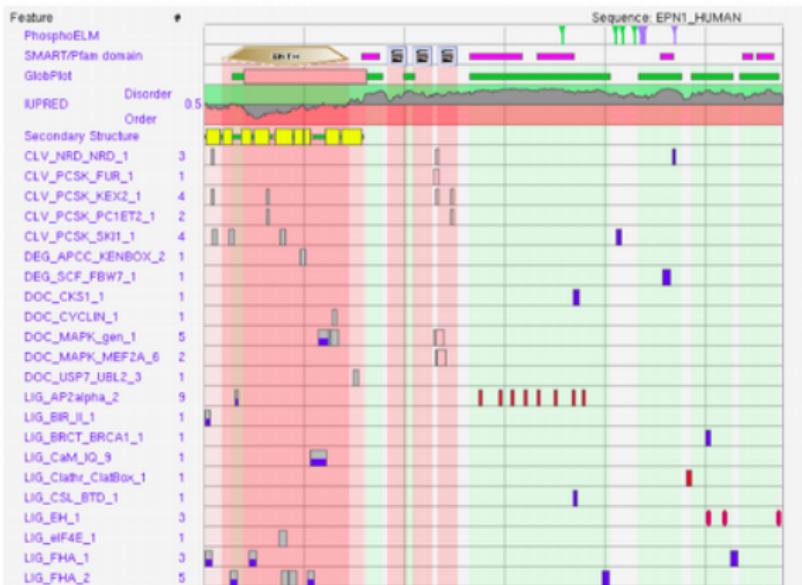
ELM

■ Summary for sequence 'EPN1_HUMAN'.

KEY

DOMAINS:	Smart/Pfam domain	Signal peptide (pred.)	Low-complexity region	Coiled-coil (pred.)	TM helix (pred.)
GLOBPLOT:	GlobDom	Disorder			
2D STRUCT:	Strand	Helix	Loop		3/10 Helix
MOTIFS:	Favourable Context	Sparse/Smart filtered	Neutral	Annotated:	TP FP TN U <= Assigned by homology
CONSCORE:	low Conservation	medium Conservation	high Conservation		
Phospho-ELM:	Phosphorylated Serine	Phosphorylated Threonine	Phosphorylated Tyrosine		

(Mouseover the matches for more details)



SLIMSearch4

About

SLIMSearch is a short, linear motif (SLIM) discovery tool. The tool searches a proteome using a motif consensus to discover putative novel motif instances. Motif attributes known to be strong discriminators of motif functionality, such as accessibility and conservation, are calculated for each instance. Instances are also annotated with experimental, proteomic and genomic data. The tool also allows instance filtering based on keyword, interaction, localisation information. In depth ontology enrichment and conservation analysis tools are also available. Find a tutorial to understand the SLIMSearch output [here](#).

Reference:

"SLIMSearch - a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions"

Izabella Krystkowiak and Norman E. Davey

Nucleic Acids Res. 2017 Apr 6. (Epub ahead of print).

Jobs are stored for two weeks.

Search details

Format: WRPW motif example [?](#)

Choose species: *Homo sapiens*

Homo sapiens

Saccharomyces cerevisiae

Xenopus tropicalis

Schizosaccharomyces pombe

Drosophila melanogaster

Arabidopsis thaliana

Caenorhabditis elegans

Choose options

Disorder score cut-off [?](#) : 0.40



Flank length [?](#) : 5



SLIMSearch Web Site <http://slim.ucd.ie/slimsearch/>

SLIMs SEARCH

Query: KDEL\$ Species: *Homo sapiens* | 12 hits in 12 proteins |

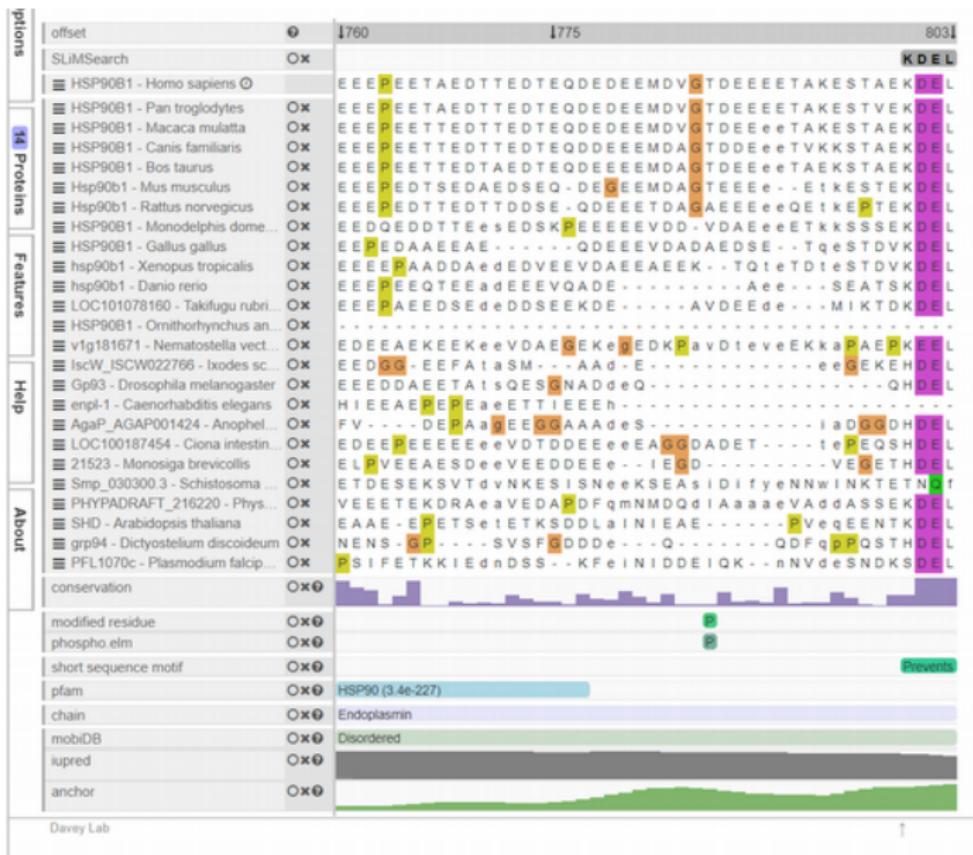
JobID: Hz591GnLhomm7Z3VFjuU1cZDJRii14QG

Download: [txt](#) | [json](#)Page 1 of 1 First | Prev | **1** | Next | Last

Protein Name	Peptide	Length	conservation		accessibility		Domain	Structure	Secondary Structure	← expand →											
			Start	End	Motif	OFO				Motif	Region	Switch	Modification	Topology	Isomform	Mutagenesis	SNP	Other			
Endoplasmic (HSP90B1) ▲	estaaKDEL	4	800	803	0.001	6.6e-5	0.973	0.7878		1											
Protein disulfide-isomerase (P4HB) ▲	dskavKDEL	4	505	508	6.1e-4	3.7e-4	0.971	0.7893		1											
DnaJ homolog subfamily C member 10 (DNAJC10)	agkrnKDEL	4	790	793	0.947	0.009	0.859	0.0379		1											1
Prolyl 3-hydroxylase 1 (LEPRE1)	seskpKDEL	4	733	736	0.013	0.043	0.984	0.7778		1	1										3
Prolyl 3-hydroxylase 2 (LEPREL1)	lningKDEL	4	705	708	0.038	0.058	0.86	0.7506		1											
78 kDa glucose-regulated protein (HSPA5)	edtaeKDEL	4	651	654	0.993	0.065	0.982	0.7381		1											
Thioredoxin domain-containing protein 5 (TXNDC5) ▲	vlsqaKDEL	4	429	432	0.059	0.075	0.695	0.7149	2	3	1	1									1
KDEL motif-containing protein 1 (KDELC1) ▲	hrkkktKDEL	4	499	502	0.621	0.088	0.766	0.6654	1		1										1
Calreticulin (CALR) ▲	vpgqaKDEL	4	414	417	2.9e-5	0.698	0.999	0.8892		1	1										
Protein disulfide-isomerase A6 (PDIA6)	lddlgKDEL	4	437	440	0.997	0.996	0.842	0.1126		1	1										1
Ubiquitin-conjugating enzyme E2L (UBE2L)	—	—	—	—	1.000	1.000	0.045	0.100													

<http://slim.ucd.ie/rest/#/slimsearch/annotations?jobId=Hz591GnLhomm7Z3VFjuU1cZDJRii14QG>

SLIMs SEARCH



SLIMs SEARCH

View: Flank conservation | Taxonomic range

Clad: Chordata | Metazoa | QFO

12 hits in 12 proteins

Page 1 of 1 First | Prev | **1** | Next | Last

Questions?