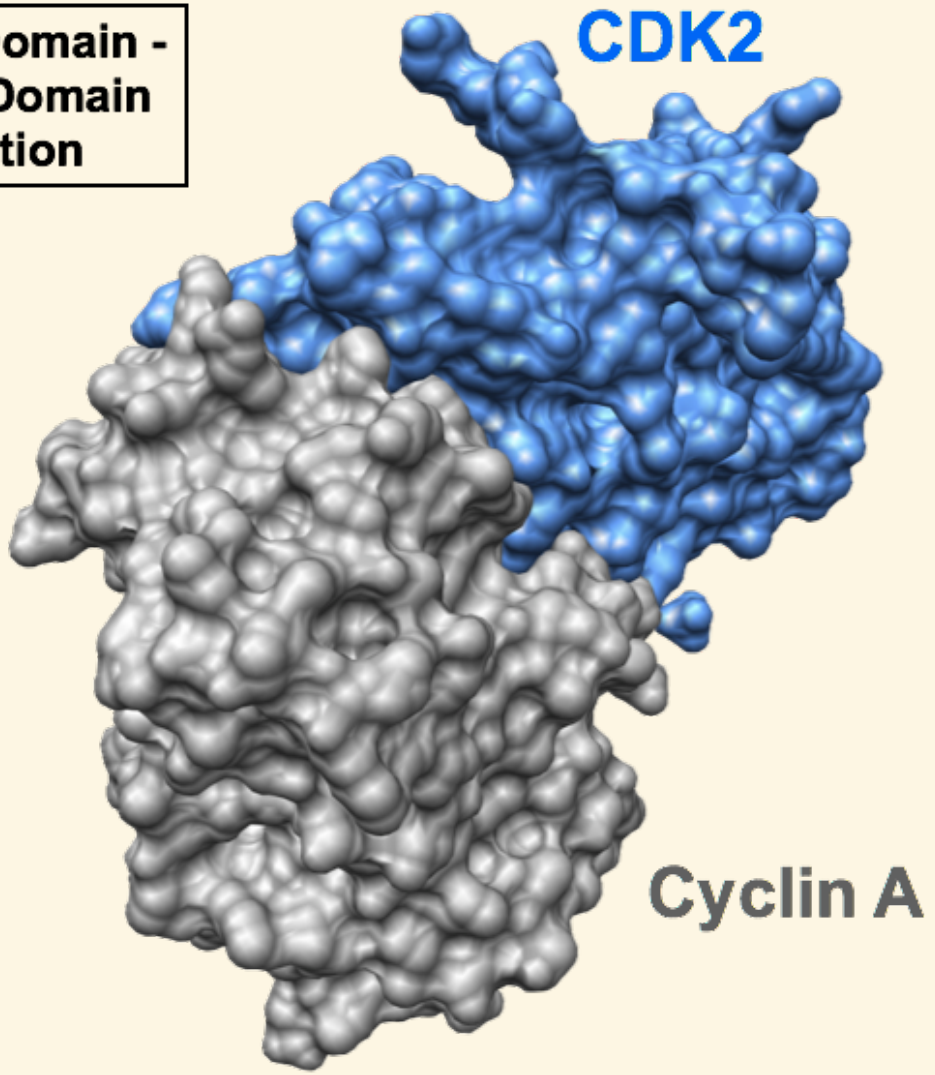
Short Linear Motifs

Holger Dinkel

EMBO Practical Course:
"Computational analysis of protein-protein interactions
in cell function and disease"
Bangalore, 05. 12. 2019

Importance of Short Linear Motifs

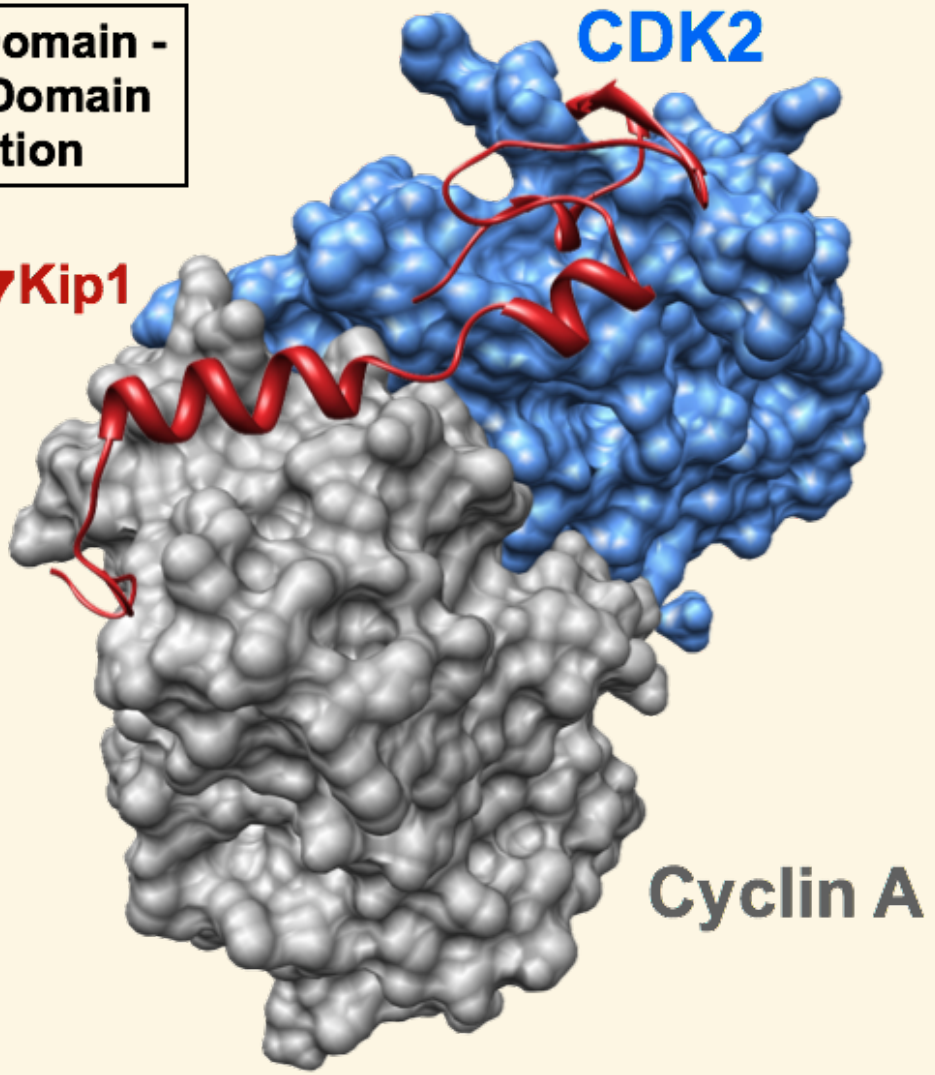
Globular Domain - Globular Domain Interaction



Importance of Short Linear Motifs

Globular Domain - Globular Domain Interaction

Globular Domain - Disordered Domain Interaction

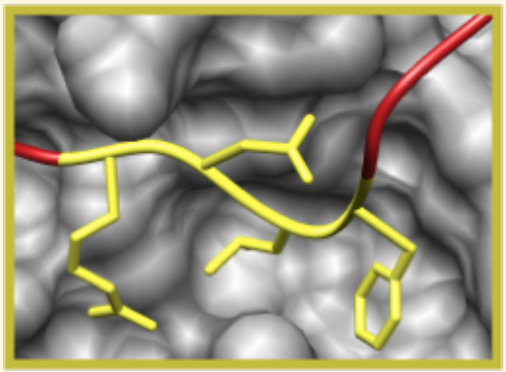


Importance of Short Linear Motifs

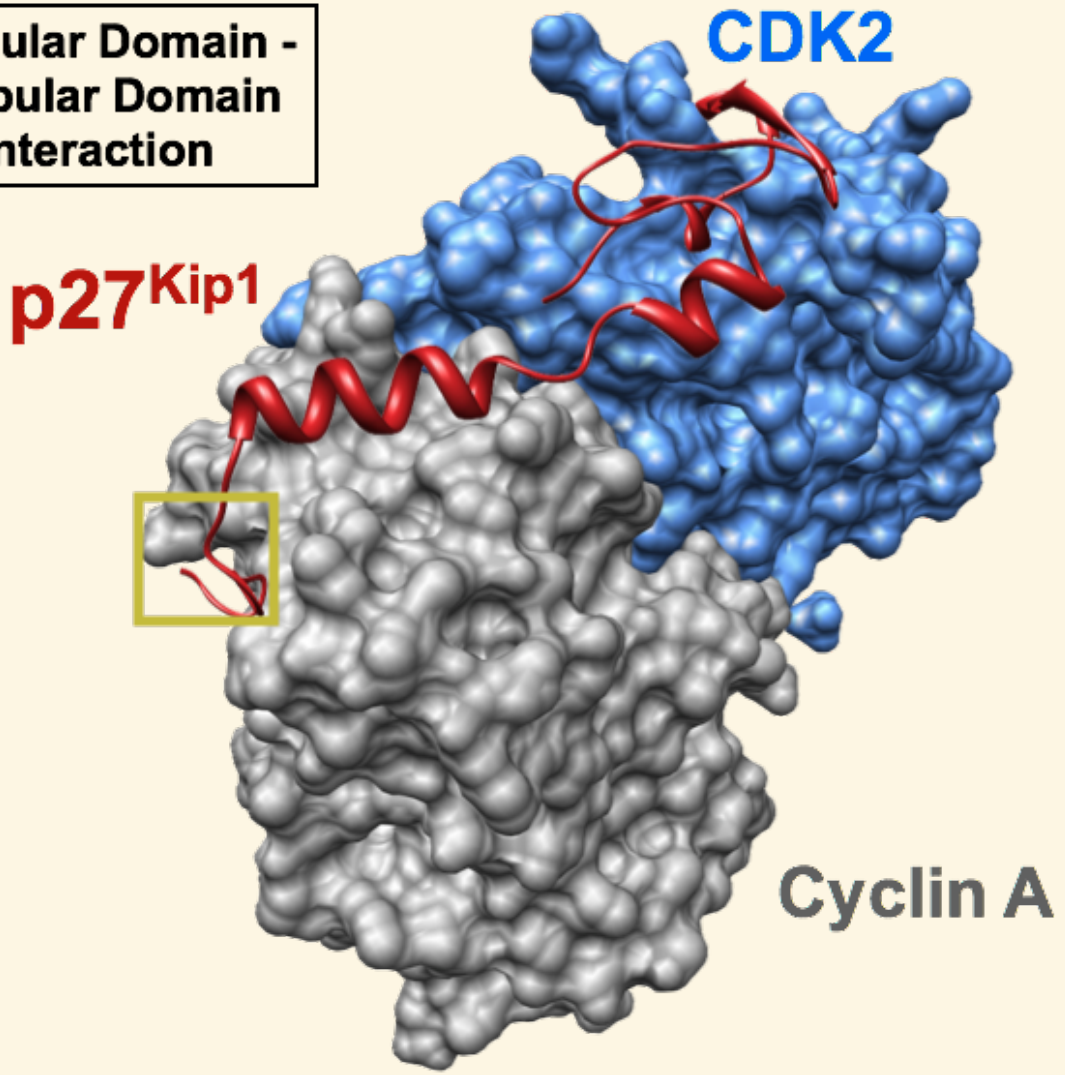
Globular Domain - Globular Domain Interaction

Globular Domain - Disordered Domain Interaction

Globular Domain - Short Linear Motif Interaction



RNLF



Importance of Short Linear Motifs

Globular Domain - Globular Domain Interaction

Globular Domain - Disordered Domain Interaction

CDK2

p27Kip1

Regular Expression:

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

Defined positions

Fixed positions

Degenerate positions

Undefined positions

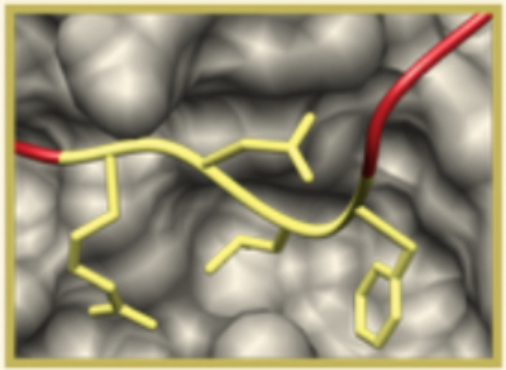
Fixed-length wildcard

Flexible-length wildcard {min,max}

Cyclin A

PDB 1JSU

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



RNLF

Regular Expressions are used to describe Short Linear Motifs

Character	Meaning
.	**Any** amino acid allowed
$[xy]$	Amino acids **listed** are allowed
$[\hat{}xy]$	Amino acids listed are **not** allowed
$\{min,max\}$	**Min** required, **max** allowed
$\hat{}$	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

Regular Expressions are used to describe Short Linear Motifs

Character	Meaning
.	**Any** amino acid allowed
$[xy]$	Amino acids **listed** are allowed
$[\hat{}xy]$	Amino acids listed are **not** allowed
$\{min,max\}$	**Min** required, **max** allowed
$\hat{}$	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

DOC_CYCLIN_1 $[RK].L.\{0,1\}[FYLIVMP]$

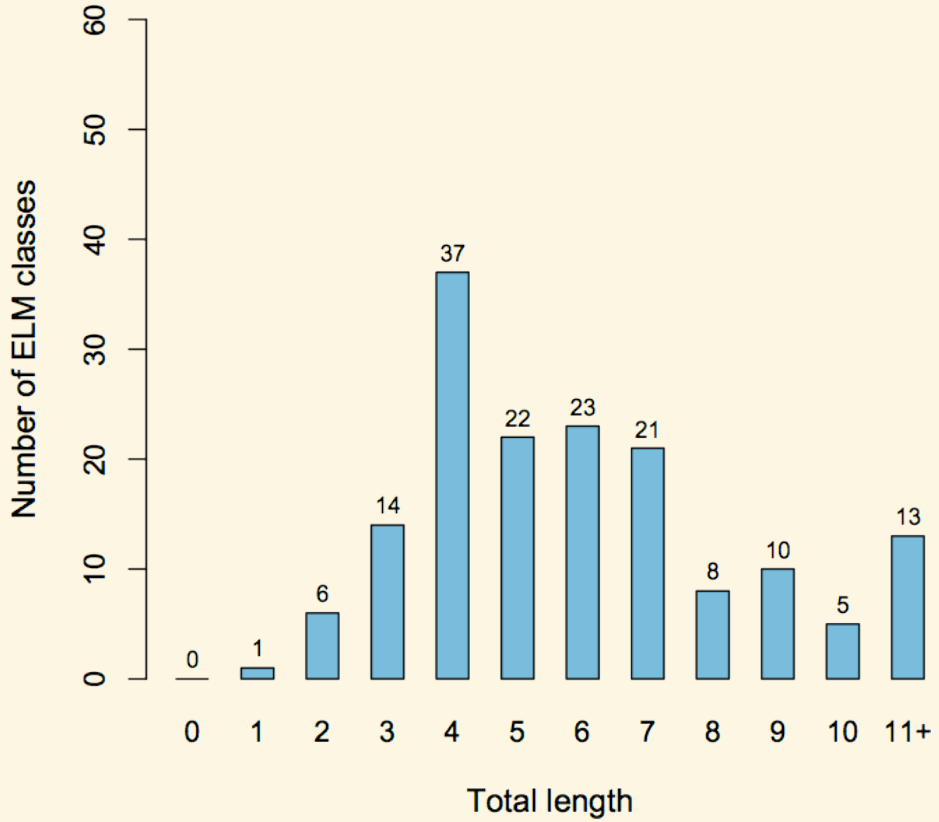
has been replaced by the more specific DOC_CYCLIN_RxL_1:

$$(.||([KRH].\{0,3\}))[\hat{}EDWNSG][\hat{}D]L.\{0,1\}[FLMP].\{0,3\}[EDST]$$

Attributes of Short Linear Motifs

Linear Motifs

- are small.
- have few defined positions.
- mediate transient, low affinity interactions.

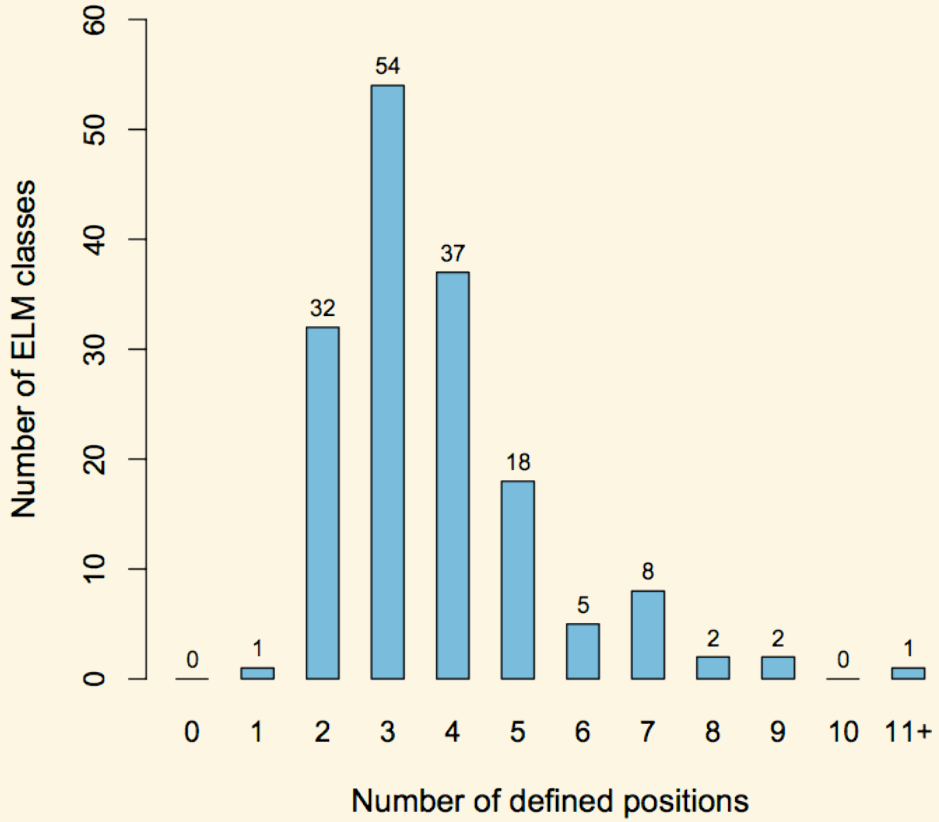


"Attributes of short linear motifs"; Davey, van Roey, Weatheritt, Toedt, Uyar, Altenberg, Budd, Diella, Dinkel & Gibson; (Mol Biosyst. 2011)

Attributes of Short Linear Motifs

Linear Motifs

- are small.
- have few defined positions.
- mediate transient, low affinity interactions.

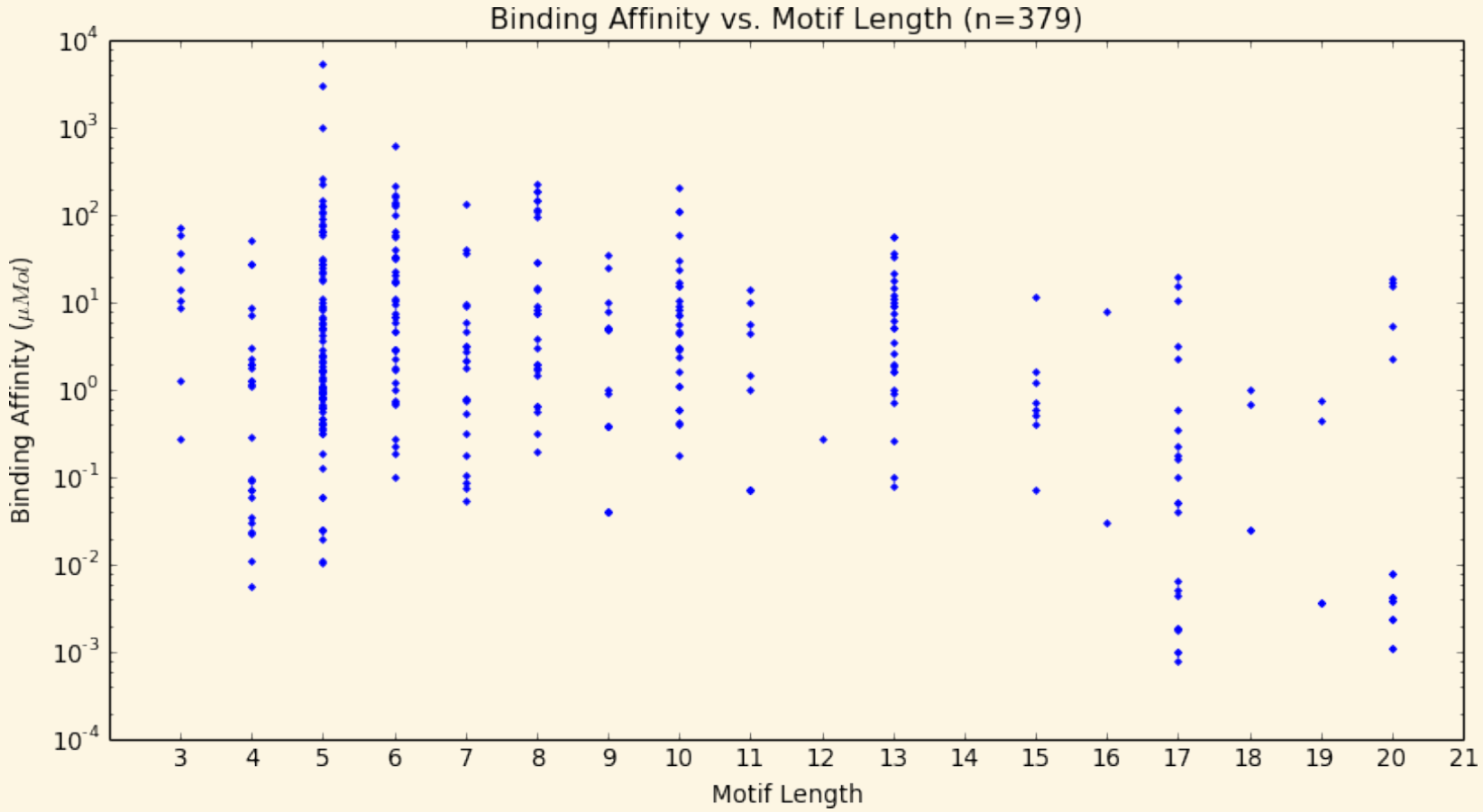


"Attributes of short linear motifs"; Davey, van Roey, Weatheritt, Toedt, Uyar, Altenberg, Budd, Diella, Dinkel & Gibson; (Mol Biosyst. 2011)

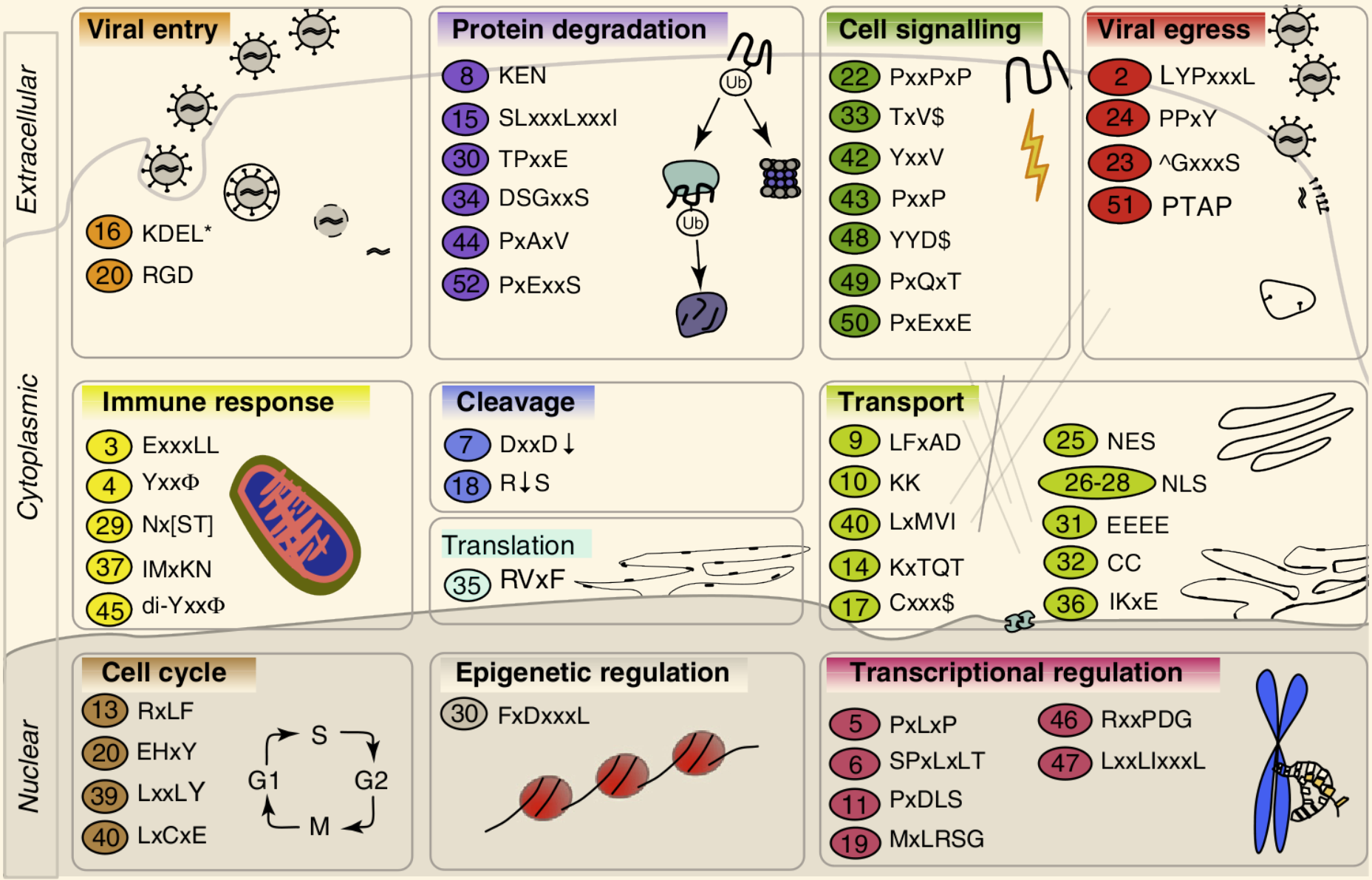
Attributes of Short Linear Motifs

Linear Motifs

- are small.
- have few defined positions.
- mediate transient, low affinity interactions.



Importance of Short Linear Motifs: Viruses



"How viruses hijack cell regulation"; Davey, Travé & Gibson; (TiBS 2010)

Importance of Short Linear Motifs: Diseases

Liddle's-Syndrome: WW-Interaction motif

has been implicated with autosomal dominant activating mutations in the WW interaction motif in the β- and γ-subunits of the epithelial sodium channel ENaC. These mutations abrogate the binding to the ubiquitin ligase NEDD4-2, ultimately resulting in increased Na^+ reabsorption, plasma volume extension and hypertension.

See `http://elm.eu.org/infos/diseases.html` for a list of motifs implicated in various diseases.

Importance of Short Linear Motifs: Diseases

Liddle's-Syndrome: WW-Interaction motif

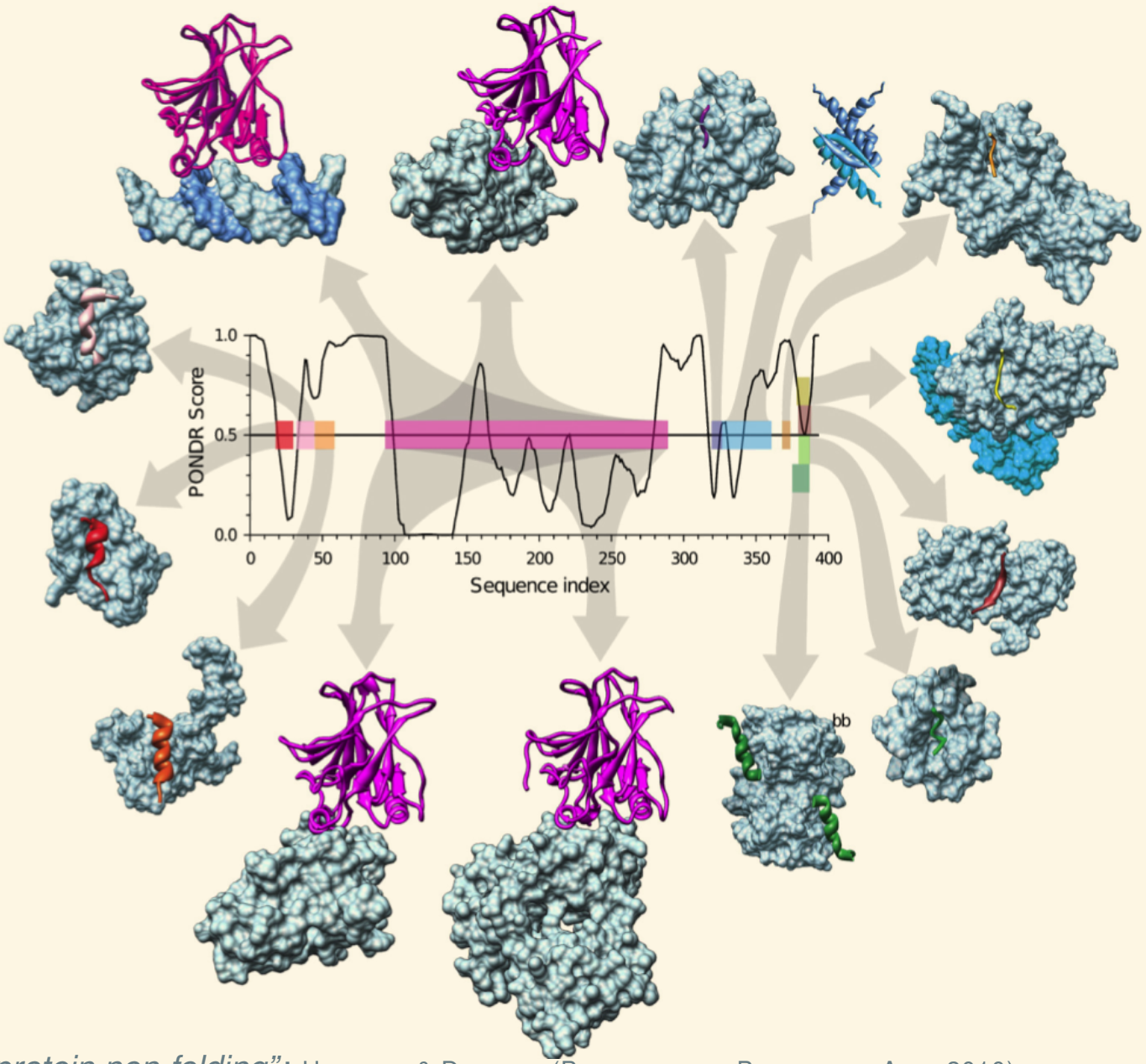
has been implicated with autosomal dominant activating mutations in the WW interaction motif in the β- and γ-subunits of the epithelial sodium channel ENaC. These mutations abrogate the binding to the ubiquitin ligase NEDD4-2, ultimately resulting in increased Na^+ reabsorption, plasma volume extension and hypertension.

Bacillus anthracis "lethal factor"

The protein LEF_BACAN is a metalloprotease (one of the three proteins composing the anthrax toxin) that specifically targets mitogen-activated protein kinase kinases (MKKs), which are important regulators of signal transduction as they phosphorylate and thus activate specific MAPKs (such as ERK1, ERK2, p38 or JNK). Bacillus anthracis' "lethal factor" cleaves its MKK substrates within or close to the MAPK docking sites, thus effectively preventing the MKK to dock to its MAPK.

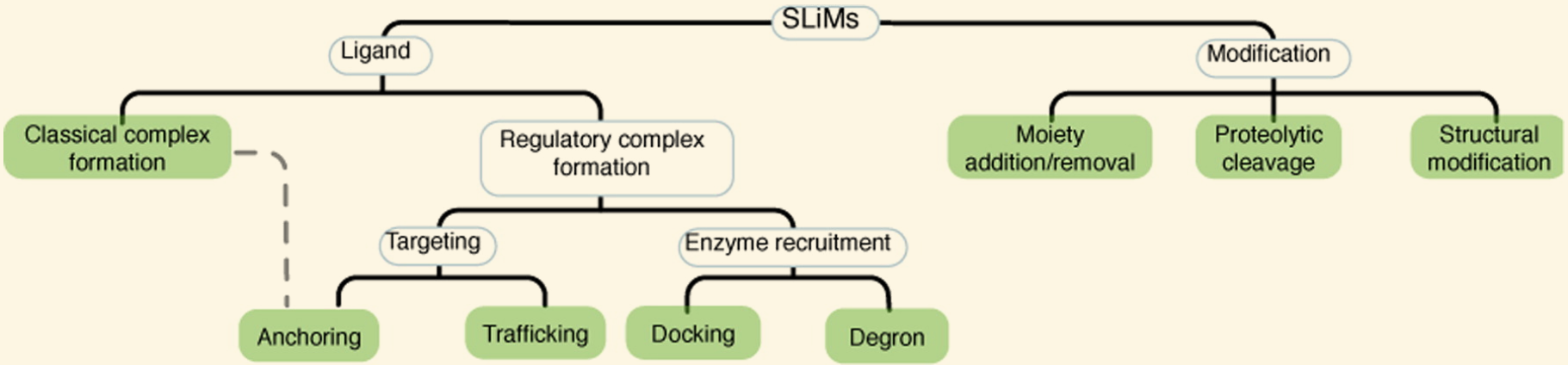
See `http://elm.eu.org/infos/diseases.html` for a list of motifs implicated in various diseases.

Importance of Short Linear Motifs: P53



"Understanding protein non-folding"; Uversky & Dunker; (Biochimica et Biophysica Acta 2010)

Classification of Motifs



Motif Classes: 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_1

RegEx $\ldots([ST])P.[KR]$

Kinase domain

`CDK site`

Motif Classes: 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_1

RegEx $\ldots([ST])P.[KR]$

Motif Classes: Modification Sites

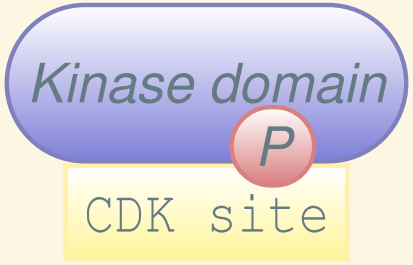
Description:

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_1

RegEx $\ldots([ST])P.[KR]$



Motif Classes: Modification Sites

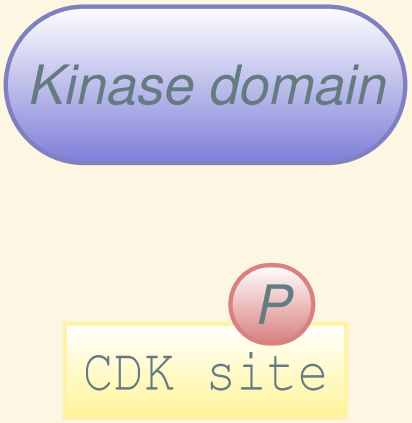
Description:

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_1

RegEx $\ldots([ST])P.[KR]$



Motif Classes: Docking Motifs

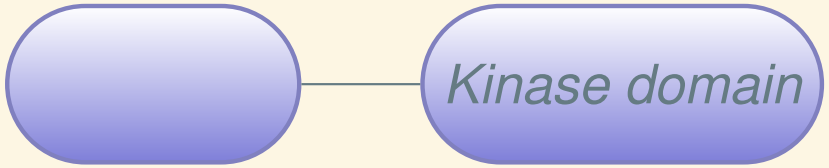
Description:

Docking motifs recruit enzymes via a surface that is distinct from the active site.

Example:

Name DOC_CYCLIN_1

RegEx $[RK].L.\{0,1\}[LFY]$



`docking site` —— `CDK site`

Motif Classes: Docking Motifs

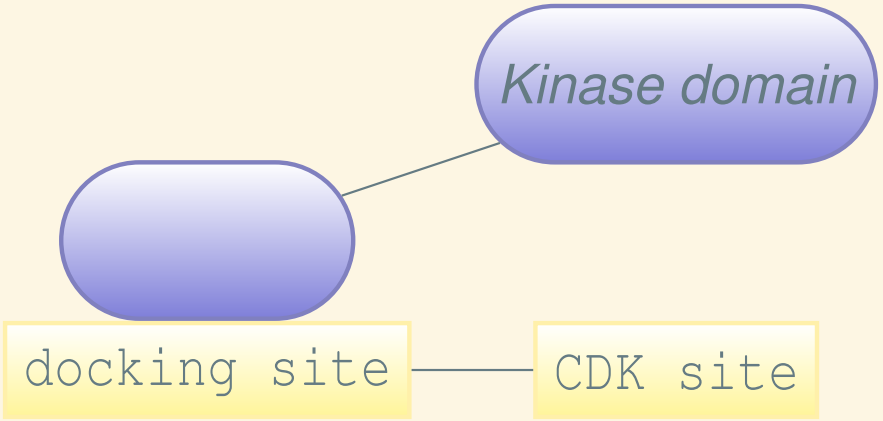
Description:

Docking motifs recruit enzymes via a surface that is distinct from the active site.

Example:

Name DOC_CYCLIN_1

RegEx $[RK].L.\{0,1\}[LFY]$



Motif Classes: Docking Motifs

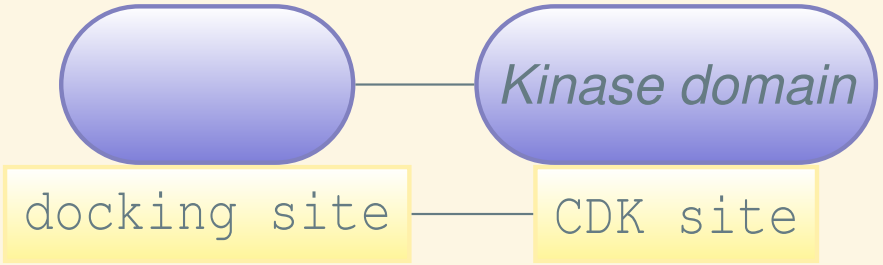
Description:

Docking motifs recruit enzymes via a surface that is distinct from the active site.

Example:

Name DOC_CYCLIN_1

RegEx $[RK].L.\{0,1\}[LFY]$



Motif Classes: Docking Motifs

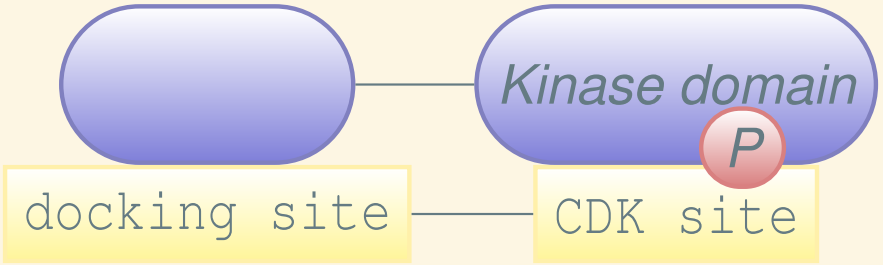
Description:

Docking motifs recruit enzymes via a surface that is distinct from the active site.

Example:

Name DOC_CYCLIN_1

RegEx $[RK].L.\{0,1\}[LFY]$



Motif Classes: Docking Motifs

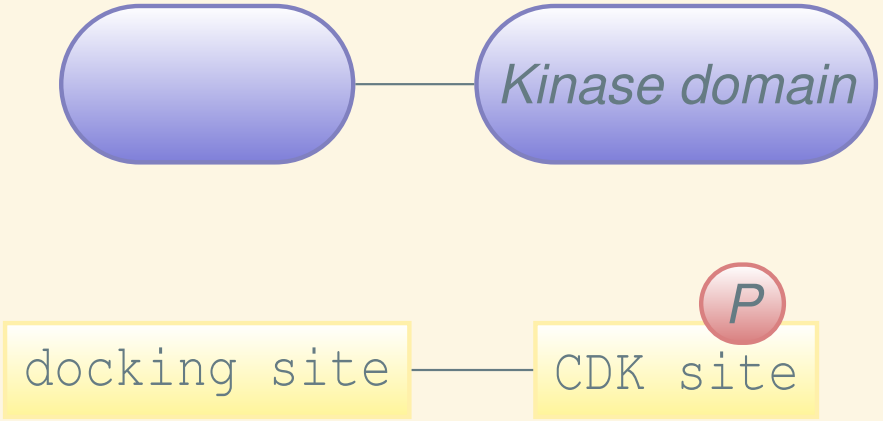
Description:

Docking motifs recruit enzymes via a surface that is distinct from the active site.

Example:

Name DOC_CYCLIN_1

RegEx $[RK].L.\{0,1\}[LFY]$



Motif Classes: Cleavage Motifs

Description:

Proteolytic processing of proteins into smaller polypeptides by protease-catalyzed hydrolysis of specific peptide bonds

Example:

Name CLV_Separin_Metazoa

RegEx $E[IMPVL][MLVP]R.$

Cleavage site

Motif Classes: Cleavage Motifs

Description:

Proteolytic processing of proteins into smaller polypeptides by protease-catalyzed hydrolysis of specific peptide bonds

Example:

Name CLV_Separin_Metazoa

RegEx $E[IMPVL][MLVP]R.$

Motif Classes: Cleavage Motifs

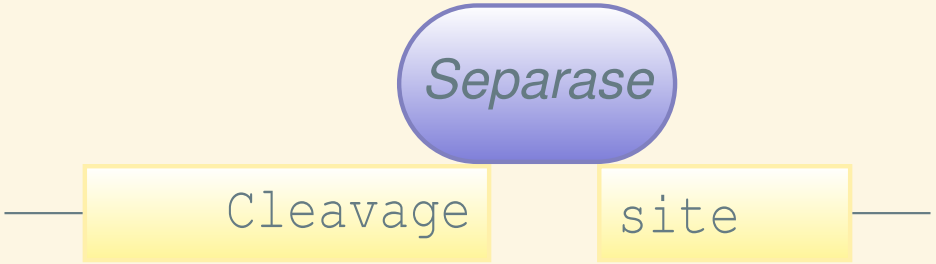
Description:

Proteolytic processing of proteins into smaller polypeptides by protease-catalyzed hydrolysis of specific peptide bonds

Example:

Name CLV_Separin_Metazoa

RegEx $E[IMPVL][MLVP]R.$



Motif Classes: Cleavage Motifs

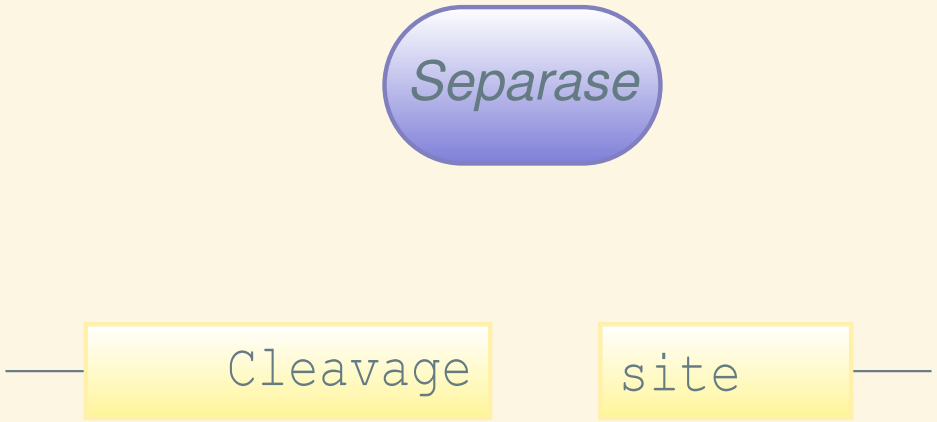
Description:

Proteolytic processing of proteins into smaller polypeptides by protease-catalyzed hydrolysis of specific peptide bonds

Example:

Name CLV_Separin_Metazoa

RegEx $E[IMPVL][MLVP]R.$



Motif Classes: Degradation Motifs

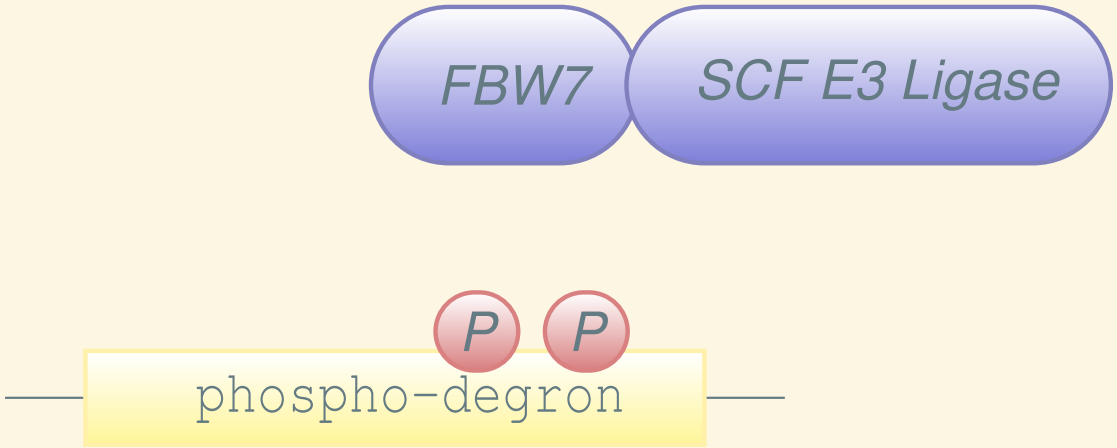
Description:

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

Example:

Name DEG_SCF_TRCP1_1

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



"Degrons in cancer"; Balint Meszaros, Manjeet Kumar, Toby J. Gibson, Bora Uyar & Zsuzsanna Dosztanyi; (Science Signalling 2017)

Motif Classes: Degradation Motifs

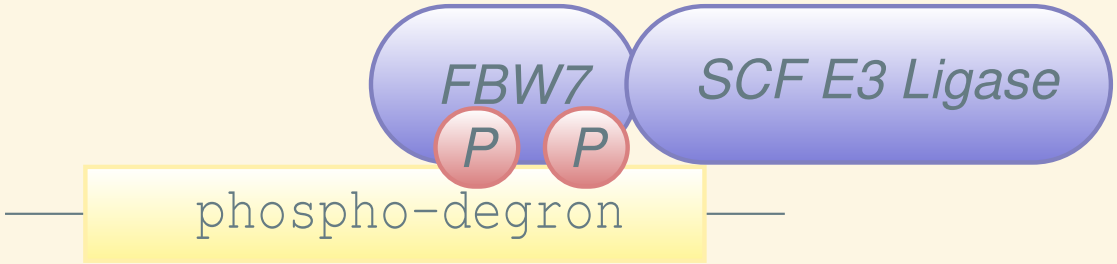
Description:

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

Example:

Name DEG_SCF_TRCP1_1

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



"Degrons in cancer"; Balint Meszaros, Manjeet Kumar, Toby J. Gibson, Bora Uyar & Zsuzsanna Dosztanyi; (Science Signalling 2017)

Motif Classes: Degradation Motifs

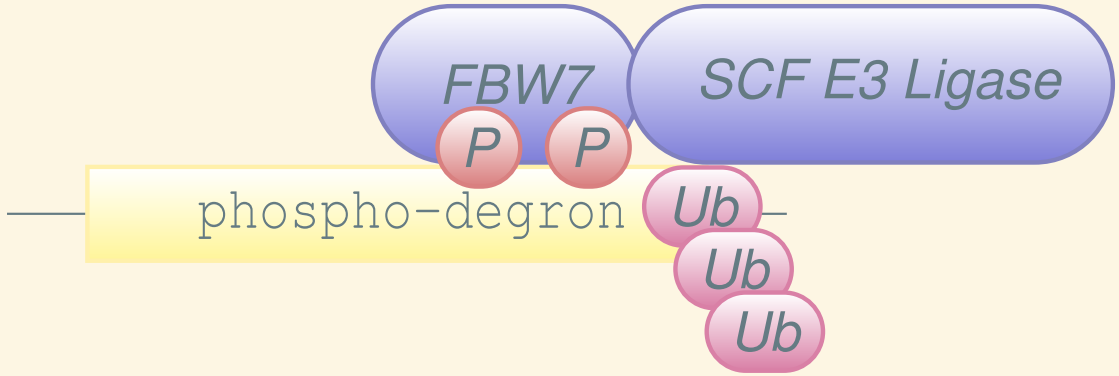
Description:

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

Example:

Name DEG_SCF_TRCP1_1

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



"Degrons in cancer"; Balint Meszaros, Manjeet Kumar, Toby J. Gibson, Bora Uyar & Zsuzsanna Dosztanyi; (Science Signalling 2017)

Motif Classes: Degradation Motifs

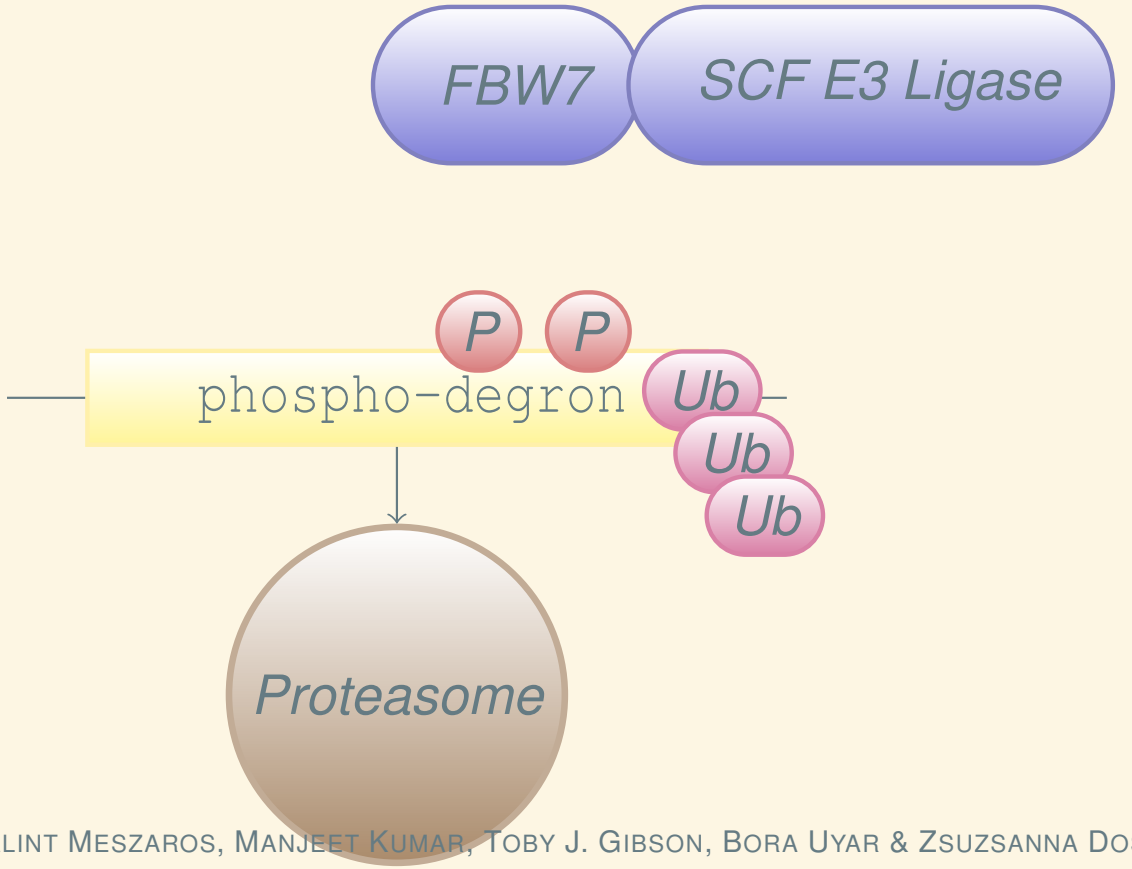
Description:

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

Example:

Name DEG_SCF_TRCP1_1

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



"Degrons in cancer"; Balint Meszaros, Manjeet Kumar, Toby J. Gibson, Bora Uyar & Zsuzsanna Dosztanyi; (Science Signalling 2017)

Motif Classes: 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 $[\hat{}DE](K[RK]|RK)[KRP][KR][\hat{}DE]$

Motif Classes: Targeting/Anchoring Motifs

Description:

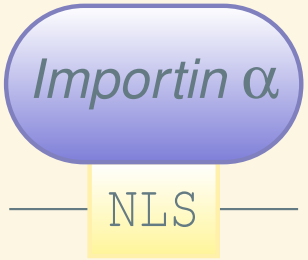
Targeting motifs allow a protein to bind to the transport machinery that relocalizes it to a particular sub-cellular location.

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 $[\hat{}DE](K[RK]|RK)[KRP][KR][\hat{}DE]$



Motif Classes: Targeting/Anchoring Motifs

Description:

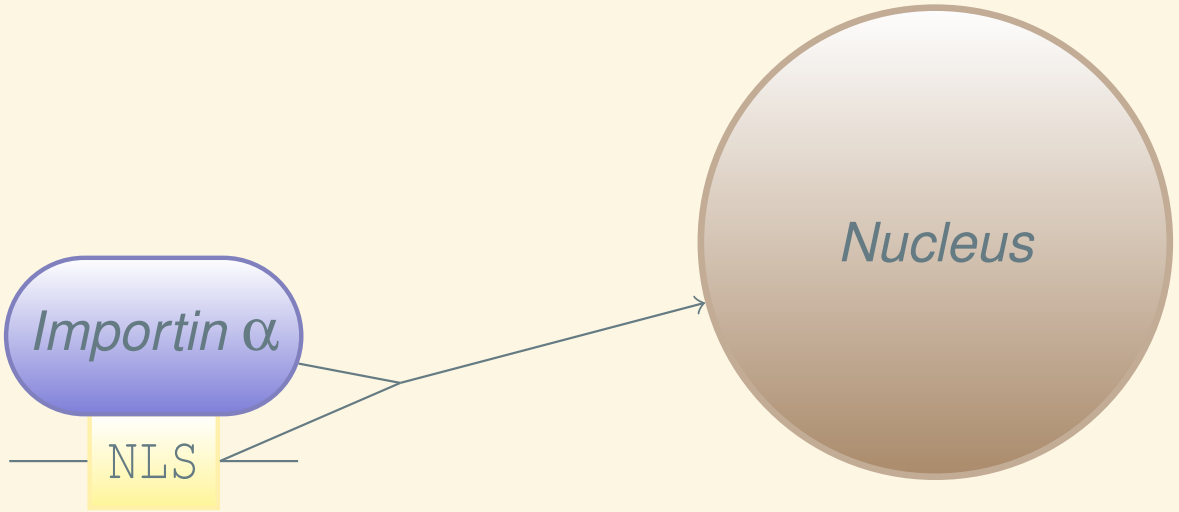
Targeting motifs allow a protein to bind to the transport machinery that relocalizes it to a particular sub-cellular location.

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

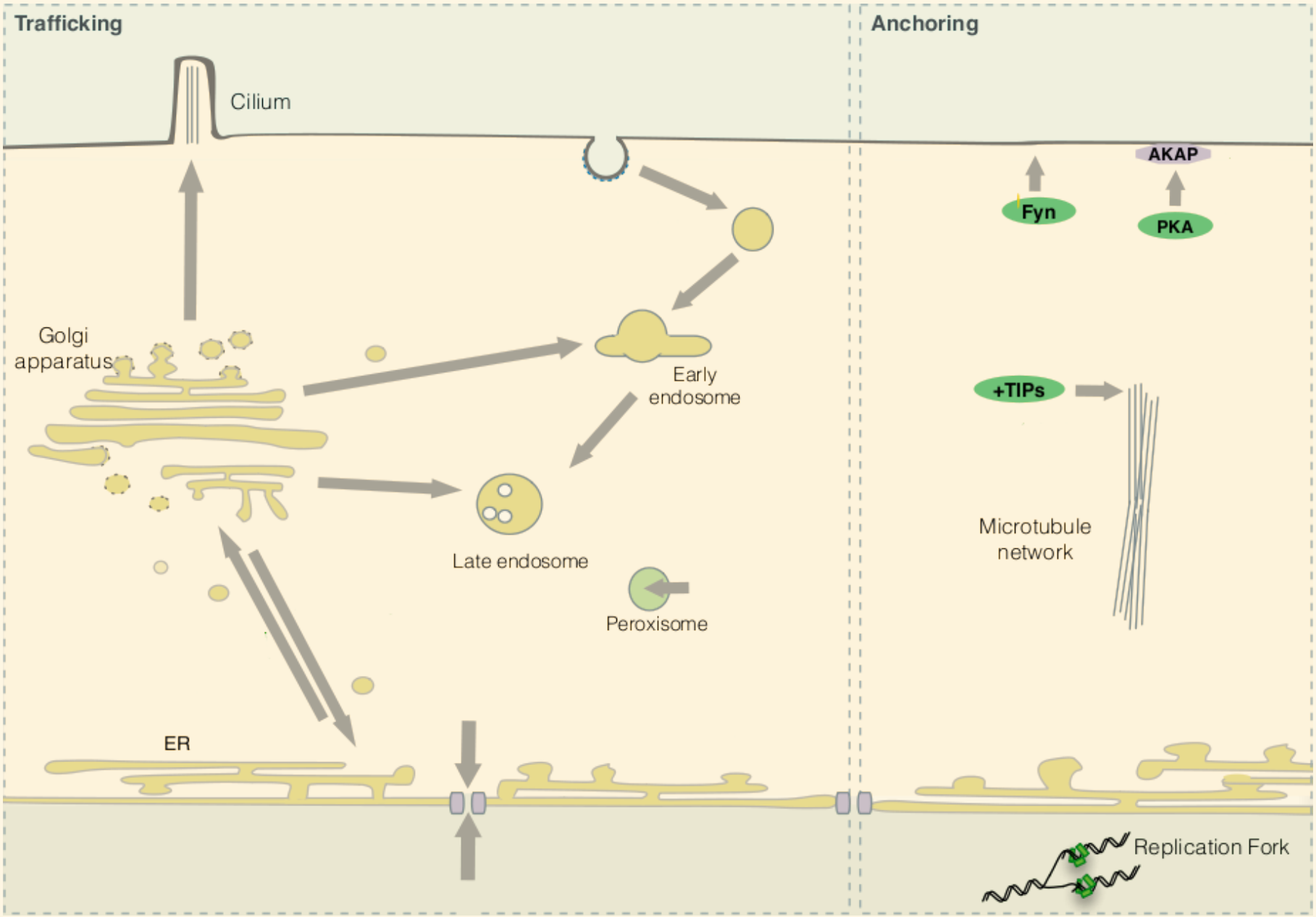
Example:

Name TRG_NLS_MonoCore_2

RegEx $[\hat{}DE](K[RK]|RK)[KRP][KR][\hat{}DE]$

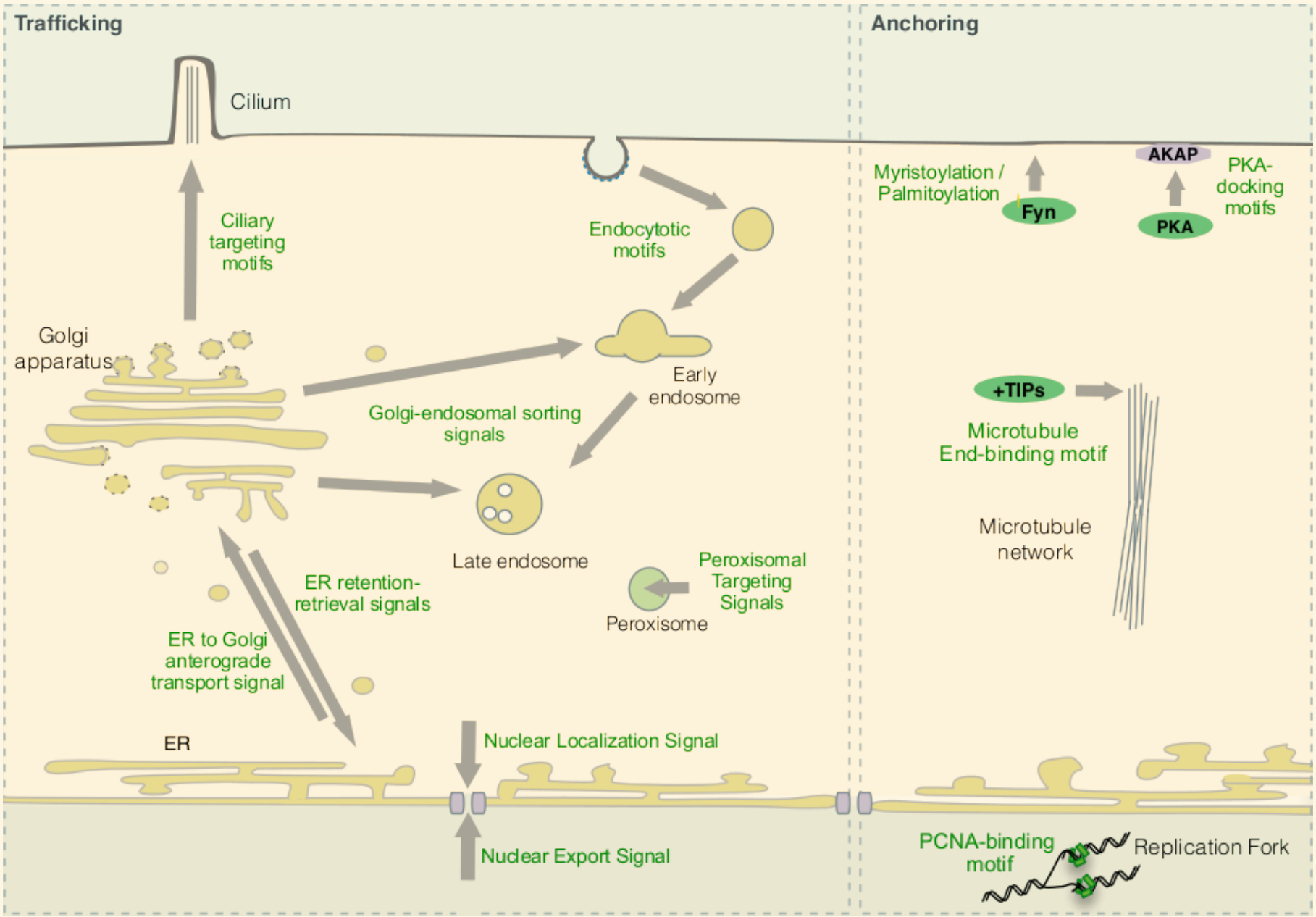


Motif Classes: Targeting/Anchoring Motifs



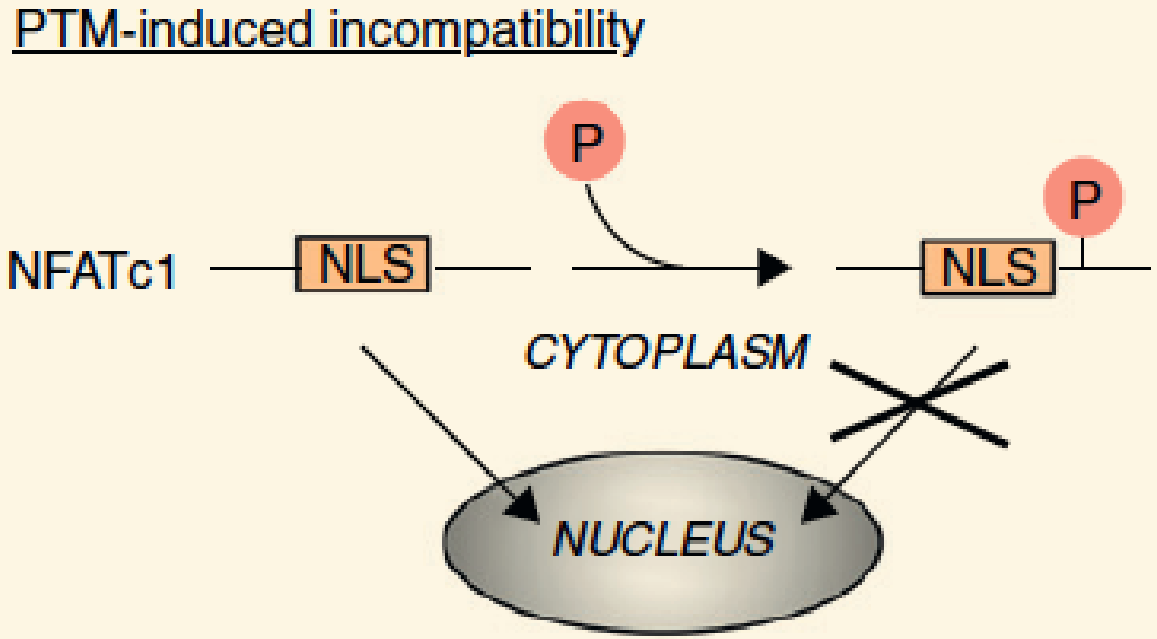
"Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation"; van Roey, Uyar, Weatheritt, Dinkel, Seiler, Budd, Gibson & Davey; (Chem. Reviews; 2014)

Motif Classes: Targeting/Anchoring Motifs



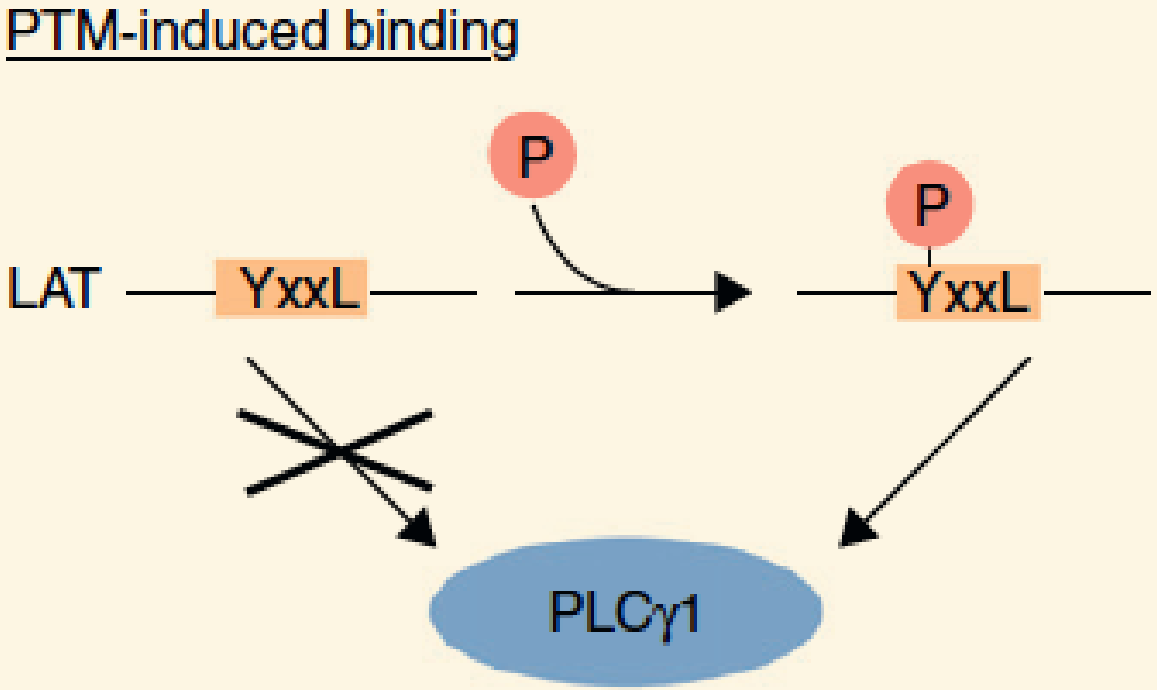
"Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation"; van Roey, Uyar, Weatheritt, Dinkel, Seiler, Budd, Gibson & Davey; (Chem. Reviews; 2014)

Short Linear Motifs



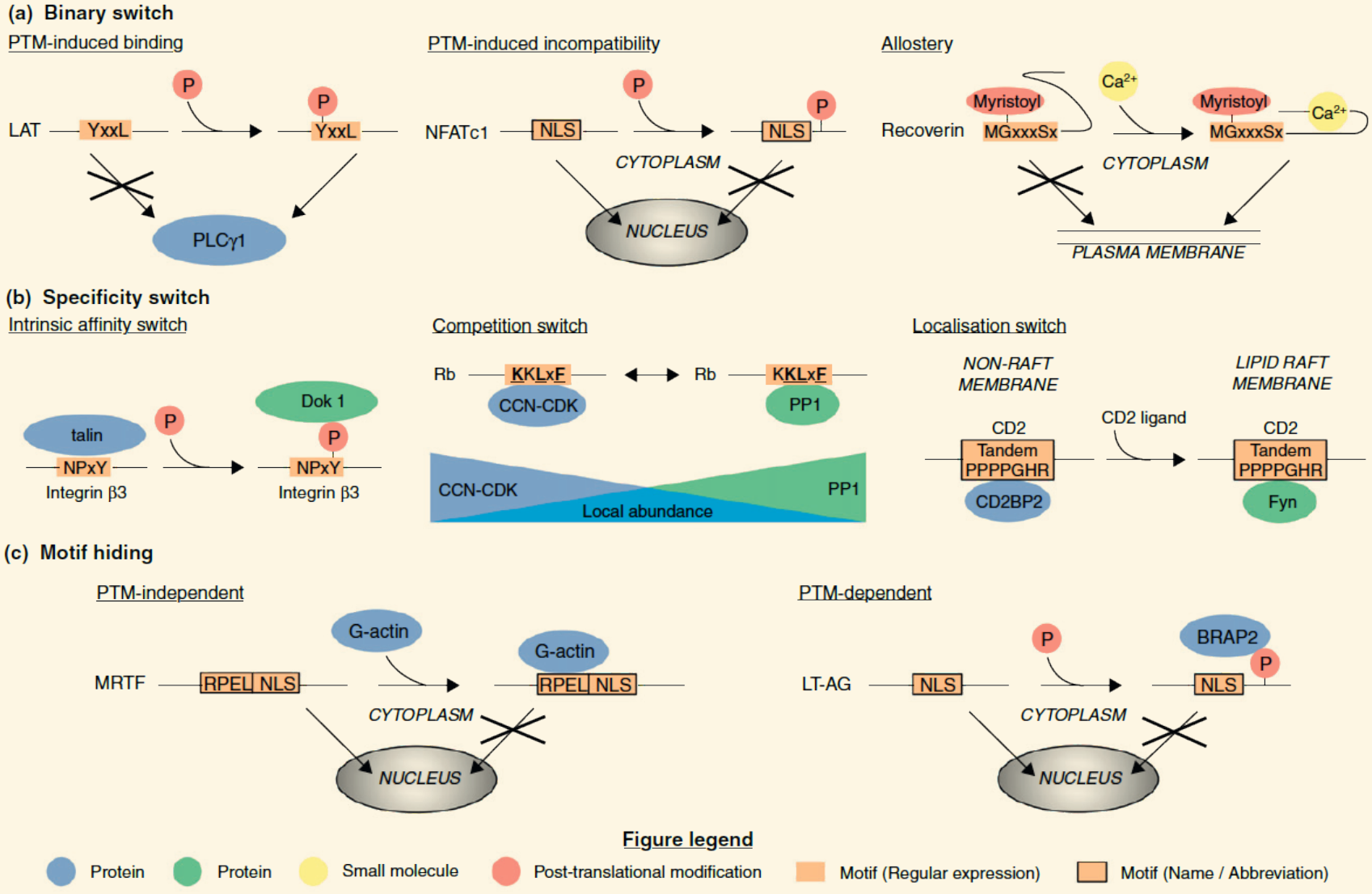
"Motif switches: decision-making in cell regulation."; van Roey, Gibson & Davey; (Curr Opin Struct Biol. 2012)

Short Linear Motifs



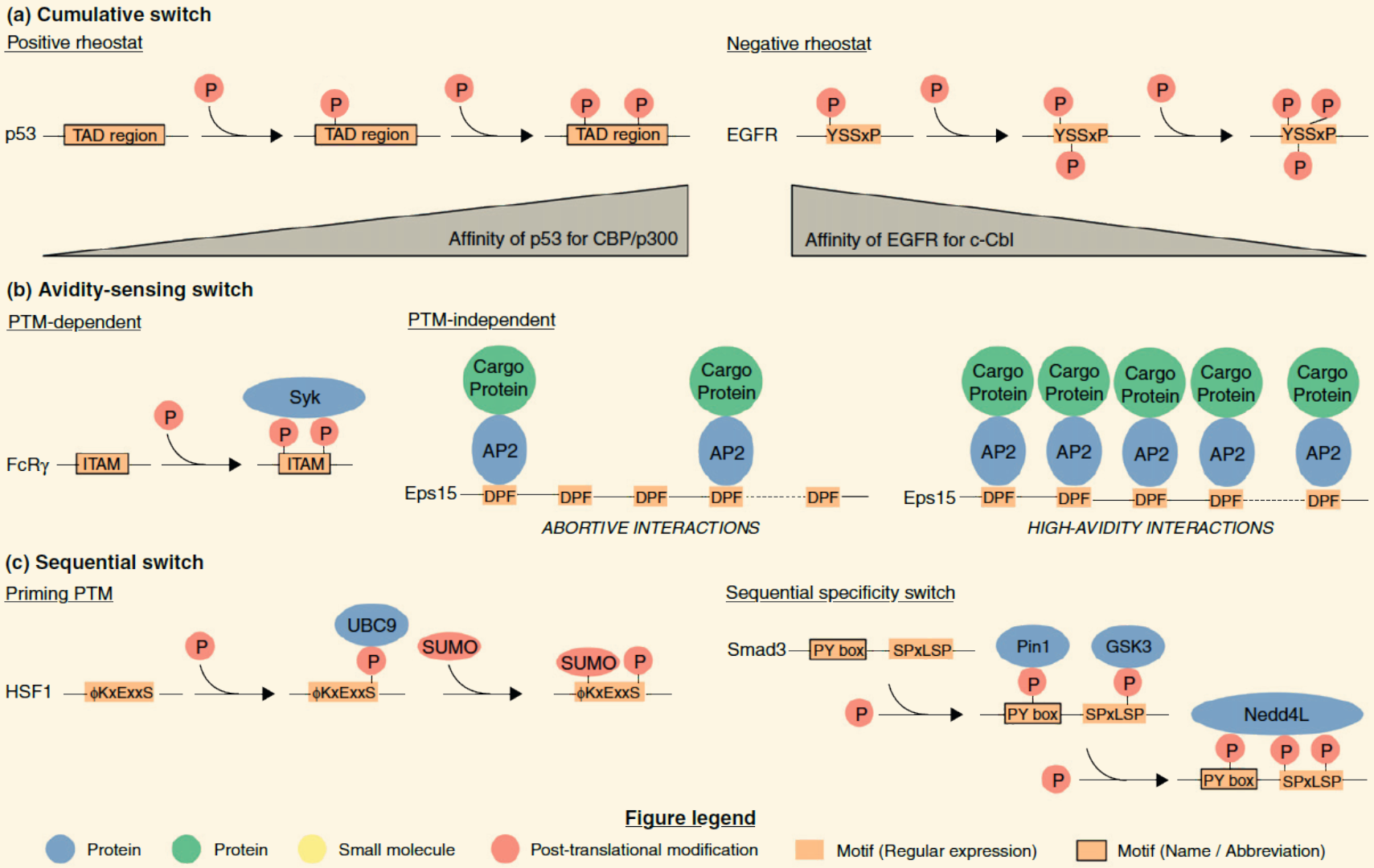
"Motif switches: decision-making in cell regulation."; van Roey, Gibson & Davey; (Curr Opin Struct Biol. 2012)

Short Linear Motifs



"Motif switches: decision-making in cell regulation."; VAN ROEY, GIBSON & DAVEY; (CURR OPIN STRUCT BIOL. 2012)

Short Linear Motifs

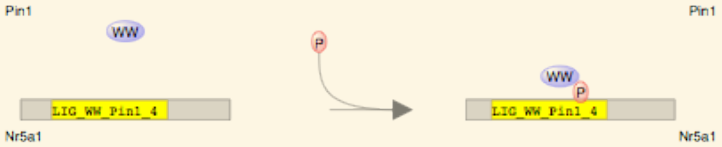


"Motif switches: decision-making in cell regulation."; VAN ROEY, GIBSON & DAVEY; (CURR OPIN STRUCT BIOL. 2012)

SHORT LINEAR MOTIFS

Switch #:	Switch type:	Switch subtype:
◇ SWTI000055 ◇	Binary ⓘ	Physicochemical compatibility ⓘ

Switch Description:
Phosphorylation of S203 in the Pin1-binding motif of *Steroidogenic factor 1 (Nr5a1)* induces binding to the *Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1)* protein.



Participants:
(1) *Steroidogenic factor 1 (Nr5a1)*
(2) *Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1)*

Interactions

Interaction #1 **Nr5a1 - Pin1**
Interfaces
(1) *LIG_WW_Pin1_4* motif ($_{200}$**PYASPP**$_{205}$) in *Steroidogenic factor 1 (Nr5a1)*
(2) *WW domain* (7-37) in *Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1)*

Interaction Regulation
PTM-dependent Induction (Phosphorylation of S203 on *Steroidogenic factor 1 (Nr5a1)*) of the *Steroidogenic factor 1 (Nr5a1) LIG_WW_Pin1_4* motif - *Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1) WW domain* interaction

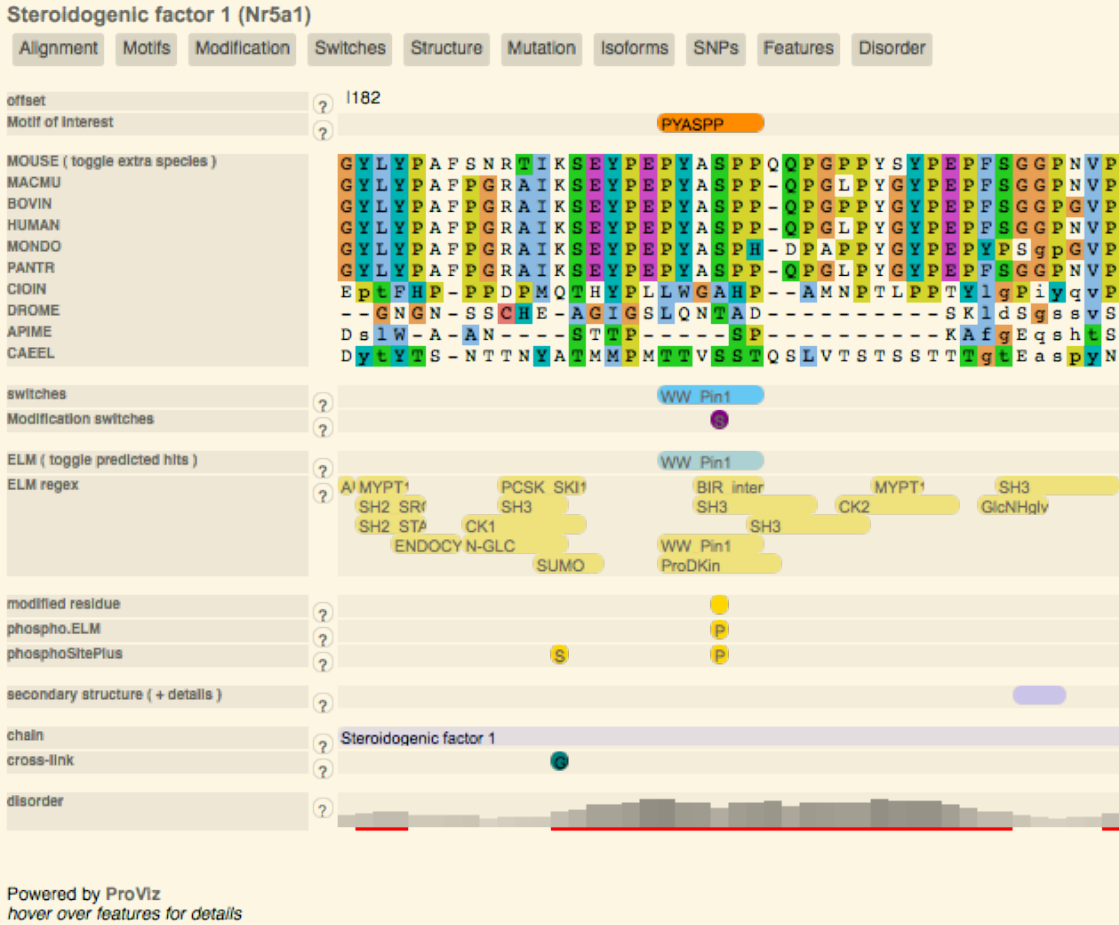
References

(1) Pin1 facilitates the phosphorylation-dependent ubiquitination of SF-1 to regulate gonadotropin beta-subunit gene transcription.
Luo *et al.* Mol. Cell. Biol. (2010)

See also

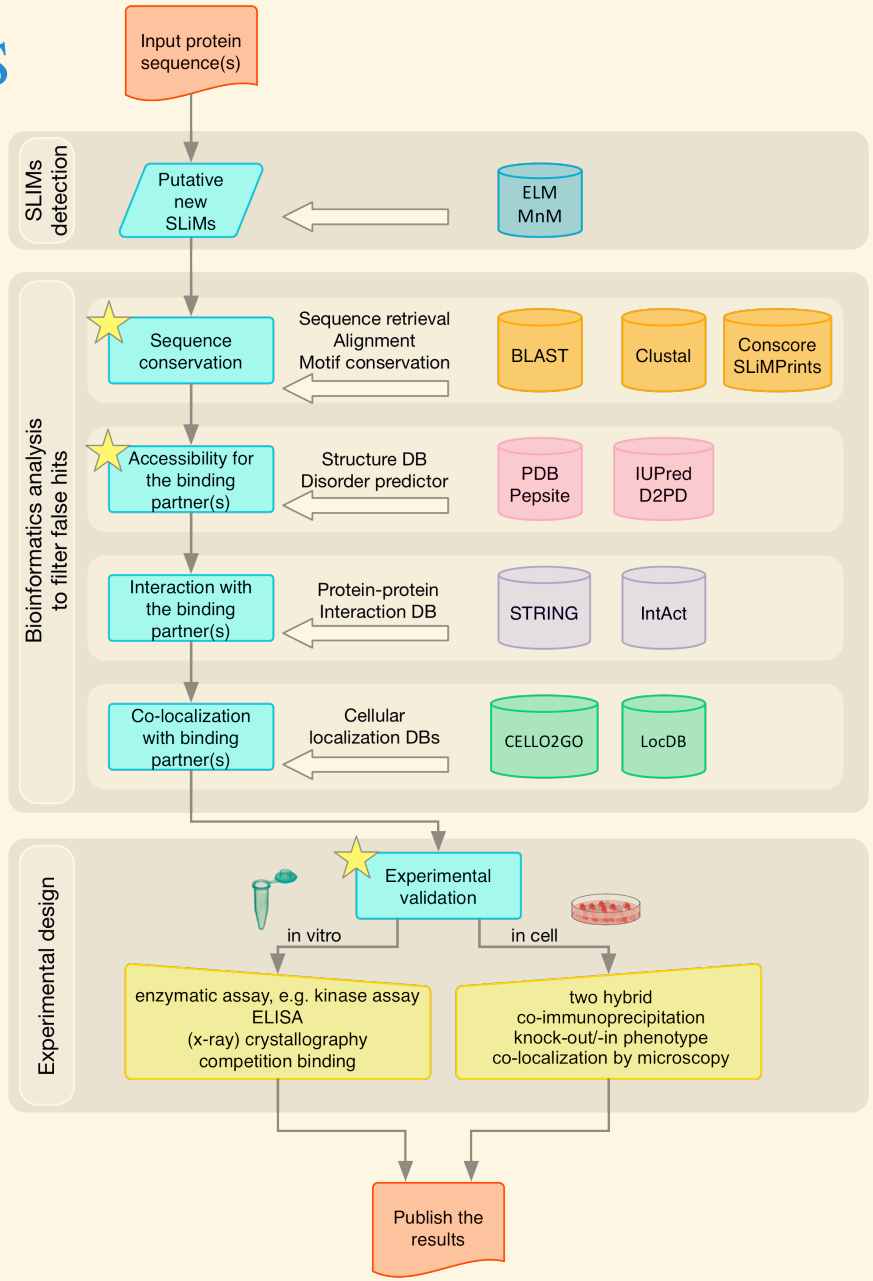
Other switches involving participants
Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1) - 28 more (view)

Other switches involving interfaces
LIG_WW_Pin1_4 - 89 more (view)
WW domain - 102 more (view)



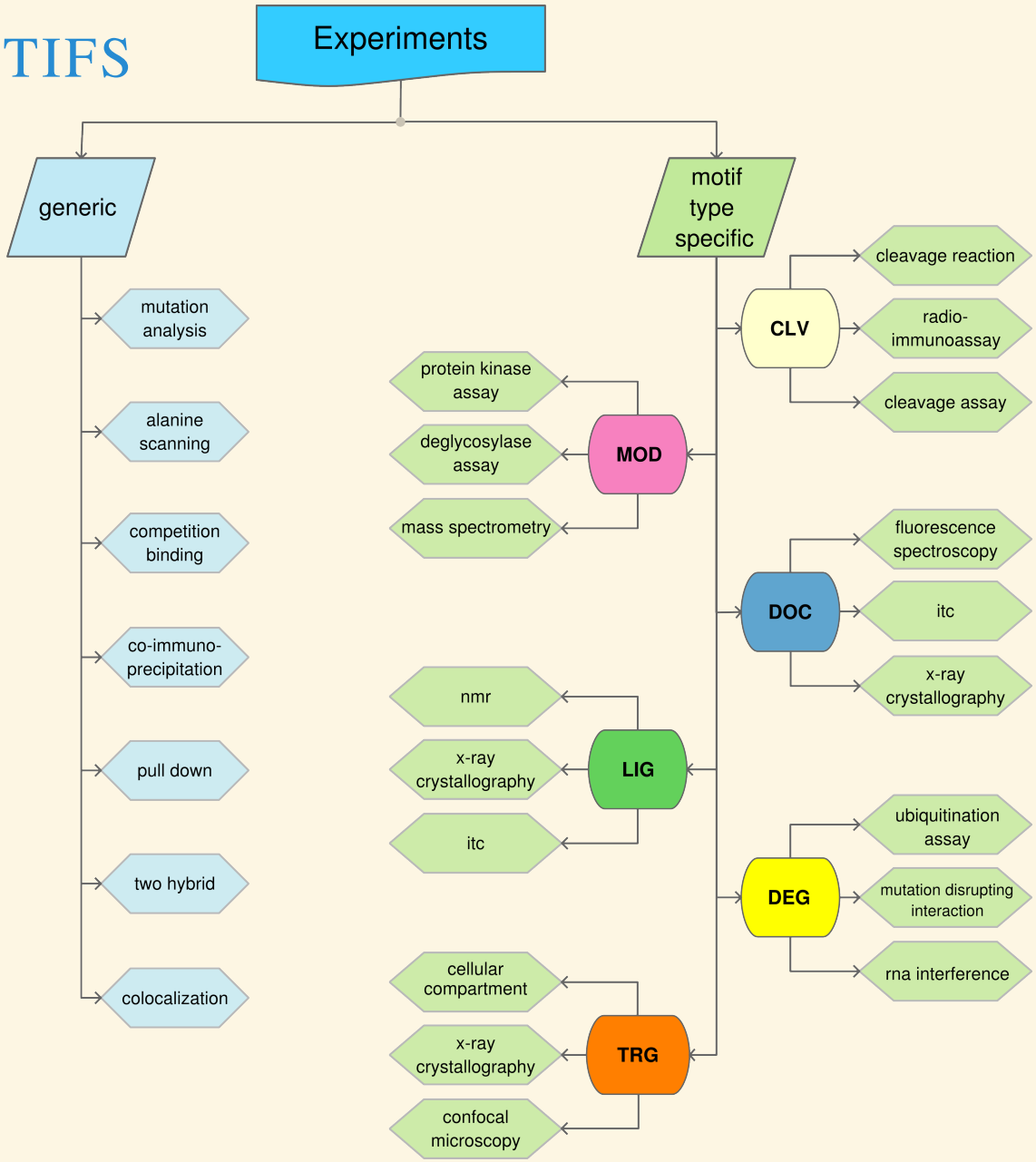
"The switches.ELM resource: a compendium of conditional regulatory interaction interfaces."; VAN ROEY, DINKEL, WEATHERITT, GIBSON & DAVEY; (SCI. SIGNAL. 2013)

GUIDELINES FOR EXPERIMENTAL DETECTION OF SHORT LINEAR MOTIFS



"Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad."; GIBSON TJ, DINKEL H, VAN ROEY K, DIELLA F.; (CELL COMMUN. SIGNAL 2015)

GUIDELINES FOR EXPERIMENTAL DETECTION OF SHORT LINEAR MOTIFS



"Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad."; GIBSON TJ, DINKEL H, VAN ROEY K, DIELLA F.; (CELL COMMUN. SIGNAL 2015)

Summary

Short Linear Motifs

- small, versatile modules which mediate transient interactions
- important regulators of cellular processes.
- “kidnapped” by viruses
- play an important role in diseases
- collected in the Eukaryotic Linear Motif Resource (ELM)

Questions?