10.
Linear motifs, post-translational modifications, molecular switches

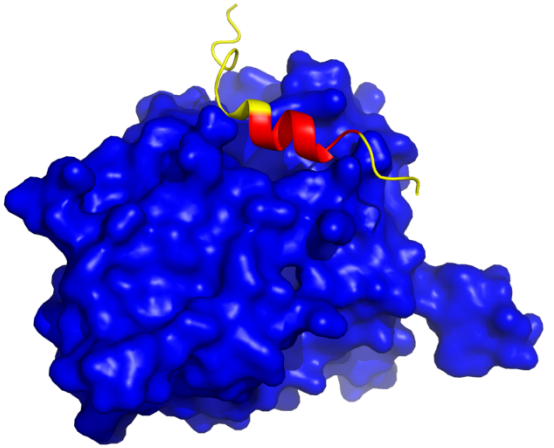
Short (3-10 residues), sequentially localized motifs that mediate the interaction with a common protein partner/domain:

Interaction partners of nuclear receptors:

```
PA2G4:  ...MEVQDAELKALLQSSASRKT...
NRIP1:  ...DSIVLTYLEGLLMHQAAGGS...
NcoA6:  ...MREAPTSLSQLLDNSGAPNV...
NcoA2:  ...DSKGQTKLLQLLTTKSDQME...
CBP:    ...AASKHKQLSELLRGGSGSSI...
```

Common interaction motif: (LIG_NRBOX)

xLxxLLx



Linear Motifs

- **SLiMs = “Short Linear Motif”:**
  - Short functional modules (3–10 aa long)
  - Mediate transient interactions
  - Usually reside within IDRs
  - Function independently of the rest of the protein

Regular expressions

- Defined positions
  - Fixed positions
  - Degenerate positions
- Un-defined positions
  - Fixed length
  - Flexiible length
- LIG_CYCLIN_1: [RK]xLx{0,1}[FYLIVMP]

  L: one type of amino acid “L”=Leucine

  [KR]: group of allowed types of amino acid in the given position

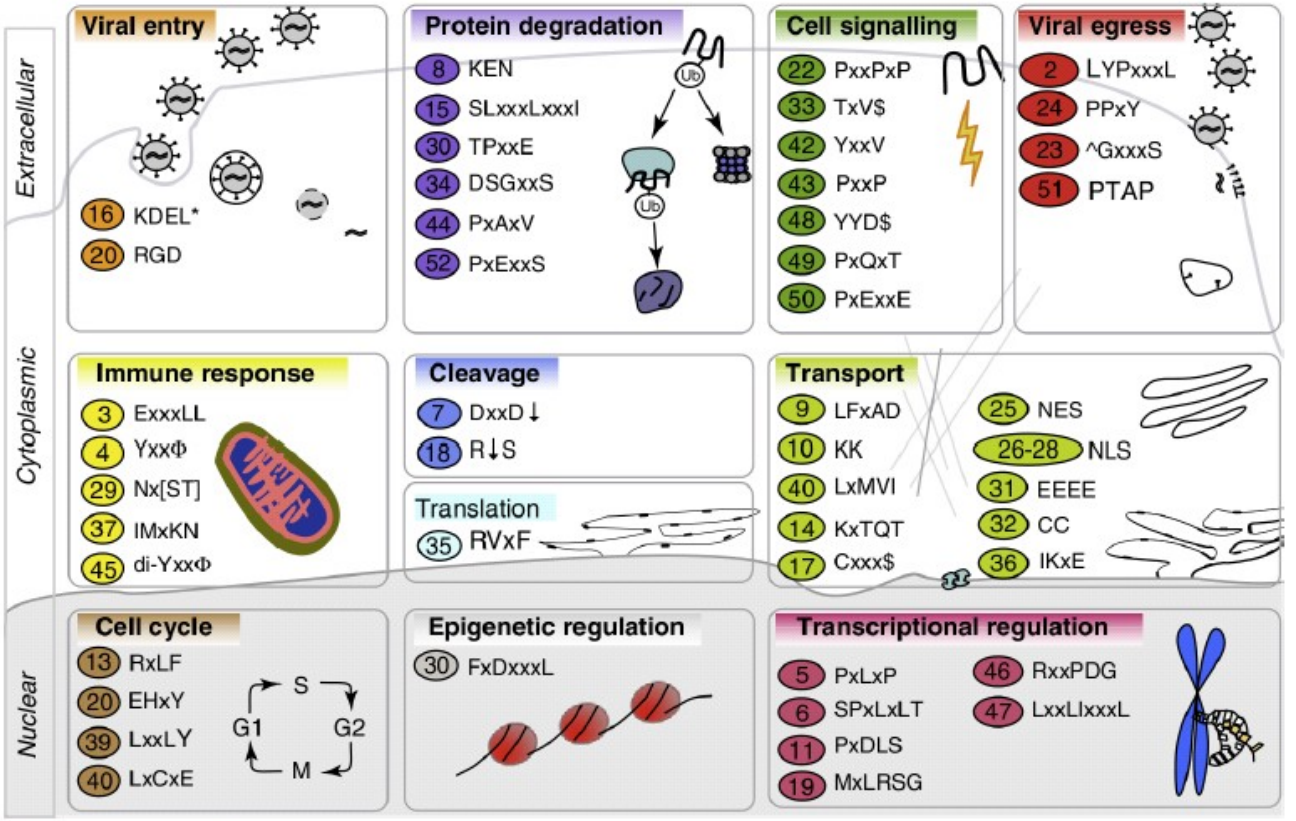
  x or . : anything (no restriction)

  {0,1}: variable length

Frequency of common linear motif binding domains

DOMAIN FREQUENCIES FROM PFAM (HUMAN PROTEOME)		
Domain Family	Frequency [Domains / Proteins]	Pattern of recognized motif
PDZ	573 / 342	[ST]x[ACVILF] −COOH
SH3	451 / 382	PxxP
SH2	237 / 219	pTxx[IV]
WW	151 / 103	PPxY
PTB	142 / 133	NPx p Y

Linear motifs in viruses



TIBS-817; No. of Pages 11 ARTICLE IN PRESS

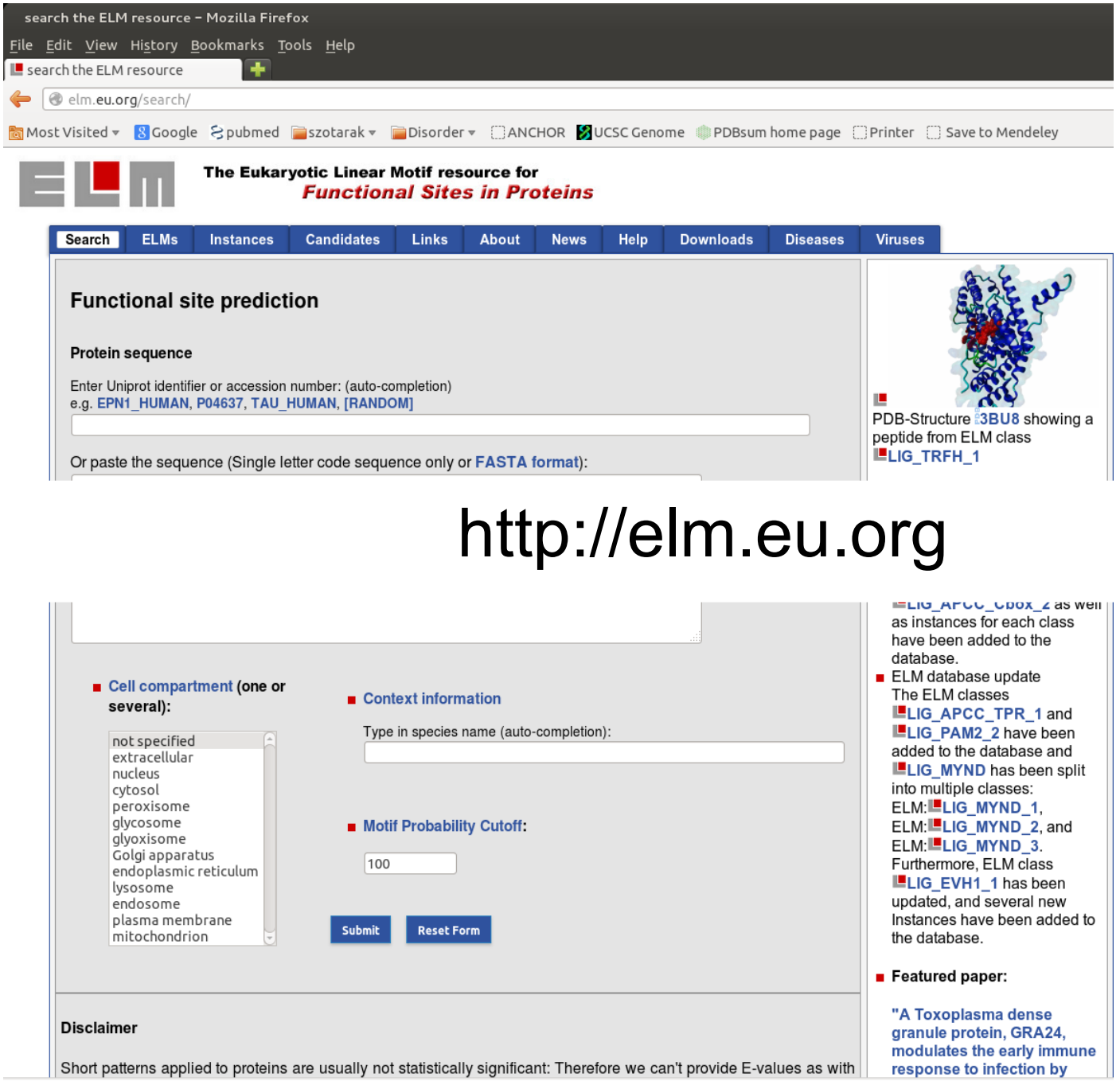
Review

Cell PRESS

How viruses hijack cell regulation

Norman E. Davey[1], Gilles Travé[2] and Toby J. Gibson[1]

[1] Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[2] Equipe Oncoproteines, FRE CNRS 3211, ESBS, 1, Bld Sébastien Brandt, BP10413, 67412 Illkirch, France


search the ELM resource – Mozilla Firefox
File Edit View History Bookmarks Tools Help
search the ELM resource
elm.eu.org/search/
Most Visited Google pubmed szotarak Disorder ANCHOR UCSC Genome PDBsum home page Printer Save to Mendeley
The Eukaryotic Linear Motif resource for
Functional Sites in Proteins
Search ELMs Instances Candidates Links About News Help Downloads Diseases Viruses
Functional site prediction
Protein sequence
Enter Uniprot identifier or accession number: (auto-completion)
e.g. EPN1_HUMAN, P04637, TAU_HUMAN, [RANDOM]
Or paste the sequence (Single letter code sequence only or FASTA format):
PDB-Structure 3BU8 showing a peptide from ELM class LIG_TRFH_1
http://elm.eu.org
Cell compartment (one or several):
not specified
extracellular
nucleus
cytosol
peroxisome
glycosome
glyoxisome
Golgi apparatus
endoplasmic reticulum
lysosome
endosome
plasma membrane
mitochondrion
Context information
Type in species name (auto-completion):
Motif Probability Cutoff:
100
Submit Reset Form
as instances for each class have been added to the database.
ELM database update
The ELM classes LIG_APCC_TPR_1 and LIG_PAM2_2 have been added to the database and LIG_MYND has been split into multiple classes: ELM:LIG_MYND_1, ELM:LIG_MYND_2, and ELM:LIG_MYND_3. Furthermore, ELM class LIG_EVH1_1 has been updated, and several new Instances have been added to the database.
Featured paper:
"A Toxoplasma dense granule protein, GRA24, modulates the early immune response to infection by
Disclaimer
Short patterns applied to proteins are usually not statistically significant: Therefore we can't provide E-values as with

The Eukaryotic Linear Motif resource for
Functional Sites in Proteins

ELM is a REPOSITORY 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

Class

Condensed information about a motif

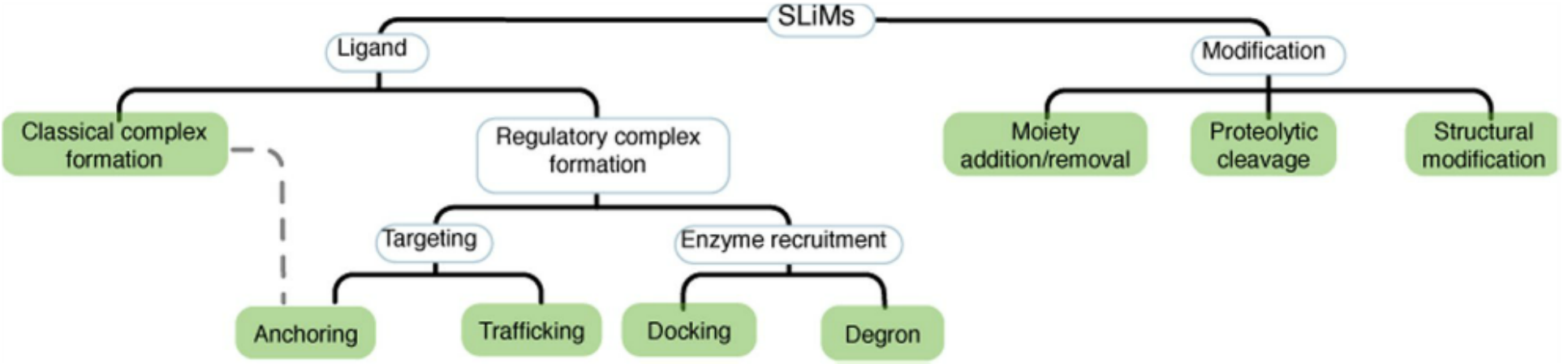
Regular expression is used to annotate the motif

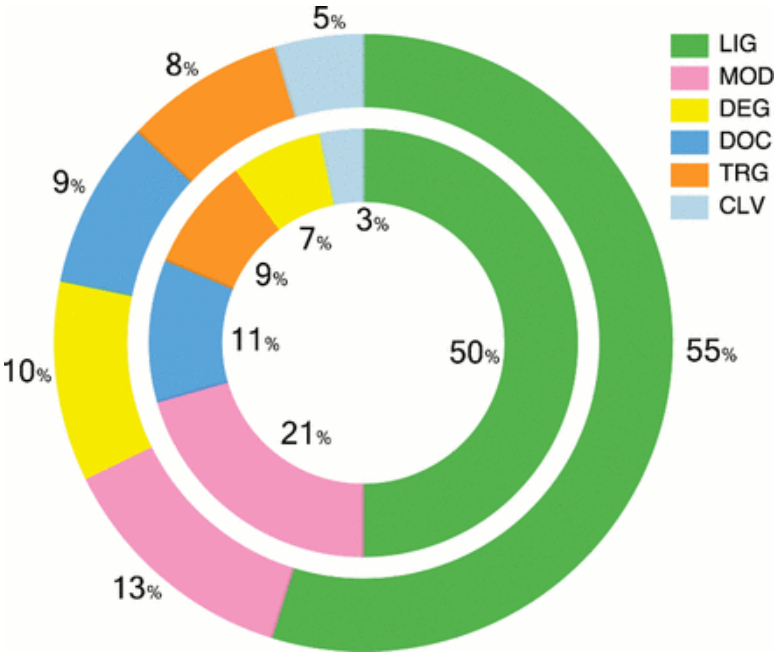
Instance

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

- Experimental Evidences
- Methods
- References
- Interactions

Types of linear motifs





Percentages of ELM classes (outer ring) and instances (inner ring) by type.

Known linear motifs can be used for predictions as well

Main problem – hits are dominated by random occurences

e.g the xLxxLLx motif can be foun in 40% of human proteins

```
>sp|P48552|NRIP1_HUMAN
MTHGEELGSDVHQDSIVLTYLEGLLMHQAA
GGSGTAVDKKSAGHNEEDQNFNISGSAFPT
CQSNGPVLNTHTYQGSGMLHLKKARLLQSS
EDWNAAKRKRLSDSIMNLNVKKEALLAGMV
DSVPKGKQDSTLLASLLQSFSSRLQTVALS
QQIRQSLKEQGYALSHDSLKVEKDLRCYGV
ASSHLKTLLKKSKVKDQKPDTNLPDVTKNL
IRDRFAESPHHVGQSGTKVMSEPLSCAARL…
```

N terminal regions of NRIP protein
(disordered)

```
>sp|Q8WZ42|TITIN_HUMAN Titin
MTTQAPTFTQPLQSVVVLEGSTATFEAHIS
GFPVPEVSWFRDGQVISTSTLPGVQISFSD
GRAKLTIPAVTKANSGRYSLKATNGSGQAT
STAELLVKAETAPPNFVQRLQSMTVRQGSQ
VRLQVRVTGIPTPVVKFYRDGAEIQSSLDF
QISQEGDLYSLLIAEAYPEDSGTYSVNATN
SVGRATSTAELLVQGEEEVPAKKTKTIVST
AQISESRQTRIEKKIEAHFDARSIATVEMV…
```

Ig domain of titin protein
(ordered)

Filtering of false positive hits

- Cellular localization
- Conservation
- Structure filter

Summary for sequence 'Q9Y6I3'.

KEY

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

GLOBPLOT: GlobDom | Disorder

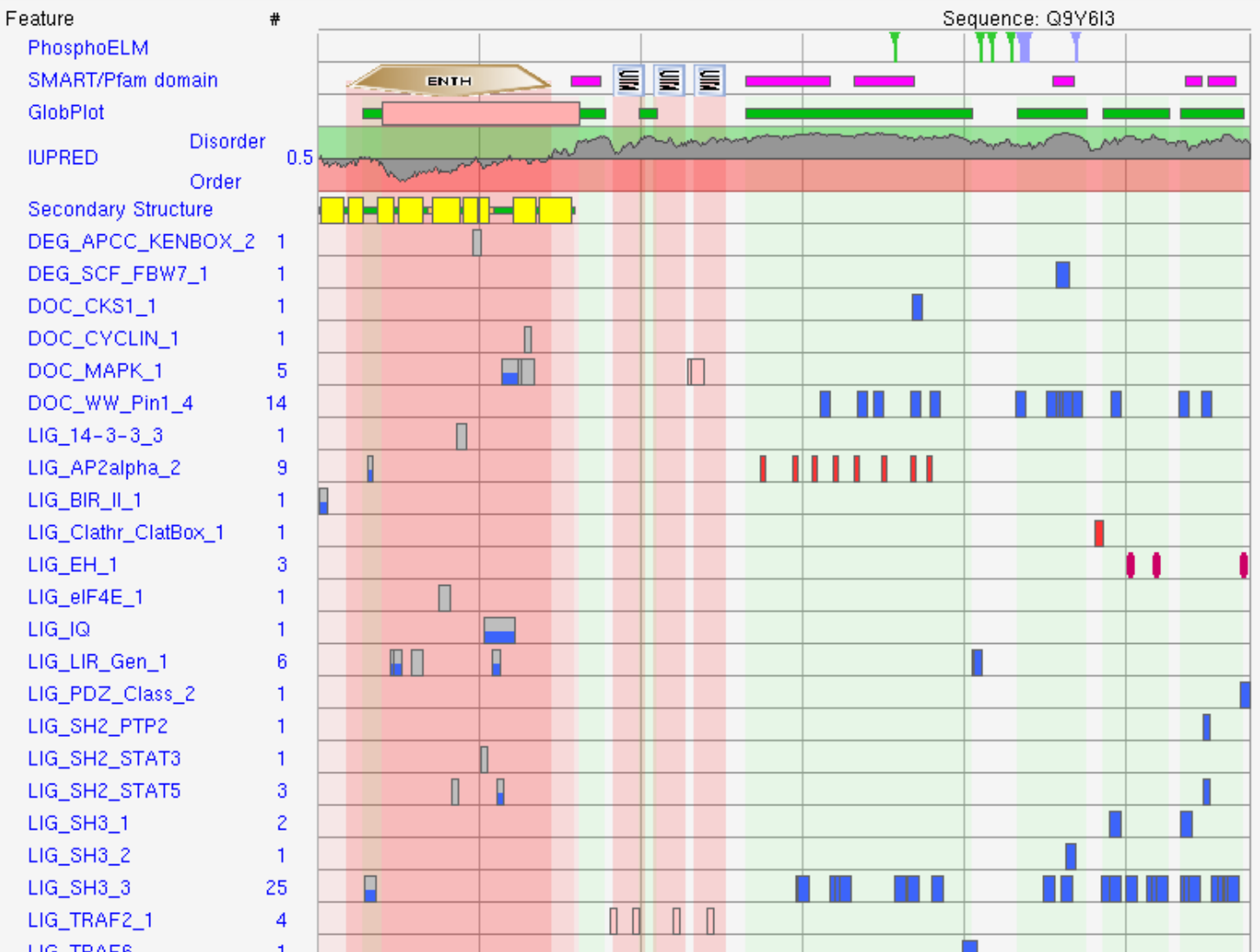
2D STRUCT: Strand | Helix | Loop | 3/10 Helix

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

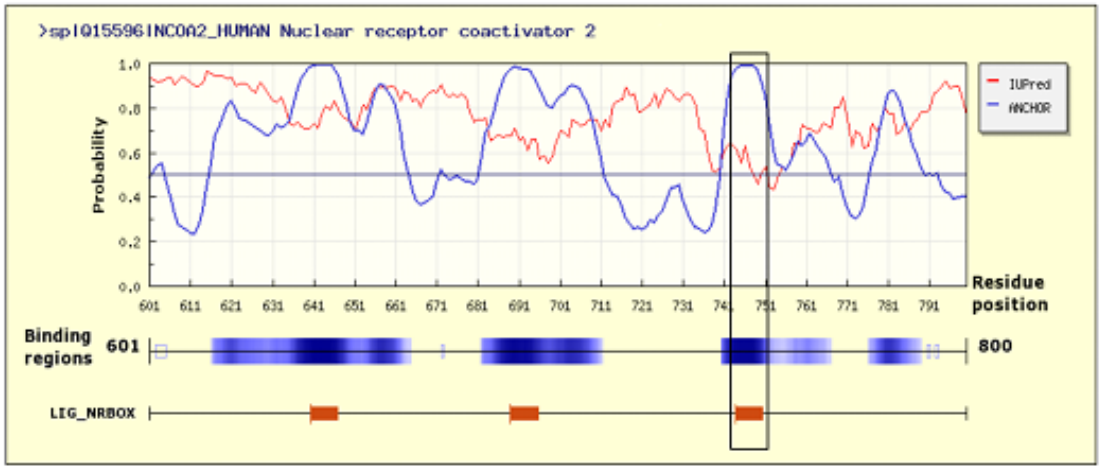
(Mouseover the matches for more details)

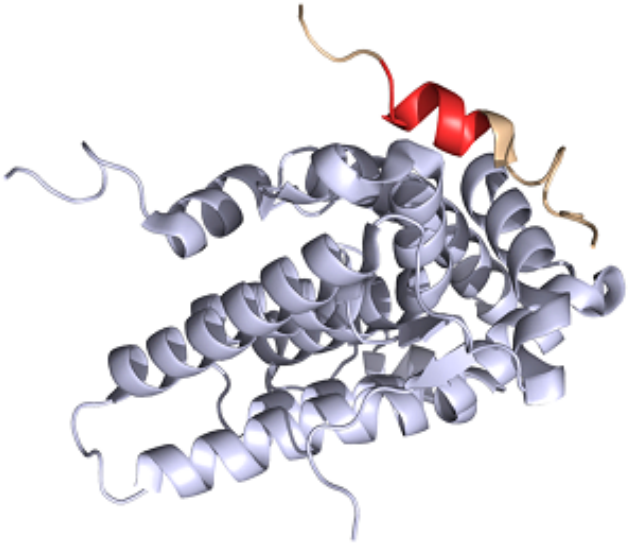


ANCHOR and linear motifs

NCOA2 transcription co-activator

A 600-800 region is completely disordered,
Contains 3 receptor linear motifs





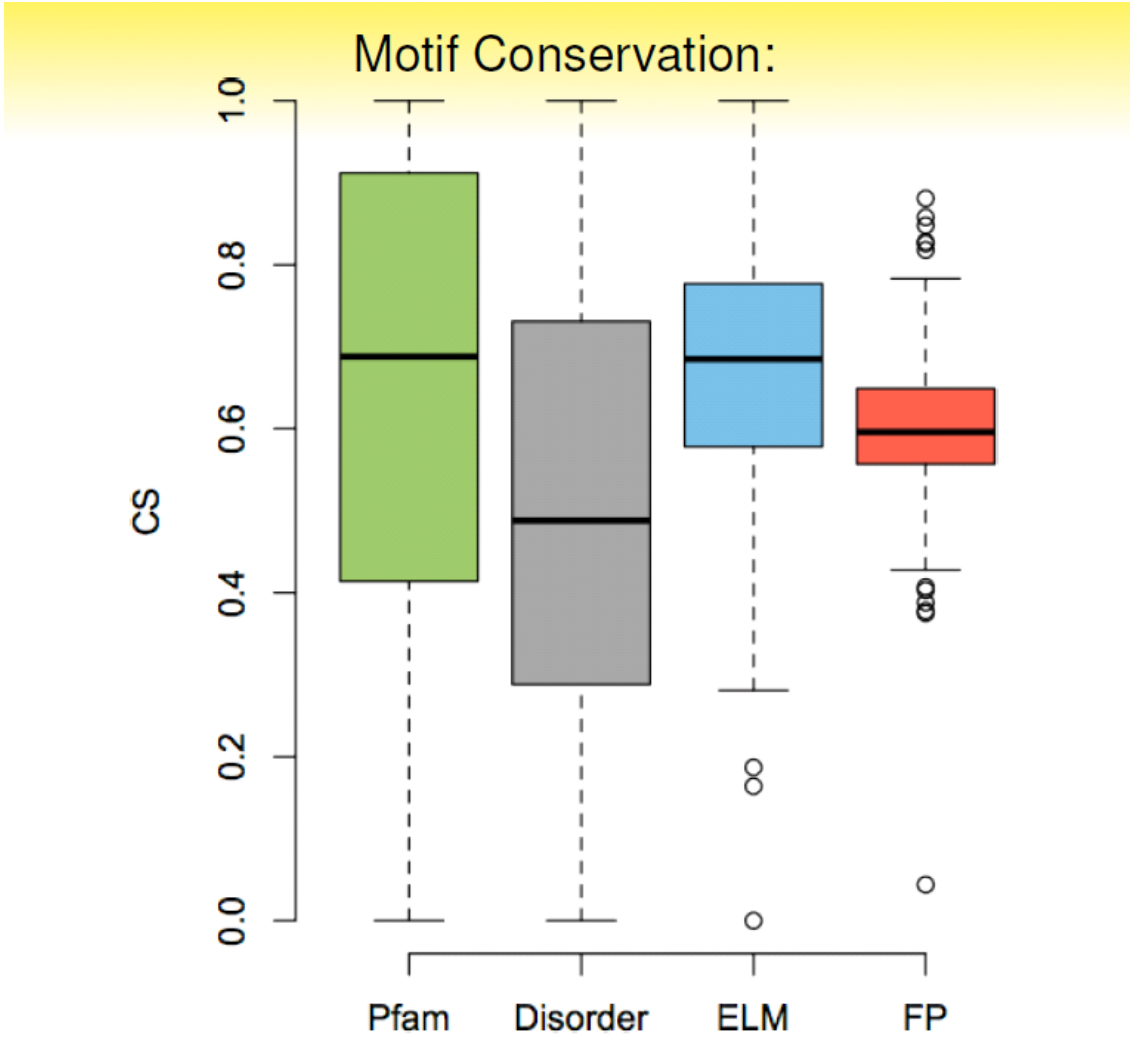
Conservation

The functionality of a protein segment is often approached by investigating the evolutionary history of its primary sequence

Can this approach used for disordered proteins?

Sometimes …

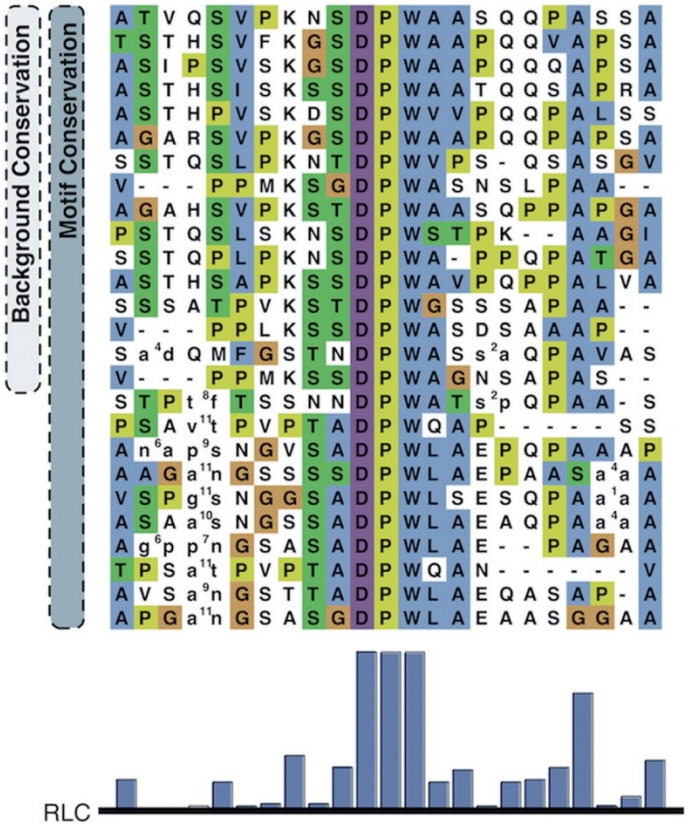
Conservation of motifs



Conservation patterns of linear motifs

No evolutionary constraints to keep the structure

Strong constraints on functional site



C

Island-like conservation

Davey et al. Nucleic Acids Res. 2012; 40:10628

SlimPrints

Generates sequence alignments of orthologous sequences

Relative conservation score per position

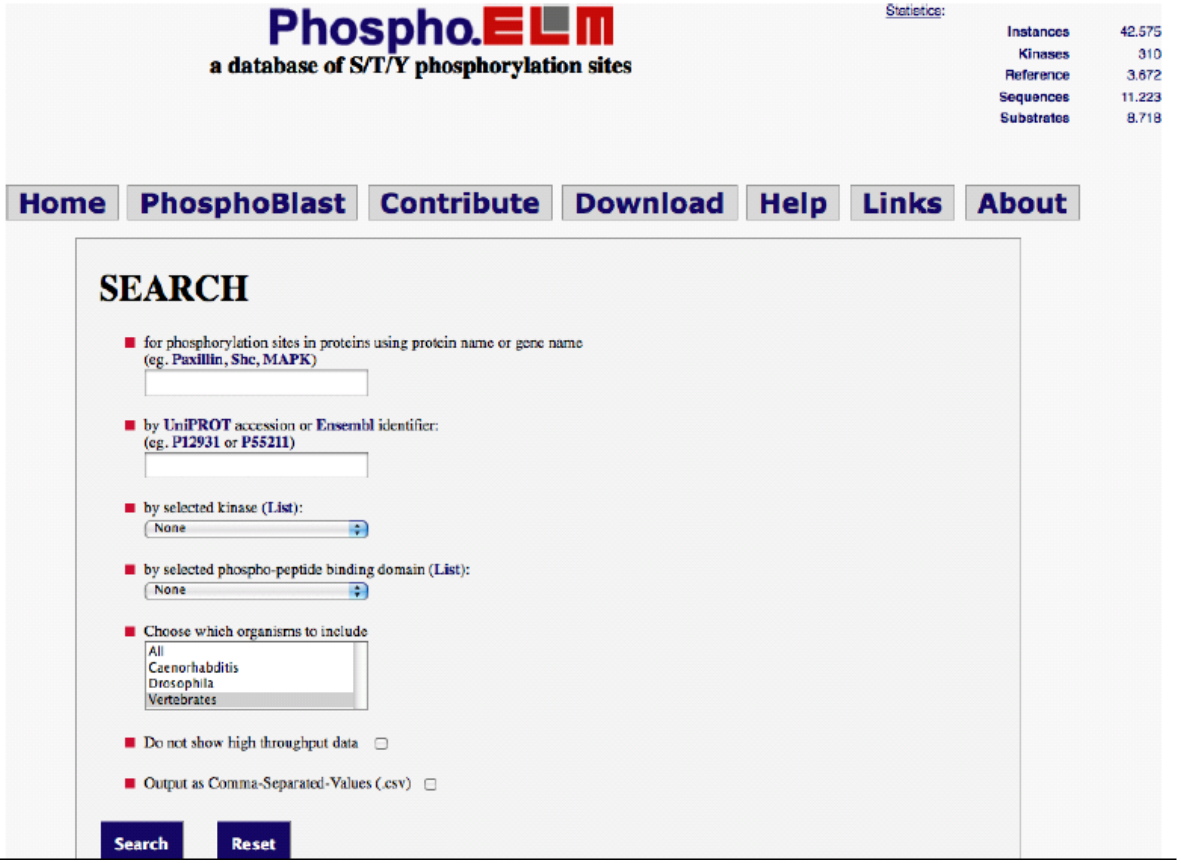
Filters out less reliable regions

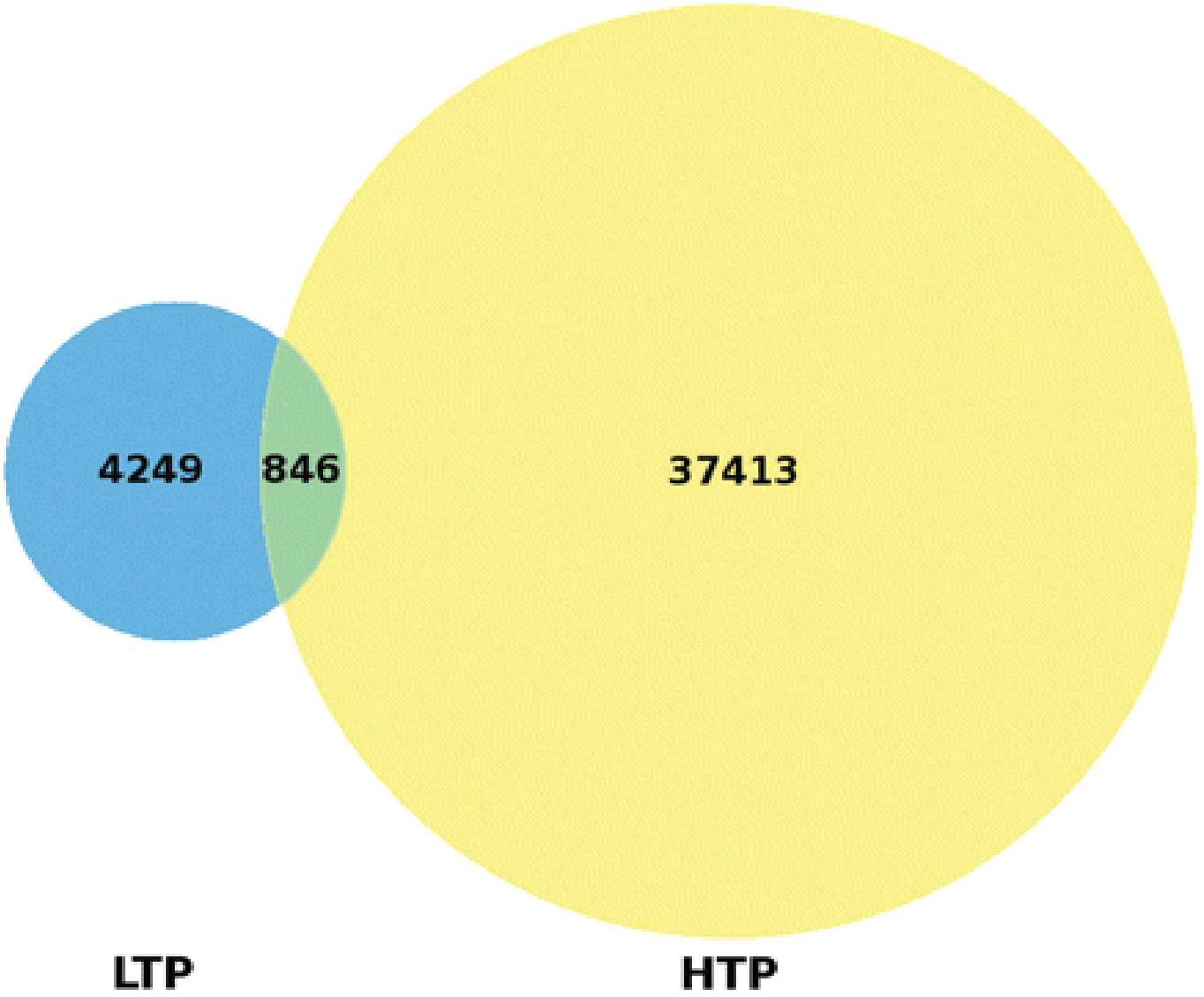
Fails if sequences are too divergent, or too similar

http://bioware.ucd.ie/slimprints.html.

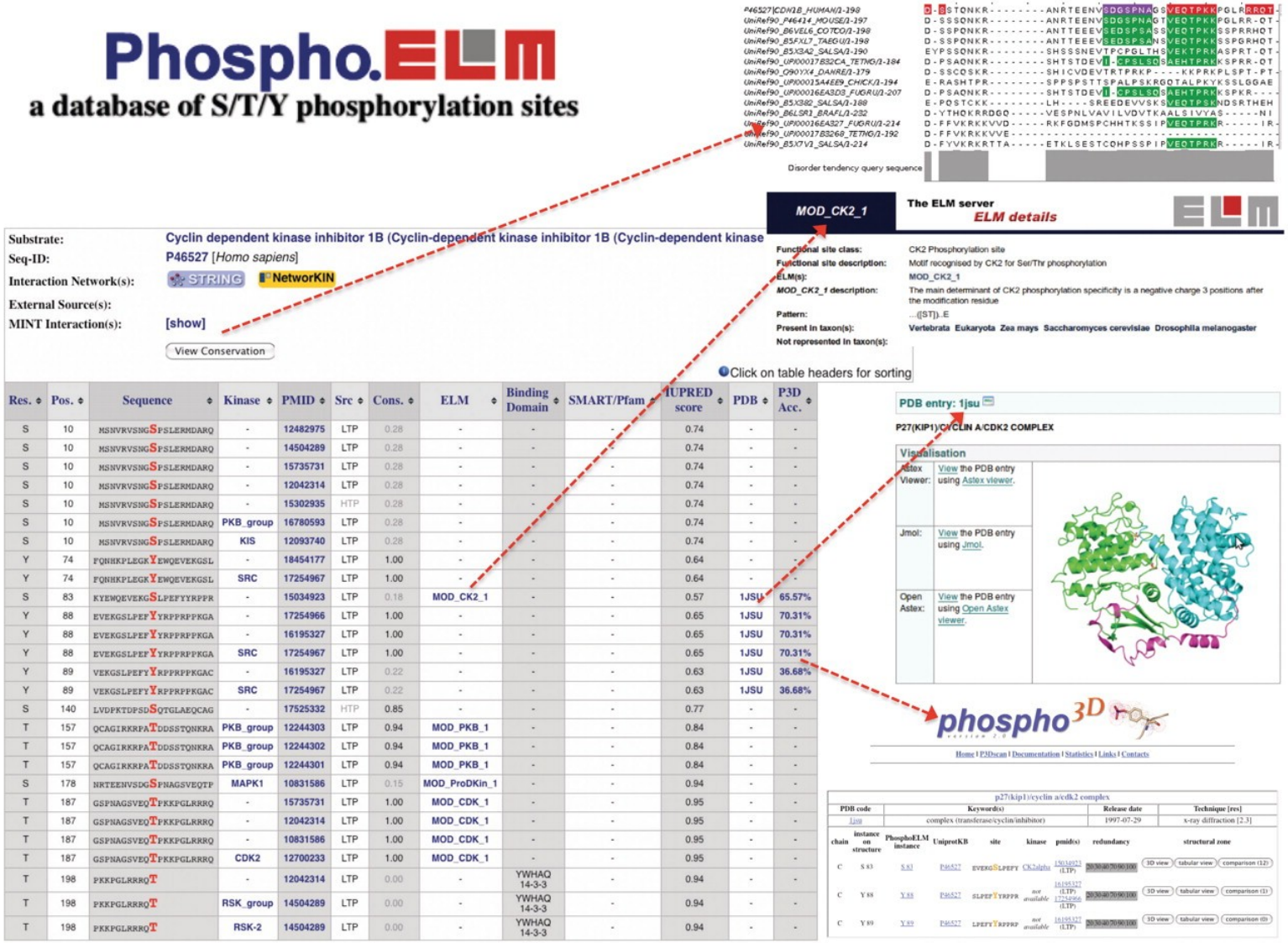
Phospho.ELM

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.)




4249
846
37413
LTP
HTP

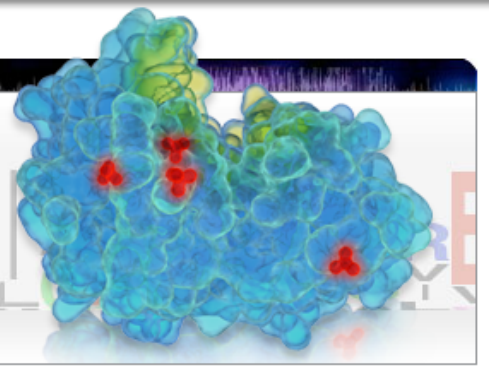
Output example of a Phospho.ELM search using the Cyclin dependent kinase inhibitor 1B (UniProt P46527) as query.



Res.	Pos.	Sequence	Kinase	PMID	Src	Cons.	ELM	Binding Domain	SMART/Pfam	IUPRED score	PDB	P3D Acc.
S	10	MSNVRVSNGSPSLERMDARQ	-	12482975	LTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	-	14504289	LTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	-	15735731	LTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	-	12042314	LTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	-	15302935	HTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	PKB_group	16780593	LTP	0.28	-	-	-	0.74	-	-
S	10	MSNVRVSNGSPSLERMDARQ	KIS	12093740	LTP	0.28	-	-	-	0.74	-	-
Y	74	FQNHKPLEGKYEWQEVEKGSL	-	18454177	LTP	1.00	-	-	-	0.64	-	-
Y	74	FQNHKPLEGKYEWQEVEKGSL	SRC	17254967	LTP	1.00	-	-	-	0.64	-	-
S	83	KYEWQEVEKGSLPEFYYRPFR	-	15034923	LTP	0.18	MOD_CK2_1	-	-	0.57	1JSU	65.57%
Y	88	EVEKGSLPEFYYRPFRPFKGA	-	17254966	LTP	1.00	-	-	-	0.65	1JSU	70.31%
Y	88	EVEKGSLPEFYYRPFRPFKGA	-	16195327	LTP	1.00	-	-	-	0.65	1JSU	70.31%
Y	88	EVEKGSLPEFYYRPFRPFKGA	SRC	17254967	LTP	1.00	-	-	-	0.65	1JSU	70.31%
Y	89	VEKGSLPEFYYRPFRPFKGAC	-	16195327	LTP	0.22	-	-	-	0.63	1JSU	36.68%
Y	89	VEKGSLPEFYYRPFRPFKGAC	SRC	17254967	LTP	0.22	-	-	-	0.63	1JSU	36.68%
S	140	LVDPKTDPSDSQTGLAEQCAG	-	17525332	HTP	0.85	-	-	-	0.77	-	-
T	157	QCAGIRKRFATDDSSTQNKRA	PKB_group	12244303	LTP	0.94	MOD_PKB_1	-	-	0.84	-	-
T	157	QCAGIRKRFATDDSSTQNKRA	PKB_group	12244302	LTP	0.94	MOD_PKB_1	-	-	0.84	-	-
T	157	QCAGIRKRFATDDSSTQNKRA	PKB_group	12244301	LTP	0.94	MOD_PKB_1	-	-	0.84	-	-
S	178	NRTEENVSDGSPNAGSVEQTP	MAPK1	10831586	LTP	0.15	MOD_ProDKin_1	-	-	0.94	-	-
T	187	GSPNAGSVEQTPKKPGLRRRQ	-	15735731	LTP	1.00	MOD_CDK_1	-	-	0.95	-	-
T	187	GSPNAGSVEQTPKKPGLRRRQ	-	12042314	LTP	1.00	MOD_CDK_1	-	-	0.95	-	-
T	187	GSPNAGSVEQTPKKPGLRRRQ	-	10831586	LTP	1.00	MOD_CDK_1	-	-	0.95	-	-
T	187	GSPNAGSVEQTPKKPGLRRRQ	CDK2	12700233	LTP	1.00	MOD_CDK_1	-	-	0.95	-	-
T	198	PKKPGLRRRQT	-	12042314	LTP	0.00	-	YWHAQ 14-3-3	-	0.94	-	-
T	198	PKKPGLRRRQT	RSK_group	14504289	LTP	0.00	-	YWHAQ 14-3-3	-	0.94	-	-
T	198	PKKPGLRRRQT	RSK-2	14504289	LTP	0.00	-	YWHAQ 14-3-3	-	0.94	-	-

Holger Dinkel et al. Nucl. Acids Res. 2010;nar.gkq1104

Nucleic Acids Research



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 PhosphoSite above for more information.

Please cite the following reference for this resource: Hornbeck PV, et al (2015) *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. Nucleic Acids Res. 43:D512-20. [reprint]

A PROTEIN MODIFICATION RESOURCE

PROTEIN OR SUBSTRATE SEARCH

Protein Name: SEARCH

ADVANCED SEARCH AND BROWSE OPTIONS

- Protein, Sequence, or Reference Search
- Site Search
- Comparative Site Search
- Browse MS2 Data By Disease
- Browse MS2 Data by Cell Line
- Browse MS2 Data by Tissue

DOWNLOADS, LINKS & APPLICATIONS

- Reprints, References, Supplemental Tables
- Downloadable Datasets
- Motif Analysis Tools

WHAT'S NEW

Dec 2014 Download **PhosphoSitePlus, 2014: mutations, PTMs and re-calibrations**. Nucleic Acids Res.(2015) 43:D512-20.

Aug 2014 **Download PTMVar 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**.

Dec 2011 **Download "PhosphoSitePlus: a comprehensive resource..."** in January 2012 issue of *Nucleic Acids Research*.

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

SITE STATISTICS

	TOTAL	NON-REDUNDANT
Proteins:	52,894	20,102
Sites, all types:	469,617	372,630
Low throughput (LTP) sites:	21,665	16,840
High throughput (HTP) MS sites:	459,259	364,806
MS peptides:	2,145,451	430,687
Number curated papers:	19,769	

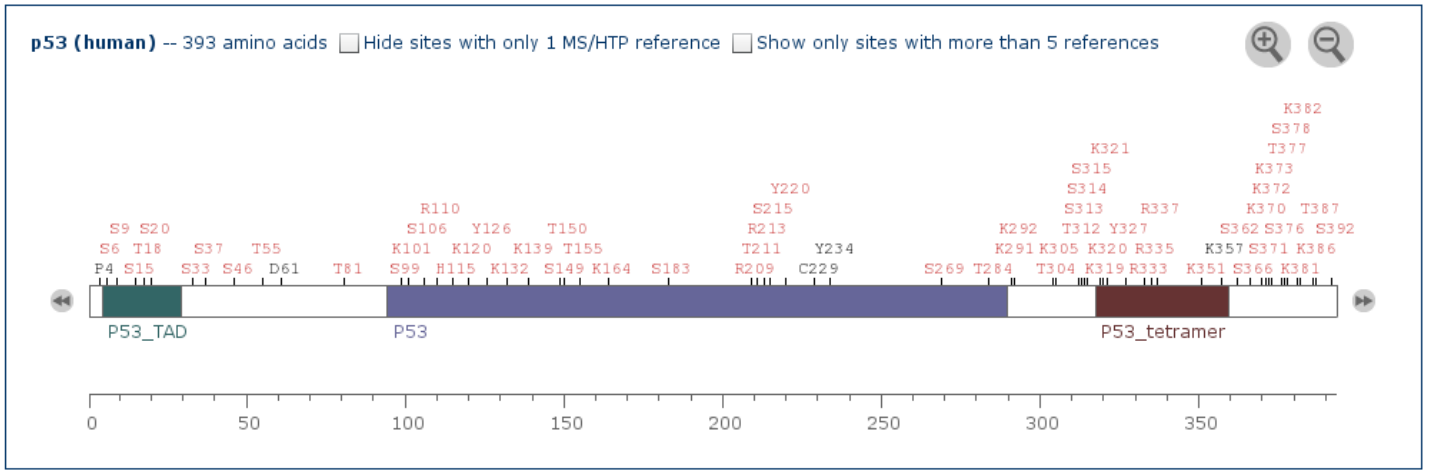
MODIFICATION SITE STATISTICS, NON-REDUNDANT:

Acetylation:	36,280	Caspase cleavage:	482
Di-Methylation:	2,583	Methylation:	192
Mono-Methylation:	4,999	O-Galnac:	2,118
O-Glcnac:	1,456	Phospho-Ser:	156,203
Phospho-Thr:	64,379	Phospho-Tyr:	42,164
Succinylation:	4,628	Sumoylation:	852
Tri-Methylation:	322	Ubiquitylation:	56,725

Modification for p53

Modification Sites and Domains — Show Modification Legend

Click here to view phosphorylation modifications only



Modification Sites in Parent Protein, Orthologs, and Isoforms — Show Modification Legend

Show Multiple Sequence Alignment

LTP	HTP	human		mouse (▼ Show Isoforms)		rat		rabbit		monkey	
6	0	P4	____MEEPQsDPsVE	S4-p	____MEEsQsDIsLE	S4-p	____MEDsQsDMsIE	S4	____MEESQSDLSLE	P4	____MEEPQSDPSIE
31	4	S6-p	__MEEPQsDPsVEPP	S6-p	__MEEsQsDIsLELP	S6-p	__MEDsQsDMsIELP	S6	__MEESQSDLSLEPP	S6	__MEEPQSDPSIEPP
34	3	S9-p	EEPQsDPsVEPPLsQ	S9-p	EEsQsDIsLELPLsQ	S9-p	EDsQsDMsIELPLsQ	S9	EESQSDLSLEPPLSQ	S9	EEPQSDPSIEPPLsQ
374	4	S15-p	PsVEPPLsQEtFsDL	S15-p	IsLELPLsQEtFsGL	S15-p	MsIELPLsQEtFsCL	S15	LSLEPPLSQETFSDL	S15-p	PSIEPPLsQETFSDL
30	0	T18-p	EPPLsQEtFsDLWKL	T18-p	ELPLsQEtFsGLWKL	T18-p	ELPLsQEtFsCLWKL	T18	EPPLSQETFSDLWKL	T18	EPPLsQETFSDLWKL
114	1	S20-p	PLsQEtFsDLWKLLP	S20-p	PLsQEtFsGLWKLLP	S20-p	PLsQEtFsCLWKLLP	S20	PLSQETFSDLWKLLP	S20	PLsQETFSDLWKLLP

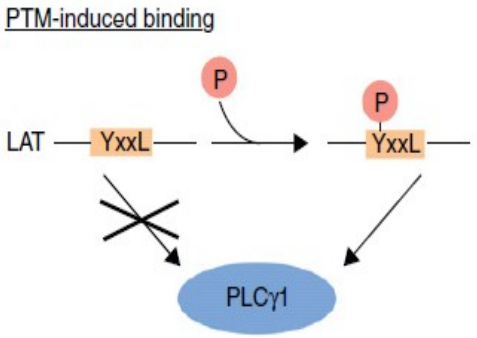
SLiMs

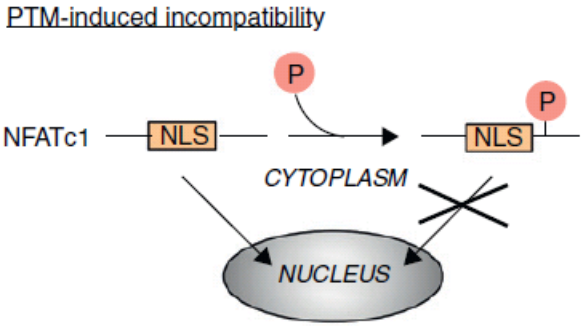
are compact, degenerate protein interaction interfaces (in IDRs)

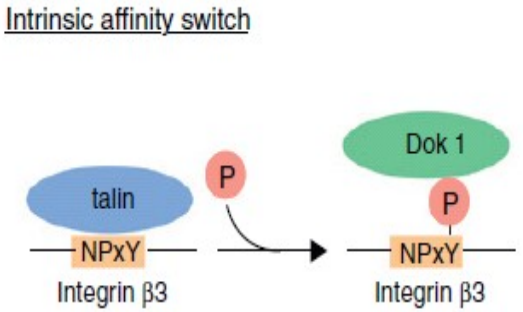
are ubiquitous in eukaryotic proteomes and mediate many regulatory functions:

- directing ligand binding
- providing docking sites for modifying enzymes
- controlling protein stability
- acting as signals to target proteins to specific subcellular locations

Molecular switches







Available online at www.sciencedirect.com
SciVerse ScienceDirect
Current Opinion in Structural Biology

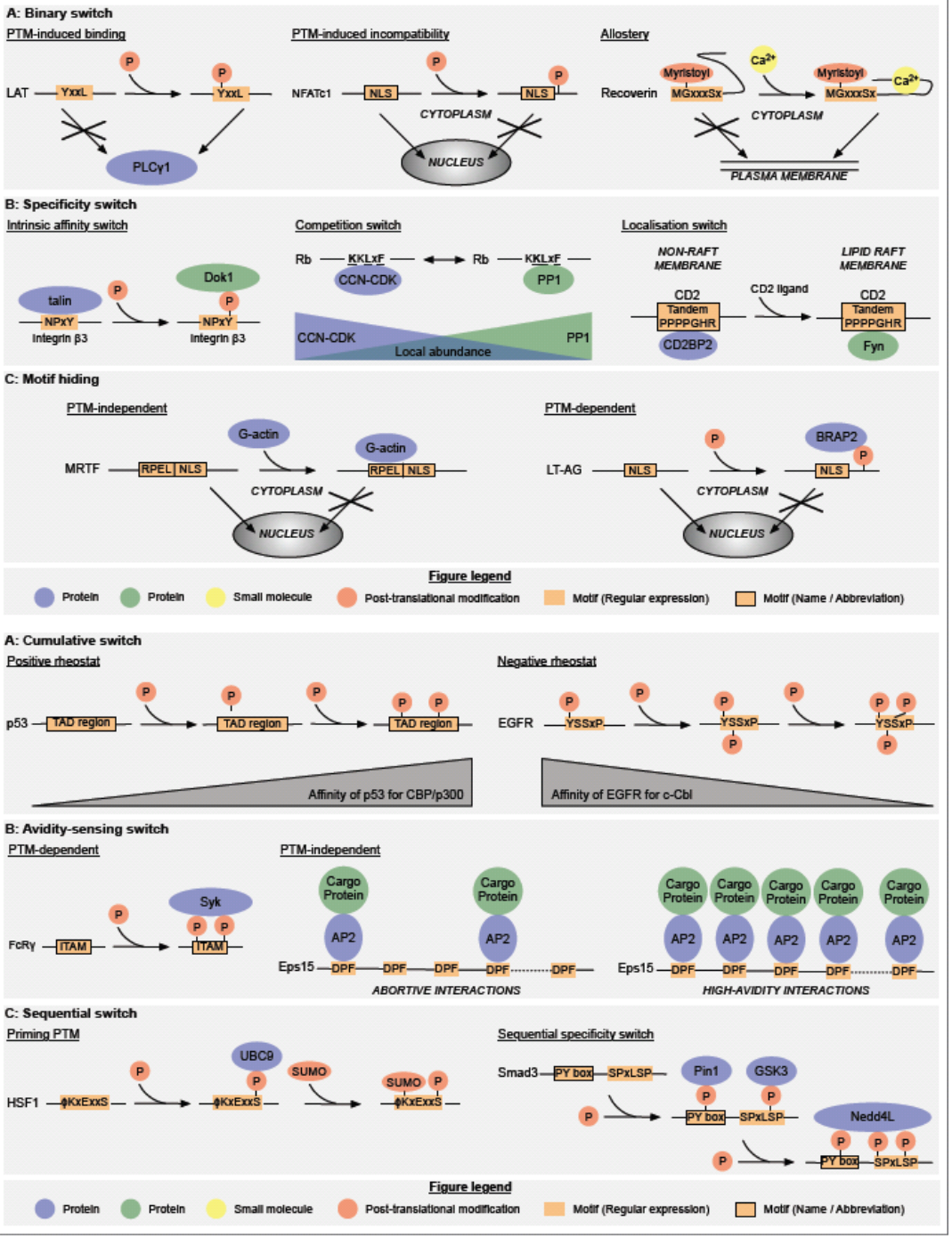
Switches.elm.eu.org

Motif switches: decision-making in cell regulation

Kim Van Roey[1], Toby J Gibson[1] and Norman E Davey[1,2]

Six classes of molecular switch involving IDP

- Binary Switch
  - Simple On-Off
- Specificity Switch
  - Multiple On states
- Motif-Hiding Switch
  - Conditional motif accessibility
- Cumulative Switch
  - Graduated rheostat-like behaviour
- Avidity sensing
  - Sharp, cooperative affinity shift
- Sequential Switch
  - Strict logical dependence of execution



http://switches.elm.eu.org