

ELM—the Eukaryotic Linear Motif resource—2024 update

Manjeet Kumar^{1,*}, Sushama Michael¹, Jesús Alvarado-Valverde^{1,2}, Andrés Zeke³,
Tamas Lazar^{4,5}, Juliana Glavina^{6,7}, Eszter Nagy-Kanta⁸, Juan Mac Donagh⁹,
Zsofia E. Kalman^{1,8}, Stefano Pascarelli^{10,11,12}, Nicolas Palopoli⁹, László Dobson^{1,13},
Carmen Florencia Suarez^{6,7}, Kim Van Roey¹⁴, Izabella Krystkowiak¹⁵, Juan Esteban Griffin⁹,
Anurag Nagpal¹⁶, Rajesh Bhardwaj^{17,18}, Francesca Diella¹, Bálint Mészáros¹⁹,
Kellie Dean²⁰, Norman E. Davey¹⁵, Rita Pancsa^{3,*}, Lucía B. Chemes^{6,7,*} and
Toby J. Gibson¹

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

²Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, Germany

³Institute of Enzymology, HUN-REN Research Centre for Natural Sciences, Budapest 1117, Hungary

⁴VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Pleinlaan 2, 1050 Brussels, Belgium

⁵Structural Biology Brussels, Department of Bioengineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

⁶Instituto de Investigaciones Biológicas, Universidad Nacional de San Martín (UNSAM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CP 1650, Buenos Aires, Argentina

⁷Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín, Av. 25 de Mayo y Francia, CP1650 San Martín, Buenos Aires, Argentina

⁸Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, Budapest 1083, Hungary

⁹Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Bernal, Buenos Aires, Argentina

¹⁰Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

¹¹Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

¹²Swiss Institute of Bioinformatics, Lausanne, Switzerland

¹³Department of Bioinformatics, Semmelweis University, Tűzoltó u. 7, Budapest 1094, Hungary

¹⁴Health Services Research, Sciensano, Brussels, Belgium

¹⁵Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Rd, Chelsea, London SW3 6JB, UK

¹⁶Department of Biological Sciences, BITS Pilani, K. K. Birla Goa campus, Zuarinagar, Goa 403726, India

¹⁷Inselspital, University of Bern, Freiburgstrasse 15, CH-3010 Bern, Switzerland

¹⁸Present Address: Signal Transduction Laboratory, National Institute of Environmental Health Sciences, 111 TW Alexander Drive, Durham, NC 27713, USA

¹⁹Department of Structural Biology and Center of Excellence for Data Driven Discovery, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

²⁰School of Biochemistry and Cell Biology, 3.91 Western Gateway Building, University College Cork, Cork, Ireland

*To whom correspondence should be addressed. Tel: +49 6221 387 8530; Email: manjeet.kumar@embl.de

Correspondence may also be addressed to Lucía B. Chemes. Tel: +54 11 40061500 (Ext 2133); Email: lchemes@iib.unsam.edu.ar

Correspondence may also be addressed to Rita Pancsa. Tel: +36 30 398 0557; Email: pancsa.rita@ttk.hu

Abstract

Short Linear Motifs (SLiMs) are the smallest structural and functional components of modular eukaryotic proteins. They are also the most abundant, especially when considering post-translational modifications. As well as being found throughout the cell as part of regulatory processes, SLiMs are extensively mimicked by intracellular pathogens. At the heart of the Eukaryotic Linear Motif (ELM) Resource is a representative (not comprehensive) database. The ELM entries are created by a growing community of skilled annotators and provide an introduction to linear motif functionality for biomedical researchers. The 2024 ELM update includes 346 novel motif instances in areas ranging from innate immunity to both protein and RNA degradation systems. In total, 39 classes of newly annotated motifs have been added, and another 17 existing entries have been updated in the database. The 2024 ELM release now includes 356 motif classes incorporating 4283 individual motif instances manually curated from 4274 scientific publications and including >700 links to experimentally determined 3D structures. In a recent development, the InterPro protein module resource now also includes ELM data. ELM is available at: <http://elm.eu.org>.

Received: September 15, 2023. Revised: October 22, 2023. Editorial Decision: October 23, 2023. Accepted: October 24, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

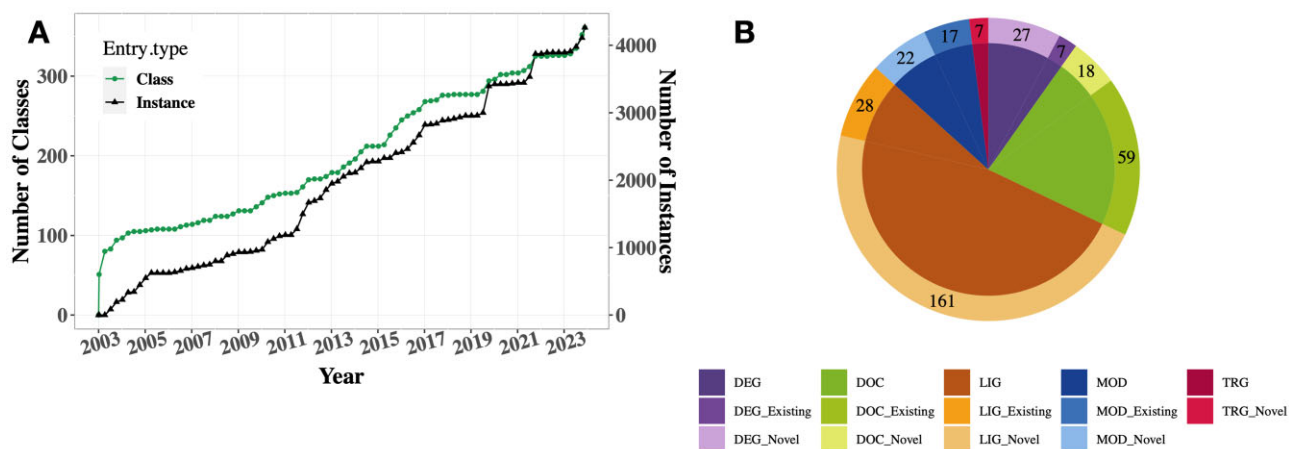


Figure 1. Data growth trends within the ELM resource since the last update paper (37). **(A)** This depicts the cumulative count of both motif classes and instances inserted into ELM in the last two decades. **(B)** The pie chart illustrates the allocation of new instances among both new and updated ELM classes (lighter and darker shades in the outer circle, respectively) belonging to 5 different functional categories. Note that the pie chart focuses only on the classes where new instances were added in this update. Visualisations were done using ggplot2 in RStudio (<http://www.rstudio.com/>).

Table 1. Concise summary of data captured in the ELM resource

Functional sites	ELM classes	ELM instances	GO terms	PDB structures	PubMed links	
Total	220	356	1018	718	4274	
By Category	LIG	201	Human	2459	Biological Process	567
	MOD	40	Mouse	404		
	DOC	43	Rat	174	Cellular Component	220
	DEG	33	Yeast	380		
	TRG	28	Viruses	290	Molecular Function	231
	CLV	11	Others	576		

docking (DOC), ligand (LIG), modification (MOD) and targeting (TRG). Within themselves, each of these categories comprises high-level functionally similar motif classes/entries.

The core of each motif entry is built from a detailed compilation of motif instances, which are rigorously extracted from existing scientific studies. In the ELM DB, these instances hold fine-grained details of annotated evidence, including methods used to characterise the motif as well as the curator's inference on the reliability of annotated data (38). The instances are also analysed from various aspects, such as sequence, structure, function, localisation, evolution and interaction context, to define the primary amino acid residues contributing to the specificity and binding strength of the motif-mediated protein complex. This systematic analysis by the ELM curators enables defining the motif sequence pattern in a standard POSIX regular expression format (detailed in Wikipedia (https://en.wikipedia.org/wiki/Regular_expression) and the ELM website for an in-depth view of their consensus definitions). A motif class also includes local sequence context and molecular interaction features from the core motif positions. Moreover, each class also assimilates critical insights into the biological role as well as the contextual basis for supporting the presence of a motif in the given protein sequence. Overall, every class distributes several key biological insights for the motif such as its functional attributes, cellular context, interacting domain knowledge, cellular locations, and curated knowledge from the scientific literature.

ELM curations are systematically organised in the PostgreSQL relational database (<http://www.postgresql.org/>) backend and can be updated as new knowledge or findings become available for any existing motif entry. The ELM

DB can be queried through the Django web framework (<https://djangoproject.com/>), and output is presented via its front-end interface. The ELM resource also provides access to a toolkit for motif exploration, allowing users to identify potential motifs in proteins of their interest. Once the query sequence matches with the known motif patterns in the ELM DB, they are displayed graphically on the protein sequence. The motif matches are overlaid with contextual features, which include accessibility and conservation information, among other functional insights.

The curated knowledge within the ELM repository can be accessed and downloaded for free in various formats. Detailed specifications on the available formats and datasets are available at <http://elm.eu.org/downloads.html>. Furthermore, a REST-API is available to facilitate automated queries of ELM class consensus matches against proteins of interest. The access details for utilising the motif search REST-API can be found at <http://elm.eu.org/api/manual.html>.

Data updates in ELM

Since its 2022 update (37), ELM curation has primarily focused on motifs pivotal to RNA and protein degradation, innate immunity, kinase biology, synapse systems, the cell cycle and pathogen-host interactions, among other motif-governed areas. With this update, ELM catalogues 356 motif classes (Figure 1; Table 1)—an addition of 39 since the last release (Table 2)—with a total of 4283 annotated motif instances, resulting in a net increase of 346 instances. Additionally, 17 existing motif classes and 111 instances have been updated

Table 2. New and revised ELM entries in the current ELM release

ELM class identifier	#Instances	ELM class (short) description
New ELM Classes		
DEG_Cend_DCAF12_1	6	C-terminal diGlu degrons recognised by the DCAF12 E3 ligase
DEG_Cend_FEM1AC_1	-	C-terminal degrons recognised by the FEM1A and FEM1C E3 ligases
DEG_Cend_FEM1B_2	2	C-terminal degrons recognised by the FEM1B E3 ligase
DEG_Cend_KLHDC2_1	2	C-terminal GG degrons targeted by the KLHDC2 E3 ligase
DEG_Cend_TRIM7_1	6	C-terminal Q degrons targeted by the TRIM7 E3 ligase
DEG_CRBN_cyclicCter_1	-	Degron recognised by the thalidomide-binding domain (TBD) of cereblon for ubiquitination and subsequent proteasomal degradation
DEG_Kelch_KLHL12_1	11	Proline-rich degron of the Wnt signalling pathway recognised by the E3 ligase KLHL12 with moderate affinity
DOC_CYCLIN_RevRxL_6	1	The hydrophobic patch (hp) of Cyclin A2 can bind an RxL-like motif in the reverse orientation
DOC_MIT_MIM_2	5	MIT- interacting motif (MIM2) present in CHMP6, Vfa1, Vps20 docks at MIT domain present in VPS4
DOC_PUB_PIM_1	2	The PIM motif binds to the PUB domain of proteins involved in the regulation of ubiquitination
DOC_RSK_DDVF_1	5	This short, partly acid [DE]-[DE]-V-F motif binds to the surface region of the RSK N-terminal kinase domain
DOC_TBK1_STING_1	5	A TBK1 kinase recruitment and activation docking motif conserved in STING
LIG_Arc_Nlobe_1	11	Binding motif for the N-lobe part of the C-terminal domain in Arc. The N-lobe domain has structural homology with HIV virus capsid, but the binding appears to be unique to higher vertebrates
LIG_ARS2_EDGEI_1	21	Several proteins functioning in RNA decay/processing/splicing bind a positively charged patch on the C-terminal leg domain (a znF) of ARS2 through their negatively charged EDGEI motif
LIG_CaM_Basic-1-5-8-14_1	4	Amphipathic α -helical calmodulin binding motif
LIG_CtBP_RRT_2	5	The RRT motif binds to a groove on the nucleotide binding domain of CtBP proteins functioning as NAD-dependent transcriptional corepressors
LIG_FERM_MyoX_1	4	Motif specifically recognised by the MyTH4-FERM domain of MyosinX that is important for cargo recognition
LIG_Integrin_collagen_1	7	A collagen-specific binding motif that is recognized by the I-domain of collagen binding integrins $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$, $\alpha 11\beta 1$
LIG_Integrin_RGDSP_6	6	A variant of the canonical RGD motif minimizing flanking interactions with the integrin, thus achieving weak selectivity and comparable binding strength over all RGD-binding integrins
LIG_IRF7_LxLS_2	1	A binding site for IRF-7 present in the protein itself, in various innate adaptor proteins, and in rotaviral NSP1 which triggers the innate immune responsive pathways
LIG_LEDGF_IBM_1	9	A bipartite motif recognised by the integrase binding domain of LEDGF/p75
LIG_Menin_MBM1_1	4	High affinity motif recognised by the palm region of Menin protein
LIG_MSH2_SHIPbox_1	10	The aromatic SHIP box motif is employed by several positive regulators of DNA mismatch repair (MMR) to interact with MSH2
LIG_MTR4_AIM_1	8	Diverse nuclear exosome adaptors employ the so-called arch-interacting motif (AIM) to recruit the MTR4-exosome complex to different RNA species for facilitating their efficient degradation
LIG_RBL1_LxSxE_2	1	The LxSxE motif is a suboptimal variant of the LxCxE motif found in the LIN52 protein that interacts with two members of the pocket protein family (p107 and p130). Binding requires phosphorylation of an adjacent residue creating a phosphoswitch
LIG_SH2_SFK_CTtail_3	4	Internal SH2 binding motif in SRC family kinase C-terminal tails
LIG_SH3_PxRPPK_7	6	This PxRPPK subtype of the RxxK motifs is typically employed by GRB2-associated-binding proteins (GABs) 1, 2 and 3, and by other proteins, such as THEMIS, for binding of the C-terminal SH3 domains of GRB2 or GADS
LIG_SH3_PxxPPRxxK_8	6	This HPK1 subtype of the RxxK motifs, where the RxxK is preceded by an upstream PxxP, is used by HPK1 and some other proteins mainly in T-cell receptor signalling to interact with the C-terminal SH3 domain of GADS or GRB2
LIG_SH3_PxxxRxxKP_6	16	The C-terminal SH3 domains of GADS and GRB2, and the SH3s of STAM1 and STAM2 have been described to bind this canonical RxxK motif
LIG_TRAF3_MATH_PxP_3	2	A motif that specifically binds the TRAF3 E3 ligase
LIG_TRAF4_MATH_1	3	A TRAF4 MATH domain binding motif present in some platelet receptors
LIG_TrF4_IWRxY_1	4	TRAMP complex subunits Air1/2 bind subunits Trf4/Trf5 through an extended interaction surface, involving the IWRxY motif
LIG_VCP_SHPBox_1	17	The SHP box motif is a VCP-binding ligand present in some adaptors that bind to the C-terminal NTD subdomain of VCP

Table 2. Continued

ELM class identifier	#Instances	ELM class (short) description
LIG_VCP_VBM_3	3	The VCP Binding Motif (VBM) binds to the N-terminal domain of VCP and is present in some of its cofactors.
LIG_VCP_VIM_2	9	VCP Interacting Motif (VIM) binds to the N-terminal domain of VCP and is present in various cofactors
MOD_AAK1BIKe_LxxQxTG_1	17	Motif that is phosphorylated by the endocytic kinases AAK1 and/or BIKe
MOD_LOK_YxT_1	5	The optimal phosphorylation site of the basophilic lymphocyte-oriented kinase (LOK) and SLK that mainly regulate cell shape and motility
TRG_NESrev_CRM1_2	2	Reverse NES binding the CRM1 groove in the minus direction. The spacing of the initial two hydrophobic residues $\Phi x \Phi$ dictates the reverse orientation
TRG_Oom_RxLR_1	5	Oomycete host targeting signal
Extensively Revised ELM Classes		
DEG_APCC_DBOX_1	18	An RxxL-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
DOC_PP2B_LxvP_1	50	Docking motif in calcineurin substrates that binds at the interface of the catalytic CNA and regulatory CNB subunits.
DOC_PP2B_PxIxIT_1	27	PxIxIT docking motif in calcineurin substrates that binds the catalytic CNA subunit
LIG_CaMK_CASK_1	7	Motif that mediates binding to the calmodulin-dependent protein kinase (CaMK) domain of the peripheral plasma membrane protein CASK/Lin2
LIG_IRFs_LxIS_1	9	A binding site for proteins IRF-3, IRF-5 and IRF-6 present in the protein themselves, in various innate adaptor proteins, and in rotaviral NSP1 which triggers the innate immune responsive pathways
LIG_Nrd1CID_NIM_1	5	The CTD-interacting domain (CID) of Nrd1 interacts with the NIM motif in Trf4 and some other proteins in ncRNA degradation
LIG_RB_LxCxE_1	45	The LxCxE motif is found in multiple host and viral interactors of the pocket protein family (Rb, p107 and p130)
LIG_RB_pABgroove_1	7	The LxDLFD motif binds in a deep groove between the A and B subdomains of the Retinoblastoma (Rb), p107 and p130 pocket domains
LIG_SH2_NCK_1	17	NCK Src Homology 2 (SH2) domain binding motif
LIG_SH2_SFK_2	20	Phosphotyrosine motifs bound by SH2 domains of Src family kinases (SFKs)
LIG_SH3_PxxDY_5	7	The PxxDY motif is recognized by some SH3 domains including in Nck and Eps8
LIG_TRAF2like_MATH_loPxQ_2	5	Long (or minor) TRAF2 binding motif. Members of the tumour necrosis factor receptor (TNFR) superfamily initiate intracellular signaling by recruiting the C-domain of the TNFR-associated factors (TRAFs) to their cytoplasmic tails
LIG_TRAF2like_MATH_shPxQ_1	13	Short (or major) TRAF2 binding motif. Members of the tumour necrosis factor receptor (TNFR) superfamily initiate intracellular signaling by recruiting the C-domain of the TNFR-associated factors (TRAFs) to their cytosolic tails
LIG_TRAF6_MATH_1	34	TRAF6 binding site. Members of the tumour necrosis factor receptor (TNFR) superfamily initiate intracellular signalling by recruiting the C-domains of the TNFR-associated factors (TRAFs) using motifs in their cytoplasmic tails
MOD_NEK2_1	2	Strict version of the motif targeted by NEK2 for phosphorylation
MOD_NEK2_2	-	The more tolerant version of the Nek2 phosphosite motif
MOD_PIKK_1	48	(ST)Q motif which is phosphorylated by PIKK family members

based on recent research findings (Table 2). Major advances in the updated motif classes were focused on improvements of the regular expressions defining the motif specificity and on the inclusion of structural and affinity data for each motif. ELM now maintains links to 4274 scientific publications. The growth in protein structure information in ELM has reached a new high and with this update 102 structures have been integrated, bringing the total cross-referenced structures with databases such as PDBe and RCSB-PDB to 718. Additionally, the ELM dataset now defines 2749 motif-partner interactions,

of which 688 have curated binding affinities. In-depth insights into ELM's data are provided in Table 1.

A selection of newly added motif biology insights covered in the current ELM update

SLiMs in RNA processing and decay

In eukaryotic cells, numerous RNA processing and modification events are required to produce RNAs in their final

functional forms. To remove defective, unstable or unwanted transcripts from the nucleus, RNAs must first be identified and then degraded by the nuclear exosome (39), a multi-component protein complex of RNA processing enzymes.

The RNA helicase Mtr4 is an important cofactor of the nuclear exosome (20,40,41). Mtr4 is not just a helicase but rather a multi-domain, modular hub protein (20) that contributes to the processing of ribosomal RNAs (42) and the targeting of the exosome to different degradable RNA species (40). Mtr4 is part of the TRAMP complex, together with the Trf4 poly(A)polymerase and the Air2 RNA-binding adaptor protein (40,43). Most of the hitherto discovered SLiMs associated with nuclear RNA surveillance are either mediating the inter-subunit interactions of the TRAMP complex or facilitating the recruitment of the complex by diverse nuclear exosome cofactors (Figure 2).

A highly conserved motif within the N-terminal tail of yeast Trf4 interacts with the helicase core of Mtr4 (see the `LIG_Mtr4_Trif4_1` motif class; (40)). This interaction is conserved in vertebrates, although with slightly different motif sequence preferences (see `LIG_Mtr4_Trif4_2`). Within the N-terminal tail of Air2 there are two distinct Mtr4-interacting motifs, one interacts with the helicase core (`LIG_Mtr4_Air2_1`) (40), while the other can bind to the KOW domain within the arch of Mtr4 (`LIG_Mtr4_AIM_1`) (40). This newly added, so-called arch-interacting motif (AIM; `LIG_Mtr4_AIM_1`) is particularly important, because many nuclear exosome cofactors, such as Nop53, Utp18 and Sqs1 also employ this motif for binding to and recruiting Mtr4 and the exosome to specific RNA transcripts to be degraded (41,44). Furthermore, this mode of interaction is also conserved in vertebrates (20,45). In the TRAMP complex, the Air proteins, Air1 or Air2, interact with Trf4 through a conserved motif within the linker between their fourth and fifth zinc knuckles (`LIG_Trif4_IWRxY_1`) plus their fifth zinc knuckle itself (46,47). Trf4 also contributes to the targeting of complexes for the degradation of incorrectly folded non-coding RNAs. The C-terminal Nrd1-interacting motif (NIM) of Trf4 binds to the NNS (Nrd1-Nab3-Sen1) complex subunit Nrd1 (48). This NIM motif class (`LIG_Nrd1CID_NIM_1`) has now been extensively revised, since the NNS subunit Sen1 (49) and exosome-associated RNA-binding protein Mpp6 (50) also turned out to employ NIM motifs for the binding of Nrd1.

Besides the TRAMP complex, additional RNA processing complexes have been characterised. In the yeast *S. pombe*, the MTREC (Mtl1-Red1 core) complex is responsible for directing the degradation of meiotic, non-coding and unspliced RNAs (51–53). Mtl1 (Mtr4-like protein 1) is a homolog of Mtr4. Red1 protein (54) interacts with Mtr1 and additional proteins in the MTREC complex, including zinc-finger containing protein Ars2 (52). Red1 contains an EDGEI motif (see newly annotated class `LIG_ARS2_EDGEI_1`) that is responsible for its interaction with Ars2.

In humans, there are a myriad of binding partners that interact with the human homolog of yeast Ars2, the Serrate RNA effector molecule homolog protein, through one or more EDGEI motifs. The binding of these partners facilitates different RNA processing activities (55–57). Interestingly, there are examples of EDGEI-mediated interactions even in *A. thaliana* and *D. melanogaster*; therefore they seem to be highly conserved across evolution.

Protein degradation

Protein quality control mechanisms are of quintessential importance to all living cells and their representation is becoming increasingly important for the ELM resource. Targeted degradation of proteins can remove unwanted, mis-translated polypeptides, improperly localised, fragmented or misfolded proteins, proteins from viral intruders, and many more (Figure 3) (58). Protein degradation within the eukaryotic cytoplasm and nucleus (and even of proteins translocated from the endoplasmic reticulum) is performed by the ubiquitin-proteasome system (59). Ubiquitination is orchestrated by at least three families of enzymes, E1 ubiquitin activating and E2 and E3 ubiquitin conjugating enzymes, which act sequentially, in distinct complexes (60) to add lysine 48 (K48)-linked polyubiquitin chains to the target protein. E3 ligases recognise their target proteins through molecular features called degrons. The majority of known degrons are SLiMs, and many of them are located at the protein amino (N-degrons) or carboxy (C-degrons) terminus (61). Most N-terminal degrons are well known (explaining the so-called ‘N-end-rule’ determining protein half-life in eukaryotic cells) (62). However, C-degron pathways that selectively degrade proteins depending on the amino acids preceding their C-termini, also referred to as DesCEND (Destruction via C-End Degron) pathways, have also been identified in recent years (58). The existence of C-terminal degrons explains why the distribution of C-terminal amino acids is unevenly biased in the proteome, as protein evolution actively selects against destabilising residues unless functionally required (63). So far, terminal Gly, Ala, Arg and Glu residues (‘destabilising residues’) are known to form parts of C-end degrons, being recognised by specific ubiquitin ligases (62,63).

C-degron SLiMs consist of a few (typically only 2) conserved residues. A C-terminal location and high structural accessibility (local disorder) are essential for C-degrons to trigger degradation (58). C-terminal degrons can sometimes be present at the native C-termini of proteins or can be introduced by alternative mechanisms including premature translation termination or post-translational modifications (64,65). Terminal degron signals can also initially be internal, being latently present within the protein sequence and becoming exposed following proteolytic cleavage. E3 ligases recognising C-end degrons can be single- or multi-subunit molecules, like cullin-RING ligases (CRLs). The list of ubiquitin ligases and recognition subunits known to be associated with DesCEND pathways is extensive: KLHDC1/2/3/10, FEM1A/B/C, DCAF12, APPBP2, TRPC4AP, TRIM7, PIRH2, etc. (63,64,66,67). Unfortunately, not all E3 ligase substrate preferences have been sufficiently explored. The current ELM update includes new motif class definitions for the C-end recognition motifs of the DCAF12, FEM1A, FEM1B, FEM1C, KLHDC2 and TRIM7 ubiquitin E3 ligases (Table 2). Below we briefly discuss the DCAF12 and KLHDC2 cases.

The cullin 4-RING (CRL4) ubiquitin ligases are highly modular and associate to various substrate recognition subunits (termed DCAFs) through their shared subunit DDB1. DCAF12 has been found to recruit proteins with a C-terminal di-Glu motif (Glu-Glu-COOH). Known substrates of CRL4^{DCAF12} Cullin-RING E3 ligase include CCT5, a member of the T-complex protein ring (TRiC) chaperonin complex (68), as well as the MAGE-A3 and MAGE-A6 melanoma-associated antigens, involved in the regulation of

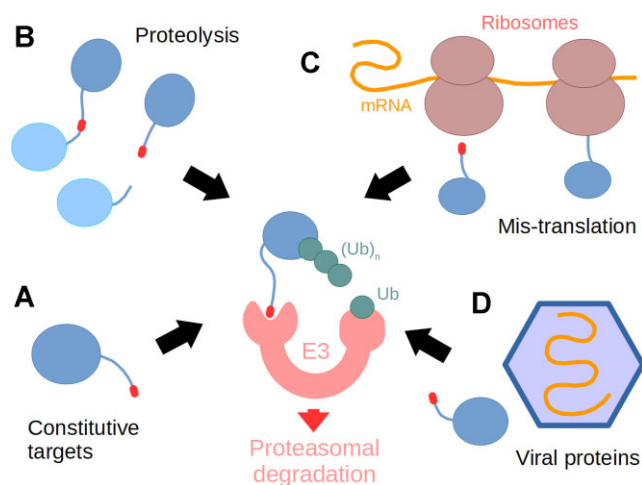


Figure 3. The diverse functions of C-terminal degrons. These motifs (red) can be part of native, endogenous protein C-termini (A), or be at an internal location and exposed by proteolysis (B). Degron SLiMs may also originate from erroneously translated mRNA sequences (C) or be present in virus-encoded proteins (D). These degrons are subsequently recognized by specific ubiquitin ligases (E3) and the proteins are polyubiquitylated, targeting them into the proteasome.

ubiquitylation and autophagy (69). In all these examples, the extreme C-terminus of the substrate protein is bound tightly at the center of the WD40 domain of DCAF12, by a positively charged coronet of lysine (K91, K108, K254 in PDB: 8AJM/8AJO) and arginine (R203, R256, R344 in PDB: 8AJM/8AJO) residues (68). However, other, yet-unknown recognition modes might also exist for DCAF12, explaining its interaction with IAPs helping cell survival (70) or MOV10 involved in multiple cellular roles (71).

The KLHDC1, KLHDC2, KLHDC3 and KLHDC10 proteins are closely related vertebrate cullin 2 (Cul2) or cullin 5 (Cul5) RING (CRL2/CRL5) ubiquitin ligase substrate recognition subunits, binding C-terminal degrons ending in glycine with distinct but overlapping specificities. Out of this group, KLHDC2 is the best characterised protein, associated with Cul2, responsible for the recognition of substrates ending in a di-Gly (Gly-Gly-COOH) motif. Structural studies reveal that the degron is bound to the center of the Kelch repeat domain, with the twin arginines (R236 and R241 in PDB: 6DO3/6DO4/6DO5) coordinating the C-terminal carboxylate (65). The size of this pocket is restricted to binding either GG\$ or (rarely seen) GA\$ motifs (72). The last 6 amino acids also contact the surface, but there is relatively less restriction imposed on them, except for a preference for small side chains.

Interestingly, Di-Gly degrons (or Gly-Ala degrons), substrates of CRL2^{KLHDC2} can be generated by at least three distinct physiological processes. Some rare proteins natively end in two glycines, such as EPHB2 (73). These motifs can also be generated by proteolysis, as they are compatible with the substrates processed by certain ubiquitin-specific proteases, including an autoprocessing site in USP1 (65). Several selenoproteins can also suffer premature termination in case of selenium starvation, precluding the physiological overriding of the stop codons by selenocysteinyl-tRNAs (74). Selenoprotein K (SelK) and Selenoprotein S (SelS), when terminated prematurely, will end in a di-Gly motif, that is targeted by KLHDC2 (as well as other, related subunits) for ubiquitylation and destruction (65,74). Finally, KLHDC2 has also been

shown to recognize its C-terminus (Gly-Ser-COOH) due to its resemblance to a degron, leading to the formation of inactive oligomers (73), adding a further layer of regulation to this already complicated system.

Endoplasmic reticulum (ER)-associated degradation (ERAD)

Similar to degron-mediated protein destruction, the endoplasmic reticulum (ER)-associated degradation (ERAD) pathway functions as a guardian of exported and vesicular protein quality. This pathway differentiates between folded and misfolded proteins inside the ER lumen (75) and shuttles problematic proteins into the cytosol through membrane channels in an energy-dependent manner. Therefore, malfunctioning of the ERAD pathway can lead to the accumulation of defective proteins, a condition implicated in >60 diseases that include neurological disorders, cancer, and cystic fibrosis (76). Moreover, many viruses, including several Nidovirales species such as coronaviruses, hijack the ERAD machinery during viral replication while simultaneously evading immune detection by the host. Viruses exploit ER membrane-derived structures, such as double-membrane vesicles (DMVs), to conceal viral RNA from cytosolic cellular sensors responsible for initiating interferon production and innate immunity responses (77,78).

Prior to their elimination, most ERAD substrates are tagged with ubiquitin chains, which then serve as docking sites for VCP (Valosin-Containing Protein, also known as P97/TERA). VCP is an abundant protein (among the top 5% of cellular proteins) of the AAA+-ATPase group. It regulates various cellular processes, ranging from protein quality control to supporting DNA damage response and membrane trafficking. Importantly, VCP mutations have been linked to diseases characterised by protein aggregation, including Multisystem Proteinopathy (MSP), Familial Amyotrophic Lateral Sclerosis (fALS), and Charcot-Marie-Tooth Disease Type2Y (CMT2Y) (79). Essential to VCP's interactions is its N-terminal domain (VCPN), which acts as a hub for binding various adaptor proteins. These interactions happen through three known motifs: VIM (VCP-Interacting Motif), VBM (VCP-Binding Motif), and the SHP box. These three motifs have been added as new ELM classes in the current update (Table 2). Notably, the VIM motif is hallmarked by a sequence module flanked on both sides with arginine preferences and a hydrophobic centre with conserved alanine residues. The VBM motif shows similarities with VIM in its arginine-rich amino acid content and interacts with the exact same VCP surface. In contrast, the SHP box forms a short, antiparallel β -strand, which augments the β -sheet of VCP's NTD subdomain at a site distinct from the two other VCP-binding motifs. VCP itself has the newly annotated Pub domain Interacting (PIM) motif that docks to the E3 ligase HOIL-1-interacting protein (HOIP), a part of the Linear ubiquitin chain assembly complex (LUBAC) which modulates the NF- κ B pathway (80).

Update to the Kinase substrate specificities in ELM

In the current ELM update, we have added two new classes that define the phosphorylation site specificity of two kinases. One class represents the substrates of two Ser/Thr directed kinases, Adaptor-Associated kinase 1 (AAK1) and BMP-2 Inducible Kinase (BIKe/BMP2K). The second class defines the

specificity of the Lymphocyte-oriented kinase (LOK), also named STK10 (Table 2).

AAK1 and BIKe/BMP2K are endocytic kinases that regulate diverse cellular functions, including receptor-mediated endocytosis, Notch pathway regulation, and dendrite morphogenesis. These kinases preferentially phosphorylate threonine residues and phosphorylate the QxTG motif present in the adaptor-related protein complex mu2 subunit (AP2M1) (81). Endocytic functions of these kinases also enable them to play a crucial role in controlling intracellular trafficking during the SARS-CoV-2 infection and entry process (82). Hepatitis C virus (HCV) and dengue virus (DENV) are other examples where these kinases are involved in viral entry and assembly/egress processes (83,84). The motif preference of the AAK1 and BIKe kinases is known from peptide SPOT arrays and other low-throughput in-vitro kinase assays (28,81,84–87) and indicates a strong preference for a Gly residue at the T + 1 position, Gln at the T – 2 position and Ile at the T – 5 position. Prk1p, the yeast homolog of AAK1/BIKe kinases, phosphorylates the Pan1p, Sla1p, and Scd5p substrates which all contain a similar phosphorylation motif (88–90).

LOK/STK10 regulates cytoskeletal dynamics in addition to affecting the cell shape and movements. An example of the latter includes its role in lymphocyte migration and microvilli formation in epithelial cells, which is achieved via phosphorylating the ERM proteins (Ezrin, Radixin and Moesin). Defining the specificity determinants of the LOK phosphosites is challenging due to the high sequence conservation of the C-terminal ERM protein regions, which contain the target phosphosites. Nevertheless, experiments from Belkina *et al.* (91) and SPOT arrays from Johnson JL *et al.* (28) reveal that LOK preferentially phosphorylates Thr over Ser residues. LOK sequence preferences around the phosphosite include Tyr at the T-2 position, large hydrophobic or aromatic amino acids at the T + 1 position and positively charged amino acids in the T – 3 and T + 2 positions flanking the phosphosite.

In addition to the two kinase classes with Thr site-specificity, a substantial update has been done to three classes describing the specificity for NEK (MOD_NEK2_1 and MOD_NEK2_2) and DNA damage signalling (MOD_PIKK_1) kinases. (Table 2). The ELM kinome update was assisted by the availability of a recent comprehensive assessment of the human Ser/Thr kinome using SPOT array technology (28), where the phosphosite preferences for 303 kinases have been profiled. This provides a unique opportunity to further improve kinase motif definitions in ELM by integrating these SPOT arrays with existing experimentally identified phosphosites.

Receptor tyrosine kinases (RTKs) regulate proliferation and actin cytoskeleton remodelling in response to extracellular signals. RTKs become autophosphorylated in tyrosine residues (pTyr) and relay the signal to non-receptor Src-family kinases (SFKs) (92) and adaptor proteins such as NCK1/2 (93), among others. This first step is achieved through binding of pTyr motifs to Src-homology 2 (SH2) domains present in SFKs and NCK1/2. The human proteome contains over 100 SH2 domains which bind defined ligands but present a significant degree of cross specificity (94,95). The present ELM update adds an extensive revision of the NCK (LIG_SH2_NCK_1) and SFK (LIG_SH2_SFK_2) specificity classes, as well as a new class describing the autoinhibitory interaction involving the C-tail of SFKs (LIG_SH2_SFK_CTail_3). The loss of the C-tail interaction leads to oncogenic activation in v-SRC. Viral

and bacterial pathogens also mimic SFK and NCK SH2 motifs to induce cell proliferation and actin cytoskeleton remodelling (96).

SLiMs in innate immunity

The human cell must be able to respond to many different types of pathogens and the tell-tale signs of infection that they provide (97). Within the cytosol of the cell, these can include both DNA (cGAS-STING) or dsRNA detectors (RIGI-MAVS) that control whether interferon regulatory factors (IRFs) can activate transcription (98,99). However, cells that are not already primed to expect pathogen invasion are most likely to succumb to the invader. Additionally, therefore, the cell surface is equipped with Tumour necrosis family receptors, Toll-like receptors, Interferon receptors and so forth so that it can be alerted to the presence of pathogens in the body and rapidly initiate defensive signalling internally, including transcriptional activation of critical defence genes such as interferons and cytokines by transcription factors such as NF- κ B (97,100).

The canonical NF- κ B activation pathway begins by an activated TNF superfamily receptor trimerising and recruiting a combination of TRAF family E3 ligases via TRAF-binding motifs such as PxQ variants in the cytosolic tails (Table 2). Unlike most E3 ligases, the Lysine-63 ubiquitin chains generated by the TRAFs are activating, rather than destructive signals. The pathway proceeds via IKK protein kinase activation and I κ B protein kinase degradation, ultimately releasing the NF- κ B RelA component for nuclear entry and gene activation. The non-canonical pathway involves TNFR recruitment of a partially distinct subset of TRAFs and ultimately releasing the NF- κ B component RelB into the nucleus for gene activation (101). These pathways involve communication between several multiprotein complexes but are typically shown as rather linear.

The RIGI-MAVS detection systems for cytosolic pathogen dsRNA are present on the mitochondrial membrane's outer surface (102). The cGAS-STING detection systems for cytosolic pathogen DNA are also present on the mitochondrial outer membrane (but also the endoplasmic reticulum) (99). Both of these systems control interferon regulatory factors, using the phosphorylated LxI (pS) SLiM (Table 2) to keep the IRFs in the cytosol as long as there are no pathogenic nucleic acid signals. Most of the IRF protein family have the same SLiM and, when phosphorylated by TBK1 or IKK ϵ at the SLiM itself, they homodimerise and escape the cytosol into the nucleus where they can bind and oligomerise the general transcription factors P300/CBP and together stimulate interferon gene transcription (103).

There are many complications to the innate immunity systems simplistically described above - for example, they connect to each other via some of their SLiM-mediated interactions. One of the strongest connections is through the TANK Binding Kinase (TBK1). Currently, the only biophysically identified SLiM binder to TBK1 is STING, which interacts via a TBK1 docking motif (DOC_TBK1_STING_1) (104). The acronym TANK is derived directly from NF- κ B regulation (TRAF family member-associated NF-kappa-B activator), but TANK and TBK1 are reported to be core components of an IRF3 activation complex (105). Figure 4 shows a network of twenty close interactors with TBK1: Multiple interactions unite the NF- κ B and IRF gene activation pathways.

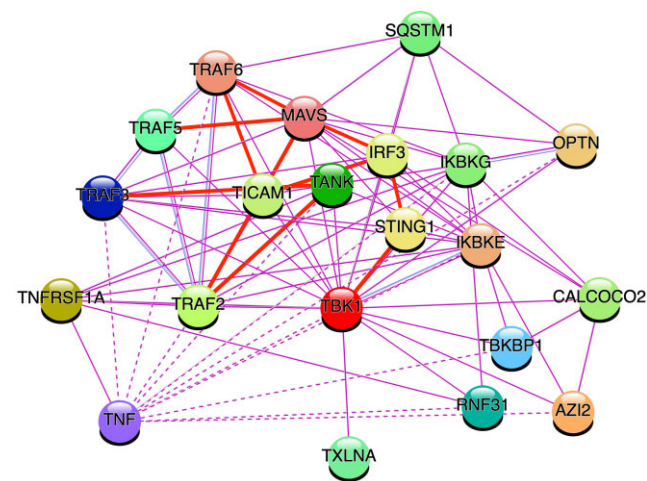


Figure 4. Protein interaction network centred on the key anti-pathogen and inflammatory kinase TBK1 generated by the STRING resource (122). The network shows how TBK1 plays a core role in bringing together the TRAF and IRF3 signalling systems. Edges with SLiMs annotated in ELM are shown by thick red lines. TNF (extracellular) interactions with cytosolic proteins are indirect (dotted lines). STRING settings for building the interaction network were as follows: only experimentally identified interactors of TBK1 and the 1st shell of 'no more than 20 interactors' in max number of interactors.

These closely connected anti-pathogen systems (98) illustrate why it is essential to think in terms of regulatory networks (where vertical pathways represent specific routes of information flow that can be controlled by any number of lateral interactions subject to other cell state conditions), as well as the fundamental importance of different SLiM classes in how network state is being modulated (2,106).

Impact of AlphaFold on SLiM bioinformatics

AlphaFold has proved to be an effective predictor of IDRs (107). ELM annotators now routinely use available structure predictions from the AlphaFold database (108,109) as they aid assessing SLiM instance plausibility, taking advantage of properties such as estimated disorderliness (a low pLDDT score) (109). AlphaFold prediction results have recently been incorporated into LeishMANIAdb (110) to more precisely score the probability of putative SLiMs in *Leishmania* species proteomes, that might interact with host proteins.

In addition there is increasing interest in using AlphaFold for *de novo* modelling of putative SLiM interactions with known or potential interaction domains. For example, in a recent C degron study, AlphaFold was used to model the interaction site of a POLD2 Proline C degron from FEM1B, predicting a distinct binding site to the better studied Arginine C-degron (72). In a benchmarking study of a protein-peptide interaction dataset, AlphaFold2-Multimer achieved only 40% success rate in modelling the correct binding site and interface by default protocol. However, the accuracy was enhanced with a combination of prediction modes, which included different protocols to incorporate the peptide alignment information while modelling the protein-peptide complex (111). In a second study, an ELM-derived dataset was used to test the accuracy of AlphaFold-Multimer in modelling SLiM interactions, finding that a fragment-based approach around the interacting SLiM region can enhance the performance of modelling SLiMs (112). We expect the application of AlphaFold

for SLiM prediction will be a growing area of bioinformatics research.

Conclusions and future perspectives

Research into the biology of SLiMs continues its steady expansion. SLiMs underlie much of the complexity in cell regulation (106,113,114). Additional classes of SLiMs are being revealed while the depth of understanding accrues for the more established SLiMs. The importance of SLiMs in disease and infection continues to be better appreciated (61,115–118). Despite these advances, we consider that the current set of experimentally defined SLiMs only represent a small fraction of the 'SLiMome' in eukaryotic proteomes (12,36). The adoption of more expansive experimental technologies can identify novel SLiMs and feed them into more directed experimentation allowing further refinement of the recognition patterns for incompletely defined motifs. In addition, artificial intelligence approaches such as AlphaFold hold out the tantalising promise of playing a pivotal role in *de novo* computational prediction of new SLiM candidates together with their interacting protein domains (119,120). This might all mean that resource providers and funders need to work together to further enhance existing tools, develop new ones and better integrate the SLiM resources with each other and with other bioinformatic and biomedical platforms in the way that, for example, InterPro has recently connected with ELM (121). Only then can the research community fully benefit from knowledge bases such as ELM and the enormous efforts that so many have contributed to their development.

Data availability

ELM is available at: <http://elm.eu.org>.

Acknowledgements

We express our gratitude to the ELM user community for valuing our efforts. Our appreciation extends to everyone who contributed in developing the previous versions of this resource. We are also thankful to our colleagues in the SLiM and IDP fields for their continuous collaboration on improving ELM.

Funding

T.L. is holder of a postdoctoral innovation mandate [HBC.2022.0194] by the Flanders Innovation & Entrepreneurship Agency (VLAIO); European Union's Horizon 2020 research and innovation programme provided funding under the Marie Skłodowska-Curie action [101028908 (SMILE) to L.D.]; N.E.D is supported by Cancer Research UK Senior Cancer Research Fellowship [C68484/A28159]; B.M. is financially supported by ALSAC; Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund [FK-142285 to R.P.]; R.P. is a holder of the János Bolyai Research Fellowship of the Hungarian Academy of Sciences [BO/00174/22]; L.B.C. and N.P. are National Research Council Investigators (CONICET, Argentina); L.B.C. has received funding from Grants Agencia Nacional de Promoción Científica y Tecnológica [PICT #2019-02119, #2021-001027]; J.G., N.P., C.F.S., L.B.C. and T.J.G. also received support from the European Union's

Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange (MSCA-RISE) programme IDPfun [778247]. Funding for open access charge: EMBL.

Conflict of interest statement

None declared.

References

- Pawson,T. and Nash,P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
- Gibson,T.J. (2009) Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, **34**, 471–482.
- Wright,P.E. and Dyson,H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
- Quaglia,F., Mészáros,B., Salladini,E., Hatos,A., Pancsa,R., Chemes,L.B., Pajkos,M., Lazar,T., Peña-Díaz,S., Santos,J., *et al.* (2022) DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.*, **50**, D480–D487.
- Bah,A. and Forman-Kay,J.D. (2016) Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.*, **291**, 6696–6705.
- Pajkos,M. and Dosztányi,Z. (2021) Functions of intrinsically disordered proteins through evolutionary lenses. *Prog. Mol. Biol. Transl. Sci.*, **183**, 45–74.
- Van Roey,K., Uyar,B., Weatheritt,R.J., Dinkel,H., Seiler,M., Budd,A., Gibson,T.J. and Davey,N.E. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.*, **114**, 6733–6778.
- Zanzoni,A., Ribeiro,D.M. and Brun,C. (2019) Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cell. Mol. Life Sci.*, **76**, 4407–4412.
- Kliche,J. and Ivarsson,Y. (2022) Orchestrating serine/threonine phosphorylation and elucidating downstream effects by short linear motifs. *Biochem. J.*, **479**, 1–22.
- Weikl,T.R. and Paul,F. (2014) Conformational selection in protein binding and function. *Protein Sci.*, **23**, 1508–1518.
- Lambrughi,M., Maiani,E., Aykac Fas,B., Shaw,G.S., Kragelund,B.B., Lindorff-Larsen,K., Teilum,K., Invernizzi,G. and Papaleo,E. (2021) Ubiquitin interacting motifs: duality between structured and disordered motifs. *Front. Mol. Biosci.*, **8**, 676235.
- Tompa,P., Davey,N.E., Gibson,T.J. and Babu,M.M. (2014) A million peptide motifs for the molecular biologist. *Mol. Cell*, **55**, 161–169.
- Hunt,T. (1990) Protein sequence motifs involved in recognition and targeting: a new series. *Trends Biochem. Sci.*, **15**, 305.
- Munro,S. and Pelham,H.R. (1987) A C-terminal signal prevents secretion of luminal ER proteins. *Cell*, **48**, 899–907.
- Dingwall,C. and Laskey,R.A. (1991) Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.*, **16**, 478–481.
- Kalderon,D., Roberts,B.L., Richardson,W.D. and Smith,A.E. (1984) A short amino acid sequence able to specify nuclear location. *Cell*, **39**, 499–509.
- Gould,S.J., Keller,G.A. and Subramani,S. (1988) Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. *J. Cell Biol.*, **107**, 897–905.
- Van Roey,K., Gibson,T.J. and Davey,N.E. (2012) Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.*, **22**, 378–385.
- Robinson,M.S. (2004) Adaptable adaptors for coated vesicles. *Trends Cell Biol.*, **14**, 167–174.
- Lingaraju,M., Johnsen,D., Schlundt,A., Langer,L.M., Basquin,J., Sattler,M., Heick Jensen,T., Falk,S. and Conti,E. (2019) The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs. *Nat. Commun.*, **10**, 3393.
- Giangreco,G., Malabarba,M.G. and Sigismund,S. (2021) Specialised endocytic proteins regulate diverse internalisation mechanisms and signalling outputs in physiology and cancer. *Biol. Cell*, **113**, 165–182.
- Bugge,K., Brakti,L., Fernandes,C.B., Dreier,J.E., Lundsgaard,J.E., Olsen,J.G., Skriver,K. and Kragelund,B.B. (2020) Interactions by disorder - a matter of context. *Front. Mol. Biosci.*, **7**, 110.
- Hossain,M., Bhalla,K. and Stillman,B. (2021) Multiple, short protein binding motifs in ORC1 and CDC6 control the initiation of DNA replication. *Mol. Cell*, **81**, 1951–1969.
- González-Foutel,N.S., Glavina,J., Borcherds,W.M., Safranchik,M., Barrera-Vilarmau,S., Sagar,A., Estaña,A., Barozet,A., Garrone,N.A., Fernandez-Ballester,G., *et al.* (2022) Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nat. Struct. Mol. Biol.*, **29**, 781–790.
- Case,L.B., Ditlev,J.A. and Rosen,M.K. (2019) Regulation of transmembrane signaling by phase separation. *Annu. Rev. Biophys.*, **48**, 465–494.
- Zeng,M., Chen,X., Guan,D., Xu,J., Wu,H., Tong,P. and Zhang,M. (2018) Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell*, **174**, 1172–1187.
- Leung,G.C., Murphy,J.M., Briant,D. and Sicheri,F. (2009) Characterization of kinase target phosphorylation consensus motifs using peptide SPOT arrays. *Methods Mol. Biol.*, **570**, 187–195.
- Johnson,J.L., Yaron,T.M., Huntsman,E.M., Kerelsky,A., Song,J., Regev,A., Lin,T.-Y., Liberatore,K., Cizin,D.M., Cohen,B.M., *et al.* (2023) An atlas of substrate specificities for the human serine/threonine kinome. *Nature*, **613**, 759–766.
- Münger,K., Werness,B.A., Dyson,N., Phelps,W.C., Harlow,E. and Howley,P.M. (1989) Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J.*, **8**, 4099–4105.
- Davey,N.E., Travé,G. and Gibson,T.J. (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.*, **36**, 159–169.
- Madhu,P., Davey,N.E. and Ivarsson,Y. (2022) How viral proteins bind short linear motifs and intrinsically disordered domains. *Essays Biochem.*, **66**, 935–944.
- Ali,M., Simonetti,L. and Ivarsson,Y. (2020) Screening intrinsically disordered regions for short linear binding motifs. *Methods Mol. Biol.*, **2141**, 529–552.
- Vincentelli,R., Luck,K., Poirson,J., Polanowska,J., Abdat,J., Blémont,M., Turchetto,J., Iv,F., Ricquier,K., Straub,M.-L., *et al.* (2015) Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods*, **12**, 787–793.
- Nguyen,H.Q., Roy,J., Harink,B., Damle,N.P., Latorraca,N.R., Baxter,B.C., Brower,K., Longwell,S.A., Kortemme,T., Thorn,K.S., *et al.* (2019) Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads. *eLife*, **8**, e40499.
- Bandyopadhyay,S., Bhaduri,S., Örd,M., Davey,N.E., Loog,M. and Pryciak,P.M. (2020) Comprehensive Analysis of G1 Cyclin Docking Motif Sequences that Control CDK Regulatory Potency In Vivo. *Curr. Biol.*, **30**, 4454–4466.
- Davey,N.E., Simonetti,L. and Ivarsson,Y. (2023) The next wave of interactomics: mapping the SLiM-based interactions of the intrinsically disordered proteome. *Curr. Opin. Struct. Biol.*, **80**, 102593.
- Kumar,M., Michael,S., Alvarado-Valverde,J., Mészáros,B., Sámano-Sánchez,H., Zeke,A., Dobson,L., Lazar,T., Örd,M., Nagpal,A., *et al.* (2022) The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res.*, **50**, D497–D508.
- Gouw,M., Alvarado-Valverde,J., Čalyševa,J., Diella,F., Kumar,M., Michael,S., Van Roey,K., Dinkel,H. and Gibson,T.J. (2020) How to annotate and submit a short linear motif to the eukaryotic linear motif resource. *Methods Mol. Biol.*, **2141**, 73–102.

39. Houseley, J., LaCava, J. and Tollervey, D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.
40. Falk, S., Weir, J.R., Hentschel, J., Reichelt, P., Bonneau, F. and Conti, E. (2014) The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol. Cell*, **55**, 856–867.
41. Thoms, M., Thomson, E., Baßler, J., Gnädig, M., Griesel, S. and Hurt, E. (2015) The exosome is recruited to RNA substrates through specific adaptor proteins. *Cell*, **162**, 1029–1038.
42. Vanáčová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G. and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.*, **3**, e189.
43. Losh, J.S., King, A.K., Bakelar, J., Taylor, L., Loomis, J., Rosenzweig, J.A., Johnson, S.J. and van Hoof, A. (2015) Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. *Nucleic Acids Res.*, **43**, 1848–1858.
44. Falk, S., Tants, J.-N., Basquin, J., Thoms, M., Hurt, E., Sattler, M. and Conti, E. (2017) Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA*, **23**, 1780–1787.
45. Wang, J., Chen, J., Wu, G., Zhang, H., Du, X., Chen, S., Zhang, L., Wang, K., Fan, J., Gao, S., et al. (2019) NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction. *Genes Dev.*, **33**, 536–549.
46. Hamill, S., Wolin, S.L. and Reinisch, K.M. (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15045–15050.
47. Fasken, M.B., Leung, S.W., Banerjee, A., Kodani, M.O., Chavez, R., Bowman, E.A., Purohit, M.K., Rubinson, M.E., Rubinson, E.H. and Corbett, A.H. (2011) Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.*, **286**, 37429–37445.
48. Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell*, **55**, 467–481.
49. Zhang, Y., Chun, Y., Buratowski, S. and Tong, L. (2019) Identification of three sequence motifs in the transcription termination factor Sen1 that mediate direct interactions with Nrd1. *Structure*, **27**, 1156–1161.
50. Kim, K., Heo, D.-H., Kim, I., Suh, J.-Y. and Kim, M. (2016) Exosome cofactors connect transcription termination to RNA processing by guiding terminated transcripts to the appropriate exonuclease within the nuclear exosome. *J. Biol. Chem.*, **291**, 13229–13242.
51. YAMAMOTO, M. (2010) The selective elimination of messenger RNA underlies the mitosis–meiosis switch in fission yeast. *Proc. Jpn. Acad., Ser. B*, **86**, 788–797.
52. Egan, E.D., Braun, C.R., Gygi, S.P. and Moazed, D. (2014) Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA*, **20**, 867–881.
53. Zhou, Y., Zhu, J., Schermann, G., Ohle, C., Bendrin, K., Sugioka-Sugiyama, R., Sugiyama, T. and Fischer, T. (2015) The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat. Commun.*, **6**, 7050.
54. Sugiyama, T. and Sugioka-Sugiyama, R. (2011) Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *EMBO J.*, **30**, 1027–1039.
55. Schulze, W.M., Stein, F., Rettel, M., Nanao, M. and Cusack, S. (2018) Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat. Commun.*, **9**, 1701.
56. Foucher, A.-E., Touat-Todeschini, L., Juarez-Martinez, A.B., Rakitch, A., Laroussi, H., Karczewski, C., Acajjaoui, S., Soler-López, M., Cusack, S., Mackereth, C.D., et al. (2022) Structural analysis of Red1 as a conserved scaffold of the RNA-targeting MTREC/PAXT complex. *Nat. Commun.*, **13**, 4969.
57. Rouvière, J.O., Salerno-Kochan, A., Lykke-Andersen, S., Garland, W., Dou, Y., Rathore, O., Molska, E.Š., Wu, G., Schmid, M., Bugai, A., et al. (2023) ARS2 instructs early transcription termination-coupled RNA decay by recruiting ZC3H4 to nascent transcripts. *Mol. Cell*, **83**, 2240–2257.
58. Yeh, C.-W., Huang, W.-C., Hsu, P.-H., Yeh, K.-H., Wang, L.-C., Hsu, P.W.-C., Lin, H.-C., Chen, Y.-N., Chen, S.-C., Yeang, C.-H., et al. (2021) The C-degron pathway eliminates mislocalized proteins and products of deubiquitinating enzymes. *EMBO J.*, **40**, e105846.
59. Varshavsky, A. (2019) N-degron and C-degron pathways of protein degradation. *P. Natl. Acad. Sci. U.S.A.*, **116**, 358–366.
60. Chen, X., Liao, S., Makarov, Y., Guo, Q., Zhu, Z., Krizelman, R., Dahan, K., Tu, X., Yao, X., Koren, I., et al. (2021) Molecular basis for arginine C-terminal degron recognition by Cul2FEM1 E3 ligase. *Nat. Chem. Biol.*, **17**, 254–262.
61. Mészáros, B., Kumar, M., Gibson, T.J., Uyar, B. and Dosztányi, Z. (2017) Degrons in cancer. *Sci. Signal.*, **10**, eaak9982.
62. Timms, R.T. and Koren, I. (2020) Tying up loose ends: the N-degron and C-degron pathways of protein degradation. *Biochem. Soc. Trans.*, **48**, 1557–1567.
63. Koren, I., Timms, R.T., Kula, T., Xu, Q., Li, M.Z. and Elledge, S.J. (2018) The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*, **173**, 1622–1635.
64. Sitron, C.S. and Brandman, O. (2019) CAT tails drive degradation of stalled polypeptides on and off the ribosome. *Nat. Struct. Mol. Biol.*, **26**, 450–459.
65. Rusnac, D.-V., Lin, H.-C., Canzani, D., Tien, K.X., Hinds, T.R., Tsue, A.F., Bush, M.F., Yen, H.-C.S. and Zheng, N. (2018) Recognition of the diglycine C-end degron by CRL2KLHDC2 ubiquitin ligase. *Mol. Cell*, **72**, 813–822.
66. Okumura, F., Fujiki, Y., Oki, N., Osaki, K., Nishikimi, A., Fukui, Y., Nakatsukasa, K. and Kamura, T. (2020) Cul5-type ubiquitin ligase KLHDC1 contributes to the elimination of truncated SELENOS produced by failed UGA/Sec decoding. *iScience*, **23**, 100970.
67. Ru, Y., Yan, X., Zhang, B., Song, L., Feng, Q., Ye, C., Zhou, Z., Yang, Z., Li, Y., Zhang, Z., et al. (2022) C-terminal glutamine acts as a C-degron targeted by E3 ubiquitin ligase TRIM7. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2203218119.
68. Pla-Prats, C., Cavadini, S., Kempf, G. and Thomä, N.H. (2023) Recognition of the CCT5 di-Glu degron by CRL4DCAF12 is dependent on TRiC assembly. *EMBO J.*, **42**, e112253.
69. Ravichandran, R., Kodali, K., Peng, J. and Potts, P.R. (2019) Regulation of MAGE-A3/6 by the CRL4-DCAF12 ubiquitin ligase and nutrient availability. *EMBO Rep.*, **20**, e47352.
70. Jiao, D., Chen, Y., Wang, Y., Sun, H., Shi, Q., Zhang, L., Zhao, X., Liu, Y., He, H., Lv, Z., et al. (2022) DCAF12 promotes apoptosis and inhibits NF- κ B activation by acting as an endogenous antagonist of IAPs. *Oncogene*, **41**, 3000–3010.
71. Lidak, T., Baloghova, N., Korinek, V., Sedlacek, R., Balounova, J., Kasperek, P. and Cermak, J. (2021) CRL4-DCAF12 ubiquitin ligase controls MOV10 RNA helicase during spermatogenesis and T cell activation. *Int. J. Mol. Sci.*, **22**, 5394.
72. Timms, R.T., Mena, E.L., Leng, Y., Li, M.Z., Tchasovnikarova, I.A., Koren, I. and Elledge, S.J. (2023) Defining E3 ligase-substrate relationships through multiplex CRISPR screening. *Nat. Cell Biol.*, **25**, 1535–1545.
73. Scott, D.C., King, M.T., Baek, K., Gee, C.T., Kalathur, R., Li, J., Purser, N., Nourse, A., Chai, S.C., Vaithiyalingam, S., et al. (2023) E3 ligase autoinhibition by C-degron mimicry maintains C-degron substrate fidelity. *Mol. Cell*, **83**, 770–786.
74. Lin, H.-C., Ho, S.-C., Chen, Y.-Y., Khoo, K.-H., Hsu, P.-H. and Yen, H.-C.S. (2015) SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/Sec decoding. *Science*, **349**, 91–95.

75. Hwang, J. and Qi, L. (2018) Quality control in the endoplasmic reticulum: crosstalk between ERAD and UPR pathways. *Trends Biochem. Sci.*, **43**, 593–605.
76. Guerriero, C.J. and Brodsky, J.L. (2012) The delicate balance between secreted protein folding and endoplasmic reticulum-associated degradation in human physiology. *Physiol. Rev.*, **92**, 537–576.
77. Zhang, J., Lan, Y. and Sanyal, S. (2020) Membrane heist: coronavirus host membrane remodeling during replication. *Biochimie*, **179**, 229–236.
78. Noack, J., Bernasconi, R. and Molinari, M. (2014) How viruses hijack the ERAD tuning machinery. *J. Virol.*, **88**, 10272–10275.
79. Ye, Y., Tang, W.K., Zhang, T. and Xia, D. (2017) A Mighty “Protein Extractor” of the Cell: structure and Function of the p97/CDC48 ATPase. *Front. Mol. Biosci.*, **4**, 39.
80. Schaeffer, V., Akutsu, M., Olma, M.H., Gomes, L.C., Kawasaki, M. and Dikic, I. (2014) Binding of OTULIN to the PUB domain of HOIP controls NF- κ B signaling. *Mol. Cell*, **54**, 349–361.
81. Conner, S.D. and Schmid, S.L. (2002) Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J. Cell Biol.*, **156**, 921–929.
82. Karim, M., Saul, S., Ghita, L., Sahoo, M.K., Ye, C., Bhalla, N., Lo, C.-W., Jin, J., Park, J.-G., Martinez-Gualda, B., et al. (2022) Numb-associated kinases are required for SARS-CoV-2 infection and are cellular targets for antiviral strategies. *Antiviral Res.*, **204**, 105367.
83. Neveu, G., Barouch-Bentov, R., Ziv-Av, A., Gerber, D., Jacob, Y. and Einav, S. (2012) Identification and targeting of an interaction between a tyrosine motif within hepatitis C virus core protein and AP2M1 essential for viral assembly. *PLoS Pathog.*, **8**, e1002845.
84. Pu, S., Schor, S., Karim, M., Saul, S., Robinson, M., Kumar, S., Prugar, L.I., Dorosky, D.E., Brannan, J., Dye, J.M., et al. (2020) BIKE regulates dengue virus infection and is a cellular target for broad-spectrum antivirals. *Antiviral Res.*, **184**, 104966.
85. Sorensen, E.B. and Conner, S.D. (2008) AAK1 regulates Numb function at an early step in clathrin-mediated endocytosis. *Traffic*, **9**, 1791–1800.
86. Ricotta, D., Conner, S.D., Schmid, S.L., von Figura, K. and Honing, S. (2002) Phosphorylation of the AP2 mu subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *J. Cell Biol.*, **156**, 791–795.
87. Liu, Q., Bautista-Gomez, J., Higgins, D.A., Yu, J. and Xiong, Y. (2021) Dysregulation of the AP2M1 phosphorylation cycle by LRRK2 impairs endocytosis and leads to dopaminergic neurodegeneration. *Sci. Signal.*, **14**, eabg3555.
88. Zeng, G. and Cai, M. (1999) Regulation of the actin cytoskeleton organization in yeast by a novel serine/threonine kinase Prk1p. *J. Cell Biol.*, **144**, 71–82.
89. Zeng, G., Yu, X. and Cai, M. (2001) Regulation of yeast actin cytoskeleton-regulatory complex Pan1p/Sla1p/End3p by serine/threonine kinase Prk1p. *Mol. Biol. Cell*, **12**, 3759–3772.
90. Henry, K.R., D’Hondt, K., Chang, J.S., Nix, D.A., Cope, M.J.T.V., Chan, C.S.M., Drubin, D.G. and Lemmon, S.K. (2003) The actin-regulating kinase Prk1p negatively regulates Scd5p, a suppressor of clathrin deficiency, in actin organization and endocytosis. *Curr. Biol.*, **13**, 1564–1569.
91. Belkina, N.V., Liu, Y., Hao, J.-J., Karasuyama, H. and Shaw, S. (2009) LOK is a major ERM kinase in resting lymphocytes and regulates cytoskeletal rearrangement through ERM phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4707–4712.
92. Shah, N.H., Amacher, J.F., Nocka, L.M. and Kuriyan, J. (2018) The Src module: an ancient scaffold in the evolution of cytoplasmic tyrosine kinases. *Crit. Rev. Biochem. Mol. Biol.*, **53**, 535–563.
93. Buday, L., Wunderlich, L. and Tamás, P. (2002) The Nck family of adapter proteins: regulators of actin cytoskeleton. *Cell. Signal.*, **14**, 723–731.
94. Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., et al. (2008) Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics*, **7**, 768–784.
95. Sámano-Sánchez, H., Gibson, T.J. and Chemes, L.B. (2023) Using linear motif database resources to identify SH2 domain binders. *Methods Mol. Biol.*, **2705**, 153–197.
96. Frese, S., Schubert, W.-D., Findeis, A.C., Marquardt, T., Roske, Y.S., Stradal, T.E.B. and Heinz, D.W. (2006) The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains. *J. Biol. Chem.*, **281**, 18236–18245.
97. Li, D. and Wu, M. (2021) Pattern recognition receptors in health and diseases. *Signal Transduct. Target. Ther.*, **6**, 291.
98. Zevini, A., Olagnier, D. and Hiscott, J. (2017) Crosstalk between cytoplasmic RIG-I and STING sensing pathways. *Trends Immunol.*, **38**, 194–205.
99. Chen, Q., Sun, L. and Chen, Z.J. (2016) Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat. Immunol.*, **17**, 1142–1149.
100. Hildebrand, J.M., Yi, Z., Buchta, C.M., Poovassery, J., Stunz, L.L. and Bishop, G.A. (2011) Roles of tumor necrosis factor receptor associated factor 3 (TRAF3) and TRAF5 in immune cell functions. *Immunol. Rev.*, **244**, 55–74.
101. Bishop, G.A. (2016) TRAF3 as a powerful and multitiered regulator of lymphocyte functions. *J. Leukoc. Biol.*, **100**, 919–926.
102. Liu, B. and Gao, C. (2018) Regulation of MAVS activation through post-translational modifications. *Curr. Opin. Immunol.*, **50**, 75–81.
103. Ortega, E., Rengachari, S., Ibrahim, Z., Houghoughi, N., Gaucher, J., Holehouse, A.S., Khochbin, S. and Panne, D. (2018) Transcription factor dimerization activates the p300 acetyltransferase. *Nature*, **562**, 538–544.
104. Zhao, B., Du, F., Xu, P., Shu, C., Sankaran, B., Bell, S.L., Liu, M., Lei, Y., Gao, X., Fu, X., et al. (2019) A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1. *Nature*, **569**, 718–722.
105. Al Hamrashdi, M. and Brady, G. (2022) Regulation of IRF3 activation in human antiviral signaling pathways. *Biochem. Pharmacol.*, **200**, 115026.
106. Guharoy, M., Lazar, T., Macossay-Castillo, M. and Tompa, P. (2022) Degron masking outlines degrons, co-degrading functional modules in the proteome. *Commun. Biol.*, **5**, 445.
107. Akdel, M., Pires, D.E.V., Pardo, E.P., Jänes, J., Zalevsky, A.O., Mészáros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2022) A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.*, **29**, 1056–1067.
108. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
109. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
110. Tusnády, G.E., Zeke, A., Kalman, Z.E., Fatoux, M., Ricard-Blum, S., Gibson, T.J. and Dobson, L. (2023) LeishMANIAdb: a comparative resource for Leishmania proteins. *Database*, **2023**, baad074.
111. Bret, H., Andreani, J. and Guerois, R. (2023) From interaction networks to interfaces: Scanning intrinsically disordered regions using AlphaFold2. bioRxiv doi: <https://doi.org/10.1101/2023.05.25.542287>, 25 May 2023, preprint: not peer reviewed.
112. Lee, C.Y., Hubrich, D., Varga, J.K., Schäfer, C., Welzel, M., Schumbera, E., Đokić, M., Strom, J.M., Schönfeld, J., Geist, J.L., et al. (2023) Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. bioRxiv doi: <https://doi.org/10.1101/2023.08.07.552219>, 09 August 2023, preprint: not peer reviewed.

113. Diella,F, Haslam,N, Chica,C, Budd,A, Michael,S, Brown,N.P, Trave,G and Gibson,T.J. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.*, **13**, 6580–6603.
114. Van Roey,K. and Davey,N.E. (2015) Motif co-regulation and co-operativity are common mechanisms in transcriptional, post-transcriptional and post-translational regulation. *Cell Commun. Signal.*, **13**, 45.
115. Uyar,B, Weatheritt,R.J., Dinkel,H., Davey,N.E. and Gibson,T.J. (2014) Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? *Mol. Biosyst.*, **10**, 2626–2642.
116. Mészáros,B, Sámano-Sánchez,H, Alvarado-Valverde,J, Čalyševa,J, Martínez-Pérez,E, Alves,R, Shields,D.C., Kumar,M, Rippmann,F, Chemes,L.B., *et al.* (2021) Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. *Sci. Signal.*, **14**, eabd0334.
117. Pittner,N.A., Solomon,R.N., Bui,D.-C. and McBride,J.W. (2023) Ehrlichia effector SLiM-icry: artifice of cellular subversion. *Front. Cell. Infect. Microbiol.*, **13**, 1150758.
118. Goswami,S, Samanta,D. and Duraivelan,K. (2023) Molecular mimicry of host short linear motif-mediated interactions utilised by viruses for entry. *Mol. Biol. Rep.*, **50**, 4665–4673.
119. Tunyasuvunakool,K, Adler,J, Wu,Z, Green,T, Zielinski,M, Židek,A, Bridgland,A, Cowie,A, Meyer,C, Laydon,A, *et al.* (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590–596.
120. Tsaban,T, Varga,J.K., Avraham,O, Ben-Aharon,Z., Khrumushin,A. and Schueler-Furman,O. (2022) Harnessing protein folding neural networks for peptide-protein docking. *Nat. Commun.*, **13**, 176.
121. Paysan-Lafosse,T, Blum,M., Chuguransky,S, Grego,T, Pinto,B.L., Salazar,G.A., Bileschi,M.L., Bork,P, Bridge,A., Colwell,L., *et al.* (2023) InterPro in 2022. *Nucleic Acids Res.*, **51**, D418–D427.
122. Szklarczyk,D, Kirsch,R., Koutrouli,M., Nastou,K., Mehryary,F, Hachilif,R., Gable,A.L., Fang,T, Doncheva,N.T, Pyysalo,S., *et al.* (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.*, **51**, D638–D646.