

ELM—the eukaryotic linear motif resource in 2020

Manjeet Kumar^{1,*}, Marc Gouw^{1,†}, Sushama Michael^{1,†}, Hugo Sámano-Sánchez^{1,2}, Rita Pancsa³, Juliana Glavina⁴, Athina Diakogianni¹, Jesús Alvarado Valverde¹, Dayana Bukirova^{1,5}, Jelena Čalyševa^{1,2}, Nicolas Palopoli⁶, Norman E. Davey⁷, Lucía B. Chemes^{4,*} and Toby J. Gibson^{1,*}

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, ³Institute of Enzymology, Research Centre for Natural Sciences, Budapest 1117, Hungary, ⁴Instituto de Investigaciones Biotecnológicas (IIBio) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de San Martín. Av. 25 de Mayo y Francia, CP1650, Buenos Aires, Argentina, ⁵Nazarbayev University, Nur-Sultan 010000, Kazakhstan, ⁶Department of Science and Technology, Universidad Nacional de Quilmes - CONICET, Bernal B1876BXD, Buenos Aires, Argentina and ⁷The Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Rd, Chelsea, London SW3 6JB, UK

Received September 25, 2019; Revised October 18, 2019; Editorial Decision October 21, 2019; Accepted October 23, 2019

ABSTRACT

The eukaryotic linear motif (ELM) resource is a repository of manually curated experimentally validated short linear motifs (SLiMs). Since the initial release almost 20 years ago, ELM has become an indispensable resource for the molecular biology community for investigating functional regions in many proteins. In this update, we have added 21 novel motif classes, made major revisions to 12 motif classes and added >400 new instances mostly focused on DNA damage, the cytoskeleton, SH2-binding phosphotyrosine motifs and motif mimicry by pathogenic bacterial effector proteins. The current release of the ELM database contains 289 motif classes and 3523 individual protein motif instances manually curated from 3467 scientific publications. ELM is available at: <http://elm.eu.org>.

INTRODUCTION

Short linear motifs (SLiMs), eukaryotic linear motifs (ELMs), MoRFs and miniMotifs, are a distinct class of protein interaction interface that is central to cell physiology (1,2). In the original 1990 definition, SLiMs were described as ‘linear, in the sense that 3D organization is not required to bring distant segments of the molecule together to make the recognizable unit.’ (3). This unexpected structural property was later explained by their frequent occurrence within

intrinsically disordered regions of proteins or in exposed flexible loops within folded domains (1,4). The preference for flexible regions and their lack of tertiary structural constraints allows them to be accessible for protein–protein interaction and adopts the bound structure required for interaction with their binding partner.

The cell uses transient and reversible SLiM-mediated interactions to build dynamic complexes, control protein stability and direct proteins to the correct cellular compartment. Post-translational modification SLiMs act like switches that allow the transmission of cell state information to the wider protein population (5) and integrate different signaling inputs to allow decision-making on the protein level (6,7). Given the central regulatory role of SLiMs, they are now understood to be at the interface between biology and medicine. SLiMs are mutated in many human diseases including the degrons of tumor promoters in cancer (8,9) and are pervasively mimicked by pathogens through convergent evolution to hijack and deregulate host cellular functions (10–13). This understanding of the therapeutic relevance of SLiMs has resulted in an increased interest in drugging SLiM-mediated interactions (14).

Based on estimates obtained from high-throughput screening (HTS) experiments and computational studies, the human proteome is predicted to contain over 100 000 binding motifs and vastly more post-translational modification sites (PTMs) (4). However, motif discovery and characterization are hampered by computational and experimental difficulties (15) and only a small fraction of these anticipated sites have been discovered to date, which

*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 x 2133; Email: lchemes@iib.unsam.edu.ar
Correspondence may also be addressed to Toby J. Gibson. Tel: +49 6221 387 8398; Email: toby.gibson@embl.de
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.
Present address: Marc Gouw, Intomics, Lottenborgvej 26, DK-2800 Lyngby (Copenhagen), Denmark.

is underscored by the fact that we currently ignore the interaction partners for ~75% of structural domain families (4). Because of the time consuming nature of literature curation, only a fraction of the experimentally discovered SLiM instances and classes are currently represented in the ELM resource. Therefore, improving the curation coverage of both known and novel motif classes is an important task for the the motif biology field.

The current census of SLiMs has been characterized over 30 years of small steps using cell biology and biophysical approaches. These advances are often limited by our inability to characterize SLiMs *in vivo* in the context of complex multiprotein assemblies and the difficulty of reproducing these assemblies *in vitro*. Nevertheless, the reductionist approach favored in motif biology has still resulted in numerous fundamental insights in cell biology. The application of medium and high-throughput approaches for the discovery of motifs, such as proteomic phage display (ProP-PD) (16) and peptides attached to Microspheres with Ratiometric Barcode Lanthanide Encoding (MRBLE-pep) (17), is now on the cusp of revolutionizing the field of motif biology. Consequently, a large body of motif data is on the verge of becoming available.

The ELM resource has an important role in guiding the development of these novel experimental approaches, as it is the only existing resource where motif definitions are described in the context of the underlying biology and evolution. SLiM curation remains the gold standard for motif data and the ELM instances will provide benchmarking data for these novel approaches and help define discriminatory motif attributes that will drive the discovery of novel motifs. This is in addition to the existing roles of the ELM resource in the molecular biology community as a repository of motif information, a server for exploring candidate motifs in protein sequences and a source of training data for bioinformatics tool development. As the 20th anniversary of ELM approaches, the resource remains a foundational hub for the motif community, and new tools such as *articles.ELM* (<http://slim.icr.ac.uk/articles/>) have been developed to assist the curation process in the face of the increased data that will become available in the near future.

THE ELM RESOURCE

The ELM resource (<http://www.elm.eu.org>) contains two services: the ELM server for exploring candidate motifs and, the main focus of the current update, the ELM database. The ELM relational database is a repository that collects, classifies and curates experimental information on SLiMs. The ELM database has been under development for almost 20 years and has shown steady growth in the number of curated articles, collected motif instances and motif class definitions (18–23) (Figure 1). The ELM database classifies motif instances into class entries based on shared function, specificity determinants or binding partner. For each motif class, ELM provides a comprehensive report analogous to a short review describing the motif's function, interacting domains, binding determinants and taxonomic range. Related motif classes such as those interacting with the same protein domain are grouped under a unique functional site class. Motif classes are also grouped by type based on their high

level function as ligand (LIG), targeting (TRG), docking (DOC), degradation (DEG), modification (MOD) or cleavage (CLV) motifs. Each ELM motif class entry also provides a list of experimentally validated motif instances manually curated from the literature. For each instance, ELM curates the binding peptide (mapped to the protein entry in UniProt (24), the protein information, the relevant publication, the methods used to characterize the motif and information on the binding partner(s). If available, the binding affinity (typically as dissociation constants) and structural information are also collected. With the current release, ELM encompasses 3523 motif instances, 289 motif classes, 516 structures containing SLiM peptides and 3467 scientific publications. Table 1 provides a breakdown of the main data types in the ELM resource.

Programmatic access to the ELM resource is available through the REST API (for instructions see <http://elm.eu.org/api/manual.html>). For example, motif matches for the human p53 protein (UniProt accession:P04637) can be retrieved using the REST request http://elm.eu.org/start_search/P04637.tsv. Other features of the ELM resource have been outlined in the 2018 ELM paper (23) or earlier.

ELM motif data will become linked from PDBe-KB (25) and structures in ELM are now linked to PDBe (26) from the ELM structure page (<http://elm.eu.org/pdbs/>).

NOVEL AND UPDATED ELM CLASSES

As novel aspects of motif biology have appeared, the ELM resource has at times changed curation focus to populate high profile or underpopulated biological pathways. In previous releases, this has included curation drives for SLiMs in viral proteins, conditionally regulated motif switches and motifs regulating cell cycle progression. The current release of ELM has continued this approach by focusing curation on DNA damage, the cytoskeleton, kinase specificity, SH2 domains and mimicry by pathogenic effectors. The current ELM release includes 21 new classes (Table 2), >400 new instances and 67 added structures. In addition, 12 existing motif classes have been updated to reflect advances in our understanding of those motifs (Table 2).

DNA damage and repair

In the new release of ELM, we have expanded our encoding of DNA damage and DNA repair motifs, providing a comprehensive picture of this large and diverse motif group (Figure 2). We have included several novel classes of proliferating cell nuclear antigen (PCNA)-interacting protein (PIP) box-like motifs including the APIM, and the related RIR and MIP motifs. We have expanded the definition of the PIP Box motif creating two classes that reflect the variation observed in metazoan versus fungal motifs. A variant motif representing the translesion synthesis polymerases is in preparation. The inclusion of the novel PIP-like motif classes has led to addition of 2 APIM (New class: LIG_PCNA_APIM_2), 10 RIR (New class: LIG_REV1ctd_RIR_1) and 6 MIP (New class: LIG_MLH1_MIPbox_1) motif instances. In addition, we updated the metazoan PIP Box (LIG_PCNA_PIPBox_1)

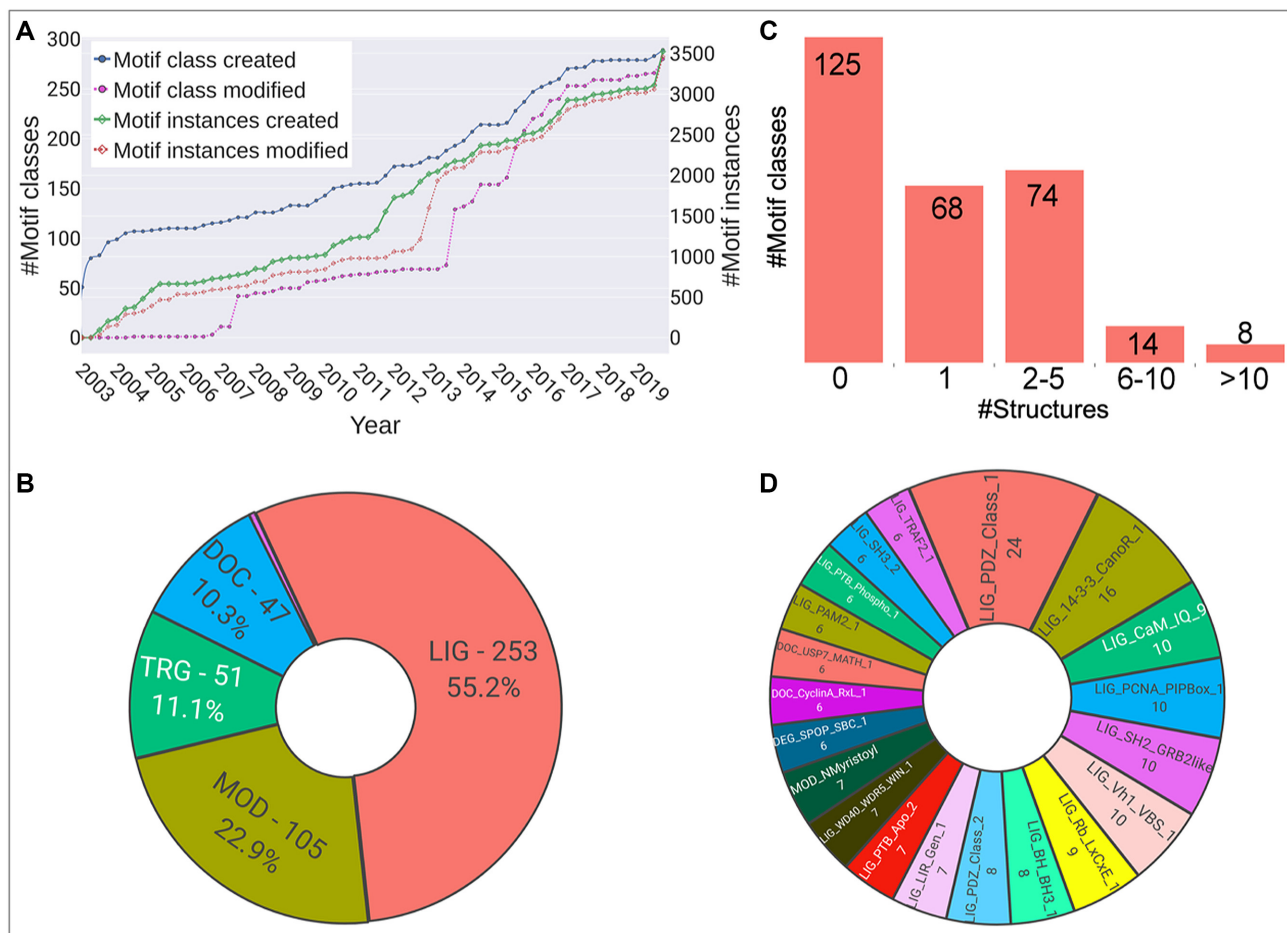


Figure 1. (A) Progression of the motif classes and instances integrated in the ELM resource. (B) Pie-chart showing count and proportion of new instance addition from each motif class type in the current ELM release. (C) Barplot showing the motif classes grouped according to the coverage of their instances by PDB structures, only one structure per instance has been considered for showing the coverage. In total, 164 ELM classes are covered by at least one structure. (D) Top 20 motif classes in terms of the number of representative PDB structures are shown. The plots were generated using plotly chart studio (<https://chartstudio.plot.ly>).

Table 1. Overview of the data stored in the ELM database

| Functional sites | ELM classes | ELM instances | GO terms | PDB structures | ELM instances with affinity values | PubMed Links | |
|------------------|-------------|---------------|----------|----------------|------------------------------------|--------------|------|
| Total | 176 | 289 | | 791 | 516 | 265 | 3467 |
| By category | LIG | 163 | Human | 2090 | Biological process | 430 | |
| | MOD | 37 | Mouse | 341 | | | |
| | DOC | 31 | Rat | 150 | Cellular component | 163 | |
| | DEG | 25 | Yeast | 110 | | | |
| | TRG | 22 | Fly | 98 | Molecular function | 198 | |
| | CLV | 11 | Others | 734 | | | |

with 19 instances and the fungal PIP Box (New class: LIG_PCNA_yPIPBox.3) with 12 instances. In total, the PIP-like motif classes have been expanded with 49 novel instances and 24 additional structures.

The accurate replication of DNA is essential for genome stability and for the faithful transmission of genetic information from mother to daughter cells. Successful DNA replication depends on the DNA synthesis machinery and on the efficient sensing of DNA damage in order to initiate the repair of DNA lesions or activate tolerance mechanisms that allow the replicative bypass of damaged DNA. The

ability of cells to tolerate DNA damage is a key determinant of cancer therapy response, making DNA repair and damage proteins attractive drug candidates (27). PCNA, Mh1 and Rev1 are hubs of genome maintenance networks responsible for the sensing and integration of DNA replication stress signaling. Protein partners interact with these hubs via PIP Box, MIP Box and RIR motifs, respectively.

Several DNA replication and repair pathways cooperate to ensure the reliable repair of different DNA damage types. The Mh1 protein acts as a major signal integrator of the mismatch repair pathway. Partners from other repair path-

Table 2. Novel and revised ELM classes since the last ELM publication

| Novel ELM classes | | |
|----------------------------------|---------------------|--|
| ELM class identifier | Number of instances | ELM class (short) description |
| LIG.SH2.CRK | 34 | CRK family SH2 domain binding motif |
| LIG.PDZ.Wminus1.1 | 27 | The C-terminal Trp-1 PDZ-binding motif is represented by a pattern like W(ACGILV)S. |
| LIG.SH2.STAP1 | 22 | STAP1 Src Homology 2 (SH2) domain Class 2 binding motif |
| LIG.SH2.NCK.1 | 17 | NCK Src Homology 2 (SH2) domain binding motif |
| LIG.PROFILIN.1 | 16 | The polyproline profilin-binding motif is found in regulators of actin cytoskeleton. |
| LIG.PCNA.yPIPBox.3 | 12 | The PCNA binding motifs include the PIP Box, PIP degron and the APIM motif, and are found in proteins involved in DNA replication, repair, methylation and cell cycle control. This is the variant for the yeast PIPbox. |
| LIG.REV1ctd.RIR.1 | 10 | Several DNA repair proteins interact with the C-terminal domain of the Rev1 translesion synthesis scaffold through the Rev1-Interacting Region RIR motif that is centered around two neighboring Phe residues. |
| LIG.IBAR.NPY.1 | 7 | A short NPY motif present in the bacterial effector protein Tir binds the I-BAR domain and is involved in actin polymerization. |
| LIG.MLH1.MIPbox.1 | 6 | Proteins involved in DNA repair and replication employ conserved MIP-box motifs to bind the C-terminal domain of mismatch repair protein MLH1. |
| LIG.FXIDFP.1 | 5 | The DFP motif enables binding to the 2nd apple domain of coagulation factor XI (FXI) and plasma kallikrein heavy chain. |
| LIG.deltaCOP1.diTrp.1 | 5 | Tryptophan-based motifs enable targeting of the tethering and (dis)assembly factors to the C-terminal mu homology domain (MHD) of the coatomer subunit delta, delta-COP. |
| LIG.CaM.NSCaTE.8 | 3 | Short motif recognized by CaM that is only present in the Cav1.2 and Cav1.3 L-type calcium channels. |
| LIG.ARL.BART.1 | 2 | The ligand motif present in N-terminus region of ARL2 and ARL3 proteins ensures GTD-dependent binding to BART and BARTL1. |
| LIG.PCNA.APIM.2 | 2 | The PCNA-binding APIM motif is found in proteins involved in DNA repair and cell cycle control. |
| MOD.PRMT.GGRGG.1 | 24 | A GGRGG motif recognized by the arginine methyltransferase for arginine methylation. |
| MOD.DYRK1A.RPxSP.1 | 22 | Serine/Threonine residue phosphorylated by Arginine and Proline directed DYRK1A kinase. |
| DOC.PP4.FxxP.1 | 15 | The FxxP-like docking motif recognized by the EVH1 domains of the PPP4R3 regulatory subunits of the PP4 holoenzyme. |
| DOC.PP4.MxPP.1 | 2 | The MxPP-like docking motif recognized by the EVH1 domains of the PPP4R3 regulatory subunits of the PP4 holoenzyme. |
| DOC.MAPK.GRA24.9 | 2 | A kinase docking motif that mediates interaction toward the ERK1/2 and p38 subfamilies of MAP kinases. |
| TRG.Pf-PMV.PEXEL.1 | 24 | Plasmodium Export Element, PEXEL, is a trafficking signal for protein cleavage by PMV protease and export from Plasmodium parasites to infected host cells. |
| TRG.ER.FFAT.2 | 7 | A variant of the classic MSP-domain binding FFAT (diphenylalanine [FF] in an Acidic Tract) motif. |
| ELM Classes with major revisions | | |
| LIG.CaM.IQ.9 | 75 | Helical peptide motif responsible for Ca ²⁺ -independent binding of the CaM. |
| LIG.SH2.GRB2like | 35 | GRB2-like Src Homology 2 (SH2) domain binding motif. |
| LIG.LIR.Gen.1 | 21 | Canonical LIR motif that binds to Atg8 protein family members to mediate processes involved in autophagy. |
| LIG.PCNA.PIPBox.1 | 19 | The PCNA-binding PIP Box motif is found in proteins involved in DNA repair and cell cycle control. |
| LIG.Vh1.VBS.1 | 15 | An amphipathic α -helix recognized by the head domain of vinculin that is required for vinculin activation and actin filament attachment. |
| LIG.IRF3.LxIS.1 | 7 | A binding site for IRF-3 protein present in various innate adaptor proteins and the viral protein NSP1 to trigger the innate immune responsive pathways. |
| MOD.CK2.1 | 34 | Casein kinase 2 (CK2) phosphorylation site. |
| MOD.CK1.1 | 27 | CK1 phosphorylation site. |
| MOD.CDK.SPxK.1 | 26 | Canonical version of the CDK phosphorylation site that shows specificity toward a lysine/arginine residue at the [ST]+3 position. |
| MOD.CAAXbox | 17 | Generic CAAX box prenylation motif. |
| DOC.CyclinA.RxL.1 | 28 | This motif is mainly based on cyclin A binding peptides and may not apply to all cyclins. |
| TRG.ER.FFAT.1 | 29 | MSP-domain binding FFAT (diphenylalanine [FF] in an Acidic Tract) motif. |

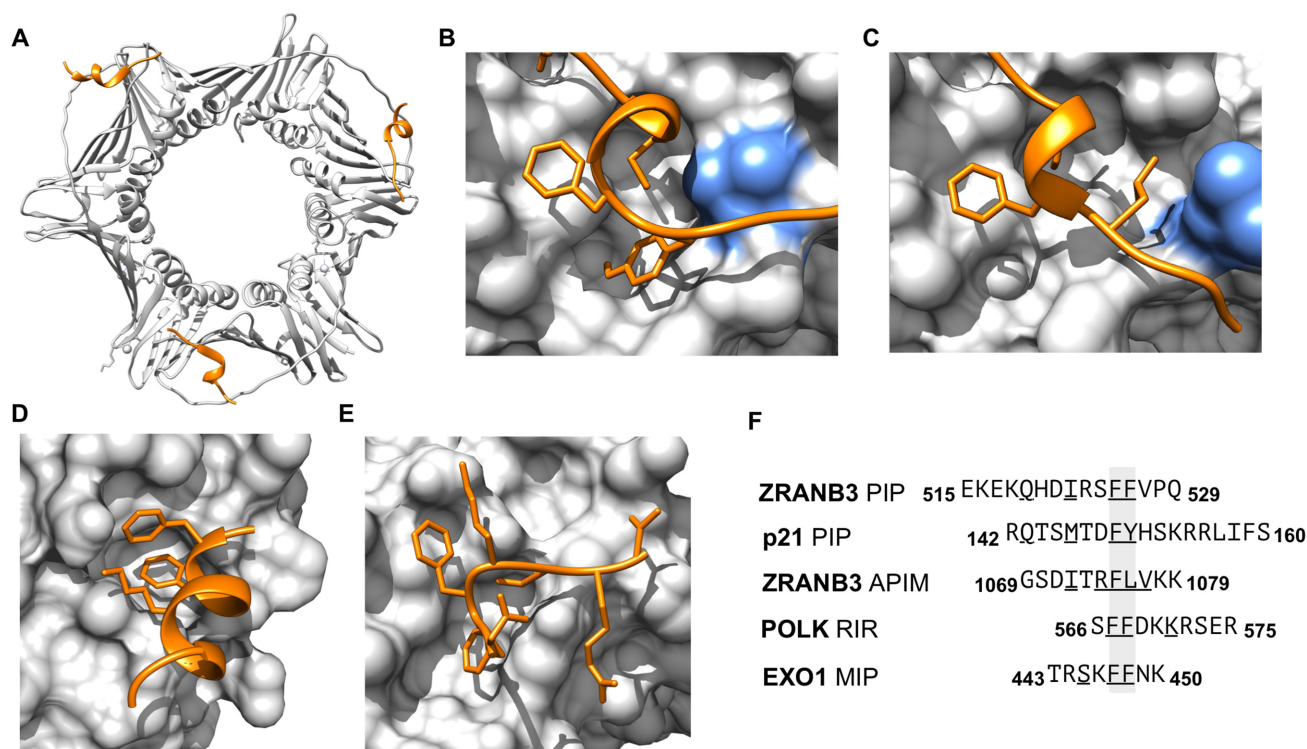


Figure 2. Structural information on representative DNA damage and repair motif instances and classes added in the current ELM update. (A) Structure of PCNA trimer in complex with PIP box of ZRANB3 [PDB ID: 5MLO] (77). (B) Closeup of the structure of PCNA PIP-binding pocket in complex with the PIP box of p21 [PDB ID: 1AXC] (34). (C) Close-up of the structure of PCNA PIP-binding pocket in complex with the APIM of ZRANB3 [PDB ID: 5YD8] (33). The blue residue in panels (B) and (C) shows the rearrangement of a leucine 126 in the PIP-binding pocket to accommodate the APIM peptide. (D) Close-up of the structure of the Rev1 C-terminal domain with the RIR motif of DNA polymerase kappa [PDB ID: 4FJO] (78). (E) Close-up of the structure of the C-terminal domain of the yeast MUTL alpha (MLH1/PMS1) bound to MIP box motif of Exo1 [PDB ID: 4FMO] (79). (F) Peptides from the structures of panels (A–E) aligned around their core hydrophobic residues. Underlined residues define the motif consensus residues in the peptide. Structural figures were prepared using the UCSF Chimera software (80).

ways communicate with Mlh1 through the widely conserved MIP box motif (New class: `LIG_MLH1_MIPbox_1`) (28). The replicative bypass of DNA lesions is performed in a process termed translesion synthesis (TLS). Here, the Rev1 protein acts as a major scaffold that orchestrates the exchange of different polymerases. Rev1 is well suited for this job, because it can simultaneously bind Pol ζ and other TLS polymerases that have Rev1-interacting regions, so called RIR motifs (New class: `LIG_REV1ctd_RIR_1`) (29,30).

The PCNA protein is the ‘sliding clamp’ that encircles DNA at the replication fork. PCNA acts as a major scaffolding protein that orchestrates the assembly of replicative DNA polymerases, and integrates DNA damage signaling with tolerance mechanisms, working in combination with Rev1 to facilitate the recruitment of low-fidelity TLS polymerases to stalled replication forks and allow the replicative bypass of DNA lesions (31). The metazoan and fungal PIP Box (`LIG_PCNA_PIPBox_1` and New class: `LIG_PCNA_yPIPBox_3`) (31) and APIM motifs (New class: `LIG_PCNA_APIM_2`) (32,33) mediate binding of a large number of PCNA-interacting proteins to the PCNA PIP Box cleft, including p21 and the Pol η TLS polymerase. The Pol κ and Pol ν TLS polymerases use a variant PIP-like motif that binds to the same binding cleft in PCNA (34,35). DNA Damage and cell cycle signaling are integrated by the p21 cyclin-dependent kinase inhibitor, which binds PCNA

through its PIP Box and mediates cell cycle arrest in response to DNA damage to prevent cell cycle progress until replication can resume.

PIP-like motifs share a core hydrophobic helix that often contains a double-aromatic residue pair (36) (Figure 2), and several studies suggest that many PIP-like motifs are able to interact with at least two of these hub proteins (37,38). The available motif instances reveal the diversity but also the high conservation of PIP-like motifs, and point to the existence of a broader group of functionally and structurally related DNA damage and repair motifs that might show an unexpected degree of cross-functionality (37,38).

Motif mimicry in bacterial effector proteins

A major ELM focus continuing from the last release has been the curation of the available literature on human motif mimicry by bacterial effector proteins. This curation drive mirrors a previous ELM release where the curation of the complete corpus of viral motif literature added over 200 novel ELM instances in 84 different viral taxa (10,20). Pathogens have an intimate relationship with their host and often produce proteins that mimic higher eukaryotic SLiMs to hijack, deregulate or rewire host pathways. This mimicry is facilitated by the ease of *ex nihilo* motif evolution due to the degeneracy of motifs and the rapid evolu-

tion of most bacterial and viral pathogens (1,39). The available literature on bacterial motifs is not as extensive as the viral motif literature but interest in the research field is increasing. ELM now contains information on >110 bacterial motif instances from 28 bacterial species mapping to 31 ELM classes. Our focus on bacterial mimicry has required us to improve ELM annotation for several topics, notably for cytoskeleton and membrane regulation, and for SH2 domain-binding motifs because ELM lacked entries that matched some of the effector motifs. For example, enteropathogenic *Escherichia coli* (EPEC) Tir protein is tyrosine-phosphorylated and then binds to the NCK SH2 domain (40). An NCK SH2 motif class entry has now been added to ELM (discussed below). The bacterial effector annotation in ELM is now close to being comprehensive with the current literature. It is clear that motif mimicry is a common feature of bacterial effector proteins.

To use the ELM server correctly for non-Eukaryotic pathogen proteins, the input parameters have to be set up appropriately for the host organism, not for the bacterial species. Figure 3 shows correct settings for the VBS motif-containing effector TarP from *Chlamydomonas reinhardtii* that infects the guinea pig (41).

Cytoskeletal regulatory motifs

SLiM-mediated interactions play an important role in the control of the actin cytoskeleton, particularly for initiation of actin filament polymerization, and these interactions are often hijacked by bacterial pathogens. Figure 4 shows the KEGG resource (42) Actin Regulatory Pathway color-coded by ELM motif class types and with pathogen intervention sites marked. In the current release of ELM, we have added two new classes (the Profilin-binding polyproline motif and the IRSp53 I-BAR domain-binding NPY motifs) and revised an existing class (Vinculin Binding Sites) that mediate functions associated with the actin cytoskeleton.

Profilin is a key regulator of the cytoskeleton due to its actin-binding and filament-inducing activity. Several actin filament promoting proteins employ poly-proline sequence motifs (New class: `LIG_PROFILIN_1`) to interact with profilin. Sixteen of these proline-rich motif instances of profilin-binding motifs have been added, including motifs in the key actin regulators WASF1 and VASP.

The I-BAR domain of IRSp53/IRTKS binds NPY motifs (New class: `LIG_IBAR_NPY_1`) (43–46). The NPY motif was originally discovered in a bacterial pathogenic effector and cellular proteins containing the motif were predicted. The bacterial effector protein Tir of enterohemorrhagic *Escherichia coli* (EHEC) binds IRSp53 with an NPY motif (47,48) to ultimately achieve the activation of actin polymerization and actin pedestal formation. Six new instances including four human motifs and the examples of bacterial IRSp53 hijacking have been added: however, the human examples are all hypothetical motif matches that are plausible but have yet to be validated.

Finally, the Vinculin binding sites class (Revised class: `LIG_Vhl_VBS_1`) has been updated with a revised regular expression enabling inclusion of several additional instances. Vinculin primarily works as a linker that strength-

ens the association of Talin and F-Actin at sites of integrin activation, allowing stronger actin binding and stabilization of the sites of focal adhesion (49). Talin contains a long tail with several Vinculin binding sites (VBSs). *Shigella flexneri*, *Rickettsia* and *Chlamydomonas* all secrete effectors that mimic Talin VBSs to induce actin polymerization without the need for integrin activation (50–53).

Membrane-associated pathways

Two novel motif classes involved in membrane trafficking pathways have been added in the current ELM release. A novel class describing a δ -COP interacting motif (New class: `LIG_deltaCOP1_diTrp_1`) including five new instances has been added. The interaction between tryptophan-based motifs surrounded by negatively charged residues within the lasso-like loop of the Dsl1-tethering complex (54) and the C-terminal μ homology domain (MHD) of δ -COP located in the outermost layer of the coat has an important role in docking COPI vesicles to the ER (55). COPI-coated vesicles mediate the retrograde trafficking pathways from the Golgi to the endoplasmic reticulum (ER) and within the Golgi. The life cycle of COPI-coated vesicles is controlled by essential assembly/disassembly factors, including their specific multisubunit tethering complexes, SNARE complexes and the regulators of their small GTPase Arf1, the ArfGAPs. ArfGAPs (Gcs1p in yeast and ArfGAP1 in mammals) use similar tryptophan-based motifs to interact with the MHD of δ -COP (55).

The classical FFAT motif regular expression has been updated and many new instances have been curated (Revised class: `TRG_ER_FFAT_1`). A second FFAT class variant with seven instances has also been added to reflect two distinct binding modes (New class: `TRG_ER_FFAT_2`). FFAT motifs are a class of membrane-protein targeting motifs (56,57), and are important for the formation of membrane contact sites (MCSs) between the ER and cellular membranes (58). The FFAT motifs are recognized by the cytosolic N-terminal MSP domain of the highly conserved VAP integral membrane proteins of the eukaryotic ER. Numerous proteins are targeted to the ER by FFAT motifs and both viral and bacterial pathogens may use FFAT motifs to target the intracellular membrane system of the host. For example, *Chlamydia trachomatis* IncV is a membrane protein on the Chlamydia-containing vacuole, termed the inclusion, that binds host VAP proteins through a FFAT motif (59) to form MCSs that tether the vacuole to the ER.

Apicomplexan export elements

Apicomplexans are a wide group of unicellular intracellular parasites responsible for various animal and human diseases. *Plasmodium*, *Toxoplasma*, *Cryptosporidium* and *Babesia* are among the most highly studied Apicomplexa genera and they are the parasites that cause malaria, toxoplasmosis, cryptosporidiosis and babesiosis, respectively (60). Apicomplexans invade host cells, remodel them and proliferate inside them, thanks to the coordinated secretion of proteins (61). These proteins are exported using peptide export signals and protein transport complexes, and disrupt the host's signaling pathways, to sequester nutrients

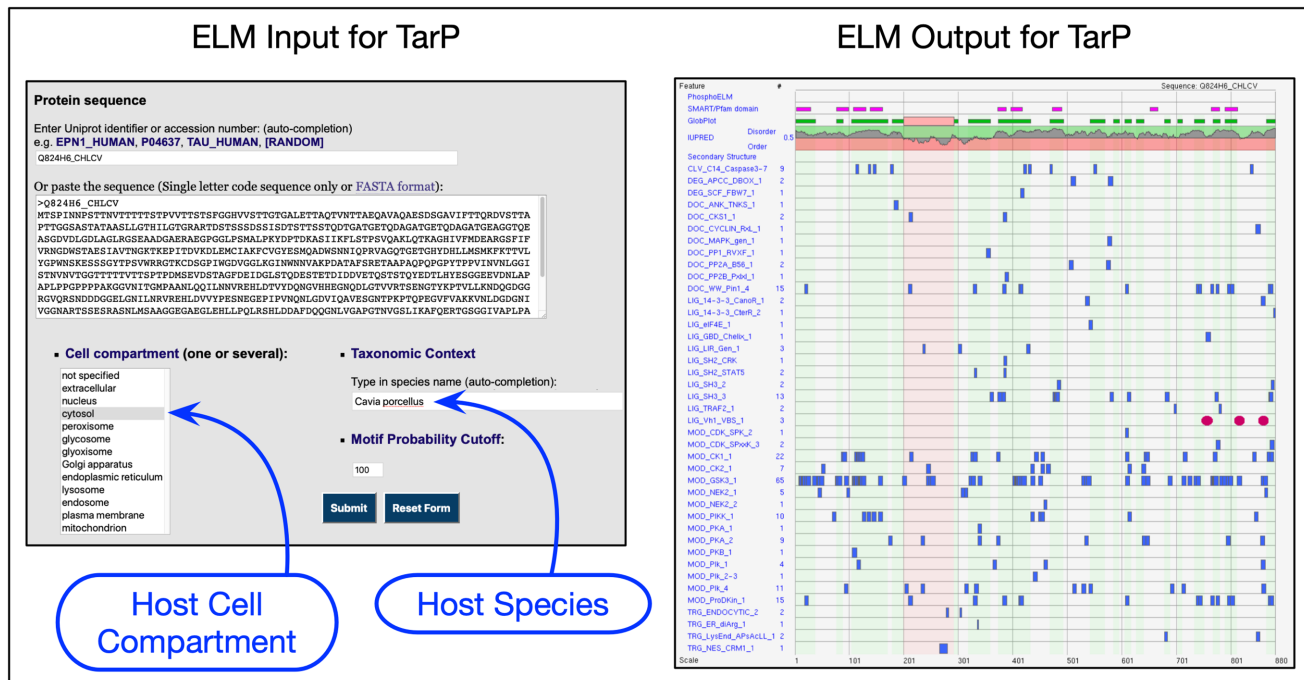


Figure 3. Setting up the ELM server correctly to query bacterial effectors for SLiM candidates using, as an example, the IDP-rich TarP effector from *Chlamydophila caviae* for which the natural host is guinea pig. TarP is extracellular for the bacterium but the correct cell compartment to use is cytosol for the host cell. The correct species is the host *Cavia porcellus*. In the output, the three recently added VBS motifs (41) are shown as red ovals. All other motif matches are hypothetical.

and to evade the immune responses. The Plasmodium Export Element (PEXEL) is the best-characterized export signal in the Apicomplexan phylum. PEXEL is a five residue motif located near the N-terminus of exported proteins following an endoplasmic reticulum (ER) targeting signal peptide (61). It has a dual function: first, as a cleavage site recognized by the aspartyl protease Plasmepsin V and, second, after processing, as a targeting signal to export proteins from the endoplasmic reticulum (ER) through the parasite and parasitophorous vacuole membrane into the infected cell cytosol (61–63). In the current release of ELM, we have added the PEXEL motif as a novel motif class (TRG_Pf-PMV_PEXEL_1). Despite the dual role of the motif, the entry has been added as a targeting motif rather than as a cleavage motif due to its essential role in protein export. We have included 24 novel instances from *Plasmodium falciparum* proteins. These instances are representative of the sequence variation among the PEXELs of other *Plasmodium* species. The regular expression is less strict than the consensus used in the literature, but it should allow the discovery of exported proteins in divergent *Plasmodium* species.

Expansion of the ELM kinome

In the current release, we present a new motif class describing the modification sites of the DYRK1A kinase (New class: MOD_DYRK1A_RPxSP_1). The dual-specificity tyrosine phosphorylation-regulated kinases (DYRK) family consists of five arginine/proline-directed kinases. The novel motif class describes the specificity of the most studied family member, DYRK1A, which is associated with Alzheimer's disease, Down syndrome and early onset neu-

rodegeneration (64,65). The optimal DYRK1A phosphorylation site has the consensus R[PSAV].[ST]P motif, however, substrates exist without the consensus proline or arginine and therefore it can act as both a proline-directed and basophilic kinase. The novel DYRK1A class includes 22 motif instances. Since the last ELM release, the modification motif classes of the CK1, CK2 and Cdk kinases have also been revised, expanding the number of instances. In total, 87 novel motif instances have been added to kinase modification site classes.

Expansion of SH2 motif classes

As a part of the current ELM update, we have significantly expanded the representation of Src homology 2 (SH2) domain binding motifs, grouped under the SH2 functional site. More than 100 SH2 domains are present in mammalian proteomes, where they relay cell state signals through binding to phosphotyrosine motifs that are created following the activation of tyrosine kinases (66). The circa 120 human SH2 domains exhibit a large degree of cross specificity (66,67). Three loops in the SH2 domain determine the accessibility of three hydrophobic pockets, defining clear specificity classes for binding motifs with Asn at position pTyr +2 or hydrophobic residues at positions pTyr +3 and +4 (68,69). We have created three new SH2 classes that reflect their different specificities (New classes: LIG_SH2_CRK, LIG_SH2_NCK_1 and LIG_SH2_STAP1) (40,67,69) and revised an existing class (Revised class: LIG_SH2_GRB2like) (68,70), adding updated structural information to all entries. In total, this has led to the curation of more than 80 individual SH2 motifs and 15 new struc-

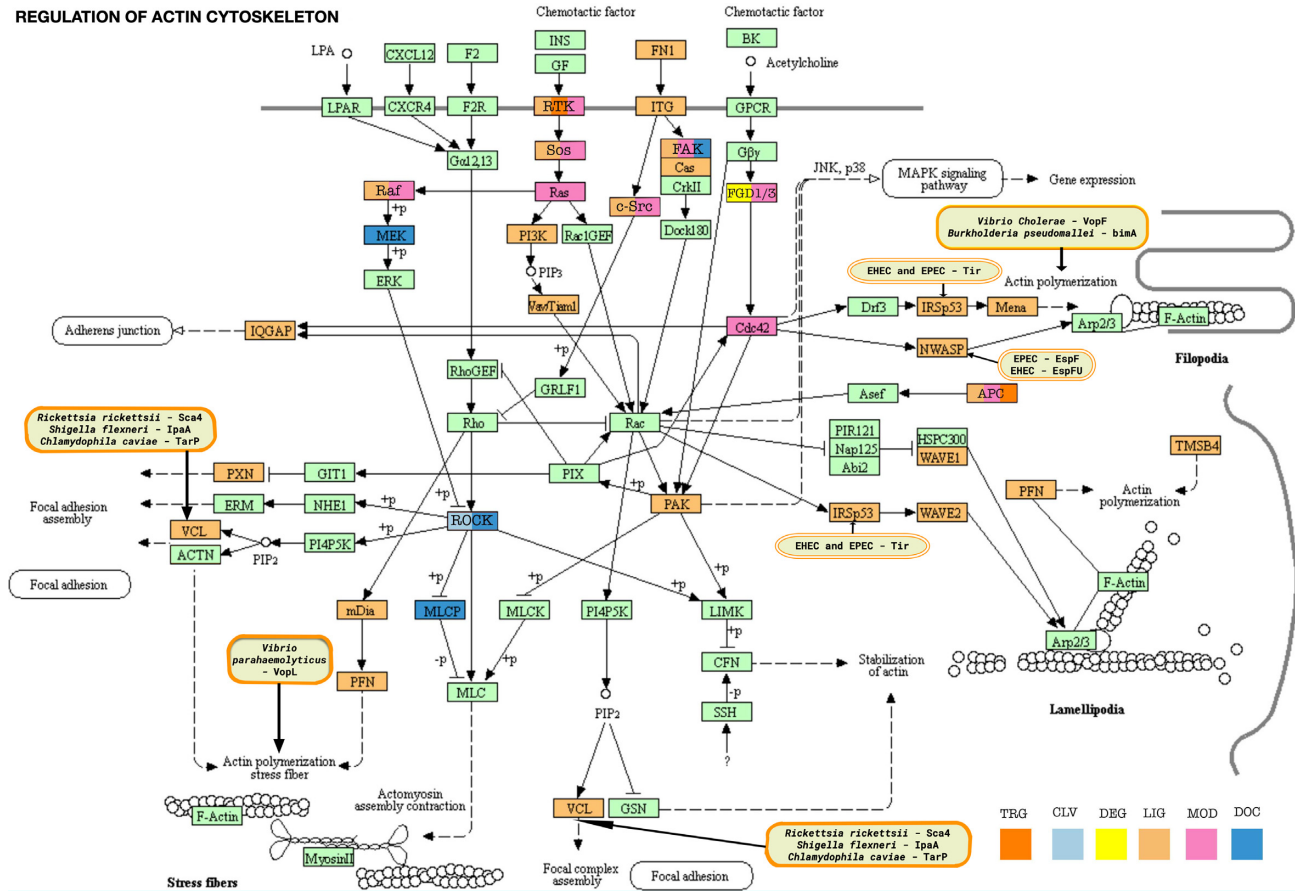


Figure 4. Motif-mediated interactions of the Actin Cytoskeleton network. The KEGG resource network for Regulation of Actin Cytoskeleton (KEGG:hsa04810) is color-coded by ELM motif classes. Proteins of the pathway have a light mint green color by default. Motif-containing proteins are re-colored as follows: DOC class (docking sites) - moderate blue; LIG class (ligand binding motifs) - vivid orange; MOD class (modification sites) - soft pink; DEG class (degradation sites) - yellow; CLV class (cleavage sites) - very soft blue; TRG class (targeting sites) - pure orange; proteins with motifs belonging to multiple classes are marked with the respective colors as described in the bottom right of the figure. ELM has instances for pathogen hijack of actin polymerization at VCL, IRSp53, NWASP and Actin itself. The pathogen proteins affecting these hotspots are shown in the rounded boxes colored with light orange background.

tures. SH2-binding motifs are not straightforward to annotate as there are many similar preferences revealed by SPOT arrays (66,67). Furthermore, there are examples of peptides that match poorly to the consensus determined by the SPOT arrays but bind with relatively high affinity, perhaps because of the three flexible loops surrounding and contributing to the binding surface (68). Nevertheless, work is ongoing to capture the major SH2 variants in ELM as they are so important in health and disease.

UPDATES IN THE ELM ANNOTATION PROCESS

SLiM curation is a complex process that requires a curator to read and interpret the relevant information in a motif-related article. New motifs are annotated for the ELM resource by completing two template documents: a text document to describe the motif class and a spreadsheet to annotate instances of a motif class. Both template documents can be downloaded from the ELM website (http://elm.eu.org/downloads/elm_template.doc and http://elm.eu.org/downloads/elm_template.xls). Typically, an annotator will alternate between reading the experimental literature,

the motif class template and the motif instances spreadsheet while annotating a new SLiM. We have updated the curation process to simplify annotation activities. We have also improved the motif instance spreadsheet to provide a better overview of the information needed to annotate a SLiM. Furthermore, we have recently prepared a detailed step-by-step protocol on how annotators should work with these templates (Gouw, M. *et al.* (2020) *Methods in Mol. Biol.*, in press). This protocol will serve as a useful guideline for annotators contributing data to ELM, and perhaps even encourage contributions from the research community. The protocol may also be used by developers of other resources to create related guidelines.

COLLECTION OF PAPERS FOR FUTURE CURATION

The curation of a motif class entry for the ELM resource is a time-consuming process, often taking over a month to complete. This difficulty means that the data in ELM is not comprehensive with regard to motif publications. However, over the past decade, ELM curation has collected over 6000 articles related to SLiMs that await curation, including nu-

merous articles describing novel motif classes. To bridge the gap between the motifs curated in the ELM resource and those awaiting curation, we have created a companion for the ELM resource called articles.ELM. The articles.ELM resource is a literature repository that contains a manually collected compendium of SLiM-related articles. The articles.ELM resource uses text-mining approaches to link novel uncurated articles with motif classes in the ELM resource. This permits a researcher to rapidly find motif literature related to their interests that awaits curation. The resource also allows the deposition of novel articles describing motif data, which are expected to be massively abundant in the upcoming years. The articles.ELM resource is available at <http://slim.icr.ac.uk/articles/> and classified articles for an ELM class are available as a link from the ELM class entry page (<http://elm.eu.org/elms>). For example, the link from DEG_APCC_DBOX_1 (http://slim.icr.ac.uk/articles/browse/?motif_class=DEG_APCC_DBOX_1) returns a total of 152 articles of which 18 are curated in ELM.

WORKING WITH LINEAR MOTIFS

Reported SLiM instances that are not considered valid are annotated in ELM as False Positives. Most commonly, this is because the suggested motif is buried in the protein fold but sometimes because the interacting protein actually works in a different cellular location. Unfortunately, new examples of False Positive motifs continue to be reported regularly. It is essential to undertake contextual analysis when preparing to investigate a new motif candidate. We have provided guidance to help researchers avoid pitfalls (15). A core set of computational tools that we ourselves use all the time include IUPred, MobiDB and DisProt for assessing intrinsically disordered polypeptide (71–73), JalView and ProViz for motif conservation plus the testing and refinement of Regular Expressions (74,75) and SLiMSearch for searching proteomes (76).

CONCLUSIONS AND PERSPECTIVES

ELM is a fundamental source of information for the dynamically developing motif biology field. The ELM database is the major resource of quality information on motif-mediated interactions and, thanks to the effort of the motif community, ELM has been continuously developed for almost 20 years. SLiM-mediated interactions constitute a significant and growing fraction of cellular protein–protein interactions (4). They are implicated in diverse human diseases (8,9) and often hijacked by viral, bacterial and eukaryotic pathogens (10–12,62). Therefore, their discovery and characterization is crucial to our understanding of both the physiological and disease states of the cell. We are committed to maintaining, improving and expanding the ELM resource in the future. A key goal for ELM in the coming years will be the addition of new tools to help researchers deal with the anticipated imminent explosion of motif biology information. As ELM approaches its third decade, we believe the resource will continue to support researchers elucidating the key role of motifs in cell biology.

ACKNOWLEDGEMENTS

We thank the ELM resource users for their interest and the value it places on our work. We are grateful to our collaborators and colleagues in the SLiM and IDP fields for their help, support and extensive interactions. Tim Levine (UCL) is thanked for informative FFAT motif discussions.

FUNDING

European Molecular Biology Laboratory (EMBL) International PhD Program; Argentine Ministry of Science and Technology and German Academic Exchange Service (MinCyT-DAAD) grant [CyCmotif DA/16/05 to L.B.C., T.G.]; Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) [Grants PICT 2017/1924 to L.B.C. and PICT 2015/3367 to N.P.]; L.B.C. is an independent researcher, N.P. is an adjunct researcher and J.G. holds a postdoctoral fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Hungarian National Research, Development, and Innovation Office (NKFIH) [FK-128133 to R.P.]; Hungarian Academy of Sciences, [PREMIUM-2017–48 to R.P.]; Cancer Research UK Senior Cancer Research Fellowship [C68484/A28159 to N.E.D.]; European Union's (EU) Horizon 2020 research and innovation programme, Project number 778247 (IDP-fun); European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675341 (PDZnet) (JČ) (in part). Funding for open access charge: EMBL.

Conflict of interest statement. None declared.

REFERENCES

- Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H. and Gibson, T.J. (2012) Attributes of short linear motifs. *Mol. Biosyst.*, **8**, 268–281.
- 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.
- Hunt, T. (1990) Protein sequence motifs involved in recognition and targeting: a new series. *Trends Biochem. Sci.*, **15**, 305.
- 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.
- 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.
- Scott, J.D. and Pawson, T. (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science*, **326**, 1220–1224.
- Gibson, T.J. (2009) Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, **34**, 471–482.
- Mészáros, B., Kumar, M., Gibson, T.J., Uyar, B. and Dosztányi, Z. (2017) Degrons in cancer. *Sci. Signal.*, **10**, eaak9982.
- 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.
- Davey, N.E., Travé, G. and Gibson, T.J. (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.*, **36**, 159–169.
- Chemes, L.B., de Prat-Gay, G. and Sánchez, I.E. (2015) Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. *Curr. Opin. Struct. Biol.*, **32**, 91–101.
- Via, A., Uyar, B., Brun, C. and Zanzoni, A. (2015) How pathogens use linear motifs to perturb host cell networks. *Trends Biochem. Sci.*, **40**, 36–48.

13. Hrabec, P., O'Maille, P.E., Silberfarb, A., Davis-Anderson, K., Generous, N., McMahon, B.H. and Fair, J.M. (2019) Resources to discover and use short linear motifs in viral proteins. *Trends Biotechnol.*, doi:10.1016/j.tibtech.2019.07.004.
14. Corbi-Verge, C., Garton, M., Nim, S. and Kim, P.M. (2017) Strategies to develop inhibitors of motif-mediated protein-protein interactions as drug leads. *Annu. Rev. Pharmacol. Toxicol.*, **57**, 39–60.
15. Gibson, T.J., Dinkel, H., Van Roey, K. and Diella, F. (2015) Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.*, **13**, 42.
16. Davey, N.E., Seo, M.-H., Yadav, V.K., Jeon, J., Nim, S., Krystkowiak, I., Blikstad, C., Dong, D., Markova, N., Kim, P.M. *et al.* (2017) Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J.*, **284**, 485–498.
17. 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.
18. Puntervoll, P., Linding, R., Gemünd, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M.A., Ausiello, G., Brannetti, B., Costantini, A. *et al.* (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, **31**, 3625–3630.
19. Gould, C.M., Diella, F., Via, A., Puntervoll, P., Gemünd, C., Chabanis-Davidson, S., Michael, S., Sayadi, A., Bryne, J.C., Chica, C. *et al.* (2010) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.*, **38**, D167–D180.
20. Dinkel, H., Michael, S., Weatheritt, R.J., Davey, N.E., Van Roey, K., Altenberg, B., Toedt, G., Uyar, B., Seiler, M., Budd, A. *et al.* (2012) ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.*, **40**, D242–D251.
21. Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M. *et al.* (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.*, **42**, D259–D266.
22. Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A. *et al.* (2016) ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.*, **44**, D294–D300.
23. Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeke, A., Lang, B., Bely, B., Chemes, L.B., Davey, N.E., Deng, Z. *et al.* (2018) The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.*, **46**, D428–D434.
24. UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
25. PDBE-KB consortium (2019) PDBE-KB: a community-driven resource for structural and functional annotations. *Nucleic Acids Res.*, doi:10.1093/nar/gkz853.
26. Mir, S., Alhroub, Y., Anyango, S., Armstrong, D.R., Berrisford, J.M., Clark, A.R., Conroy, M.J., Dana, J.M., Deshpande, M., Gupta, D. *et al.* (2018) PDBE: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, **46**, D486–D492.
27. Bertolin, A.P., Mansilla, S.F. and Gottifredi, V. (2015) The identification of translesion DNA synthesis regulators: Inhibitors in the spotlight. *DNA Repair (Amst)*, **32**, 158–164.
28. Dherin, C., Gueneau, E., Francin, M., Nunez, M., Miron, S., Liberti, S.E., Rasmussen, L.J., Zinn-Justin, S., Gilquin, B., Charbonnier, J.-B. *et al.* (2009) Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. *Mol. Cell Biol.*, **29**, 907–918.
29. Wojtaszek, J., Liu, J., D'Souza, S., Wang, S., Xue, Y., Walker, G.C. and Zhou, P. (2012) Multifaceted recognition of vertebrate Rev1 by translesion polymerases ζ and κ . *J. Biol. Chem.*, **287**, 26400–26408.
30. Pustovalova, Y., Bezsonova, I. and Korzhnev, D.M. (2012) The C-terminal domain of human Rev1 contains independent binding sites for DNA polymerase η and Rev7 subunit of polymerase ζ . *FEBS Lett.*, **586**, 3051–3056.
31. Moldovan, G.-L., Pfander, B. and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665–679.
32. Gilljam, K.M., Feyzi, E., Aas, P.A., Sousa, M.M.L., Müller, R., Vågbo, C.B., Catterall, T.C., Liabakk, N.B., Slupphaug, G., Drablos, F. *et al.* (2009) Identification of a novel, widespread, and functionally important PCNA-binding motif. *J. Cell Biol.*, **186**, 645–654.
33. Hara, K., Uchida, M., Tagata, R., Yokoyama, H., Ishikawa, Y., Hishiki, A. and Hashimoto, H. (2018) Structure of proliferating cell nuclear antigen (PCNA) bound to an APIM peptide reveals the universality of PCNA interaction. *Acta Crystallogr. F Struct. Biol. Commun.*, **74**, 214–221.
34. Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M. and Kuriyan, J. (1996) Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell*, **87**, 297–306.
35. Hishiki, A., Hashimoto, H., Hanafusa, T., Kamei, K., Ohashi, E., Shimizu, T., Ohmori, H. and Sato, M. (2009) Structural basis for novel interactions between human translesion synthesis polymerases and proliferating cell nuclear antigen. *J. Biol. Chem.*, **284**, 10552–10560.
36. Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays*, **22**, 997–1006.
37. Boehm, E.M., Powers, K.T., Kondratick, C.M., Spies, M., Houtman, J.C.D. and Washington, M.T. (2016) The proliferating cell nuclear antigen (pcna)-interacting protein (pip) motif of dna polymerase η mediates its interaction with the c-terminal domain of rev1. *J. Biol. Chem.*, **291**, 8735–8744.
38. Boehm, E.M. and Washington, M.T. (2016) R.I.P. to the PIP: PCNA-binding motif no longer considered specific: PIP motifs and other related sequences are not distinct entities and can bind multiple proteins involved in genome maintenance. *Bioessays*, **38**, 1117–1122.
39. Davey, N.E., Cyert, M.S. and Moses, A.M. (2015) Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun. Signal.*, **13**, 43.
40. 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.
41. Whitewood, A.J., Singh, A.K., Brown, D.G. and Goult, B.T. (2018) Chlamydial virulence factor TarP mimics talin to disrupt the talin-vinculin complex. *FEBS Lett.*, **592**, 1751–1760.
42. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
43. Campellone, K.G., Brady, M.J., Alamares, J.G., Rowe, D.C., Skehan, B.M., Tipper, D.J. and Leong, J.M. (2006) Enterohaemorrhagic Escherichia coli Tir requires a C-terminal 12-residue peptide to initiate EspF-mediated actin assembly and harbours N-terminal sequences that influence pedestal length. *Cell. Microbiol.*, **8**, 1488–1503.
44. Brady, M.J., Campellone, K.G., Ghildiyal, M. and Leong, J.M. (2007) Enterohaemorrhagic and enteropathogenic Escherichia coli Tir proteins trigger a common Nck-independent actin assembly pathway. *Cell. Microbiol.*, **9**, 2242–2253.
45. Weiss, S.M., Ladwein, M., Schmidt, D., Ehinger, J., Lommel, S., Städing, K., Beutling, U., Disanza, A., Frank, R., Jänsch, L. *et al.* (2009) IRSp53 links the enterohemorrhagic E. coli effectors Tir and EspFU for actin pedestal formation. *Cell Host Microbe*, **5**, 244–258.
46. de Groot, J.C., Schlüter, K., Carius, Y., Quedenau, C., Vingadassalom, D., Faix, J., Weiss, S.M., Reichelt, J., Standfuss-Gabisch, C., Lesser, C.F. *et al.* (2011) Structural basis for complex formation between human IRSp53 and the translocated intimin receptor Tir of enterohemorrhagic E. coli. *Structure*, **19**, 1294–1306.
47. Aitio, O., Hellman, M., Kazlauskas, A., Vingadassalom, D.F., Leong, J.M., Saksela, K. and Permi, P. (2010) Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21743–21748.
48. Cheng, H.-C., Skehan, B.M., Campellone, K.G., Leong, J.M. and Rosen, M.K. (2008) Structural mechanism of WASP activation by the enterohaemorrhagic E. coli effector EspF(U). *Nature*, **454**, 1009–1013.
49. Bouvard, D., Pouwels, J., De Franceschi, N. and Ivaska, J. (2013) Integrin inactivators: balancing cellular functions in vitro and in vivo. *Nat. Rev. Mol. Cell Biol.*, **14**, 430–442.
50. Izard, T., Tran Van Nhieu, G. and Bois, P.R.J. (2006) Shigella applies molecular mimicry to subvert vinculin and invade host cells. *J. Cell Biol.*, **175**, 465–475.

51. Hamiaux, C., van Eerde, A., Parsot, C., Broos, J. and Dijkstra, B.W. (2006) Structural mimicry for vinculin activation by IpaA, a virulence factor of *Shigella flexneri*. *EMBO Rep.*, **7**, 794–799.
52. Park, H., Lee, J.H., Gouin, E., Cossart, P. and Izard, T. (2011) The rickettsia surface cell antigen 4 applies mimicry to bind to and activate vinculin. *J. Biol. Chem.*, **286**, 35096–35103.
53. Park, H., Valencia-Gallardo, C., Sharff, A., Tran Van Nhieu, G. and Izard, T. (2011) Novel vinculin binding site of the IpaA invasin of *Shigella*. *J. Biol. Chem.*, **286**, 23214–23221.
54. Ren, Y., Yip, C.K., Tripathi, A., Huie, D., Jeffrey, P.D., Walz, T. and Hughson, F.M. (2009) A structure-based mechanism for vesicle capture by the multisubunit tethering complex Dsl1. *Cell*, **139**, 1119–1129.
55. Suckling, R.J., Poon, P.P., Travis, S.M., Majoul, I.V., Hughson, F.M., Evans, P.R., Duden, R. and Owen, D.J. (2015) Structural basis for the binding of tryptophan-based motifs by δ -COP. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14242–14247.
56. Loewen, C.J.R., Roy, A. and Levine, T.P. (2003) A conserved ER targeting motif in three families of lipid binding proteins and in Opi1p binds VAP. *EMBO J.*, **22**, 2025–2035.
57. Kaiser, S.E., Brickner, J.H., Reilein, A.R., Fenn, T.D., Walter, P. and Brunger, A.T. (2005) Structural basis of FFAT motif-mediated ER targeting. *Structure*, **13**, 1035–1045.
58. Phillips, M.J. and Voeltz, G.K. (2016) Structure and function of ER membrane contact sites with other organelles. *Nat. Rev. Mol. Cell Biol.*, **17**, 69–82.
59. Stanhope, R., Flora, E., Bayne, C. and Derré, I. (2017) IncV, a FFAT motif-containing *Chlamydia* protein, tethers the endoplasmic reticulum to the pathogen-containing vacuole. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 12039–12044.
60. Arisue, N. and Hashimoto, T. (2015) Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol. Int.*, **64**, 254–259.
61. Marti, M. and Spielmann, T. (2013) Protein export in malaria parasites: many membranes to cross. *Curr. Opin. Microbiol.*, **16**, 445–451.
62. de Koning-Ward, T.F., Dixon, M.W.A., Tilley, L. and Gilson, P.R. (2016) Plasmodium species: master renovators of their host cells. *Nat. Rev. Microbiol.*, **14**, 494–507.
63. Boddey, J.A., Hodder, A.N., Günther, S., Gilson, P.R., Patsiouras, H., Kapp, E.A., Pearce, J.A., de Koning-Ward, T.F., Simpson, R.J., Crabb, B.S. *et al.* (2010) An aspartyl protease directs malaria effector proteins to the host cell. *Nature*, **463**, 627–631.
64. Soundararajan, M., Roos, A.K., Savitsky, P., Filippakopoulos, P., Kettenbach, A.N., Olsen, J.V., Gerber, S.A., Eswaran, J., Knapp, S. and Elkins, J.M. (2013) Structures of Down syndrome kinases, DYRKs, reveal mechanisms of kinase activation and substrate recognition. *Structure*, **21**, 986–996.
65. Sitz, J.H., Baumgärtel, K., Hämmerle, B., Papadopoulos, C., Hekerman, P., Tejedor, F.J., Becker, W. and Lutz, B. (2008) The Down syndrome candidate dual-specificity tyrosine phosphorylation-regulated kinase 1A phosphorylates the neurodegeneration-related septin 4. *Neuroscience*, **157**, 596–605.
66. Tinti, M., Kiemer, L., Costa, S., Miller, M.L., Sacco, F., Olsen, J.V., Carducci, M., Paoluzi, S., Langone, F., Workman, C.T. *et al.* (2013) The SH2 domain interaction landscape. *Cell Rep.*, **3**, 1293–1305.
67. 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.
68. Kaneko, T., Huang, H., Zhao, B., Li, L., Liu, H., Voss, C.K., Wu, C., Schiller, M.R. and Li, S.S.-C. (2010) Loops govern SH2 domain specificity by controlling access to binding pockets. *Sci. Signal.*, **3**, ra34.
69. Liu, B.A., Jablonowski, K., Shah, E.E., Engelmann, B.W., Jones, R.B. and Nash, P.D. (2010) SH2 domains recognize contextual peptide sequence information to determine selectivity. *Mol. Cell. Proteomics*, **9**, 2391–2404.
70. Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J. and Grütter, M.G. (1996) Structural basis for specificity of Grb2-SH2 revealed by a novel ligand binding mode. *Nat. Struct. Biol.*, **3**, 586–589.
71. Mészáros, B., Erdos, G. and Dosztányi, Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
72. Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Micetic, I., Camilloni, C., Davey, N., Dosztányi, Z., Mészáros, B., Monzon, A.M. *et al.* (2018) MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.*, **46**, D471–D476.
73. Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C.J., Aspromonte, M.C., Davey, N.E., Davidović, R., Dosztányi, Z. *et al.* (2017) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.*, **45**, D219–D227.
74. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
75. Jehl, P., Manguy, J., Shields, D.C., Higgins, D.G. and Davey, N.E. (2016) ProViz—a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Res.*, **44**, W11–W15.
76. Krystkowiak, I. and Davey, N.E. (2017) SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res.*, **45**, W464–W469.
77. Sebesta, M., Cooper, C.D.O., Ariza, A., Carnie, C.J. and Ahel, D. (2017) Structural insights into the function of ZRANB3 in replication stress response. *Nat. Commun.*, **8**, 15847.
78. Wojtaszek, J., Lee, C.-J., D’Souza, S., Minesinger, B., Kim, H., D’Andrea, A.D., Walker, G.C. and Zhou, P. (2012) Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) ζ , and Pol κ . *J. Biol. Chem.*, **287**, 33836–33846.
79. Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F., Quemener, C., Le Du, M.-H., Márquez, J.A. *et al.* (2013) Structure of the MutL α C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.*, **20**, 461–468.
80. Huang, C.C., Meng, E.C., Morris, J.H., Pettersen, E.F. and Ferrin, T.E. (2014) Enhancing UCSF Chimera through web services. *Nucleic Acids Res.*, **42**, W478–W484.