FEBS Letters

FEBS PRESS science publishing by scientists

REVIEW

Uncovering domain motif interactions using high-throughput protein–protein interaction detection methods

Sobia Idrees[1,2], Keshav Raj Paudel[2], Tayyaba Sadaf[2] and Philip M. Hansbro[2]

1 School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
2 Centre for Inflammation, Centenary Institute and Faculty of Science, School of Life Sciences, University of Technology Sydney, Australia

Correspondence
S. Idrees, Centre for Inflammation, Centenary Institute and Faculty of Science, School of Life Sciences, University of Technology Sydney, 15 Broadway, Ultimo, Sydney 2007, NSW, Australia
Tel: +61-02-9565203
E-mail: sobia.idrees@uts.edu.au

(Received 17 October 2023, revised 9 January 2024, accepted 18 February 2024)

doi:10.1002/1873-3468.14841

Edited by Christian Griesinger

Protein–protein interactions (PPIs) are often mediated by short linear motifs (SLiMs) in one protein and domain in another, known as domain–motif interactions (DMIs). During the past decade, SLiMs have been studied to find their role in cellular functions such as post-translational modifications, regulatory processes, protein scaffolding, cell cycle progression, cell adhesion, cell signalling and substrate selection for proteasomal degradation. This review provides a comprehensive overview of the current PPI detection techniques and resources, focusing on their relevance to capturing interactions mediated by SLiMs. We also address the challenges associated with capturing DMIs. Moreover, a case study analysing the BioGrid database as a source of DMI prediction revealed significant known DMI enrichment in different PPI detection methods. Overall, it can be said that current high-throughput PPI detection methods can be a reliable source for predicting DMIs.

Keywords: domain–motif interactions; high-throughput methods; PPI databases; protein–protein interactions

Proteins interact with their partners through different modules mediating different cell functions [1,2]. The two main modules are known as Short Linear Motifs (SLiMs), which mediate domain–motif interactions (DMIs), and globular domains, which are engaged in domain–domain interactions (DDIs) and DMIs. In addition to these modules, other critical features include coiled-coil domains involved in protein–protein interactions (PPIs), transmembrane domains facilitating membrane associations, and proteins that fold upon binding. These modules collectively contribute to the diverse repertoire of protein interaction mechanisms, which is essential in controlling a wide range of cellular mechanisms within organisms [3,4]. While many proteins require a well-defined structure for their functionality, a significant portion of an organism's proteome comprises polypeptide segments that are unlikely to adopt a defined three-dimensional structure yet remain functional. These segments are termed intrinsically disordered regions (IDRs) [5]. Because IDRs typically lack bulky hydrophobic amino acids, they are unable to form the well-organised hydrophobic core characteristic of structured domains. Consequently, their functionality differs from the classical structure–function paradigm associated with globular, structured proteins [6,7]. Moreover, proteins, thriving in a cellular milieu with high concentrations, manifest diverse states extending beyond the commonly acknowledged native and amyloid configurations [8]. The discovery of proteins undergoing liquid–liquid

Abbreviations
AP-MS, affinity purification coupled mass spectrometry; CoFrac-MS, co-fraction coupled mass spectrometry; DDIs, domain-domain interactions; DMIs, domain motif interactions; ELM, eukaryotic linear motif; *E*-score, enrichment score; hPPIs, human protein–protein interactions; IDRs, intrinsically disordered regions; PPIs, protein–protein interactions; SLiMs, short linear motifs; Y2H, yeast two-hybrid.

phase separation is rapidly growing. Given the likelihood that, at the high concentrations within cells, most proteins would adopt a liquid condensed state, this state should be recognised as a fundamental protein state, alongside the native and amyloid states [9]. For example, RNA-binding proteins like TDP43 and FUS undergo liquid-like phase separation, forming droplets where RNA concentration is high in the nucleus. In the cytoplasm, characterised by low RNA concentration, these proteins give rise to solid-like pathological condensates [10]. SLiMs are recurrent linear peptide microdomains of 2–15 consecutive amino acid residues, primarily found within IDRs [3,11,12,13]. According to an estimate, there are approximately 100 000 motifs involved in protein binding, most of which are found in IDRs. The flexibility of SLiMs allows them to function in disordered and structured regions, contributing to the diverse mechanisms by which proteins interact and carry out specific functions within cells [14-16]. Alternative exons exhibit a notable enrichment of IDRs, showing their importance in functional diversity [17,18]. Interactions facilitated by SLiMs are transient and exhibit low affinity, typically within 1–150 μM [19,20]. It is important to note that SLiMs usually comprise only 2–5 precisely defined positions, posing a challenge for their identification using experimental and computational approaches [21]. SLiMs mediate various cellular processes through PPIs, specifically DMIs, where rapid response is necessary for transmissions [22,23]. SLiMs serve as intriguing molecular switches within proteins, orchestrating rapid and reversible transitions between different functional states. These compact and evolutionarily conserved peptide sequences act like regulatory modules, enabling proteins to respond dynamically to cellular signals or environmental changes. With the introduction of a single mutation, these SLiMs can induce substantial alterations in the protein's behaviour, allowing it to toggle between distinct functionalities. This inherent flexibility underscores the importance of SLiMs as critical elements in cellular control mechanisms, contributing to the dynamic regulation of diverse biological processes. However, it is essential to note that the effectiveness of SLiMs as molecular switches can vary depending on the specific context, the nature of the protein, and the type of mutation introduced. Some SLiMs may be highly sensitive to single mutations, while others might require more complex alterations. Understanding the nuanced roles of SLiMs as molecular switches provides valuable insights into the intricacies of protein regulation and cellular signalling pathways [24]. The presence of motif residues interacting with the domain suggests that these positions of residues are likely to undergo evolutionary conservation. Notably, a substantial portion of SLiMs includes two or more conserved hydrophobic residues, as exemplified by the nuclear export sequence (NES), which contains four residues [25], and a single mutation of TQG to TQT can result in synaptic transport in the neuronal cells [26]. The small size and evolutionary plasticity of SLiMs suggest that these linear motifs are inclined to arise independently, facilitating the discovery of new motifs that share interaction partners [3,21]. The impact of SLiMs on protein functionality was first suggested in the 1970s and confirmed in 1990 when the KDEL motif was studied in conjunction with the ERD2 receptor. The identification and experimental validation of these motifs remain challenging. Additionally, the proportion of validated motifs compared to the total number is likely to be relatively small [27]. Different new computational tools and methods have been developed to ease the process of SLiM prediction from the protein sequence data. The main repositories maintaining motif data include eukaryotic linear motif database (ELM database) [28], PROSITE [29], Linear Motif mediated Protein Interaction Database (LMPID) [30], Minimotif-Miner [31], PepCyber [32] and Scansite [33]. During recent years, SLiMs have gained popularity because of their key residues that have shown involvement in subcellular localisation, post-translational modifications (PTMs), regulatory functions, protein trafficking, signal transduction, controlling cell cycle, and stabilising scaffolding process [34,35].

SLiM discovery/prediction tools

The potential for false-positive outcomes has always made the development of new bioinformatics techniques for SLiM prediction difficult [36]. Significant advancements have been made in motif discovery in recent years, and various new computational methods have been devised to discover SLiMs, but most of the SLiM prediction tools rely on sequence information and known data (Table 1). Nowadays, structure-based methods are also gaining attention as these can play a crucial role in predicting SLiM-based interactions by leveraging information about the three-dimensional arrangement of proteins. These tools utilise experimental or predicted structures of protein complexes to analyse the spatial orientation and interactions of SLiMs within the binding interfaces. By considering the structural context, these tools can provide insights into the specificity and affinity of SLiM interactions, aiding in the identification of potential binding partners. Additionally, structure-based approaches contribute to

Table 1. Available SLiM/DMI prediction tools.

	Tool	Description	Availability	Reference
SLiM discovery tools based on known motifs	PROSITE	The first linear motif catalogue was called PROSITE. These days, this database focuses on globular domains and protein signatures	http://prosite.expasy.org/	[40,41]
	MnM (Minimotif Miner)	Predicts SLiMs in proteins based on functional information of other proteins	http://mnm.engr.uconn.edu	[31,42]
	Scansite	Scansite predicts important motif predictions for cell signalling. It compares user sequences to recognised motifs using profile-based searches	http://scansite.mit.edu	[33,43]
	SLiMProb (Short Linear Motif Probability)	SLiMProb searches local protein data against user-defined patterns	http://bioware.ucd.ie/slimsearch2.html	[44,45]
	ScanProsite	Scanprosite analyses public databases and user-defined protein sequences for patterns using regular expression searches	http://prosite.expasy.org/scanprosite/	[46,47]
	ELM database (eukaryotic linear motif database)	The SLiMs of eukaryotes in this database have been manually selected and verified through experimentation. It is one of the most extensive resources for functional SLiM analysis	http://elm.eu.org	[24,28]
	iSPOT (Infer Sequence Prediction of Target)	iSPOT discovers PDZ, SH3, and WW SLiMs based on structural information	http://cbm.bio.uniroma2.it/ispot	[48]
	iELM (Interactions of Eukaryotic Linear Motif)	Predicts the occurrence of known motifs using PPI data	http://i.elm.eu.org	[37]
	AMS (AutoMotifServer)	Trained support vector machine (SVM) is used to predict motifs	http://code.google.com/p/automotifserver/	[49]
De novo SLiM discovery tools	qPMS7	LDMS patterns are used without homology correction	http://pms.engr.uconn.edu/downloads/qPMS7.zip	[50]
	SLIDER (LDMS CMM tool)	Uses PPIs to finds correlated motifs	http://bioinformatics.uhasselt.be	[51,52]
	TEIRESIAS	Motifs are searched using text patterns	http://code.google.com/p/teiresias	[53]
	Pratt	Predicts motifs without homology correction using over-representation regex	http://www.ebi.ac.uk/Tools/pfa/pratt/	[54,55]
	NestedMICA (Nested Motif Independent Component Analysis)	It uses reference proteins to identify enriched motifs	http://www.sanger.ac.uk/Software/analysis/nmica/	[56]
	SLiMMaker (Short Linear Motif Maker)	Aligns peptide sequences and generates regex consensus sequences	http://slimsuite.blogspot.com.au/	[57]
	PepSite	Predicts DMI using structural information	http://pepsite2.russelllab.org/	[58]
	ANCHOR	Uses regex in context to disorder regions	http://anchor.enzim.hu	[58]
	MotifCluster	Predicts correlated motif	http://bmf.colorado.edu/motifcluster	[59]
	MoRFpred (MoRF predictor)	Finds areas inside the IDR	http://biomine.ece.ualberta.ca/MoRFpred/	[60]
	SLiMScape (SLiM plugin for Cytoscape)	Analyses SLiMs as Cytoscape plugin	http://apps.cytoscape.org/apps/slimscape	[61]
	D-MIST (Domain–Motif Interaction from Structural Topology)	Uses PDB's structural information to predict DMI	N/A	[62]
	D-MOTIF (LDMS CMM tool)	Finds correlated motifs using PDB information	http://meme-suite.org/ http://bioware.ucd.ie/~compass/biowareweb//	[63,64] [65]

Table 1. (Continued).

	Tool	Description	Availability	Reference
	SLiMDisc (Short Linear Motif Discovery)	Using a heuristic method, this tool ranks over-represented motifs in unrelated proteins		
	SLiMFinder (Short Linear Motif Finder)	Using this method, one can find overrepresented motifs in unrelated proteins	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/slimfinder.php	[66]
	QSLiMFinder (Query SLiMFinder)	SLiM discovery tool based on queries that are more sensitive and specific	http://bioware.soton.ac.uk	[66]
	SLiMPrints (Short Linear Motif Fingerprints)	Finds conservation footprints using statistical models	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/slimprints.php	[67]
	MFSPSSMPred (Masked, Filtered and Smoothed Position Specific Scoring Matrix-based Predictor)	Predicts motifs from sequence characteristics or evolutionary conservation	http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html	[68]
	MEME (Multiple Em for Motif Elicitation)	Finds DNA/protein motifs using EM	http://meme.nbcr.net	[69,70]
	SLiMPred (Short Linear Motif Predictor)	Predicts SLiMs in protein sequences	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/slimpred_legacy.php	[71]
	FIRE-pro (Finding Informative Regulatory Elements in proteins)	Uses mutual information to find correlated motifs	https://tavazoielab.c2b2.columbia.edu/FIRE-pro/	[72]
	motif-x	Creates fixed position motifs by combining background amino acid frequencies with overrepresented peptides	http://motif-x.med.harvard.edu/	[73]
	DILIMOT (Discovery of Linear MOTifs)	Identifies overrepresented motifs in a group of proteins interacting with the target protein	http://dilimot.russelllab.org/	[74]
	GLAM2 (Gapped Local Alignment of Motifs)	Finds over-represented motifs using Gibbs sampling	http://bioinformatics.org.au/glam2	[75]
	MOTIPS (MOTIf analysis pipeline)	Uses DMI in conjunction with short peptides to forecast over-represented profiles	http://motips.gersteinlab.org/	[76]
User-defined tools	FIMO (Find Individual Motif Occurrences)	MEME profiles are used to query public databases or proteins specified by the user	http://meme.sdsc.edu	[77]
	3of5 (3of5 regex search tool)	Finds sequences using regex	http://dkfz.de/mga2/3of5/3of5.html	[78]
	PRESTO (Protein Regular Expression Search Tool)	Finds SLiMs using regex searches against local protein data	http://slimsuite.blogspot.com.au/	[57]
	SLiMSearch 2.0	Finds pre-defined motifs in the proteome	http://bioware.ucd.ie/slimsearch2.html	[44]
	MAST (Motif Alignment and Search Tool)	Utilises multiple motifs to identify new ones	http://meme-suite.org/	[63]
	SLiMEnrich 1.5.1	Identifies and assesses enrichment of DMIs	http://shiny.slimsuite.unsw.edu.au/SLiMEnrich/	[36]
	SLiMAn	Predicts DMIs from interactomes	https://sliman.cbs.cnrs.fr	[79]
	DoReMe	Motif occurrence predictions by using cellular contextual information	http://doremi.jensenlab.org/	[80]

understanding the dynamic nature of SLiM-mediated interactions, capturing conformational changes and flexibility in the binding sites. To date, several methods including iELM server [37], and MoRFchibi SYSTEM [38] have been developed to utilise structural information to accurately predict DMIs. Moreover, Artificial Intelligence (AI)-driven prediction tools, utilising machine learning (ML) algorithms, analyse PPI datasets to decode complex patterns associated with SLiMs. These tools can minimise false positives, a common limitation in traditional bioinformatic approaches, thereby enhancing prediction specificity. The integration of AI can contribute to the identification of novel SLiMs and refines our understanding of their functional relevance. One recent advancement, the MotSASi method, enhanced the prediction of authentic functional SLiMs. MotSASi integrates sequence variant information and structural analysis of the energetic impact of single amino acid substitutions (SAS) in SLiM-Receptor complex structures. This involves constructing a SAS tolerance matrix, indicating the tolerance of each position to one of the 19 possible SAS. Focusing on three SLiMs related to intracellular protein trafficking (phospho-independent tyrosine-based motif (NPx[Y/F]), type 1 PDZ-binding motif ([S/T]x[V/I/L]COOH), and tryptophan-acidic motif ([L/M]xW[D/E])), MotSASi demonstrated that the inclusion of variant and structure information improves the prediction of authentic SLiMs, diminishes false positives, and enhances the categorisation of variants within SLiMs. Therefore, it's important to develop new computational methods that can utilise sequence, structure, and variant information [39].

SLiM-mediated protein–protein interactions

Protein interactions commonly involve globular domains, known as DDIs, where one protein's domain interacts with that of another. These interactions frequently entail substantial surface contacts between the interacting protein domains. [19,81]. Another type of interaction involves DMIs, facilitated by SLiMs, which constitute a distinct subset of PPIs. Due to their short length, SLiMs are well-suited for mediating processes that demand quick reactions [19]. SLiMs are particularly pivotal for PTMs, a specialised subset of transient PPIs, by serving as specific target sites. Research by Neduva *et al.* suggests that a significant portion, ranging from 15% to 40% of protein interactions, are based on SLiMs. However, it is worth noting that only a fraction of such interactions have been identified to date [82]. The ELM database categorises SLiMs into six main classes, each serving distinct functions. These classes include cleavage sites (CLV), ligand binding sites (LIG), subcellular targeting sites (TRG), sites susceptible to post-translational modifications (MOD), degradation sites (DEG), and docking sites (DOC) [83,84]. Merely a small portion (2–3 a.a.) of the SLiM residues are discovered to be involved in binding interactions or cellular functioning [13]. A study claims that these SLiMs can mediate protein binding. For instance, the plant plasma membrane H+-ATPase C-terminal phosphorylated peptide interacts with the 14-3-3 regulatory complex [85].

Public protein–protein interaction repositories

The correct functioning of cells depends on interactions between proteins (PPIs). The precise number of PPIs is still unknown, while estimates place it between 130 000 and 650 000 [86]. A more thorough understanding of the protein network is made possible by collecting these PPIs in specialised databases. The Database of Interacting Proteins (DIP) was the first database designed to preserve PPI data [87]. Public repositories of PPIs are now expanding quickly and assist in finding new SLiMs and storing PPI data. There are already several PPI repositories available, helping in studying protein networks, as shown in Table 2.

High-throughput PPI detection methods and DMIs

As we explore the intricate world of cellular organisation and activities, it is clear that protein interactions play a vital role as the dynamic elements. DMIs have attracted much interest as they provide an opportunity for finding new SLiMs. Unfortunately, low-throughput studies have primarily shaped our understanding of DMIs, providing a limited window into this complex network. In the last several years, a revolutionary change has happened by introducing high-throughput techniques to identify DMIs. It has shifted our focus to techniques including protein/peptide arrays, affinity purification, yeast two-hybrid experiments, and phage platforms [108,109]. These cutting-edge, high-throughput methods have produced significant PPI data essential for understanding the function of SLiMs and protein complex prediction. Accurate predictions may become more difficult because of false positives and negatives, which are a constant challenge that comes with efficiency. Numerous studies have employed high-throughput experimental methods to

Table 2. Public PPI resources/repositories.

	Repository	Types of data	Species	Availability	Reference
Primary databases (curation based on experimental data/ literature)	DIP (Database of Interacting Proteins)	PPIs	All	http://dip.doe-mbi.ucla.edu/	[87]
	HPRD (Human Protein Reference Database)	PTMs, PPIs, subcellular localisation	Human	http://www.hprd.org/	[88]
	BioGRID (Biological General Repository for Interaction Datasets)	PPIs, transcription data	All	http://wiki.thebiogrid.org/doku.php/statistics	[89]
	MINT (Molecular INTeraction database)	DNA/RNA and protein interactions	All	http://mint.bio.uniroma2.it/mint/	[90]
	DOMINO	DMIs	Mouse, human, yeast and rat	http://mint.bio.uniroma2.it/domino/	[91]
	BIND (Biomolecular Interaction Network Database)	PPIs	All	http://bond.unleashedinformatics.com/	[92,93]
	IntAct	PPIs	All	http://www.ebi.ac.uk/intact/	[94]
Meta databases (curations based on experimental data and integration with other interaction repositories)	HINT (High-quality INTeractomes)	PPIs	All	http://hint.yulab.org/	[95]
	PINA (Protein Interaction Network Analysis)	PPIs	All	http://cbg.garvan.unsw.edu.au/pina/	[96]
	MPIDB	PPIs	Microbial	http://www.jcvi.org/mpidb/	[97,98]
	APID	PPIs	All	http://bioinfow.dep.usal.es/apid/	[99,100]
	Cpdb (ConsensusDB)	Different sorts of interactions	Human, Yeast, Mouse	http://consensuspathdb.org/	[101]
	iRefWeb	PPIs	All	http://wodaklab.org/iRefWeb/	[102]
Prediction databases (curations based on experimental and predicted data)	PIPs (Protein–protein interaction prediction)	PPIs	Human	http://www.compbio.dundee.ac.uk/www-pips/index.jsp	[103]
	STRING	PPIs	All	http://string-db.org/	[104]
	UniHI (Unified Human Interactome)	PPIs	Human	http://www.unihi.org/	[105]
	OPHID (Online Predicted Human Interaction Database)	PPIs	Human	http://ophid.utoronto.ca	[106,107]

generate extensive PPI datasets across various domain families [110-112]. In recent years, there has been a collective effort within the SLiM field to devise large-scale approaches tailored for the comprehensive characterisation of SLiMs (Table 3) [109]. In a recent study, researchers compared the efficacy of biotinylated peptide pulldown and the protein interaction screen on a peptide matrix (PRISMA) coupled with mass spectrometry. Through testing eight distinct peptide sequences with varying affinities for three specific protein domains (KEAP1 Kelch, MDM2 SWIB, and TSG101 UEV), the study revealed that biotin-peptide pulldown outperformed PRISMA in validating SLiMs. Notably, tandem peptide repeats enhanced interaction capture, underscoring the need to consider method development parameters for effective affinity capture MS-based validation of SLiM-based interactions from cell lysates [20]. Moreover, a new method, thermal proximity coaggregation (TPCA), has been introduced to monitor the dynamics of native protein complexes in living cells. TPCA leverages the coaggregation of proteins within a complex during heat denaturation, employing a cellular shift assay to produce melting curves for numerous proteins. Validated through the detection of known protein complexes, TPCA unveiled cell-specific interactions across diverse cell lines. This high-throughput and system-wide approach to studying protein complex dynamics demonstrates its promise in identifying complexes influenced by diseases [113]. High-throughput PPI data availability has made it possible to construct cutting-edge computational methods for SLiM predictions. Although these tools are a huge advancement, it is essential to recognise their inherent limitations because they might occasionally yield false positives. One potential solution to this problem is to apply gene ontology (GO), gene expression profiles, and high-throughput data to the computational algorithms used to improve their accuracy and consistency. This comprehensive method can improve our understanding of SLiMs and their function in forming the complex web of cellular interactions [111]. In this review, we will discuss some common high-throughput methods to identify PPIs and SLiM-mediated interactions.

Yeast two-hybrid

The yeast two-hybrid (Y2H) technique stands out as a robust method for identifying PPIs within yeast cells. Designating interacting proteins as "bait" and "prey," this approach activates reporter genes upon interaction, leading to distinct colour reactions or growth on specific media. Y2H has proven effective for exploring genome-wide interactions in diverse organisms, including *Saccharomyces cerevisiae*, bacteriophage T7, humans, and *Caenorhabditis elegans* [114,115]. Serving as a powerful tool in systems biology, it facilitates the investigation of large interactomes and enhances our understanding of disease aetiology by examining protein interactions within a system [116]. Y2H offers two primary screening methods: the array method and the library method. The library approach involves searching for pairwise interactions between proteins of interest (bait and prey) within cDNA libraries. However, it is susceptible to false positives, where proteins are incorrectly identified. Moreover, interaction partner identification necessitates colony PCR and sequencing methods, rendering this approach expensive and time-consuming. Conversely, the array strategy, also known as the matrix approach, entails the direct mating of a pool of baits and a pool of prey with various yeast mating types to identify interactions. Although this method is automated, facilitating the exploration of genome-wide interactomes, it may overlook certain interactions, termed false negatives [117,118] (Fig. 1A). In recent years, Y2H has demonstrated exceptional sensitivity, proficient in detecting subtle or fleeting PPIs, specifically in context of tissue-specific interactions and dynamic assemblies, particularly those mediated by SLiMs. Functioning inside live yeast cells, the Y2H technique creates an *in vivo* setting, enhancing the relevance of studying SLiM function in various cellular contexts, including tissue-specific interactions. Its applicability to large-scale screening enables the identification of multiple interactions simultaneously. However, researchers need to be cautious of potential false positives in Y2H results, which might have arisen from artefacts or nonspecific interactions, necessitating additional validation steps. Furthermore, Y2H might have limitations in capturing interactions mediated by PTMs, a crucial aspect for some SLiMs. Additionally, it is essential to recognise that Y2H, taking place in yeast cells, might not have fully recapitulated the complexity of interactions observed in higher eukaryotic cells [108,109]. Recently, a variation of Y2H has been developed known as protein domain mapping using Yeast 2 Hybrid-Next-Generation Sequencing (DoMY-Seq) to address the limitations associated with conventional methods. By integrating Y2H and next-generation sequencing, DoMY-Seq constructs a fragment library derived from a specified open reading frame. Leveraging a Y2H reporter system alongside next-generation sequencing, DoMY-Seq facilitates the swift identification of interacting domains, offering a detailed plot of the binding interface with high resolution. The optimisation of this method, exemplified with established protein pairs like KRAS-CRAF, underscores

Table 3. A comparative summary of high-throughput methods used to capture SLiM-mediated interactions.

Category	Method	Description	Pros	Cons
Binding assay methods	Protein array	This method aims to immobilise the target protein on the surface and probe it with a tagged peptide or protein	Large-scale PPI detection and PPIs connected to PTMs Sample consumption is less compared to other techniques Quantitative information	Labour intensive set-up Stability of proteins
	Peptide array	The peptides with known sequences are chemically synthesised using peptide arrays	Known peptide sequences are used	Peptide arrays rely on known peptide sequences, and their performance is sensitive to the specific sequence context. If the chosen peptides do not fully capture the variability and nuances of the actual binding motifs, false positives or negatives may occur. If the peptides on the array do not account for PTMs relevant to the binding domain, it can result in false negatives. Additionally, PTMs introduced during the array preparation may contribute to false positives
	Protein interaction screen on peptide matrix (PRISMA)	PRISMA utilises extended peptide array protocols, incorporating AP-MS for exploratory analysis and *de novo* discovery of novel SLiM-binding partners, with applications in analysing disease mutations and microprotein interactions	Identification of proteins that interact with SLiMs and the convenience of investigating the effects of PTMs	Reduced detection of low-affinity interactions Constrained by limited Mass Spectrometry accessibility, and potential noise in the data
Display methods	Peptide phage display	Examines the binding specificities of peptide binding domains	Finds non-binding peptides	The design of the peptide library can introduce bias Affinity maturation encompasses iterative rounds of selection and amplification of phage-displayed peptides, with the goal of enhancing their binding affinity to the target. The efficiency of this process can vary, and achieving significant improvements in affinity may necessitate numerous rounds, rendering the procedure time-consuming During the amplification and mutagenesis steps, unintended mutations can occur
			It can incorporate modified/non-natural amino acids	Limited coverage
	Yeast surface display	Yeast cells with plasmid DNA-expressing peptides are used in this technique. The cell surface is covered in these peptides	Direct mapping of interactions regulated by PTMs between displayed peptides or proteins and their binding partners	High cost
			The method is versatile, accommodating both peptides and proteins as bait	Fluorescence-Activated Cell Sorting (FACS) might restrict the screening capabilities in terms of throughput, efficiency, and potential noise in the data

Table 3. (Continued).

Category	Method	Description	Pros	Cons
	Bacterial-surface display	Bacterial display assays involve linking peptides to bacterial outer membrane proteins for multivalent display	Scalability	The cost of the designed oligonucleotide library constrains the size of the library
			Comprehensive insights into both binders and non-binders	FACS sorting limits the screening capabilities
				Bacterial cells do not possess the machinery for eukaryotic PTMs. This constraint can impact the folding and modification of proteins, especially those dependent on post-translational modifications specific to eukaryotic systems
	mRNA display	mRNA display involves screening peptide-RNA fusions in a cell-free format, wherein *in vitro* translated peptides are covalently linked to their encoding mRNA *via* a puromycin linkage, facilitating selections against an immobilised bait followed by reverse transcription and sequencing	Enables the creation of extensive peptide libraries (10^{13})	The size of the library is constrained by the expense of the designed oligonucleotide library
				Enriching binders necessitates multiple rounds of selections
In cell binding assays	Pheromone signalling competitive growth assay	A yeast pheromone signalling-based assay explores the specificity of LP motif docking interactions with yeast G1 cyclin Cln2, utilising deep mutational scanning	Generates a list of peptides ranked by affinity	Limited by the transduction of cells
				Next-generation sequencing (NGS) incurs associated costs
	MAPK competitive growth assay	A yeast growth assay utilising exogenous mitogen-activated protein kinase (MAPK) signalling components to investigate MAPK docking motifs	Occurs in physiological conditions	Intricate biases may be introduced
			Generates a list of peptides ranked by affinity	Constrained to the studied MAPK kinase
				Restricted by cell transduction during library construction
Functional assays	Degradation assays	Large-scale functional assays can provide novel insights into degradation motifs, or degrons, a class of SLiMs that promote protein degradation	Provides functional insights	Lacks details on the binding partner
			Offers a semi-quantitative readout	Constrained by cell transduction during library construction and/or FACS sorting
	Transactivation assays	Functional studies on Transactivation Domains (TADs), a type of SLiM interacting with transcriptional coregulators to regulate transcription, often employ these assays measuring cell viability or fluorescent reporter production as indicators of transcriptional activity	Provides functional insights	Lacks details on the binding partner
			Offers a semi-quantitative readout	Constrained by cell transduction during library construction and/or FACS sorting
Quantitative binding assays	MRBLE-pep	MRBLE-pep assays use spectrally encoded lanthanide beads with synthesised peptides for medium-throughput quantitative measurement of protein–peptide interactions	Simultaneous determination of multiple affinities	Constrained by the synthesis of peptide-linked beads
				Necessitates specialised equipment
	Hold-up assay	Hold-up assays involve immobilising peptides on beads, distributing them in a filter plate, allowing proteins to bind at equilibrium, and collecting unbound proteins through filtration	Determining affinities	Peptide synthesis required for using a peptide bait.
				Production of bait protein needed when using purified proteins
			Numerous detection options. Versatile, accommodating both peptides and proteins as bait	Constrained by the costs associated with peptide synthesis and detection

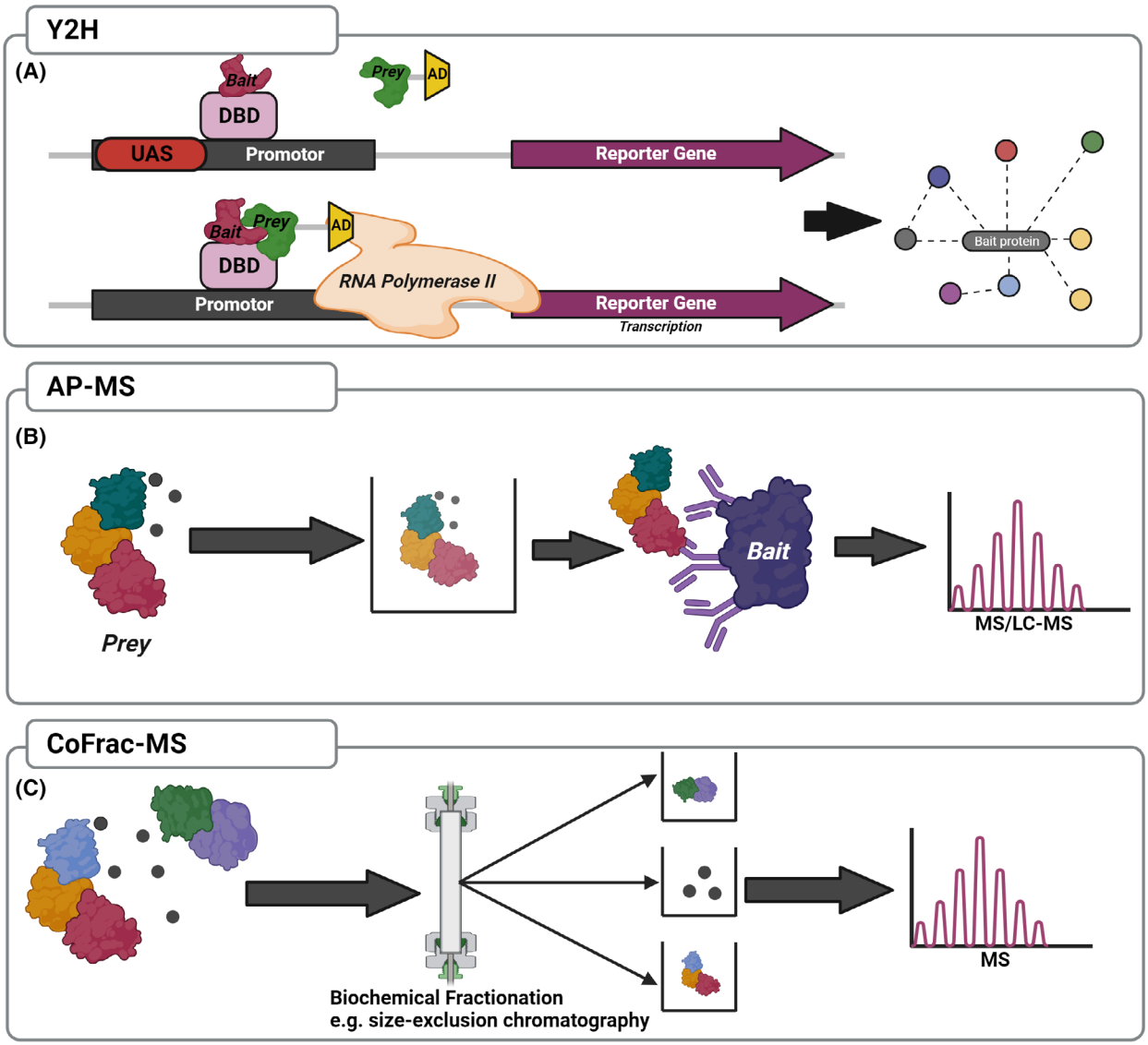

Fig. 1. High-throughput PPI detection methods. (A) The yeast two-hybrid (Y2H) system needs yeast's transcription factor activation, DNA binding domains (DBD), and activation domains (AD). A prey (possible interacting protein) is fused with the AD, and a bait (protein of interest) is fused with the DBD. The bait protein binds with a binding site in the reporter gene's promoter region after fusing with the DBD. When the prey protein fuses with the AD, it then binds to the bait protein to initiate the expression of the gene. (B) In Affinity purification coupled mass spectrometry (AP-MS) a tag protein is attached to the protein of interest (bait). Co-purified proteins are those that bind to the tagged protein. Mass spectrometry (MS) is then used to identify these proteins. (C) In co-fractionation followed by mass spectrometry (CoFrac-MS), protein extract undergoes thorough fractionation to separate protein complexes, subsequently identified by MS (created using BioRender.com).

its effectiveness in unbiased and efficient domain mapping for diverse protein interactions [119].

Affinity purification coupled mass spectrometry

Affinity purification coupled mass spectrometry (AP-MS) stands as a formidable technology in systems biology, transforming the landscape of PPI discovery. This method harnesses mass spectrometry (MS) to enable biochemical techniques, such as affinity purification and chemical cross-linking, providing invaluable insights into proteome-wide interactions across diverse biological systems [120]. In the AP-MS workflow, the protein of interest is genetically fused to a specific tag, allowing for its identification through an affinity

column designed for tag association or a specialised antibody [94]. The flexibility of the AP-MS methodology accommodates different approaches, ranging from a single-step purification using tags like the Flag tag to the more intricate two-step purification method, often preferred for its efficacy. This method involves dual tagging of proteins at the C- or N-terminal ends or with tags such as 6xHis- and Strep-tags, and the use of tandem affinity purification (TAP), featuring a cleavage site between tags, which facilitates the isolation of multiprotein complexes containing the tagged protein. Subsequent MS analysis is employed to identify the constituents of these complexes, making the two-step purification approach renowned for its sensitivity and specificity, rendering it an excellent choice for in-depth PPI investigations [121,122]. However, this method is not without its challenges. The two-step purification approach, while effective, introduces additional complexity to the experimental workflow, requiring careful planning and execution. The resource-intensive nature of AP-MS, demanding specialised equipment and expertise for optimal results, may limit its accessibility to some research settings. Furthermore, like any technique, AP-MS may introduce potential artefacts or biases, necessitating rigorous validation steps to ensure the reliability of the obtained results. AP-MS is proficient at isolating and studying stable protein complexes, rendering it suitable for investigating enduring interactions within tissues. Nonetheless, its effectiveness in probing dynamic assemblies is contingent upon the stability of the interactions involved. In instances where interactions are transient or context-dependent, AP-MS may encounter limitations in efficiently capturing these fleeting and context-specific events. Therefore, researchers should carefully design experiments to address the dynamics of the protein interactions they aim to study [114] (Fig. 1B).

Co-fractionation coupled mass spectrometry

Co-fractionation coupled mass spectrometry (CoFrac-MS) has emerged as a potent strategy for in-depth exploration of intricate protein interactions, encompassing both direct and indirect relationships [71]. In this methodology, protein extracts undergo a sophisticated fractionation process, often employing biochemical techniques such as size exclusion chromatography, followed by meticulous evaluation of the resulting fractions using mass spectrometry. Like AP-MS, CoFrac-MS has the capacity to unveil proteome-wide connections, providing a comprehensive view of cellular relationships [123]. The strength of CoFrac-MS lies in its ability to elucidate complex protein associations. It enables the differentiation between direct and indirect interactions among protein pairs, shedding light on true physical connections and those facilitated by intermediary partners. However, navigating this complexity presents a substantial challenge, particularly in distinguishing between the various types of interactions within intricate mixtures. While CoFrac-MS serves as a robust tool for identifying complex protein networks, interpreting the nature of these associations necessitates meticulous data analysis and integration with additional contextual information. Just like AP-MS, the effectiveness of CoFrac-MS in capturing dynamic events is influenced by interaction stability. In cases of transient or context-dependent interactions, CoFrac-MS effectiveness in capturing dynamic assemblies may be limited. Therefore, careful consideration of interaction characteristics is crucial when utilising CoFrac-MS for studying tissue-specific or dynamic protein assemblies. CoFrac-MS, on the positive side, offers a holistic view of protein interactions, providing valuable insights into the intricacies of cellular relationships. However, the interpretative challenge in distinguishing between direct and indirect interactions requires sophisticated computational tools and careful consideration of experimental conditions. Additionally, like other mass spectrometry-based techniques, CoFrac-MS demands specialised equipment and expertise, potentially limiting its accessibility in certain research settings [124] (Fig. 1C).

Five primary types of high-throughput experiments are now available to identify SLiM-mediated interactions (Table 3).

A case study to assess enrichment of DMIs in PPIs available in BioGrid database

To determine the viability of utilising data from public repositories for predicting DMIs, we evaluated their potential to capture interactions mediated by SLiMs effectively. Here, we have analysed human interactome retrieved from BioGrid 4.4.2 [retrieved: 2023-07-01] [89] database, known to be the most comprehensive PPI database. Most of the PPI repositories curate the same information from experimental evidence or literature. The quality of the interaction data has always been questioned. In one study, it was seen that despite curating the same data, there was a significant difference between different databases, raising questions about the interaction data's quality and reliability [95]. Therefore, to see if the PPI data stored in the BioGrid database is useful, we mapped proteins involved in

Table 4. PPI detection methods retrieved from the BioGrid database.

Method	Category	MI ontology	Interaction type	PPI[a]	DMI[b]	DMI enrichment[c] (4 s.f)
Two-hybrid PCA	Protein complementation assay	MI:0090	Direct interactions	44 862	70	71.57[d]
Affinity capture-MS Affinity capture-western Affinity capture-luminescence Biochemical activity Far western Affinity capture-RNA	Biochemical	MI:0401	Direct and indirect interactions	140 264	217	26.28[e]
Co-crystal structure	Structure-based predictions	MI:0105	Physical interactions	1174	11	86.61[d]

[a]Number of symmetric and non-redundant PPIs with protein pairs examined by Uniprot; [b]Known SLiM-Protein interactions from the ELM database; [c]Observed enrichment of known DMIs captured from PPIs; [d]False Discovery Rate (FDR) < 0.01; [e]FDR < 0.05.

PPIs to their protein sequences, including splice variants. The total number of human PPIs analysed was 302 854, involving 12 053 unique proteins. The protein names retrieved from the BioGrid database were mapped to their sequence data using Uniprot, and analysis was restricted to reviewed proteins only. 11 942 out of 12 053 proteins were successfully mapped to their protein sequences, which means most of the PPIs stored in BioGrid have their sequence data available. PPI data was then used to identify known SLiM-mediated interactions using the ELMi method of SLiMEnrich v1.5.1 [36], which works based on known interactions in the ELM database. BioGrid currently stores data generated by different techniques; we divided the dataset into four broad categories based on the Molecular Interactions (MI) ontologies [125]. These subsets of data were then used to see the enrichment of DMIs. Protein complementation assay and structure-based methods showed promising results with the highest enrichment scores [False Discovery Rate (FDR) < 0.05]. All PPI techniques, taken together, captured DMIs and demonstrated a noteworthy enrichment of DMIs as compared to random protein pairs. This was also observed in our recent study where different high-throughput methods showed significant enrichment [126] (Table 4).

Current challenges and outstanding questions

Protein–protein interaction detection techniques are powerful sources of identifying DMIs, but researchers often face several challenges and limitations. In this review, we discussed several challenges associated with capturing DMIs using PPI databases and detection methods. As we know, several PPI repositories store data related to different species. The main challenge is the quality of the PPI in these repositories, as they are typically inconsistent, and there might be interactions with high false-positive rates. This, in turn, can lead to false and inaccurate DMI predictions. Therefore, it is essential to assess the quality of the PPI in available repositories to ensure their reliability for DMI analysis [127]. PPI databases often lack 3D structural information on the interactions, which can also influence understanding DMIs and insights into biological mechanisms [128]. Furthermore, the conundrum of understudied proteins, whose biological functions remain enigmatic, poses a significant obstacle to effective research. The prevailing 'streetlight effect' in scientific inquiry tends to concentrate efforts on well-explored proteins, thereby restricting the exploration of potentially pivotal yet neglected ones. For instance, extensively studied proteins like the tumour suppressor p53 exemplify this bias. To overcome this challenge, community engagement is imperative through a survey aimed at identifying understudied proteins and determining the essential information required for in-depth mechanistic investigations. Such community-driven initiatives hold the promise of mitigating biases, fostering collaborative endeavours, and methodically advancing the molecular characterisation of understudied proteins [129,130]. Another big challenge is the possibility of significant noise and false positives in PPIs detected by high-throughput methods, making it difficult to distinguish true interactions from artefacts [131]. More importantly, integrating PPIs from different resources can be quite challenging due to differences in data formats, identifiers and curation methods [126,131,132]. Few databases have PPI interactions that are not experimentally validated, leading to uncertainty about their biological relevance. Relying on such PPI interactions can lead to invalid results and false-positive

DMIs [133]. Moreover, PPI data sometimes do not account for PTMs, which can significantly impact the identification of DMIs. Therefore, it is important to account for PTMs while uncovering DMIs [99]. Developing more comprehensive protein networks requires better integration methods/strategies to integrate data from different PPI resources efficiently. Cross-species analysis also needs to focus on finding highly conserved DMIs involved in different diseases. AI (e.g., machine learning) can offer a comprehensive assessment of SLiM interactions by leveraging sequence variant and structural information, providing an efficient approach in exploring this critical aspect of molecular biology. This emphasises the need to develop new machine learning-based computational methods to predict DMIs from PPIs in combination with structural information accurately. In short, a critical approach to PPI data analysis, including cautious data selection, integration, validation, and knowledge of the context and study objectives, is necessary to overcome these challenges.

Conclusion

The creation of profuse PPI datasets has become routine in the era of high-throughput experimental techniques. However, the importance of carefully evaluating the quality of detected PPIs as a valuable source of DMIs has become apparent due to the inherent error-proneness of high-throughput methods. The primary aim of this review is to serve as a resource for PPI detection methods, repositories, and SLiM prediction tools. In addition, we evaluated PPIs obtained from several high-throughput methods, which are indexed in BioGrid, to determine DMIs. Our analysis found a significant enrichment for all the approaches we looked at showing promise in identifying DMIs.

Acknowledgement

The authors would like to acknowledge the University of New South Wales, Sydney, and Dr Richard J. Edwards for his help in carrying out this work. Open access publishing facilitated by University of Technology Sydney, as part of the Wiley - University of Technology Sydney agreement via the Council of Australian University Librarians.

References

1 Bhattacharyya RP, Remenyi A, Yeh BJ and Lim WA (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* **75**, 655–680.
2 Armingol E, Officer A, Harismendy O and Lewis NE (2021) Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* **22**, 71–88.
3 Bhowmick P, Guharoy M and Tompa P (2015) Bioinformatics approaches for predicting disordered protein motifs. *Adv Exp Med Biol* **870**, 291–318.
4 Zanzoni A, Ribeiro DM and Brun C (2019) Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cell Mol Life Sci* **76**, 4407–4412.
5 Fuxreiter M, Tompa P and Simon I (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956.
6 van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 6589–6631.
7 Idrees S and Paudel KR (2023) Bioinformatics prediction and screening of viral mimicry candidates through integrating known and predicted DMI data. *Arch Microbiol* **206**, 30.
8 Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoynov S, Mahamid J, Saha S, Franzmann TM *et al.* (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077.
9 Fuxreiter M and Vendruscolo M (2021) Generic nature of the condensed states of proteins. *Nat Cell Biol* **23**, 587–594.
10 Maharana S, Wang J, Papadopoulos DK, Richter D, Pozniakovsky A, Poser I, Bickle M, Rizk S, Guillén-Boixet J, Franzmann TM *et al.* (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921.
11 Weatheritt RJ, Davey NE and Gibson TJ (2012) Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res* **40**, 7123–7131.
12 D'haeseleer P (2006) What are DNA sequence motifs? *Nat Biotechnol* **24**, 423–425.
13 Davey NE, van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H and Gibson TJ (2012) Attributes of short linear motifs. *Mol Biosyst* **8**, 268–281.
14 Tompa P, Davey NE, Gibson TJ and Babu MM (2014) A million peptide motifs for the molecular biologist. *Mol Cell* **55**, 161–169.
15 Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF and Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635–645.
16 Sobhy H (2016) A review of functional motifs utilized by viruses. *Proteomes* **4**, 3.

17 Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **37**, 509–516.
18 Hraber P, O'Maille PE, Silberfarb A, Davis-Anderson K, Generous N, McMahon BH and Fair JM (2020) Resources to discover and use short linear motifs in viral proteins. *Trends Biotechnol* **38**, 113–127.
19 Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, Trave G and Gibson TJ (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci* **13**, 6580–6603.
20 Kassa E, Jamshidi S, Mihalic F, Simonetti L, Kliche J, Jemth P, Sara Bergstrom L and Ivarsson Y (2023) Evaluation of affinity-purification coupled to mass spectrometry approaches for capture of short linear motif-based interactions. *Anal Biochem* **663**, 115017.
21 Edwards RJ and Palopoli N (2015) Computational prediction of short linear motifs from protein sequences. *Methods Mol Biol* **1268**, 89–141.
22 Gibson TJ (2009) Cell regulation: determined to signal discrete cooperation. *Trends Biochem Sci* **34**, 471–482.
23 Pancsa R and Fuxreiter M (2012) Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB Life* **64**, 513–520.
24 Dinkel H, van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kubań M *et al.* (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res* **42**, D259–D266.
25 Gibson TJ, 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.
26 Neduva V and Russell RB (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett* **579**, 3342–3345.
27 Semenza JC, Hardwick KG, Dean N and Pelham HR (1990) ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell* **61**, 1349–1357.
28 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.
29 Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M and Sigrist CJ (2006) The PROSITE database. *Nucleic Acids Res* **34**, D227–D230.
30 Sarkar D, Jana T and Saha S (2015) LMPID: a manually curated database of linear motifs mediating protein-protein interactions. *Database (Oxford)* **2015**, bav014.
31 Balla S, Thapar V, Verma S, Luong TB, Faghri T, Huang CH, Rajasekaran S, del Campo JJ, Shinn JH, Mohler WA *et al.* (2006) Minimotif Miner: a tool for investigating protein function. *Nat Methods* **3**, 175–177.
32 Gong W, Zhou D, Ren Y, Wang Y, Zuo Z, Shen Y, Xiao F, Zhu Q, Hong A, Zhou X *et al.* (2008) PepCyber:P~PEP: a database of human protein protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res* **36**, D679–D683.
33 Obenauer JC, Cantley LC and Yaffe MB (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31**, 3635–3641.
34 Lee JM, Hammaren HM, Savitski MM and Baek SH (2023) Control of protein stability by post-translational modifications. *Nat Commun* **14**, 201.
35 Davey NE and Morgan DO (2016) Building a regulatory network with short linear sequence motifs: lessons from the degrons of the anaphase-promoting complex. *Mol Cell* **64**, 12–23.
36 Idrees S, Perez-Bercoff A and Edwards RJ (2018) SLiMEnrich: computational assessment of protein-protein interaction data as a source of domain-motif interactions. *PeerJ* **6**, e5858.
37 Weatheritt RJ, Jehl P, Dinkel H and Gibson TJ (2012) iELM – a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res* **40**, W364–W369.
38 Malhis N, Jacobson M and Gsponer J (2016) MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res* **44**, W488–W493.
39 Martin M, Brunello FG, Modenutti CP, Nicola JP and Marti MA (2022) MotSASi: functional short linear motifs (SLiMs) prediction based on genomic single nucleotide variants and structural data. *Biochimie* **197**, 59–73.
40 Bairoch A (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Res* **21**, 3097–3103.
41 Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L and Xenarios I (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, D344–D347.
42 Lyon KF, Cai X, Young RJ, Mamun AA, Rajasekaran S and Schiller MR (2018) Minimotif Miner 4: a million peptide minimotifs and counting. *Nucleic Acids Res* **46**, D465–D470.
43 Ehrenberger T, Cantley LC and Yaffe MB (2015) Computational prediction of protein-protein interactions. *Methods Mol Biol* **1278**, 57–75.
44 Davey NE, Haslam NJ, Shields DC and Edwards RJ (2011) SLiMSearch 2.0: biological context for short linear motifs in proteins. *Nucleic Acids Res* **39**, W56–W60.
45 Edwards RJ, Paulsen K, Aguilar Gomez CM and Perez-Bercoff A (2020) Computational prediction of

disordered protein motifs using SLiMSuite. *Methods Mol Biol* **2141**, 37–72.
46 de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A and Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**, W362–W365.
47 Jakhmola S, Indari O, Kashyap D, Varshney N, Das A, Manivannan E and Jha HC (2021) Mutational analysis of structural proteins of SARS-CoV-2. *Heliyon* **7**, e06572.
48 Brannetti B and Helmer-Citterich M (2003) iSPOT: a web tool to infer the interaction specificity of families of protein modules. *Nucleic Acids Res* **31**, 3709–3711.
49 Plewczynski D, Basu S and Saha I (2012) AMS 4.0: consensus prediction of post-translational modifications in protein sequences. *Amino Acids* **43**, 573–582.
50 Dinh H, Rajasekaran S and Davila J (2012) qPMS7: a fast algorithm for finding (l, d)-motifs in DNA and protein sequences. *PLoS One* **7**, e41425.
51 Boyen P, Van Dyck D, Neven F, van Ham RC and van Dijk AD (2011) SLIDER: a generic metaheuristic for the discovery of correlated motifs in protein-protein interaction networks. *IEEE/ACM Trans Comput Biol Bioinform* **8**, 1344–1357.
52 Liu Y, Woods NT, Kim D, Sweet M, Monteiro AN and Karchin R (2011) Yeast two-hybrid junk sequences contain selected linear motifs. *Nucleic Acids Res* **39**, e128.
53 Rigoutsos I and Floratos A (1998) Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm. *Bioinformatics* **14**, 55–67.
54 Jonassen I, Collins JF and Higgins DG (1995) Finding flexible patterns in unaligned protein sequences. *Protein Sci* **4**, 1587–1595.
55 Ye K, Kosters WA and Ijzerman AP (2007) An efficient, versatile and scalable pattern growth approach to mine frequent patterns in unaligned protein sequences. *Bioinformatics* **23**, 687–693.
56 Dogruel M, Down TA and Hubbard TJ (2008) NestedMICA as an ab initio protein motif discovery tool. *BMC Bioinformatics* **9**, 19.
57 Edwards RJ, Paulsen K, Aguilar Gomez CM and Perez-Bercoff A (2020) Computational prediction of disordered protein motifs using SLiMSuite. *Methods Mol Biol* **1**, 37–72.
58 Dosztanyi Z, Meszaros B and Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746.
59 Leung HC, Siu MH, Yiu SM, Chin FY and Sung KW (2009) Clustering-based approach for predicting motif pairs from protein interaction data. *J Bioinform Comput Biol* **7**, 701–716.
60 Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN and Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75–i83.
61 O'Brien KT, Haslam NJ and Shields DC (2013) SLiMScape: a protein short linear motif analysis plugin for Cytoscape. *BMC Bioinformatics* **14**, 224.
62 Betel D, Breitkreuz KE, Isserlin R, Dewar-Darch D, Tyers M and Hogue CW (2007) Structure-templated predictions of novel protein interactions from sequence information. *PLoS Comput Biol* **3**, 1783–1789.
63 Bailey TL and Gribskov M (1997) Score distributions for simultaneous matching to multiple motifs. *J Comput Biol* **4**, 45–59.
64 Xiao P, Schiller M and Rajasekaran S (2019) Novel algorithms for LDD motif search. *BMC Genomics* **20**, 424.
65 Davey NE, Shields DC and Edwards RJ (2006) SLiMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res* **34**, 3546–3554.
66 Edwards RJ, Davey NE and Shields DC (2007) SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS One* **2**, e967.
67 Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ and Edwards RJ (2012) SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res* **40**, 10628–10641.
68 Fang C, Noguchi T, Tominaga D and Yamana H (2013) MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. *BMC Bioinformatics* **14**, 300.
69 Bailey TL and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28–36.
70 Brody S (2020) A comparison of the *Neurospora* and *Drosophila* clocks. *J Biol Rhythms* **35**, 119–133.
71 Mooney C, Pollastri G, Shields DC and Haslam NJ (2012) Prediction of short linear protein binding regions. *J Mol Biol* **415**, 193–204.
72 Lieber DS, Elemento O and Tavazoie S (2010) Large-scale discovery and characterization of protein regulatory motifs in eukaryotes. *PLoS One* **5**, e14444.
73 Chou MF and Schwartz D (2011) Biological sequence motif discovery using motif-x. *Curr Protoc Bioinformatics* **Chapter 13**, 13.15.1–13.15.24.

74 Neduva V and Russell RB (2006) DILIMOT: discovery of linear motifs in proteins. *Nucleic Acids Res* **34**, W350–W355.
75 Frith MC, Saunders NF, Kobe B and Bailey TL (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* **4**, e1000071.
76 Lam HY, Kim PM, Mok J, Tonikian R, Sidhu SS, Turk BE, Snyder M and Gerstein MB (2010) MOTIPS: automated motif analysis for predicting targets of modular protein domains. *BMC Bioinformatics* **11**, 243.
77 Grant CE, Bailey TL and Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018.
78 Seiler M, Mehrle A, Poustka A and Wiemann S (2006) The 3of5 web application for complex and comprehensive pattern matching in protein sequences. *BMC Bioinformatics* **7**, 144.
79 Reys V and Labesse G (2022) SLiMAn: an integrative web server for exploring short linear motif-mediated interactions in interactomes. *J Proteome Res* **21**, 1654–1663.
80 Horn H, Haslam N and Jensen LJ (2014) DoReMi: context-based prioritization of linear motif matches. *PeerJ* **2**, e315.
81 Idrees S, Paudel KR, Sadaf T and Hansbro PM (2023) How different viruses perturb host cellular machinery via short linear motifs. *EXCLI J* **22**, 1113–1128.
82 Neduva V and Russell RB (2006) Peptides mediating interaction networks: new leads at last. *Curr Opin Biotechnol* **17**, 465–471.
83 Dinkel H, Michael S, Weatheritt RJ, Davey NE, 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.
84 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.
85 Wurtele M, Jelich-Ottmann C, Wittinghofer A and Oecking C (2003) Structural view of a fungal toxin acting on a 14-3-3 regulatory complex. *EMBO J* **22**, 987–994.
86 Baspinar A, Cukuroglu E, Nussinov R, Keskin O and Gursoy A (2014) PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res* **42**, W285–W289.
87 De Las Rivas J and Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* **6**, e1000807.
88 Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A *et al.* (2009) Human protein reference database – 2009 update. *Nucleic Acids Res* **37**, D767–D772.
89 Oughtred R, Chatr-aryamontri A, Breitkreutz BJ, Chang CS, Rust JM, Theesfeld CL, Heinicke S, Breitkreutz A, Chen D, Hirschman J *et al.* (2016) BioGRID: a resource for studying biological interactions in yeast. *Cold Spring Harb Protoc* **2016**, pdb.top080754.
90 Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* **40**, D857–D861.
91 Ceol A, Chatr-aryamontri A, Santonico E, Sacco R, Castagnoli L and Cesareni G (2007) DOMINO: a database of domain-peptide interactions. *Nucleic Acids Res* **35**, D557–D560.
92 Bader GD, Betel D and Hogue CW (2003) BIND: the biomolecular interaction network database. *Nucleic Acids Res* **31**, 248–250.
93 Miryala SK, Anbarasu A and Ramaiah S (2018) Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **642**, 84–94.
94 Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N *et al.* (2014) The MIntAct project – IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358–D363.
95 Das J and Yu H (2012) HINT: high-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* **6**, 92.
96 Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, Biankin AV, Hautaniemi S and Wu J (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Res* **40**, D862–D865.
97 Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT and Uetz P (2008) MPIDB: the microbial protein interaction database. *Bioinformatics* **24**, 1743–1744.
98 Calderon-Gonzalez KG, Hernandez-Monge J, Herrera-Aguirre ME and Luna-Arias JP (2016) Bioinformatics tools for proteomics data interpretation. *Adv Exp Med Biol* **919**, 281–341.
99 Alonso-Lopez D, Campos-Laborie FJ, Gutierrez MA, Lambourne L, Calderwood MA, Vidal M and De Las Rivas J (2019) APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* **2019**, baz005.
100 Prieto C and De Las Rivas J (2006) APID: agile protein interaction DataAnalyzer. *Nucleic Acids Res* **34**, W298–W302.

101 Kamburov A, Stelzl U, Lehrach H and Herwig R (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* **41**, D793–D800.

102 Turner B, Razick S, Turinsky AL, Vlasblom J, Crowdy EK, Cho E, Morrison K, Donaldson IM and Wodak SJ (2010) iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)* **2010**, baq023.

103 McDowall MD, Scott MS and Barton GJ (2009) PIPs: human protein-protein interaction prediction database. *Nucleic Acids Res* **37**, D651–D656.

104 Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–D452.

105 Kalathur RKR, Pinto JP, Hernandez-Prieto MA, Machado RSR, Almeida D, Chaurasia G and Futschik ME (2014) UniHI 7: an enhanced database for retrieval and interactive analysis of human molecular interaction networks. *Nucleic Acids Res* **42**, D408–D414.

106 Brown KR and Jurisica I (2005) Online predicted human interaction database. *Bioinformatics* **21**, 2076–2082.

107 Pastrello C, Kotlyar M and Jurisica I (2020) Informed use of protein-protein interaction data: A focus on the integrated interactions database (IID). *Methods Mol Biol* **2074**, 125–134.

108 Blikstad C and Ivarsson Y (2015) High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Commun Signal* **13**, 38.

109 Davey NE, 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.

110 Rajagopala SV, Sikorski P, Kumar A, Mosca R, Vlasblom J, Arnold R, Franca-Koh J, Pakala SB, Phanse S, Ceol A *et al.* (2014) The binary protein-protein interaction landscape of *Escherichia coli*. *Nat Biotechnol* **32**, 285–290.

111 Zhang Y, Lin H, Yang Z and Wang J (2015) Integrating experimental and literature protein-protein interaction data for protein complex prediction. *BMC Genomics* **16** (Suppl 2), S4.

112 Li X, Wu M, Kwoh CK and Ng SK (2010) Computational approaches for detecting protein complexes from protein interaction networks: a survey. *BMC Genomics* **11** (Suppl 1), S3.

113 Tan CSH, Go KD, Bisteau X, Dai L, Yong CH, Prabhu N, Ozturk MB, Lim YT, Sreekumar L, Lengqvist J *et al.* (2018) Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. *Science* **359**, 1170–1177.

114 Bruckner A, Polge C, Lentze N, Auerbach D and Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* **10**, 2763–2788.

115 Bacon K, Blain A, Bowen J, Burroughs M, McArthur N, Menegatti S and Rao BM (2021) Quantitative yeast-yeast two hybrid for the discovery and binding affinity estimation of protein-protein interactions. *ACS Synth Biol* **10**, 505–514.

116 Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE *et al.* (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **125**, 801–814.

117 Elmore JM, Velasquez-Zapata V and Wise RP (2023) Next-generation yeast two-hybrid screening to discover protein-protein interactions. *Methods Mol Biol* **2690**, 205–222.

118 Gnanasekaran P and Pappu HR (2023) Yeast two-hybrid technique to identify protein-protein interactions. *Methods Mol Biol* **2690**, 1–8.

119 Castel P, Holtz-Morris A, Kwon Y, Suter BP and McCormick F (2021) DoMY-seq: a yeast two-hybrid-based technique for precision mapping of protein-protein interaction motifs. *J Biol Chem* **296**, 100023.

120 Lee PY and Low TY (2023) Identification and quantification of affinity-purified proteins with MaxQuant, followed by the discrimination of nonspecific interactions with the CRAPome interface. *Methods Mol Biol* **2690**, 299–310.

121 Zhong B, An Y, Gao H, Zhao L, Li X, Liang Z, Zhang Y, Zhao Q and Zhang L (2023) In vivo cross-linking-based affinity purification and mass spectrometry for targeting intracellular protein-protein interactions. *Anal Chim Acta* **1265**, 341273.

122 Mehta A, Kamal AHM, Cornelius S and Chowdhury SM (2023) Protein-protein interaction network mapping by affinity purification cross-linking mass spectrometry (AP-XL-MS) based proteomics. *Methods Mol Biol* **2690**, 255–267.

123 Low TY, Syafruddin SE, Mohtar MA, Vellaichamy A, Rahman NSA, Pung YF and Tan CSH (2021) Recent progress in mass spectrometry-based strategies for elucidating protein-protein interactions. *Cell Mol Life Sci* **78**, 5325–5339.

124 Luck K, Sheynkman GM, Zhang I and Vidal M (2017) Proteome-scale human interactomics. *Trends Biochem Sci* **42**, 342–354.

125 Jupp S and Burdett T (2015) A new ontology lookup service at EMBL-EBI. *Workshop on Semantic Web Applications and Tools for Life Sciences* 1546, 1–2.

126 Idrees S and Paudel KR (2024) Proteome-wide assessment of human interactome as a source of

capturing domain–motif and domain-domain interactions. *J Cell Commun Signal* e12014. https://doi.org/10.1002/ccs3.12014

127 Bader GD and Hogue CW (2002) Analyzing yeast protein-protein interaction data obtained from different sources. *Nat Biotechnol* **20**, 991–997.

128 Murray D, Petrey D and Honig B (2021) Integrating 3D structural information into systems biology. *J Biol Chem* **296**, 100562.

129 Kustatscher G, Collins T, Gingras AC, Guo T, Hermjakob H, Ideker T, Lilley KS, Lundberg E, Marcotte EM, Ralser M *et al*. (2022) Understudied proteins: opportunities and challenges for functional proteomics. *Nat Methods* **19**, 774–779.

130 Kustatscher G, Collins T, Gingras AC, Guo T, Hermjakob H, Ideker T, Lilley KS, Lundberg E, Marcotte EM, Ralser M *et al*. (2022) An open invitation to the understudied proteins initiative. *Nat Biotechnol* **40**, 815–817.

131 Peng X, Wang J, Peng W, Wu FX and Pan Y (2017) Protein-protein interactions: detection, reliability assessment and applications. *Brief Bioinform* **18**, 798–819.

132 Idrees S, Paudel KR and Hansbro PM (2024) Prediction of motif-mediated viral mimicry through the integration of host–pathogen interactions. *Arch Microbiol* **206**, 1–12.

133 Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M and Wiuf C (2008) Estimating the size of the human interactome. *Proc Natl Acad Sci USA* **105**, 6959–6964.