UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1877

Evolution and viral mimicry of short linear motif-mediated interactions

FILIP MIHALIČ

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1651-6206
ISBN 978-91-513-1628-4
URN urn:nbn:se:uu:diva-486764

Dissertation presented at Uppsala University to be publicly examined in A1:111a, BMC, Husargatan 3, 75237 Uppsala, Friday, 2 December 2022 at 13:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Leo James (MRC Laboratory of Molecular Biology, Cambridge, Great Britain).

Abstract

Mihalič, F. 2022. Evolution and viral mimicry of short linear motif-mediated interactions. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1877. 66 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1628-4.

Proteins are one of the most fundamental building blocks of life and their interactions regulate every cellular process. Historically they have been conceptualized as predominantly folded entities with well-defined secondary and tertiary structures. However, in recent decades, up to 50% of the human proteome has been shown to contain long disordered sequences that are flexible and unstructured in their natural environment. These intrinsically disordered regions exhibit low levels of sequence conservation and are enriched in short linear motifs (SLiMs). SLiMs are typically less then 10 amino acids long and serve as docking sites recognized by various globular domains. They exhibit a high degree of sequence degeneracy and evolutionary plasticity, allowing for rapid de novo emergence. SLiMs play crucial roles in a variety of cellular processes including cellular signalling, trafficking, transcriptional modulation, and protein degradation. Because they are small and degenerate, located in disordered regions, and form relatively weak interactions, they are difficult to identify using conventional high-throughput methods such as mass spectrometry and the yeast two-hybrid system. The same attributes that make them difficult to identify also make them ideal targets for viral SLiM mimicry, of which several examples have been described to date.

To address the elusive nature of SLiMs, we have developed a novel approach for the discovery of motif-mediated interactions at the proteome scale using proteomic peptide phage display. We constructed two separate phage libraries with either human or viral disordered regions displayed on their surface. These libraries were then subjected to phage display selections against over 300 globular domains, resulting in the identification of more than 1,700 potential novel interactions. We validated a subset of these interactions with affinity measurements and GST-pulldown assays, solved the crystal structure of human globular domains in complex with viral linear motifs, and demonstrated that the gained knowledge can be applied to design peptidomimetic inhibitors of viral replication. In addition, we demonstrated that direct binding of viral SLiMs to the N-terminal domain of clathrin disrupts cellular trafficking and identified the C-terminal domain of polyadenylate-binding protein 1 as a novel target for viral SLiM mimicry. Furthermore, we demonstrated that SARS-CoV-2 viral proteins possess both SLiMs that bind to human proteins, and globular domains that recognize human SLiMs, showcasing the versatility of SLiM-mediated interactions. Finally, we examined the evolutionary trajectory of the interaction between the SWIB domain of MDM2 and the SLiM of p53 TAD to describe an example of the extraordinary evolutionary plasticity of SLiM-mediated interactions.

Overall, the research presented in this thesis created the basis for an atlas of human motif-mediated interactions, yielded an extensive dataset of potential and validated cases of viral SLiM mimicry, and expanded our understanding of motif-mediated interactions from an evolutionary perspective.

Keywords: phage display, protein-protein interactions, large-scale discovery, short linear motifs, viral mimcry, viral hijacking

Filip Mihalič, Department of Chemistry - BMC, Box 576, Uppsala University, SE-75123 Uppsala, Sweden. Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1628-4
URN urn:nbn:se:uu:diva-486764 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-486764)

And you run and you run to catch up with the sun but it's sinking
Pink Floyd

It's never too late to quit
Per Jemth

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Benz, C., Ali, M., Krystkowiak, I., Simonetti, L., Sayadi, A., **Mihalic, F.,** Kliche, J., Andersson, E., Jemth, P., Norman, E. D., Ivarsson, Y. (2022) Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Molecular systems biology*, 18: Pages 1–27.

II. Kruse, T., Benz, C., Garvanska, D. H., Lindqvist, R., **Mihalic, F.,** Coscia, F., Inturi, R., Sayadi, A., Simonetti, L., Nilsson, E., Ali, M., Kliche, J., Morro, A. M., Mund, A., Andersson, E., McInerney, G., Mann, M., Jemth, P., Norman, E. D., Överby, A. K., Nilsson, J., Ivarsson, Y. (2021) Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nature Communications*, (2021)12:6761

III. **Mihalic, F.,** Simonetti, L., Giudice, G., Rubin Sander, M., Lindqvist, R., Akpiroro Peters, M. B., Benz, C., Kassa, E., Badgujar, D., Inturi, R., Ali, M., Krystkowiak, I., Sayadi, A., Andersson, E., Aronsson, H., Söderberg, O., Dobritzsch, D., Petsalaki, E., Överby, A. K., Jemth, P., Norman, E. D., Ivarsson, Y. Large-scale phage-based screening reveals extensive pan-viral mimicry of host short linear motifs. *Submitted manuscript.*

IV. **Mihalic, F.,** Benz, C., Kassa, E., Lindqvist, R., Simonetti, L., Inturi, R., Aronsson, H., Andersson, E., Chi, C. N., Davey, N., Överby, A. K., Jemth, P., Ivarsson, Y. Identification of motif-based interactions between SARS-CoV-2 protein domains and human peptide ligands pinpoints antiviral targets. *Submitted manuscript.*

V. Åberg, E.,* **Mihalic, F.,*** Farkhondehkish, P., Theys, N., Andersson, E., Jemth, P. Evolution of binding affinity between p53 and MDM2 across the animal kingdom. *Manuscript.*

VI. **Mihalic, F.,** Arcila, D., Pettersson, M. E., Farkhondehkish, P., Åberg, E., Andersson, E., Andersson, L., Betaneur, R. Jemth, P. Loss-of-affinity mutations and new interaction motifs underlie a dynamic evolution of the p53/MDM2 interaction in teleosts. *Manuscript.*

**Equal contribution*

Reprints were made with permission from the respective publishers.

Contents

Abbreviations*

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
SLiM	Short Linear Motif
PPI	Protein-protein interaction
PDB	The Protein Data Bank
ELM	The Eukaryotic Linear Motif database
PAM2	PABP-interacting motif 2
LiR	LC3-interacting region
NLS	Nuclear localization signal
NES	Nuclear export signal
AP-MS	Affinity purification coupled with mass spectrometry
Y2H	Yeast two-hybrid system
ProP-PD	Proteomic peptide phage display
K_D	Equilibrium dissociation constant
FP	Fluorescence polarisation
mP	millipolarisation unit
HD2	Human disorderome 2 library
FITC	Fluorescein isothiocyanate
RiboVD	Riboviria viral disorderome library
GST	Glutathione *S*-transferase
GFP	Green fluorescent protein
ORF	Open reading frame
NMR	Nuclear magnetic resonance
HSQC	Heteronuclear Single Quantum Coherence

**In benefit of conciseness the abbreviations for gene, protein and viral names are not stated*

Introduction

If we conceptualize DNA as the instruction manual for the creation of a living 0,0organism, proteins are the functional entities responsible for interpretation, and implementation of these instructions (with some notable exceptions, *e.g.,* catalytic RNA). As such, they are responsible for every aspect of construction and maintenance of a functional, living system. In general, proteins can be divided into two categories: those that catalyse chemical reactions (enzymes) and those that do not. The latter group of proteins, the "non-enzymes", often perform a regulatory role function, by modulating the enzymatic activity of their interaction partners through a complex network of protein-protein interactions (PPI). These interaction networks allow proteins to "communicate" and lead cellular processes in an optimal direction. It is estimated that the human interactome consists of 650,000 such PPI, majority of which remain to be discovered and adequately described[1].

Protein-protein interaction interfaces can vary considerably in terms of composition and buried solvent exposed area. Large protein complexes, such as the nuclear pore complex, typically contain many protein-protein interactions in which multiple interactants interact across large, extended interfaces[2]. In contrast, protein-protein interactions can also have small modular interfaces, as is often the case when a globular domain recognizes and interacts with a specific stretch of a disordered region of an interaction partner. Different motif-recognizing globular domains can identify specific amino acid sequences in a linear, sequential manner that does not require secondary or tertiary structure to support the precise arrangement of residues in the motif, eliminating spatial constraints and "simplifying" the interaction interface. These regions of lower structural complexity, recognised by specific globular domains, are referred to as Short Linear Motifs (SLiMs) and are often mimicked by viral proteins[3,4].

The work presented in this thesis, describes our efforts to develop a large-scale screening technique to systematically identify motif-mediated interactions at the proteome level, with a particular focus on the discovery of motif mimicry by viral proteins. This work is extended by an examination of the evolutionary history of a specific motif-mediated interaction between the E3 ubiquitin ligase MDM2 and cellular tumour antigen p53 proteins. The first part of the introduction will describe the relevant concepts, followed by a more detailed presentation of the main findings of my research.

Structural disorder in proteins

After the first crystal structure of a myoglobin protein was solved in 1958[5], the field of X-ray crystallography exploded, producing close to 200,000 detailed three-dimensional structures of proteins from organisms originating in all branches of life[6]. The first proteins whose structures were solved and studied were predominantly enzymes and proteins abundant in blood as they were easier to purify and had already known functions, making them more interesting to researchers and drug developers[7]. However, this had the effect of producing a biased view of proteins, summarised by the structure-function paradigm. This postulate states that the function of a protein is directly related to and derived from its three-dimensional structure[8]. In practice, this would mean that only those proteins with a defined three-dimensional structure have a function and that all unstructured regions are just arbitrary connecting linkers produced by random evolutionary processes. In recent decades, an alternative view regarding the disordered regions of proteins has emerged. It has been recognized that the intrinsically disordered regions play a vital functional role in the context of full-length proteins and increasing number of research groups are focusing their attention on the proteins and protein regions that lack a well-defined tertiary structure[8-14].

Intrinsically disordered regions

According to standard definition, intrinsically disordered proteins are proteins that do not have a well-defined secondary or tertiary structure[15]. This description is however misleading, since it only accurately describes a minute subset of proteins who retain the disordered nature throughout the entirety of their amino acid sequence. A more realistic representation of a typical eukaryotic protein is presented in Figure 1 using as the example the histone acetyltransferase p300 (CBP/p300) protein, in which several globular domains are connected with each other through long stretches of disordered regions. CBP/p300 is a histone acetylase and functions as a transcriptional regulator affecting hundreds of transcription factors[16]. It plays an important role in several cellular processes such as cell growth, proliferation, differentiation and apoptosis[17]. The CBP/p300 protein is also one of the most thoroughly studied proteins[18]. Disordered regions of CBP/p300 have been shown to be critical for its function, not only as flexible linkers of folded domains, but also as docking sites for a variety of interaction partners[8,18-20]. CBP/p300 is not an isolated example as recent estimates suggest that between 35 and 50 percent of all human proteins contain intrinsically disordered regions of 30 amino acids or more[21,22]. This prediction was recently reinforced by a release of the protein structure-prediction tool Alphafold[23], which elegantly highlights the interplay between globular domains and long stretches of disordered regions connecting them in a typical eukaryotic protein (Figure 1B).

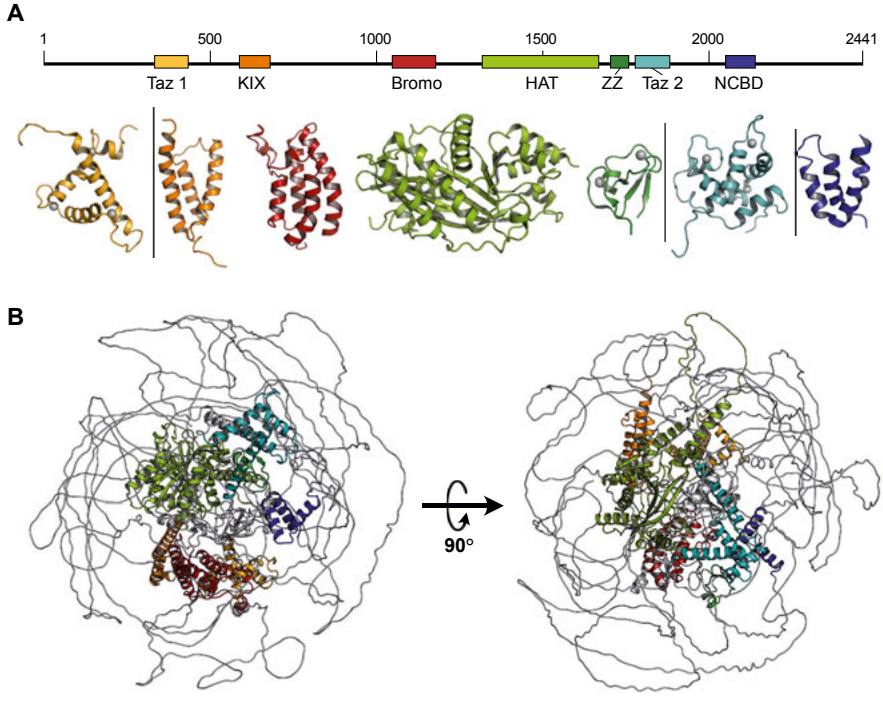


Figure 1: Structure of the CBP/p300 protein. (A) Schematic representation of CBP/p300 with folded domains color-coded to match the schematic representation. PDB accession numbers for CBP/p300 domains are 1l3e, 2kwf, 5nu5, 6v90, 6ds6, 6fgn, 6es7 in order from left to right. **(B)** Alphafold prediction for the full length p300. The relevant domains are color-coded as in **(A)** and all disordered regions are shown in grey. The high degree of disorder is evident from a grey "cloud" surrounding the folded domains.

Additionally, the concept of intrinsically disordered proteins introduces a duality that does not reflect the true nature of the phenomenon. The notion that the region is either ordered (globular) or disordered is incorrect, as different regions have different propensities to fold and essentially form a continuum between the unfolded and folded states[24]. Put simply, the folded domains have a more funnel-shaped folding energy landscape, while the energy landscape of intrinsically disordered domains is more rugged, with multiple conformers having similar folding energy minima (Figure 2). This results in disordered regions fluctuating between different, energetically similar conformers, in contrast to the folded domains, which collapse into one, globally dominant fold[25]. Since three-dimensional folding must be encoded in the amino acid sequence of proteins, so does the disorder. Sequences lacking the secondary and tertiary structure usually contain a high content of polar and charged

amino acids, as well as flexible (glycine) and secondary structure-disrupting amino acids (proline)[26-28].

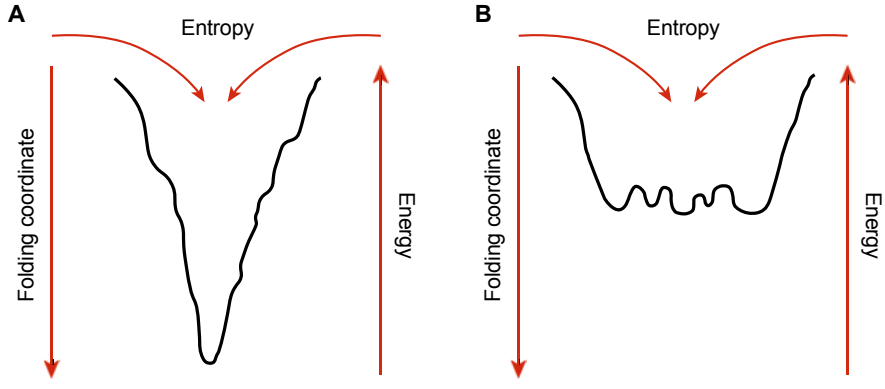


Figure 2: Energy landscape diagram of folding. (A) In the case of globular domains, the folding "funnel" has steep walls, and leads to a single, energetically stable conformation. **(B)** For intrinsically disordered regions the folding diagram shows multiple conformations with similar energetic minima.

Since stable structure is a prerequisite for enzymatic activity, disordered regions of proteins cannot perform this function. Instead, they : I) function as flexible linkers between domains; II) facilitate interactions with other proteins, nucleic acids or other ligands; and III) are enriched in post-translation modifications sites which have a regulatory role[28]. As such, they are mainly involved in cell signalling and regulation[14].

Intrinsically disordered regions of proteins are enriched with an important class of regulatory sequences termed short linear motifs.

Short Linear Motifs

Short linear motifs (SLiMs) are 3-10 amino-acid long sequences that have no clearly defined secondary or tertiary structure. As such, they are predominantly found in the disordered regions of proteins[4]. The function of SLiMs derives from their ability to form protein-protein interactions with their interaction partners. SLiMs serve as docking sites for globular domains that specifically recognize them, and the interaction can disrupt or modulate a variety of cellular functions[29]. The most comprehensive collection of experimentally validated SLiMs is the eukaryotic linear motif database (ELM; http://elm.eu.org), which contains 3934 entries of SLiMs as of October 2022[30]. However, this number is believed to be much higher, with some estimates of one million SLiMs in the human proteome[29]. Even if the estimate of all SLiMs in a given proteome is exaggerated by an order of magnitude, the discrepancy between experimentally validated SLiMs and bioinformatic predictions is substantial.

Typically, only 3-4 amino acids of a given SLiM are specifically recognised by the binding partner and contribute most of the binding affinity[4]. The recognition event is often accompanied by the induction of secondary structure in the SLiM sequence (folding upon binding)[4,31], as shown in Figure 3. The SLiM-mediated protein-protein interactions typically have medium to low micromolar affinities, which is ideal for interactions between proteins involved in signalling and regulatory cascades[29]. These cascades are organised into complex groups of interacting proteins with multiple positive and negative feedback loops designed to quickly respond to changes in the environment[14]. As such the interactions between regulatory proteins have evolved to rapidly address the changes in environmental stimuli rather than simply mediate high-affinity binding. High affinity interactions are detrimental to such regulatory processes because slow dissociation rates prevent rapid response to input signals. SLiM-mediated interactions are however not limited to signalling pathways, but also play important roles in other processes such as subcellular localization and protein stability[32].

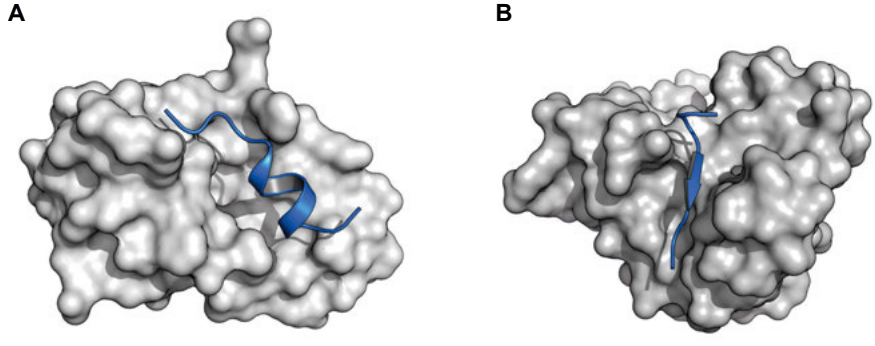


Figure 3: Two examples of SLiMs binding to globular domains. (A) A motif from p53 with the sequence ET**F**SDL**W**KL**L**PEN (residues that are part of the core motif are in bold) is bound to the SWIB domain of the MDM2 protein. Upon binding, the p53 SLiM adopts an α helical structure (PDBid: 1YCR). **(B)** The LC3 interaction motif (LiR motif) EED**I**LV**V**T of NDP52 is bound to LC3C. Upon binding, the SLiM adopts the secondary structure of a β strand (PDBid: 3VVW). In both cases, the globular domains are shown in gray, and the SLiMs in blue.

Given the small interaction interface between the two partners and the limited number of specificity-determining residue positions, SLiMs exhibit high evolutionary plasticity, as they can arise *de novo* from just one or two mutations in a disordered region of any given protein sequence[32]. Since the minimum requirement for binding is sometimes as few as 2 amino acids, they likely emerge by convergent evolution, allowing rapid SLiM proliferation and their ubiquitous occurrence in eukaryotes[3,29,30].

Another feature of SLiMs is the degenerated nature of the recognition sequences. A variety of SLiM sequences can often be accommodated at the

binding interfaces, with certain amino acids being preferred but often not obligatory[33,34]. A good example of degeneracy of SLiMs is the PABP-interacting motif 2 (PAM2), which interacts with the C-terminal domain of the polyadenylate-binding protein 1 (PABPC1 PABC). Two variants of PAM2 can be found in the ELM database. They mediate binding to overlapping but distinct regions of PABPC1 PABC (designated PAM2_1 and PAM2_2 in ELM)[35,36]. A total of 16 instances of PAM2_1 motifs from human proteins are reported in the ELM database, and the canonical consensus recognition sequence is [LFP][NS][PIVTAFL].A..(([FY].[PYLF])|(W..)), where the full stop represents any amino acid, parenthesis indicate different alternatives, and brackets indicate allowed amino acids in a given position. Clearly a wide variety of sequences accommodate these loose but vital constrains. To complicate matters further, the two versions of PAM2 are completely dissimilar, with the consensus recognition motif for the PAM2_2 being ((WPP)|([FL][PV][APQ]))EF.PG.PWKG. This demonstrates that the SLiM in this case does not have stringent specificity determinants and a variety of different sequences can facilitate binding to PABPC1 PABC. Despite the high degree of degeneracy, this motif shows a relatively high affinity in the low micromolar to nanomolar range[37,38].

Furthermore, SLiMs often serve as sites of post-translation modification. This adds another layer of control to cellular signalling and regulatory pathways. Many linear motifs can be phosphorylated, which can enhance or abolish binding of the corresponding globular domain. An example of phosphorylation affecting the interaction is LC3C protein binding to the conserved LC3-interacting region (LiR) motif [EDST].(0-2)[WFY][^RKPG][^PG][ILV]; (^ denotes prohibitory residues). The motif starts with an acidic amino acid or a serine/threonine that can be phosphorylated[39]. If the motif contains serine/threonine, phosphorylation can increase the binding affinity by up to 17-fold[40]. If the LiR motif instead contains acidic amino acids that precede the central recognition motif, the interaction cannot be modulated (this is the case with the NDP52 LiR motif shown in Figure 3). This illustrates the versatility of SLiMs, where binding of the same recognition motif can be constitutive, in the absence of phosphorylation control, or modulated by phosphorylation, in a sort of molecular on/off switch.

Viral hijacking of host cellular functions

Viruses are obligate cellular parasites. This means that they cannot reproduce without the biomolecular cellular machinery provided by host cells upon infection. As such, they are extremely specialised and constantly compete with host defence mechanisms to achieve host cell take over. To successfully infect and proliferate, viruses must enter the host and host cells, avoid or suppress host systemic and cellular immune responses, facilitate transcription and

translation of viral genomic material, stimulate the assembly and packaging of functional viral particles, and enable the egress of viral particles from infected cells[41]. All of these crucial steps of viral infection are coordinated by direct interaction, interference, and manipulation of host cellular proteins[3]. This however raises an interesting conundrum. How can viruses with a limited number of proteins, constrained by their compact genomes, orchestrate such a sophisticated takeover of so many different cellular signalling pathways and processes? In recent decades, SLiM mimicry, shown schematically in Figure 4, has emerged as at least a partial answer to this question.

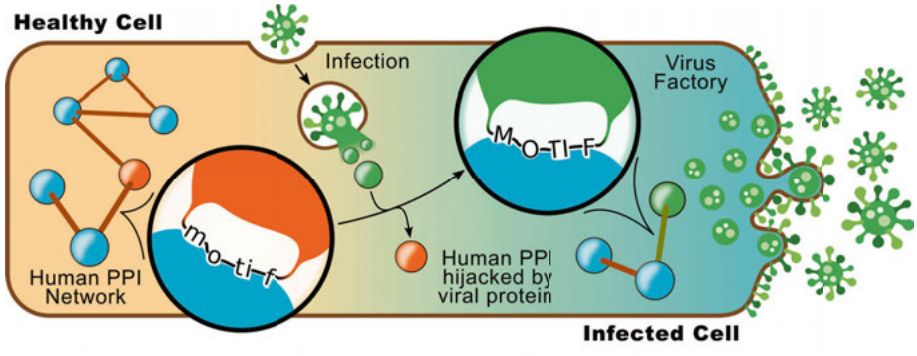


Figure 4: Schematic representation of viral SLiM mimicry. After virus entry, viral proteins interfere with endogenous signalling and regulatory pathways to tailor them to the needs of the pathogen. Viral proteins mimic SLiMs to outcompete endogenous interactions, leading to a pathological state that results in suppression of the immune response, viral genome amplification and translation, virion formation, and viral egress (illustration kindly provided by Leandro Simonetti).

Viral mimicry of SLiMs

Viral genomes are usually very compact. For example, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with one of the largest genomes among RNA viruses[42], encodes only 14 structural and non-structural proteins that, after proteolytic cleavage, result in a total number of 29 functional proteins[43]. Compared to the estimated size of the human proteome of approximately 20,000 proteins[44], the discrepancy is obvious. Given the limited resources, viral proteins have evolved the ability to perform multiple temporally and spatially separated functions[45-47]. An example of the multifunctional nature of viral proteins is the ebolavirus matrix protein VP40[48,49]. Depending on arrangement of protein subunits, VP40 serves as a regulator of viral transcription early in the life cycle. As protein monomers accumulate, VP40 reassembles and serves as a trafficking protein targeting the cell membrane. Upon reaching the membrane, electrostatic interactions again trigger restructuring, and the protein facilitates the budding process of viral particles[49]. This demonstrates both the temporally (early in the infection life cycle compared to later

stages of viral infection) and spatially (cytoplasmic localization compared to association with the cell membrane) distinct functions of the same protein.

To achieve this multifunctionality, SLiM mimicry is an elegant solution. Because SLiMs are so compact, viruses need to invest only a limited number of amino acids of a given protein sequence to achieve a significant gain of function. Additionally, SLiMs do not require a supporting secondary or tertiary structural scaffold to maintain proper folding, since they generally have none in the first place. SLiMs can arise very rapidly *de novo* from an evolutionary perspective, since the prerequisite for their emergence may be only one or two mutations. Finally, the degenerated nature of SLiM recognition sequences means that viral proteins can follow different evolutionary paths to the same functional outcome (functional SLiM), making their emergence more frequent. Accordingly, several examples of viral SLiM mimicry have been described, in which viral proteins hijack a diverse set of cellular processes[3].

Hijacking cellular transport.

Protein localization is carefully monitored by the cell, and spatial distribution may play an important role in the function of a particular protein. One of the compartments of eukaryotic cells is the nucleus, which is responsible for the physical separation between the cytoplasm and the genetic material in the form of DNA. The transcriptional machinery is located in the nucleus, and the DNA viruses alongside retroviruses and some RNA viruses need to transport their genetic material into the nucleus to achieve successful transcription and translation (as is the case for human herpesviruses (HHV)) or integrate into the host DNA (human immunodeficiency viruses (HIV))[50,51]. Viruses employ various strategies to deliver their genomes into the nucleus, but one particular mechanism involves viral proteins which possess a nuclear localization signal (NLS) that enables them to transport viral genomic DNA into the nucleus[51]. The NLS is essentially a SLiM with many variations, but the core motif consists of repeats of basic lysine and asparagine residues[52,53].

In fact, the first NLS was discovered in the large T-antigen of simian virus 40 and has been subsequently described in a plethora of other human and viral proteins[53-60]. One interesting example is Rev protein from HIV, which is responsible for transporting unspliced mRNA into the cytoplasm. The sequence of Rev contains the NLS, which overlaps with the binding site for viral mRNA. Additionally, Rev also contains another SLiM mimic, namely a nuclear export signal (NES)[61,62]. When the concentration of Rev in the nucleus is low, early in the infection process, viral mRNA is rapidly spliced to produce more Rev and other early-phase proteins. As the concentration of Rev and viral mRNA increases, Rev binds the mRNA, hiding the NLS and exposing only the NES, causing the infection to enter the late stage as Rev begins to transport the unspliced and partially spliced mRNA to the cytoplasm, starting the production of late stage proteins important for virion assembly[61]. In this

way, Rev acts as a specialised shuttling adapter for viral mRNA and facilitates the transition between the early and late stages of viral infection.

Rewiring signal transduction

As signal transduction pathways are enriched with SLiMs for the reasons described above, they represent a prime target for viral SLiM mimicry in cases where signal re-wiring is advantageous to the pathogen. One example is the nuclear factor-κB pathway, which regulates a variety of signal transduction processes, conveying the signal from receptors to the nucleus, where it triggers the transcription of genes involved in various cellular processes such as cell cycle regulation, development, immune response, and others[63,64]. Among other signals, the nuclear factor-κB pathway transduces the signal from Toll-like receptors that recognise intracellular viral double-stranded RNA, triggering an immune response in virus infected cells[65]. As it is the first line of defence against viral infections, a variety of viruses interfere with the nuclear factor-κB pathway and modulate it to their advantage[65,66]. Human T-cell leukaemia virus 1, HIV, hepatitis C virus, hepatitis B virus, Epstein-Barr virus, influenza A virus, and SARS-CoV all constitutively activate the nuclear factor-κB pathway to promote cell survival, viral replication, and prevent the organism's immune response to the invading pathogen[65,66]. This seems counter-intuitive as the nuclear factor-κB pathway is also activated as part of the immune response. Nevertheless, viruses promote the activation of the pathway while avoiding the negative consequences of immune response stimulation.

For example, the Latent membrane protein 1 (LMP1) from Epstein-Barr virus is a transmembrane protein and functions as a continuously activated Cluster of differentiation 40 (CD40) receptor mimic[67,68]. LMP1 mimics three SLiMs that interact with Tumor necrosis factor receptor associated factors TRAFs (P.Q..D), the tyrosine-protein kinase JAK3 (P..P.P), and the Tumor necrosis factor receptor type 1-associated DEATH domain (TRADD) proteins (YYD-COOH). Collective binding of these proteins is required for activation of the nuclear factor-κB pathway and promotes B-lymphocyte survival and proliferation[65,69-72]. Despite the fact that the viral protein LMP1 and the host protein CD40 have evolved convergently, the C-terminal cytoplasmic tails of CD40 and LMP1 are interchangeable in human embryonic kidney 293 cells (HEK293) cells, demonstrating the remarkable plasticity of SLiM mimicry[68].

Initiating viral egress

The endosomal sorting complexes required for transport (ESCRT) pathway is responsible for reverse topology membrane invagination and scission, which amongst other functions facilitates multivesicular body formation, cytokinesis, shedding of microvesicles from the plasma membrane, and maintenance of nuclear pore integrity[73-76]. Unlike clathrin-mediated trafficking, in which the plasma membrane invaginates inward, towards the cytoplasm, the ESCRT

pathway drives budding outward, towards the extracellular matrix[75]. The ESCRT pathway consists of five complexes, ESCRT-0, ESCRT-I, ESCRT-II and ESCRT-III, and the vacuolar protein sorting-associated protein 4 (Vps4). These complexes assemble stepwise and, together with several auxiliary proteins such as ALIX and NEDD4, sort the ubiquitinated cargo into multivesicular bodies, induce membrane curvature, and terminate the process by performing membrane scission[77]. As they facilitate membrane budding into the extracellular space, this pathway is frequently hijacked by viruses to facilitate viral particle egress from the plasma membrane[78].Three distinct examples of viral SLiM mimicry have been characterized for the ESCRT pathway[78]. First, the UEV domain of the tumour susceptibility gene 101 (TSG101) protein, which is part of the ESCRT-I complex, recognizes the P[T/S]AP motif. Mimicry of this motif by HIV and related retroviruses in their p6 late domain region of the Gag polyprotein, has been extensively characterized and shown to be critical for viral egress[79-81]. Second, the V domain of ALIX (additionally called AIP1 or PDCD6IP) is also hijacked by the p6 region of HIV-1 and the p9 region of Equine infectious anaemia virus[82,83]. The p6 region, vital for viral budding, binds to ALIX, by mimicking the host recognition sequence $LYP(.)_nL$[84]. Finally, hijacking of HECT-ubiquitin ligases such as NEDD4, Itch, or WWP1, also proved important for successful viral budding[79,85,86]. This family of ubiquitin ligases includes nine members in humans and all contain at least one WW domain, that recognizes the PP.Y recognition sequence[85]. Mimicking this recognition motif by viruses such as vesicular stomatitis virus, Rous sarcoma virus, and human T-cell leukaemia virus, leads to successful budding of virions by the recruitment of ESCRT pathway components in a manner similar to the hijacking of TSG101 and ALIX[79].

Interestingly, evolutionarily distant viruses such as HIV, hepatitis E virus[87], ebolavirus[88], Lassa mammarenavirus[89], vesicular stomatitis virus[90], and others have all evolved SLiM mimics targeting the TSG101 protein, suggesting convergent evolution and highlighting the evolutionary plasticity of SLiMs[78]. Moreover, different viral lineages have evolved distinct SLiM hijacking strategies to overtake the ESCRT machinery, further illustrating the versatility of SLiM mimicry. Both the NLS motif and the ESCRT machinery were essentially discovered through research on viruses[53,81] emphasising the importance of basic research and the multidisciplinary nature of biochemical sciences in the present era.

Despite several well-studied examples, SLiM mimicry remains largely unexplored and knowledge is limited to a few socioeconomically important viruses such as HIV and human papillomavirus[3]. Furthermore, identification of SLiMs by conventional high-throughput screening methods has proven difficult (as discussed in paper I), and new approaches for SLiM discovery and characterization are needed.

Systematic identification of novel protein-protein interactions

In recent years, significant progress has been made in the attempt to map the human protein-protein interactome. This has mainly been enabled by a staggering development of high-throughput detection techniques such as affinity purification coupled with mass spectrometry (AP-MS) and the yeast two-hybrid system (Y2H).

Mass spectrometry (MS) is a spectroscopic method that, at its essence very accurately measures the mass-to-charge ratio of small molecules and was initially used in analytical chemistry and drug discovery to accurately identify small organic molecules in a given sample[91]. Recent advances in sample processing and detection, such as electrospray ionisation and Orbitrap, have substantially increased the sensitivity, specificity, and acquisition rate, allowing for extremely rapid and accurate identification of molecular mass in the analyte[92]. Complex biological samples can be readily subjected to MS in order to identify their protein composition. In a typical experimental setup, a protein sample is prepared and fragmented (usually by a protease such as trypsin), purified to remove non-peptide impurities, ionised, and subjected to MS analysis[93] (Figure 5A). The mass-to-charge ratio of peptides in the sample is then analysed and compared to established databases to accurately determine the protein composition of the sample[94]. This approach can be coupled with the affinity purification step, prior to MS analysis by a "pulldown" with an antibody. The antibody specifically binds to the protein of interest, but also "pulls down" any additional interaction partners of it. The protein of interest and the interaction partners are then detected by MS analysis. By performing the "pulldowns" from various samples collected under different experimental conditions, one can identify the protein-protein interaction networks and how they change depending on the cell type, cell environment, or post-translational modification state of the proteins[95,96]. The MS based approaches have led to the development of several proteome-wide PPI screens[92,97] and databases such as BioPlex[98]. The disadvantage of AP-MS with respect to the identification of PPI is that the antibodies used for "pulldowns" may have low specificity, making accurate PPI annotation difficult. Furthermore, as AP relies on a crude "pulldown", the approach does not provide information on direct, binary PPI, but instead identifies all partners of a given multiprotein complex.

The latter shortcoming is mitigated by the yeast two-hybrid (Y2H) approach, first described in 1989[99]. In the original Y2H system, the transcription factor GAL4 was engineered such that the DNA-binding and transcription-activation domains were physically separated. These two domains must be in close proximity to successfully induce gene transcription of a reporter gene. The two separate domains of GAL4 were then each fused to an arbitrary protein of interest. When these two proteins of interest interact, the DNA-binding and transcription-activation domains of GAL4 are in close proximity and the

transcription of the reporter gene, facilitating the readout, is induced (Figure 5B). This initial design was further improved by use of different transcription factors and reporter genes to increase the efficiency and sensitivity of the readout[100-102]. Since the genetic manipulation of yeast cells is relatively straightforward, many different proteins and protein domains can be fused to the split transcription factor and rapidly tested for protein interactions, creating a high-throughput method for the discovery of binary interactions. As with MS, several large PPI datasets[103] and databases such as HuRI[104] have been established using the Y2H system, providing near proteome-wide coverage.

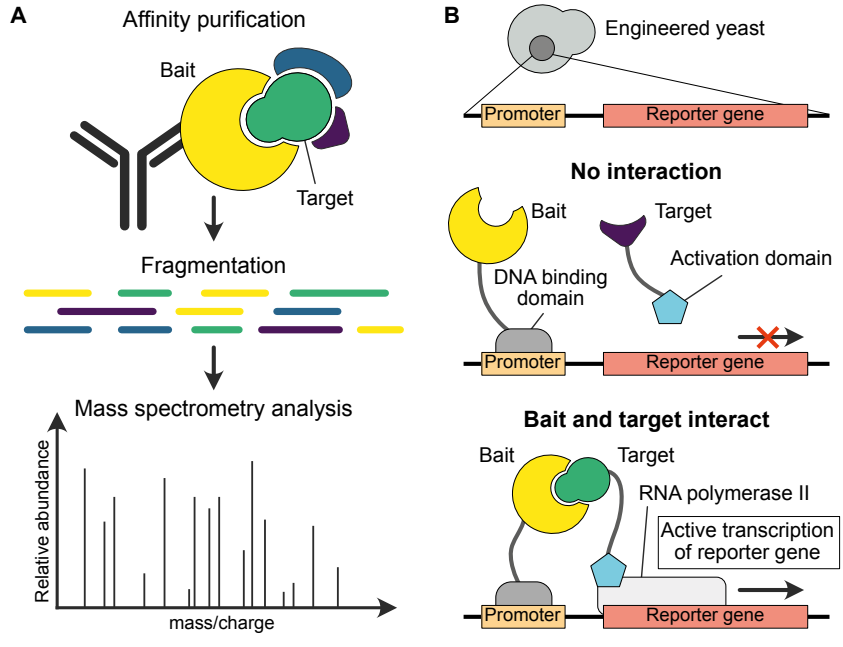


Figure 5: Schematic illustration of AP-MS and Y2H methods. (A) In the first step of AP-MS protein of interest is used as a bait to pulldown its interaction partners. This sample is then digested and subjected to MS analysis. **(B)** In the Y2H system the transcription factor is separated into DNA binding domain and Activation domain. These two fragments are in turn fused to proteins for which the interaction is tested. If they interact the activation domain will recruit RNA polymerase II which will induce the transcription of the reporter gene.

Despite its staggering versatility, the Y2H system has some shortcomings, especially in terms of false negative and false positive results. For instance, in the classical Y2H system, the membrane-bound proteins would give a false negative result as they would not translocate to the nuclei to initiate reporter gene transcription. Furthermore, the fact that the proteins under investigation are fused to the transcription factor could lead to steric hindrance and obscure the interaction interface. On the other hand, false-positive results may arise from high, non-natural expression levels of the two proteins for which the

interaction is being tested. Because of these problems, rigorous controls and a high level of subsequent validation must be employed[102].

Given the weaknesses of the two large-scale PPI identification techniques, several other approaches are currently being developed and used. One such technique is proteomic peptide phage display (ProP-PD).

Proteomic peptide-phage display as an alternative for large-scale SLiM discovery.

The phage display method was developed in 1985 by George P. Smith and awarded the Nobel Prize in 2018[105]. The method is based on the discovery that a relatively long insertion (57 amino acid residues in the first publication describing the method) into the coat protein pIII of the M13 filamentous phage, does not affect its function, while the inserted amino acid sequence is exposed on the surface of the phage. With this simple implementation, it is relatively easy to present a variety of peptides or even complete proteins such as epitope-recognizing antibody derivatives on the surface of the phage[106]. Since the insertion into the pIII protein is executed at the DNA level, it is considerably easier to manipulate and randomize compared to other peptide or protein display methods. For this reason, vast phage libraries have been constructed, displaying more than 40 million random six-residue-long peptides on the surface of the phage population, approaching the theoretical limit of chemical space (64 million 6-mers)[107]. Because the displayed sequences are presented on the surface of the phage, they can interact with all interaction partners that the phage might encounter. This is a basis for phage display selections. If we immobilize any molecule on the surface, we can expose it to the phage library and wash away the phages that do not bind to the desired molecule. After the washing step, we can elute the phages with high affinity for the molecule of interest and sequence the DNA of them to see which peptide sequence or antibody fragment was presented on the surface and bound to the immobilised bait molecule[108]. The direct genotype-phenotype link between the nucleotide insertion in the DNA sequence of the region encoding pIII and the actual amino acid insertion in pIII allows easy reverse-identification of the displayed sequence. This basic protocol can be modified in numerous ways. One of the more common modifications is the repetition of several cycles of binding, elution and amplification of the bound phages, resulting in an enrichment of particles displaying the higher affinity epitopes (Figure 6). This process is referred to as affinity maturation and can produce 10^8 - fold enrichment of phages displaying a high-affinity binding sequence[109,110]. If the initial molecule of interest, immobilized on the surface of the experimental plate, is a protein presumed to bind short peptide motifs, the exact binding motif epitope can be determined by presenting a phage library of random peptide sequences.

Over the years, several notable derivatives of the original method have been developed. Different display moieties such as the filamentous M13 phage, the lytic T7 phage[111], and the bacteriophage λ[112] have been tested and used to display different libraries. One of the benefits of using T7 or λ-phages as a display system is that they are not excreted through the periplasmic space of *E. coli*, because they escape into the extracellular space through lysis of the bacterial host cell. Excretion through the periplasm can lead to complications with unpaired cysteine residues that can form undesired disulphide bonds, which in the case of M13 phages results in the unsuccessful display of the designed sequence on the surface of the phage[113]. Despite this drawback, the filamentous M13 phage is still most commonly used as the phage of choice for phage display.

In M13 filamentous phage display, several different phage proteins can be utilised to display the desired sequences. The pIII protein is present in only five copies at the end of the elongated phage and is required for phage infectivity as it interacts with the F-pilus of *E.coli*[114]. Displaying the insert on the pIII protein ensures that only a few copies are displayed per phage, which promotes the selection of high affinity binders[106]. Alternatively, the peptides or proteins can be presented on the major coat protein pVIII, a 50 kDa protein coating the entire elongated phage "body"[115]. One advantage of displaying on the pVIII protein is that it is present in high numbers (2700 copies) on the surface of the phage, allowing capture of low-affinity interactions due to high local concentration of the peptides and the "avidity effect"[116]. This approach is also used in papers I-IV. Displaying on other coat proteins such as pVI, pVII, and pXI has been reported, but it is not widely used in modern phage display setups[106,116].

Regarding the entity displayed on the phage surface, there are two main possibilities in phage display technologies. The first option is for phages to present peptide libraries of various lengths, depending on the application and the desired outcome of screening. These phage libraries can be presented to immobilized antibodies or peptide-motif-binding proteins to pinpoint the recognition motif of the antibody or protein. Depending on the use of the pIII or pVIII protein as the library displaying method, high- or low-affinity interactions can be screened for. Alternatively, phage display libraries may display modified single-chain variable fragments (scFv) or other derivatives of antibodies[117-119]. These libraries are derived from the variable chain genes of immunoglobulins and essentially display a variety of epitope-recognizing fragments of the antibodies. When screening such libraries against toxins or other molecules of interest (such as oncogenic peptides, viral proteins, *etc.*), the human antibodies that binds to the same toxin can be reverse engineered. This approach circumvents the problematic aspects of therapeutic antibody production, such as immunization, identification of high-affinity antibody clones, and lack of consecutive screening[120]. Additionally, phage display-derived

antibodies can have affinities thousands of times higher compared to the antibodies obtained by immunization[121].

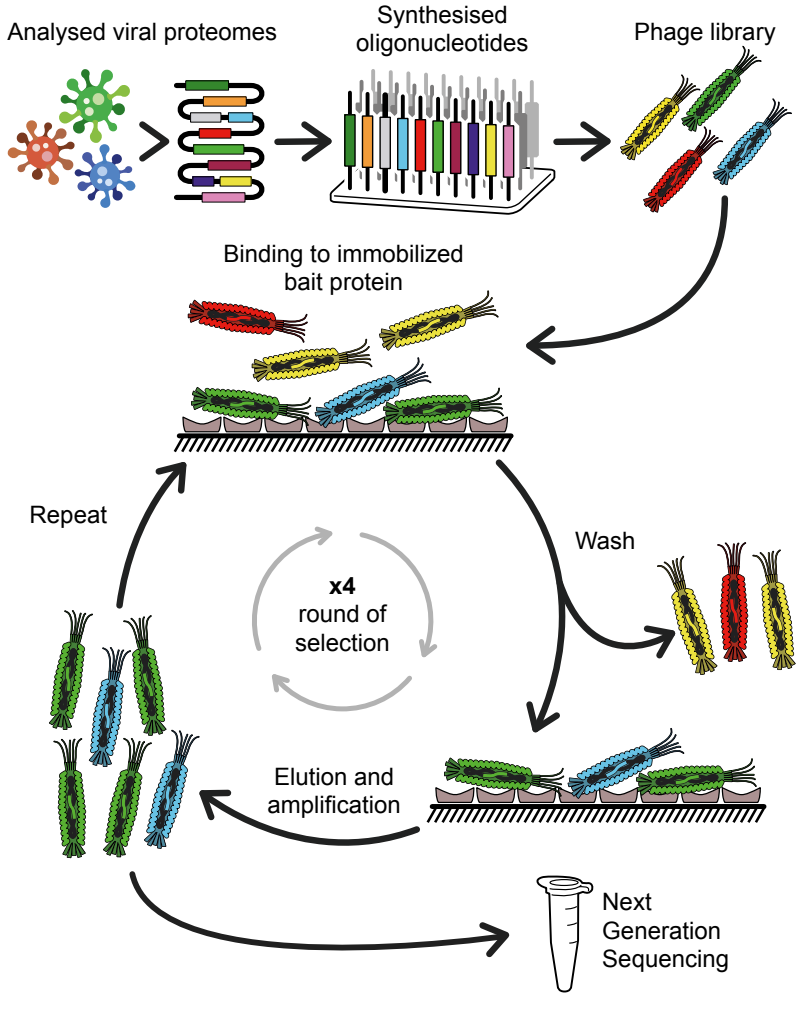


Figure 6: Schematic illustration of the phage display screening procedure. In this example, a peptide library displaying disordered regions of viral proteins is represented. Following bioinformatic design, the phage library is constructed using modern synthetic methods. In the library there can be more than 10^7 phage particles present, which display different peptide sequences (or antibody derivatives) on their surface. The next step is to present the phage library to the immobilized protein of interest. In the illustration, the protein of interest is a bait protein, but it can be any other type of molecule of interest, *e.g.*, DNA, organic compounds, *etc.* The plate with the immobilized protein is washed to remove the non-binding phages, and the tightly binding phages are eluted and amplified in *E. coli*. This procedure is repeated three to four times to enrich the phages that display the strongest binding peptides. In the final step, the phage DNA is sequenced to identify the high-affinity peptide ligands of the bait protein[122]. (Adapted from an original illustration by Leandro Simonetti)

The development of phage display thus paved the way for easier and superior identification of therapeutic antibodies, leading to a surge of phage display-derived antibodies that are being investigated for their potential as drugs[123,124]. By 2020, more than 70 such therapeutic antibodies had been tested in clinical trials, 14 of which were approved, underscoring the power of this approach.[123] The compiled information from the large-scale PPI screening methods described above led to the creation of proteome-wide interaction databases, such as IntAct[125]. The databases allow a systems approach to the analysis of interaction networks[126,127] but provide limited or no biophysical information on the affinity and specificity of interactions.

Biochemical analysis of protein-ligand interactions

Association of two molecules is the most basic chemical reaction and an initial event required for all subsequent chemical transformations. The simplest equation describing the interaction between two molecules forming a complex is:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AB \tag{1}$$

where A and B are the reactant molecules, k_1 is the association rate constant and k_{-1} is the dissociation rate constant, and AB is the complex formed between A and B. This scheme describes a reversible reaction and can be described thermodynamically or kinetically. When the reaction is in equilibrium, the ratio of reactants and products does not change and is described by the dissociation constant K_D:

$$K_D = \frac{[A]_{eq}[B]_{eq}}{[AB]_{eq}} \tag{2}$$

The dissociation constant K_D, sometimes conceptualized as affinity (or more accurately the multiplicative inverse of affinity), is a thermodynamic parameter associated with the change in Gibbs free energy upon interaction of the two reactants in equation 1:

$$\Delta G = RT \ln \frac{K_D}{c^\circ} \tag{3}$$

and

$$\Delta G = \Delta H - T\Delta S \tag{4}$$

where ΔG is the change in Gibbs free energy, R is the universal gas constant, T is the reaction temperature, c° is the standard molar reference concentration, ΔH is the change in enthalpy, and ΔS is the change in entropy of the system.

The progression of the reaction in equation 1 can also be described kinetically. If we use the law of mass action and the reaction rate constants to describe the change in concentration of the *AB* complex over time, we obtain a second order reversible reaction rate equation 5:

$$\frac{\mathrm{d}[AB]}{\mathrm{d}t} = k_1[A][B] - k_{-1}[AB] \tag{5}$$

We can combine equations 2 and 5 at equilibrium conditions:

$$\frac{\mathrm{d}[AB]}{\mathrm{d}t} = 0 \tag{6}$$

to obtain K_D, but in this case expressed with kinetic rate constant parameters:

$$K_\mathrm{D} = \frac{k_{-1}}{k_1} \tag{7}$$

Equations 2-7 form the foundation for the biophysical characterisation of chemical reactions. Since simple binding reactions involving SLiM recognition by a globular domain, which have been characterised in the present work typically follow equation 1, more complex reaction trajectories are not considered.

The appeal of using the dissociation constant K_D as a measure of affinity between two molecules stems from the fact that it can be experimentally determined with relative ease. If we consider the binding reaction in equation 1, we can set up an experiment in which we keep one of the reactants at a constant concentration (*B*), vary the concentration of the other interactor (*A*), and monitor the formation of the complex *AB*. If we ensure that the concentration of *A* is high enough to remain approximately constant upon the addition of *B*:

$$[A]_\mathrm{eq} \approx [A]_\mathrm{total} \tag{8}$$

the concentration of each species under equilibrium conditions will be:

$$K_\mathrm{D} = \frac{[A]_\mathrm{total}([B]_\mathrm{total} - [AB])}{[AB]_\mathrm{eq}} \tag{9}$$

If we rearrange the equation 9 in terms of the fraction of bound *B* in the *AB* complex we get:

$$\frac{[AB]_\mathrm{eq}}{[B]_\mathrm{total}} = \frac{[A]_\mathrm{total}}{(K_\mathrm{D} + [A]_\mathrm{total})} \tag{10}$$

This relationship follows a hyperbolic equation. If we satisfy the approximation in equation 8, and we monitor the ratio of unbound and bound *B* at increasing concentration of *A*, we will observe that the ratio does not change at high concentrations of *A* because all *B* is in bound form; the system is saturated. In this case, K_D is equivalent to the concentration of *A* at which half of all *B* is in bound form.

Colloquially, if the K_D value for the formation of the complex is low, it means that the affinity is high, implying that the formation of the complex is favourable. Over the years, many techniques have been developed to determine the K_D for binding reactions, each with its advantages and disadvantages. In the experiments presented in this thesis, the fluorescence polarisation technique described below, was predominantly used.

Fluorescence polarisation

Light is an oscillation of an electromagnetic field in which the electric and magnetic components oscillate perpendicular to the direction of light travel. In the everyday world, light hits the detector (our eyes) from different directions, is diffracted and scattered by objects, leading to depolarisation of the light wave. This means that the oscillation of the electromagnetic field takes place at all angles around the direction vector of the light wave propagation[128]. If we however pass the light through a polarizing filter, only the light waves with identical angle of both light components will pass through. In this way the oscillation occurs in the same plane, similarly to the waves on the sea. With the help of such polarizing filters, we can produce plane polarised light.

The phenomenon of fluorescence polarisation or fluorescence anisotropy, first described by Francis Perrin[129] in 1926, is based on the fact that fluorophores excited with plane polarised light emit predominantly depolarised light due to the fast tumbling of the fluorophore in solution. Emission of light by the fluorophore does not occur instantaneously after the excitation. Instead, light emission occurs with a slight delay after the excitation event, typically on the nanosecond timescale[130]. If the fluorophore is completely rigid and does not change its orientation between the excitation and emission events, all emitted light remains polarised. However, if the fluorophore rotates on the same timescale as the delay between excitation and emission (this is true for most small organic fluorophores), then the emitted light becomes depolarised due to the misalignment of the excitation and emission dipoles. The rate of rotation (rotation relaxation time) of spherical molecules is determined by the Stokes' equation:

$$\rho = \frac{3\eta V}{RT} \tag{11}$$

where ρ is the rotation relaxation time (the time required for the molecule to rotate 68.5°), V is the volume of the molecule, and η represents the viscosity of the solution. Consequently, when excited with polarised light, small and fast rotating fluorophores which tumble on sub-nanosecond timescale emit depolarised light, while large fluorophores, which rotate slower, retain polarised nature of the emitted light. The difference in the proportion of emitted polarised light, which depends on the size of the fluorescent molecule, is the foundation of the fluorescence polarisation interaction assay[131].

The two technical terms, fluorescence polarisation and fluorescence anisotropy describe the same phenomenon with slightly different definitions, but can be interconverted easily:

$$P = \frac{3r}{2+r} \tag{12}$$

where P is the fluorescence polarisation and r is the fluorescence anisotropy. For consistency, only the term fluorescence polarisation (FP) is used in the present work. During the experimental setup, a fluorophore of interest is excited with plane polarised light, and the polarisation of the emitted light is measured in parallel and perpendicular orientations with respect to the initial polarisation plane of the excitation light. The resulting measurement of FP is essentially a ratio between the polarised and depolarised light emitted by the fluorophore excited with plane polarised light, and is defined as:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{13}$$

where $I_{\parallel}$ represents the detected light intensity parallel to the plane of polarisation of the excitation light (the fraction of light that has retained the initial polarisation) and $I_{\perp}$ represents the detected light intensity in the perpendicular plane compared to the excitation polarisation light (the light that has lost its initial polarisation). Following equation 13, we can derive that the theoretical limits of FP are one when the emitted light remains 100 % polarised, and minus one when all of the light is depolarised. In reality, these values for FP are never reached, and the highest theoretical FP value in solution is 0.5[128]. Most commonly, fluorescence polarisation converted into arbitrary millipolarisation units (mP), so the maximum achievable experimental result in solution is 500 mP.

Fluorescence polarisation binding assay

For the commonly used fluorophore fluorescein isothiocyanate the delay between excitation and emission is roughly four nanoseconds[132]. The molecule itself rotates on a shorter timescale, resulting in a depolarised emission signal

(Figure 7A). However, if the fluorescein is bound to a larger molecule, such as a protein, the rate of rotation will slow down, and the emission signal will remain polarised. This principle is exploited in FP assays. In a common experimental framework, a fluorophore is covalently attached to a short peptide. The combined size of the labelled peptide is still small enough to produce a depolarised emission signal. The constant concentration of the labelled peptide is then mixed with the increasing concentrations of the protein interaction partner. When the protein and labelled peptide interact, the resulting complex rotates substantially slower than the labelled peptide alone, producing a polarised emission signal, as shown in Figure 7B. Depending on the concentration of the labelled peptide at which complex formation is achieved a dissociation constant K_D can be calculated according to equation 14:

$$y = background + \frac{x(max-background)}{K_D + x} \tag{14}$$

where y is the measured FP signal at a given protein concentration, *background* is the lower limit of the polarisation signal (FP signal of only labelled peptide in buffer solution), x is the concentration of the protein, and *max* is the upper limit of the polarisation signal (FP signal when all of the labelled peptide is bound to the protein). This approach can be used to measure the affinity between any protein and a fluorescently labelled molecule of interest. To employ equation 14, the experimental conditions must conform to the approximations described in equation 8.

Fluorescence polarisation displacement assay

A slightly modified version of the standard FP binding assay is an experimental setup in which a protein and a labelled peptide are premixed at appropriate concentrations for the formation of the complex. To this labelled peptide-protein complex, an unlabelled peptide is added at increasing concentrations. As the concentration of the unlabelled peptide increases, it outcompetes the labelled probe, and the initially polarised emission becomes depolarised (Figure 6C). The FP signal is plotted against the concentration of unlabelled competitor peptide (x), and the half-maximal inhibitory concentration IC_{50} is determined by fitting a sigmoidal dose-response equation:

$$y = background + \frac{(max-background)}{1+\left(\frac{IC_{50}}{x}\right)^n} \tag{15}$$

where n is the hill coefficient. The IC_{50} forms the basis for calculation of K_i, which corresponds to the K_D of the unlabelled peptide according to equation 16:

$$K_{\mathrm{i}} = \frac{I_{50}}{\frac{L_{50}}{K_{\mathrm{D}}} + \frac{P_0}{K_{\mathrm{D}}} + 1} \tag{16}$$

where I_{50} is the concentration of free unlabelled peptide at 50 % displacement, L_{50} is the concentration of free labelled peptide at 50 % displacement, P_0 is the concentration of free protein at 0 % displacement and K_{D} is the dissociation constant for the labelled peptide.

Several examples and designs of displacement FP assay with equations for the determination of the K_{i} of the unlabelled displacer peptide have been described in the literature[133-136].

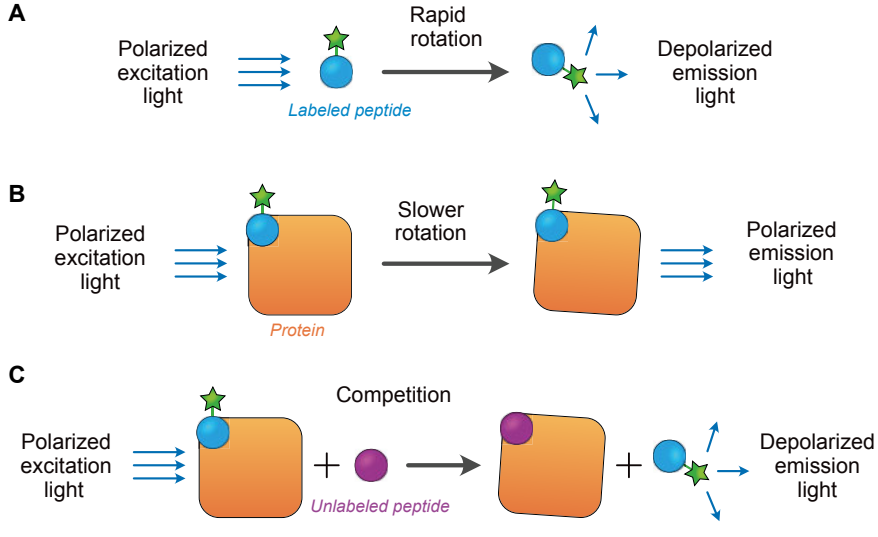


Figure 7: Schematic representation of experimental setup for FP. (A) The labelled peptide tumbles freely in solution. **(B)** Illustration of a simple binding assay. The labelled peptide forms a complex with the interaction partner, which slows the rotation of the fluorescent label, resulting in a polarised emission signal. **(C)** Illustration of the experimental setup for the displacement experiment. The unlabelled peptide displaces the labelled peptide in a concentration-dependent manner resulting in depolarisation of the emission signal.

The main advantages of using FP assays to determine affinities between proteins and peptides, compared to other methods such as isothermal titration calorimetry, stopped flow, surface plasmon resonance, and nuclear magnetic resonance are that FP requires relatively small amounts of pure protein samples, the method is “label-free” in the displacement experiment, and the experimental procedure is relatively simple allowing high experimental throughput.

Present investigation

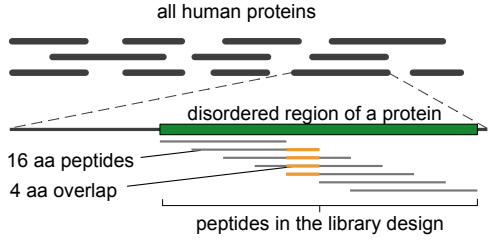
This chapter describes the main findings and conclusions of the papers I to VI. Conceptually, the research presented here can be divided into three parts. In the first part, the human disorderome library was used to identify SLiMs in the human proteome that bind to native (Paper I) or viral (Paper IV) globular domains. In papers II and III, the viral disorderome library was designed and screened against a subset of human domains to identify instances of viral SLiM mimicry. Finally, in papers V and VI a single interaction between MDM2 and p53 was examined in detail from an evolutionary perspective, analysing the evolution of affinity in relation to mutations within and in proximity of the central recognition motif.

Human disorderome library as a tool for the proteome scale discovery of SLiM-based interactions (paper I)

As discussed in the introduction, astounding progress has been made in recent years in mapping the human proteome. However, SLiM-mediated interactions have been proven hard to identify using convectional PPI identification techniques[137] and the PPI interaction databases are consequently largely devoid of motif-mediated interactions. To address this, we developed a Proteomic peptide phage display screening method in Paper I, specialised in detecting SLiM-mediated interactions. The backbone of the screening method was the construction of a phage display library, called Human disorderome 2 (HD2) library, which displayed on the surface of the phages, all predicted disordered regions of the human proteome. These regions were divided into 16 amino acid long overlapping peptide sequences and displayed on the pVIII coat protein of the M13 phage (Figure 8).

The HD2 library is a vastly improved version of the original human disorderome library[138] with a higher number of peptides, longer peptides with greater overlap, and substantially improved coverage. In theory, the HD2 library should contain all discoverable SLiMs in the human proteome. Additionally, four sub-libraries were created that subdivided the disordered regions according to the cellular localization of the proteins from which they were derived: endomembrane, cytoplasm, nucleocytoplasm, and nucleus. These additional sub-libraries were used to determine whether the cellular localization

of the proteins affected the results of library selection against the globular domain. This was largely not the case.



Cut-off IUPred:	0.4
Peptide length:	16 aa
Overlap between peptides:	12 aa
# Peptides:	938,427
# Proteins:	**16,969**
Coverage:	>95 %
Coat protein:	pIII, pVIII

Figure 8: Library design of the HD2 phage display library. All disordered regions from all human proteins were extracted and presented on the surface of phages as 16 amino acid long peptides. Some relevant information for the HD2 library is shown on the right.

To determine the utility of the HD2 library, it was tested against a set of 34 globular domains known to interact with SLiMs. This "benchmarking set" was selected to evaluate how many known interactions are rediscovered during phage display selections (referred to as "recall"). Recall was found to be ~20 %, which is a substantial improvement over other comparable methods such as the yeast two-hybrid assay (HuRI described in the introduction has a recall of ~9 %).

As with all high-throughput techniques, it is not trivial to distinguish between true and false positive hits. To mitigate this, hit confidence rates for proteomic peptide phage display (ProP-PD) selections were determined based on four quality metrics: (i) strong enrichment of a hit peptide (high NGS count), (ii) presence of a consensus motif, (iii) identification of the hit peptide in subsequent selections, and (iv) identification of overlapping peptides from the same disordered region. These quality criteria are critical to ensure stringent analysis and identification of true positive hits from all subsequent selection results against any globular domain.

To establish the quality of the selections against the benchmarking set, some of the hits were chosen for further validation. The peptide ligands identified in the ProP-PD screens against five human protein domains (KEAP1 Kelch, MDM2 SWIB, TLN1 PTB, KPNB1 HEAT, and C-terminal ARM repeat part of KPNA4) were acquired and their affinity for the globular domain was measured using FP. In addition to purifying proteins used as baits for phage display, I contributed to this publication by establishing the FP assay to measure affinity towards the MDM2 SWIB domain. First, a saturation assay was performed to determine the K_D of the fluorescein isothiocyanate (FITC)-labelled $p53_{15\text{-}26}$ peptide (FITC – SQET**F**SDL**W**KL**L**; specificity-determining residues are in bold) to be ~ 50 nM (Figure 9A). On this basis, it was possible to perform a displacement FP experiment measuring the K_D values between MDM2 SWIB and four other unlabelled peptide ligands ($p53_{15\text{-}31}$, $RNF115_{69\text{-}$

$_{84}$, KIAA1671$_{604-619}$, and NUMB$_{615-630}$) (see Figure 9B). The dissociation constant for unlabelled peptides ranged from 0.17 µM for RNF115$_{69-84}$ to 3.3 µM for NUMB$_{615-630}$ and was consistent with previously reported affinities[139,140]. Out of all four proteins the MDM2 recognition motif has, only been identified in p53 to date, however the general interaction between full-length MDM2 and both RNF115[141] and NUMB[142] has been suggested previously. This highlights the strength of the ProP-PD approach for the identification of SLiMs, as it provides resolution of the interaction on the amino acid residue level, unlike other screening approaches.

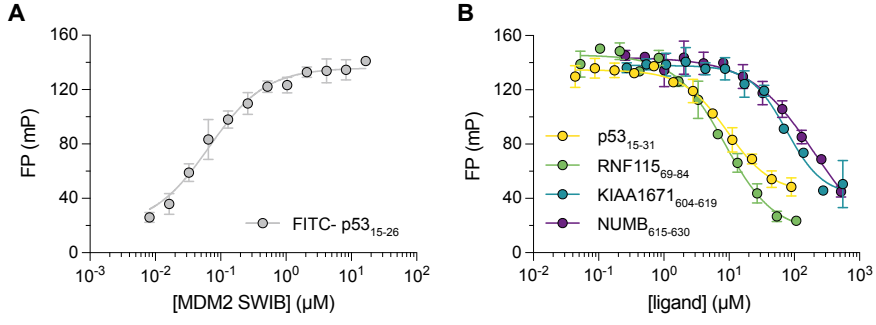


Figure 9: FP results for MDM2 SWIB domain binding to various peptides. (A) Saturation binding experiments from which the affinity of the labelled peptide was determined. **(B)** Displacement experiments for four unlabelled peptides.

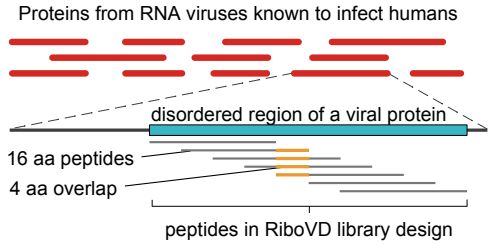
After validating the ProP-PD results in the context of affinity measurements, the interactions between KPNA4 and various NLS peptides were further investigated to show that the phage display results identified two distinct subsets of peptides binding to two different SLiM-binding pockets of KPNA4.

Overall, paper I provided an important foundation for experimental design, validation, and analysis of future ProP-PD experiments and was a crucial basis for the work described in papers II to IV.

Large-scale discovery of viral SLiM mimicry (papers II and III)

Over the years, several examples of viral SLiM mimicry have been described, often leading to the discovery of new cellular pathways and human protein functions (see Introduction). However, these discoveries have been made sporadically and from low throughput studies. Large-scale experimental identification of interactions between human proteins and viral SLiMs was never attempted due to the technical difficulties and the fact that the focus of high-throughput methods was on identifying human PPIs. In the papers II and III, we describe a novel approach to screen interactions between host globular

domains and SLiMs from viral proteins. To achieve this, we produced a new phage display library, called RiboVD, whose phage particles display on their surface the disordered regions of proteins from RNA viruses known to infect humans (Figure 10).



Cut-off IUPred:	0.4
Peptide length:	16 aa
Overlap between peptides:	12 aa
# Peptides:	19,549
# Proteins:	1483
# Species:	**229**
Coverage:	>95 %
Coat protein:	pVIII

Figure 10: Design of the RiboVD library. In this case, disordered regions of RNA virus proteins were identified and presented on the surface of phages. This library is considerably smaller and more specialised compared to the HD2 library.

The RiboVD library was screened against 139 human globular domains. These domains were selected because they are known SLiM binding domains or because they have been previously reported to be hijacked by viral proteins. Based on the ProP-PD results, several domains were chosen for further validation. This screening formed the basis for both papers II and III. The reason for splitting the results into two publications was the emergence of SARS-CoV-2, which made the publication of relevant experimental data urgent. Therefore, paper II focuses only on SLiM mimicry of SARS-COV-2 and related coronaviruses, whereas in paper III we analysed and validated several other examples of viral host protein hijacking.

The nucleoprotein of SARS-CoV-2 promotes viral infection by mimicking a G3BP-binding SLiM.

Of all interactions between potential viral SLiM-containing peptides and human globular domains screened, 13 % were from SARS-CoV-2 and related coronaviruses. The major finding of this paper was that the nucleoprotein of SARS-CoV-2 contains an alternative version of the FG[D/T]F motif which binds to the nuclear transport factor 2 domain of Ras GTPase-activating protein-binding proteins 1 and 2 (G3BP1/2 NTF2). The G3BP proteins have been identified as the main drivers of stress granule formation. Stress granules are a specific type of cellular response to stressors and appear as dense speckles in the cytoplasm where the cellular translational machinery is sequestered, resulting in an interruption of protein translation. This is of course unfavourable for efficient viral replication, and accordingly, several examples of viral interference with G3BP proteins have been described[143-147]. The specific G3BP NTF2-binding sequence in the N-terminal tail of the nucleoprotein is 15-

Φ.FG-18 and it differs from the canonical sequence. This is likely the reason for the approximately 10-fold weaker interaction compared with previously described G3BP NTF2 binders. Nevertheless, mutation of the motif in the context of the full-length protein caused the nucleoprotein to lose its ability to associate with and disrupt stress granules. Finally, we experimentally showed that the presence of the Φ.FG motif in the nucleoprotein is critical for disruption of the interaction between G3BP and other stress granule proteins mediated by the FGDF motif, implying direct competition. Taken together, our results suggest that nucleoprotein reconfigures the network responsible for stress granule induction by targeting G3BP and this process promotes viral replication.

Apart from G3BP, several other instances of SLiM mimicry were observed in coronaviral proteins. In particular, we identified two motifs in Nsp3; the LiR motif that binds to the autophagy-related family of LC3 proteins and the ubiquitin carboxyl-terminal hydrolase 7 - Meprin And TRAF-homology domain (USP7 MATH) binding motif, which was also my main contribution to this study. The observation that SARS-CoV-2 proteins hijack USP7 is intriguing because this protein serves as a protein deubiquitination enzyme removing ubiquitin from its substrates, preventing their proteasomal degradation. A reasonable hypothesis would be that recruitment of USP7 increases the longevity of viral proteins, but this requires further investigation.

Viral SLiM mimicry is an abundantly used strategy

Apart from interactions with coronavirus, we discovered a plethora of viral SLiM mimicry during our investigations. In paper III, we validated three different examples of this phenomenon.

Hijacking of the ESCRT pathway

First, we analysed the results for four members of the ESCRT and ESCRT-associated machinery. We identified several novel viral peptides containing the P[T/S]AP motif which hijacks the TSG101 UEV domain. This was however not the case for the ALIX V domain, where most hits represented already known interactions, primarily with the p6 domain of the gag polyprotein of HIV1. Nevertheless, one interesting example stood out from these two selections. We found a region in the Nsp4 protein of Banna virus that contains two motifs in close proximity targeting both UEV and V domains: 118-VDEEKG**LYP**N**L**KLTE**PTAP**MIHNYM-142. Binding of the UEV and V domains to these motifs was mutually exclusive and had 10- to 100-fold weaker affinity compared with other tight binding peptides. Interestingly, Banna virus belongs to the *Reoviridae* family, a group of double-stranded RNA viruses that lack an outer membrane and escape host cells primarily through a lytic cycle. Considering this, they do not require the ESCRT machinery for reverse topology budding of mature virions. Nevertheless, literature indicates that

bluetongue virus, a close relative of Banna virus, indeed hijacks the ESCRT pathway[148], supporting our findings. Furthermore, we identified several novel examples of viral PP.Y motif mimicry targeting the WW domains of NEDD4 and identified the CEP55 EARB domain as a novel target of viral hijacking.

Interference with clathrin-mediated endocytosis

In the second example, we turned our attention to the clathrin-mediated endocytosis machinery responsible for the internalisation and trafficking of membrane-enveloped vesicles in the cytoplasm. We investigated viral hijacking of four domains that recognise SLiMs and are responsible for sorting cargo into various membraned vesicles. Importantly, we validated the interaction between the Ebola virus nucleoprotein and the GAE domain of the ADP-ribosylation factor-binding protein GGA3 in the context of full-length proteins by a Glutathione *S*-transferase (GST)-pulldown assay. When we mutated only two residues involved in binding on the NP protein, the interaction was lost, showing that the 466-Y..Y-469 motif identified in the ProP-PD selections indeed mediates the binary interaction. We also succeeded in solving the crystal structure of the clathrin N-terminal (CLTC NTD) in complex with two peptide ligands derived from Nsp3 of eastern equine encephalitis virus (EEEV) and the mu-NS protein of mammalian orthoreovirus 1 (MRV1), allowing us to further analyse the interaction (Figure 11A). These peptides contain the LΦ.F motif and bind to the clathrin box binding pocket of CLTC NTD. We observed that the viral ligands only interact with the clathrin box binding site of CLTC NTD. Furthermore, the two viral peptides had phenylalanine at position four of the motif, which buries deeper in the hydrophobic pocket between blades one and two of CLTC NTD than the human ligand AP1B1, which has leucine at the corresponding position (PDBid: 5M5R). This could explain the 10-fold higher affinity of the viral peptides compared to the human counterpart ($K_D \approx$ 30 μM and 300 μM for viral and human peptides, respectively). Moreover, we confirmed the binding of full-length Nsp3 and CLTC NTD with GST pulldowns. Similarly, as described above for GAE domain, mutation of two amino acids, F_{1774} and D_{1775}, to alanine in the full-length sequence of Nsp3 abolished the interaction (Figure 11B).

To determine whether the Nsp3 protein interferes with the normal function of clathrin in a cellular context, we developed an assay in which we transfected HEK293 cells with negative controls, full-length Nsp3 (EEEV), or the Nsp3 (EEEV) mutant and compared the internalisation of platelet derived growth factor receptor β (PDGFRβ) upon stimulation with platelet derived growth factor (PDGF-BB). Upon stimulation, this receptor dimerises, begins phosphorylation of its downstream effectors, and is internalised by clathrin-mediated endocytosis to attenuate the signal approximately 60 minutes after activation.

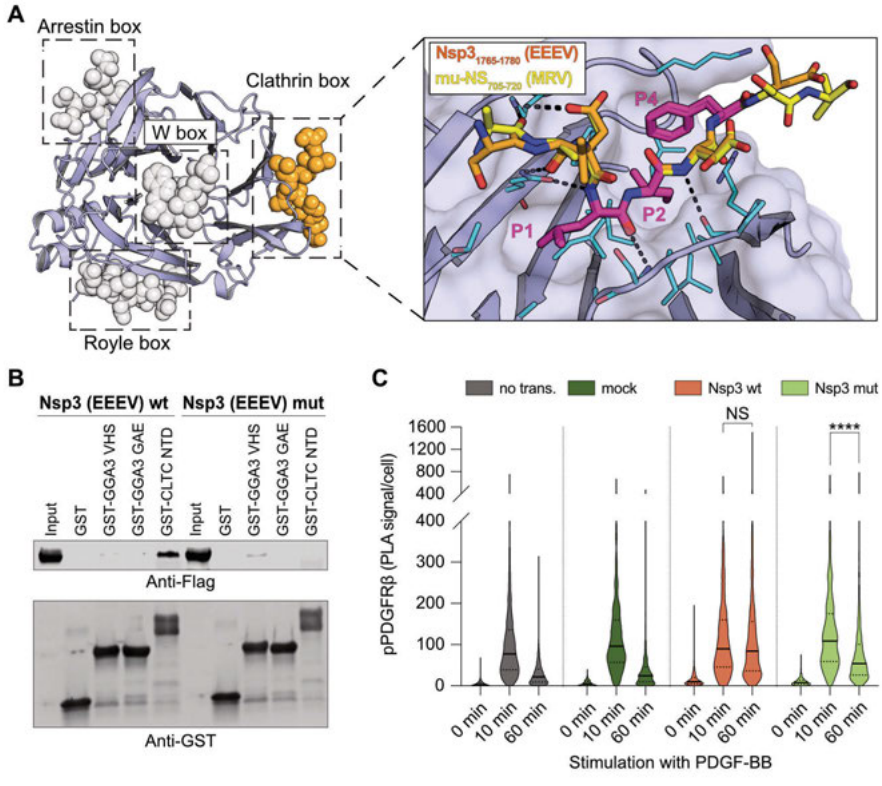


Figure 11: Viral SLiM-containing peptides bind to CLTC-NTD and disrupt its endogenous function. (A) Crystal structure model for CLTC NTD in complex with two viral peptides. The left panel shows the overall view of CLTC NTD with the peptide binding domains, and the right panel shows the close-up view of the two viral peptides binding to the clathrin box binding site. The residues vital for binding are coloured magenta. Positions one, two, and four of the motif (P1 to P4) are labelled. The blue sticks are all residues from CLTC NTD, which interact with the peptides. **(B)** The GST pulldown results for the Nsp3 (EEEV). The full-length viral protein was successfully pulled down only with the CLTC NTD domain and only in its "wild-type" form. **(C)** The presence of activated PDGFRβ after stimulation with its ligand as measured by proximity ligation assay (PLA). In all cases, except when transfected with Nsp3 wt, the signal was attenuated after 60 minutes, indicating a functional clathrin mediated endocytosis. In the presence of Nsp3 wt, the receptor was stimulated normally but remained active on the cell surface without being internalised. (Adapted from Figure 5 of paper III)

Our hypothesis was that when Nsp3 interferes with endogenous clathrin, internalisation of the receptor is also affected. This is precisely what we observed, as only when cells were transfected with wild-type Nsp3, did the receptor remain activated after 60 minutes, indicating problems with correct internalisation (Figure 11C).

Novel mechanism of host translation machinery takeover.

In our final example of SLiM mimicry, we discovered a novel mechanism of viral hijacking through interaction with the PABP1 protein. PABP1 is an RNA-binding protein that binds to the poly-A tail of mature mRNAs to prevent their degradation and regulate translation. It consists of four RNA-binding domains and a C-terminal domain (PABC) that has a regulatory role. The association of PABC with other proteins containing the PAM2 motif leads to different physiological outcomes and has been associated with the maintenance of mRNA integrity (when interacting with La-related protein 1 and 4; LARP1 or LARP4)[149], the promotion of mRNA translation (when interacting with polyadenylate-binding protein-interacting protein 1; PAIP1)[150], or its inhibition (when interacting with polyadenylate-binding protein-interacting protein 2; PAIP2)[151].

We found that several viruses such as Hendra virus (HeV), Nipah virus (NiV), and human coronavirus 229E (HCoV_229E) hijack the C-terminal domain of PABP1 (PABP1 PABC) by mimicking the PAM2_1 recognition motif described in the introduction. The affinities of the viral peptides were in the low μM range ($K_D \approx 2$ - 20 μM) and we reconfirmed the interaction in the context of full-length proteins with GST pulldowns and solved the crystal structure of PABP1 PABC in complex with viral peptides.

Next, we wanted to see if overexpression of the viral PABP1 PABC-binding peptide would have any effect on viral replication in virus infected cells. To test this, we first designed two lentiviral constructs that expressed four repeats of either the P/V/$C_{183\text{-}198}$ (HeV) peptide (183-SAG**LNP**A**A**V-P**FVP**-KNQ-198) or its negative control (SAG**AGG**A**G**VP**AAG**KNQ) and were N-terminally fused to enhanced green fluorescent protein (EGFP; Figure 12A). These constructs were designated EGFP-PABPi and EGFP-PABPi mut, respectively. To test their effects on viral replication, green monkey kidney Vero E6 cells were first incubated with one of the two constructs for 72 hours and, after successful transfection, infected with a panel of RNA viruses. The EGFP-PABPi construct significantly reduced proliferation of a broad range of viral families, including both positive- and negative-strand RNA viruses, compared with the EGFP-PABPi mut control (Figure 12B). Furthermore, we observed that cells transfected with PABPi or the PABPi control before infection with tick-borne encephalitis virus (TBEV) had a different distribution of viral replication factories, corroborating our conclusions that PABP1 impairment has deleterious effects on viral replication (Figure 12C). Finally, we confirmed that we could pull down PABP1 and its close homologue PABP4 with a GFP antibody from cells transfected with lentivirus. This served as the final confirmation that the lentiviral construct EGFP-PABPi directly targets PABP1.

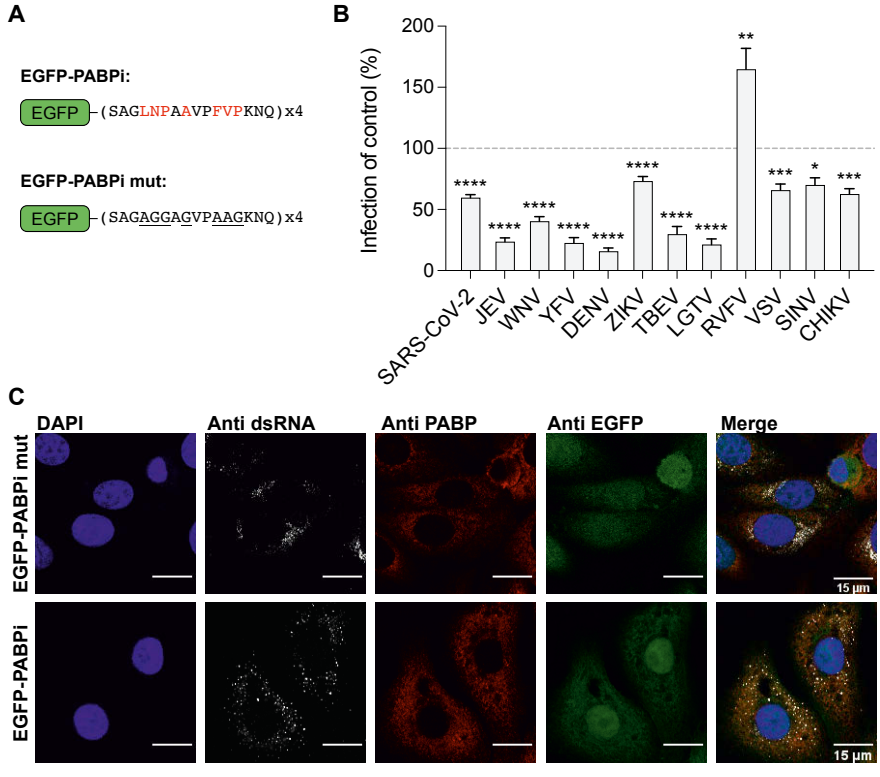


Figure 12: Viral replication is inhibited by overexpression of the PABP1-PABC binding peptide. (A) Schematic of lentiviral expression constructs. Four copies of the peptide were linked with flexible GlySerThr linkers and fused N-terminally to EGFP. **(B)** Ratio of infected cells at 16 to 24 hours post infection with the different viruses. Cells were transfected with one of the lentiviral constructs prior to the infection. The viruses used were: Japanese encephalitis virus (JEV), West Nile virus (WNV), yellow fever virus (YFV), dengue virus (DENV), Zika virus (ZIKV), tick-borne encephalitis virus (TBEV), Langat virus (LGTV), Rift Valley fever virus (RVFV), vesicular stomatitis virus (VSV), Sindbis virus (SINV), and chikungunya virus (CHIKV). Significance is indicated (*p< 0.05, **p< 0.01, ***p<0.001, and ****p<0.0001). **(C)** VeroB4 cells transfected with the EGFP-PABPi and EGFP-PABPi mut lentiviral constructs prior to infection with TBEV. In the second panel showing the localisation of double-stranded RNA, reporting the sites of active viral replication, there is a clear difference between the effects of the two lentiviral constructs. (Adapted from Figure 6 of paper III)

Overall, we found that various human targets are hijacked by viral SLiM mimicry, that viral and human SLiMs have comparable affinity for their targets, and that we can use the gained knowledge to develop potent inhibitors of viral proliferation.

SARS-CoV-2 proteins bind to human SLiMs (paper IV)

As the SARS-CoV-2 pandemic engulfed the world, forcing countries into varying degrees of imposed or voluntary lockdown, the scientific community rushed to investigate every possible aspect of the virus with unprecedented urgency and vigour. As a result, a staggering 180,000 research papers on the topic have been published in the last three years (179,649 entries in PubMed under the search term SARS-CoV-2, as of 12th of October 2022). Early in the pandemic, we realised that we could use our HD2 library described in paper I to search for the instances where human SLiMs bind to viral globular domains. This phenomenon of viral globular domains interacting with host motifs is rarer than the converse examples described in papers II and III but is not unknown[3,152-154]. SARS-CoV-2 is a positive-strand RNA virus and its large genome encodes a total of 29 proteins[155,156]. Nonstructural proteins 1-16 (Nsp1-16) are translated from an open reading frame 1ab (ORF1ab) and autoproteolytically cleaved by proteases Nsp3 and Nsp5 to form mature proteins (Figure 13). These nonstructural proteins are responsible for silencing host translation, formation of double-membrane replication vesicles, viral RNA replication and translation, and the evasion of the immune response[157-161]. In addition, thirteen minor ORFs produce four structural proteins: spike (S), envelope (E), membrane (M) and nucleoprotein (N), as well as nine smaller acessory proteins (Figure 13).

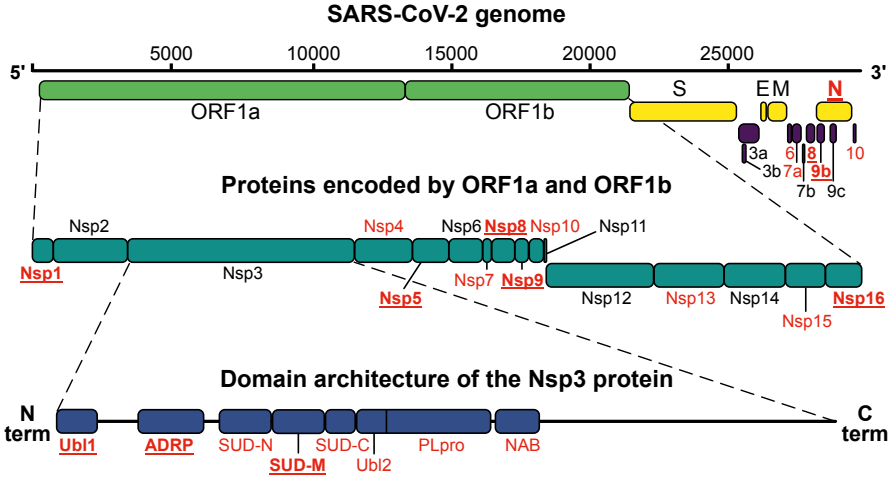


Figure 13: Organization of SARS-CoV-2 genome. The nearly 26,000 base RNA genome is translated into 29 full length proteins. ORF 1ab encodes Nsp1-16. The organisation of all globular domains of Nsp3 is shown in the last schematic. Proteins and protein domains that were successfully purified and screened against the HD2 library are coloured red, while those that were further validated are also in bold and underlined. (Adapted from Figure 1 of paper IV)

We produced 31 expression constructs of SARS-CoV-2 protein domains and were able to purify 26 of them in sufficient quantity for phage display selections against the HD2 library. Phage enrichment was observed for eleven domains, and binding of the identified potential human ligands was tested (domains and proteins shown in red, bold and underlined in Figure 13). Six of these domains and proteins did not show saturating binding for their targets, so the remaining five domains were subjected to further validation: the three Nsp3 domains Ubl1, ADRP and SUD-M, as well as Nsp9 and Nsp16.

Three domains of Nsp3 bind human SLiMs

Based on the phage display selections against Nsp3 Ubl1, ADRP and SUD-M, seven peptides were selected for affinity measurements using FP. They showed affinities in the mid to high micromolar range (30-300 μM: example for Nsp3 SUD-M in Figure 14A). To map peptide residues crucial for interaction, we performed alanine scan SPOT arrays. In these experiments, each position of the peptide was mutated to alanine, and the variants that did not facilitate binding showed negative signal (Figure 14B).

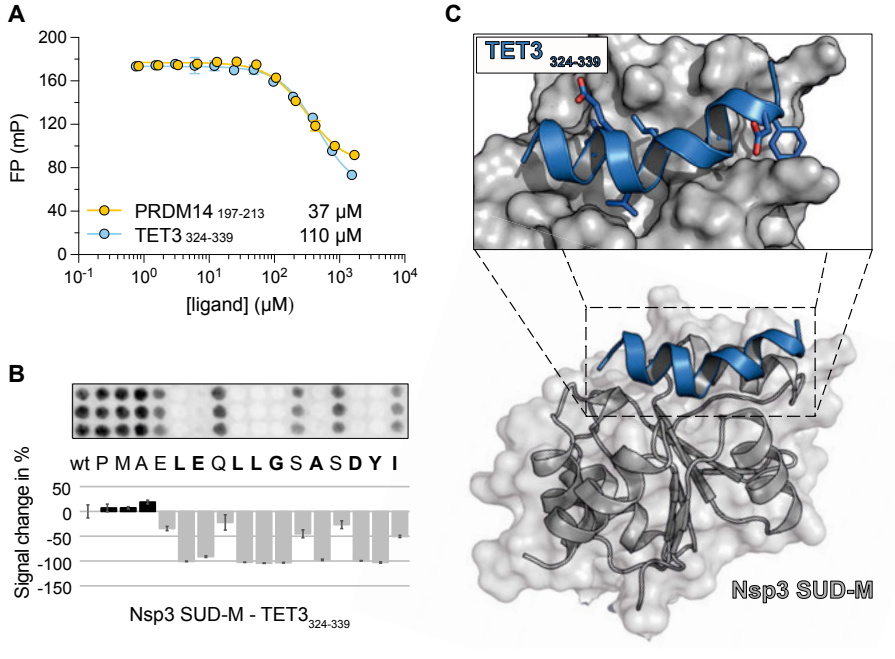


Figure 14: Analysis of the interaction between NSP3 SUD-M and TET3$_{324-339}$. (A) FP-monitored affinity measurements with K_D for each indicated peptide. **(B)** SPOT array experiments for the interaction between NSP3 SUD-M and TET3$_{324-339}$. Residues mediating the interaction are shown in bold. **(C)** Colabfold predictions for the interaction. The peptide was predicted to adopt an α-helical conformation upon binding. In the top panel, the interaction-mediating residues identified by SPOT arrays are shown as sticks. These predominantly hydrophobic residues are buried in the hydrophobic pocket of Nsp3 SUD-M. (Adapted from Figure 3 of paper IV)

To further analyse the potential binding interfaces between the three domains and the peptides, we performed ColabFold[162] predictions. ColabFold is a publicly available version of AlphaFold2[23] that accurately predicts protein folding using only primary structure information and can also be applied to predict binding interfaces between proteins. Combined information from SPOT arrays and ColabFold predictions allowed us to construct plausible interaction interfaces of Nsp3 globular domains interacting with identified peptides. The example of Nsp3 SUD-M interacting with the methylcytosine dioxygenase TET3-derived peptide (324-PMAELEQLLGSASDYI-339) is shown in Figure 14C. The largely hydrophobic LE.LLG.A.DY motif appears to mediate this particular interaction.

Peptides containing the GΦxΦ[GDP] motif interact with the C-terminal α-helix of Nsp9 and disrupt its dimerization

Nsp9 is a small RNA-binding protein that shares 98% sequence identity with its ortholog from SARS-CoV and its function was unknown until recently. In August 2022, Park *et. al.* described the function of Nsp9 as a crucial component of the viral RNA replication transcription complex, where it is involved in the early steps of 5'capping of newly synthesised viral RNA[161]. The ProP-PD results identified 147 human peptides as potential binders and showed an enrichment of GΦxΦ[GDP]. This was an interesting observation as the Nsp9 sequence itself contains the GΦxΦG motif in its C-terminal α-helix, which facilitates interaction with the RNA-dependent RNA polymerase Nsp12 and is vital for proper function of Nsp9[163]. The same C-terminal α-helix has also been proposed to mediate Nsp9 dimerization[164,165]. Considering this, we selected nine peptides for further validation using FP, with the tightest binders $LMTK3_{22\text{-}36}$ and $NKRF_{8\text{-}23}$ displaying K_D values of 34 μM and 50 μM, respectively. Interestingly, peptide 4237-LNR**GM**V**LG**SLAATVR-4251 derived from the C-terminal α-helix of Nsp9, showed no measurable affinity for full-length Nsp9.

As with the Nsp3 domains, we performed SPOT array alanine scan and ColabFold predictions to map the binding interface of Nsp9 and the peptides. While the spot arrays corroborated the importance of the GΦxΦ[GDP] motif, the ColabFold predictions did not converge to a consensus binding interface. To clarify this, we performed nuclear magnetic resonance (NMR) spectroscopy experiments. We expressed the ^{15}N-labeled Nsp9 and performed heteronuclear single quantum coherence (HSQC) experiments to obtain the chemical shifts in the ^{1}H and ^{15}N dimensions. Using the previously assigned spectra[166,167] we were able to assign the chemical shifts of specific residues of Nsp9 and observe their perturbation in the presence or absence of Nsp9-binding peptides (Figure 15A). In general, large chemical shift perturbation of a specific residue indicates that the chemical environment of that residue is substantially altered by the binding of the peptide.

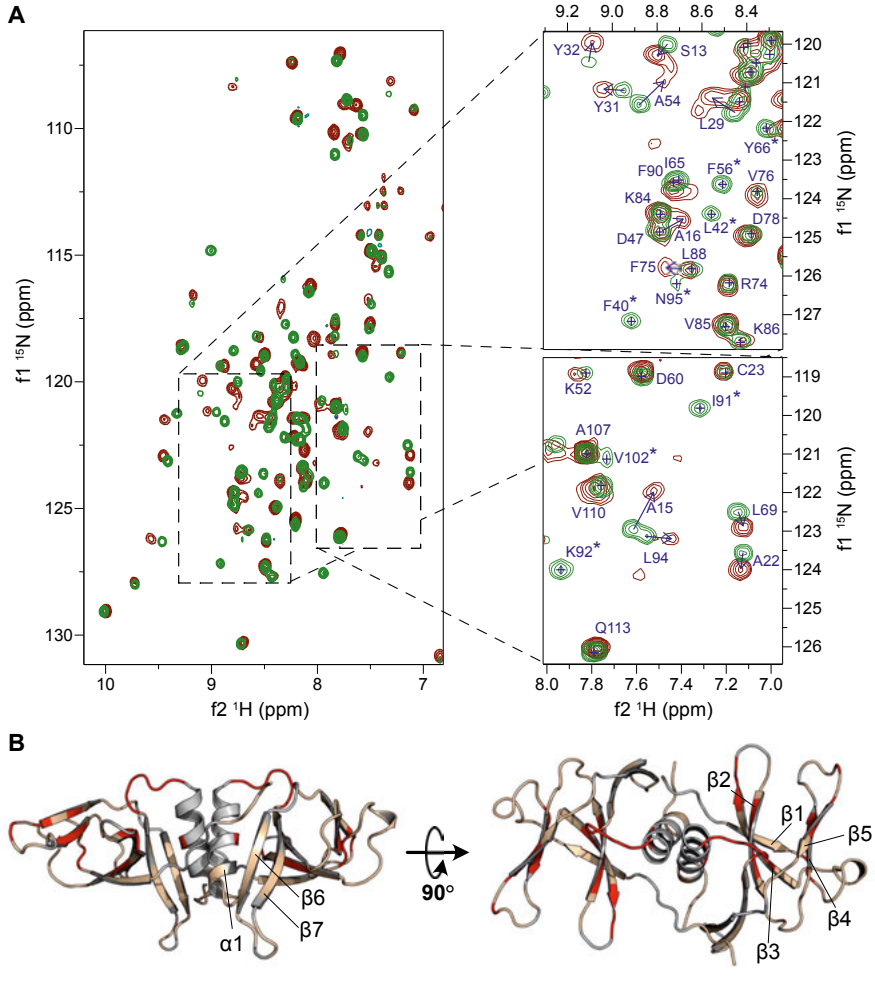


Figure 15: Binding of the peptides to the Nsp9 leads to a rearrangement of the packing in the hydrophobic core. (A) The $^{1}H^{15}N$-HSQC spectra for the free Nsp9 (green) and the Nsp9 in complex with $NKRF_{8-23}$ (red). The right panels show the detailed view of two regions showing the peak shift perturbations upon addition of the peptide. Arrows indicate the direction of the peak shift perturbation, and an asterisk marks the residues that disappeared upon addition of the peptide, suggesting large perturbations of the chemical environment. **(B)** Structure of Nsp9, where the residues showed a chemical shift perturbation larger than one standard deviation are coloured red, while the residues that were below this threshold are beige. The residues for which no chemical shifts could be assigned are shown in grey, and the major secondary structure elements are indicated. The PDBid: 6wxd model was used for the visualization. (Adapted from Figure 4 of paper IV)

Figure 15B shows the peak shift perturbations in the context of the protein structure. We found that the residues with the largest perturbations cluster in and around the hydrophobic core of the protein. Based on that, we speculated that the C-terminal α-helix residues are involved in binding to the peptides

and that the peptide ligands bind to the Nsp9 homodimer interface. As a result, the packing of the core could be disrupted, which would explain our initial observations, but the chemical shifts have not been detected for this region of the protein, preventing any concrete conclusions. To address this issue, we constructed an Nsp9Δα mutant with the C-terminal helix removed. This construct did not bind to the peptides containing GΦxΦ[GDP], indicating that the C-terminal helix is crucial for binding. Finally, we measured the rotational correlation time (τ_c) of Nsp9 in the presence and absence of peptide ligands. The rotational correlation time is used to estimate the size of the molecule based on the rate of its tumbling in solution. The results indicated that Nsp9 forms a dimer in the absence of the peptide at a concentration of ~200 μM, but this is perturbed by the addition of peptide resulting in a more monomer-like population in the presence of the ligands.

Taken together, these results suggest that the peptides with the GΦxΦ[GDP] motif interact with Nsp9 via the C-terminal α-helix. This binding disrupts the formation of the Nsp9 homodimer and should also disrupt the interaction between Nsp9 and Nsp12, as this contact is also mediated via the C-terminal α-helix of Nsp9.

Identified Nsp16-binding peptides compete with Nsp10 for binding to Nsp16

In the last example, we investigated the binding of peptides to Nsp16, an enzyme responsible for methylating the first nucleotide of newly synthesized viral RNA to form a fully functional $^{7Me}GpppA_{2'\text{-}O\text{-}Me}$-capped RNA. The ProP-PD identified three potential binders for the Nsp 16, derived from dual specificity tyrosine-phosphorylation-regulated kinase 1B (DYRK1B$_{395\text{-}410}$; PGHSPADYLRFQDLVL), islet cell autoantigen 1-like protein (ICA1L$_{450\text{-}465}$; NQDMSA**W**FNL**F**ADLDP), and uncharacterized protein FLJ43738 (FLJ43738$_{76\text{-}91}$; EDPLDSYLN**F**QALISP). All peptides displayed moderate binding affinities of ~ 30-60 μM and the SPOT arrays performed for DYRK1B$_{395\text{-}410}$ and ICA1L$_{450\text{-}465}$ implicated hydrophobic residues of the motif (bold in the previous sentence). Because Nsp16 requires association with Nsp10 to be activated, we were interested if the identified peptides bind to the same binding interface and compete with Nsp10 for binding. We purified full-length Nsp10 and demonstrated that it displaces the labeled DYRK1B$_{394\text{-}409}$, confirming the common binding interface with Nsp16[168,169]. This also allowed us to determine the affinity between Nsp16 and Nsp10 to be 5.5 μM. This was however not surprising, since Nsp10 interacts with Nsp16 through a much larger interface. Consistent with this observation, ColabFold also predicted that peptides bind to the same hydrophobic binding pocket on Nsp16 as Nsp10, formed by α-helices 3, 4, and 10, respectively. This confirmed that the identified peptide ligands compete with Nsp10 for the interaction with Nsp16 and could therefore serve as inhibitors of the viral 5'capping process.

Combined knowledge can be used to design inhibitors of viral replication

Since we have shown that identified peptides compete with viral proteins for binding to their viral interaction partners, the natural question was whether we could exploit this fact to design viral inhibitors. We designed lentiviral constructs expressing four copies of human protein-derived peptides fused to EGFP in an analogous manner as described in paper III. These constructs targeted six SARS-CoV-2 protein or protein domains, and five of them significantly inhibited SARS-CoV-2 replication in VeroE6 cells (Figure 16A-B).

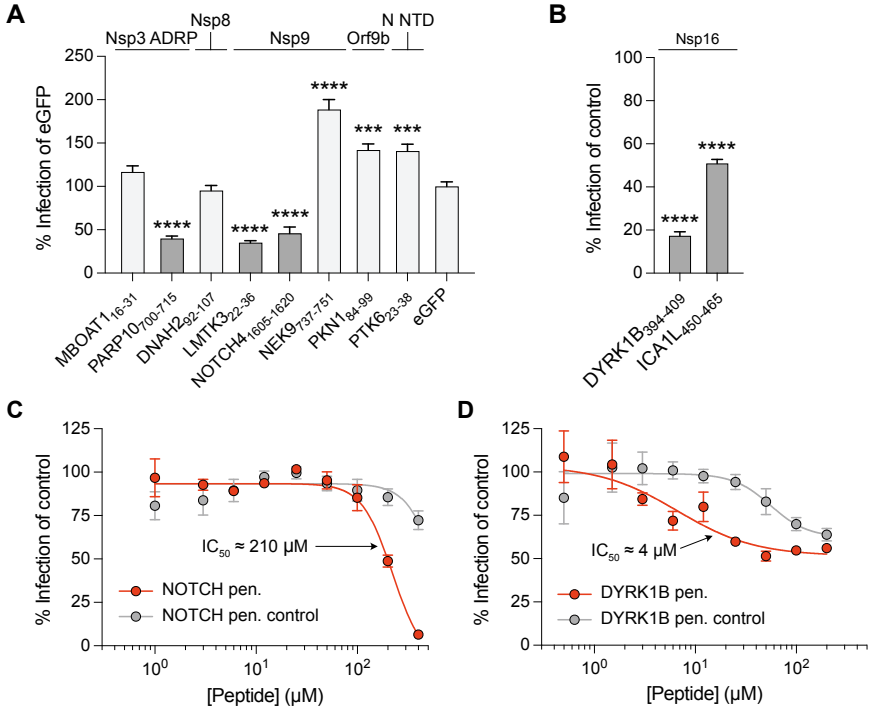


Figure 16: Lentiviral constructs and cell penetrating peptides inhibit SARS-CoV-2 replication. (A) Comparison of SARS-CoV-2 replication after treatment with lentiviral constructs targeting different proteins of SARS-COV-2. The targets are indicated above. An "empty" EGFP-expressing lentivirus was used as negative control. Cases in which viral replication was inhibited are shown in dark grey. **(B)** Inhibition of viral replication by lentiviral constructs targeting Nsp16. In this case, a construct in which the interaction motif was mutated was used as negative control. **(C)** The effect of the cell-penetrating NOTCH4 peptide targeting Nsp9. In the control constructs, residues important for binding were mutated. IC_{50} for the wt peptide is indicated. **(D)** As in (C), but with the cell-penetrating DYRK1B peptide. (Adapted from Figure 6 of paper IV)

Based on the inhibition assay, we designed three cell-penetrating peptides by fusing them N-terminally to the Tat-derived sequence (HIV1; YGRK-KRRQRRRGSG). These peptides were termed NOTCH4 pen, LMTK3 pen, targeting Nsp9 and DYRK1B pen targeting Nsp16. NOTCH4 and DYRK1B cell-penetrating peptides showed inhibitory effect on viral replication (Figure 16 C-D) establishing the proof of concept and showing that the understanding obtained from ProP-PD can lead to early-stage peptide inhibitors that could serve as a starting point for the development of peptidomimetic drugs.

Evolutionary analysis of the interaction between the SWIB domain of MDM2 and p53 (paper V and VI)

Despite the great utility of the phage display as an excellent tool for large-scale discovery of SLiMs, it provides very limited details on the biophysical characteristics of individual interactions. To fully explore PPIs, a more focused investigation and concentration on only a handful of interactions is required. With this in mind, I decided to focus my last project on exploring one particular SLiM-mediated interaction between MDM2 and p53. This extensively studied interaction was first described by Arnold J. Levine's group in 1992[170] and the same group solved its crystal structure four years later[140]. The MDM2-p53 interaction was briefly described and discussed in the introduction, and within the scope of paper I. It occurs between the MDM2 SWIB domain and the F…W..L motif found in the disordered transactivation domain (TAD) of p53 (Figure 17). Functionally the p53TAD motif serves as a docking site for MDM2 E3 ligase which recognises p53 and ubiquitinates it. This leads to degradation and low levels of p53 in healthy cells. However, in the presence of excessive stress signals, the degradation of MDM2 by p53 is impaired, leading to the accumulation of p53, which triggers the expression of stress response genes that promote cell arrest, and DNA repair or ultimately lead to apoptosis[171,172].

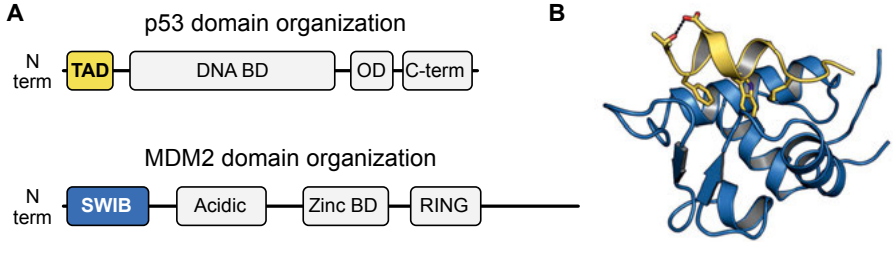


Figure 17: Overview of MDM2 SWIB-p53 TAD interaction (A) Schematic representation of both proteins. **(B)** crystal structure model of the complex. MDM2 SWIB is blue and p53 TAD in yellow (PDBid: 1ycr). Residues of recognition motif are shown as sticks. (Adapted from Figure 1 of paper V)

Ancestral MDM2-p53 interaction had high affinity but it was retained only in the tetrapod lineage.

As discussed in the introduction, rapid evolution and a degenerate nature are two important characteristics of SLiMs. In paper V, we aimed to compare the interaction between the MDM2-SWIB domain and p53 TAD in extant species that encode both proteins but in which the canonical motif has been altered substantially (Figure 18). We were also interested in whether we could unravel the evolutionary events that have modulated affinity over time. To accomplish this, we performed an ancestral reconstruction of both MDM2 SWIB and p53 TAD because that allowed us to investigate the evolutionary trajectory of this SLiM-mediated interaction.

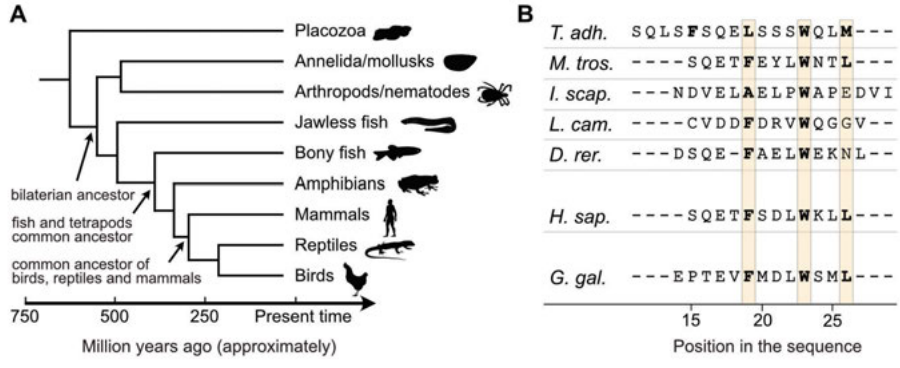


Figure 18: Variation in the present-day MDM2-binding motif of p53. (A) Phylogenetic relationship of the species from which the MDM2 and p53 sequences were obtained. The arrows indicate the nodes for which the ancestral MDM2 SWIB and p53TAD were reconstructed. **(B)** N-terminal MDM2 recognition motif of p53 from different extant species. Sequences are from placozoa (*Trichoplax adhaerens*), molluscs (bay mussel, *Mytilus trossulus*), arthropods (deer tick, *Ixodes scapularis*), jawless fish (Arctic lamprey, *Lethenteron camtschaticum*), bony fish (zebrafish, *Danio rerio*), mammals (human, *Homo sapiens*), and birds (chicken, *Gallus gallus*). The residues of the core $F_{19}...W_{23}..L_{26}$ motif are shaded yellow and are in bold when they correspond to the canonical recognition motif. (Adapted from Figure 1 of paper V)

Ancestral reconstruction of proteins attempts to recreate the most likely ancestral variant of a present-day protein based on multiple sequence alignment of protein orthologs from extant species and their phylogenetic relationship. Although this concept was proposed as early as 1963[173], it took another 30 years before the idea could be implemented and tested[174]. The reason for the delay was that high confidence resurrection required numerous sequences from various, evolutionary diverse species. Only recently have advances in sequencing technology facilitated an explosion of annotated genomes of species from all branches of life, dramatically improving the quality and reliability of ancestral reconstruction attempts.

While MDM2 is well conserved and reconstruction was relatively reliable, the p53 TAD exhibited a high degree of divergence and could not be reliably reconstructed outside the core MDM2 recognition motif. We succeeded in reconstructing the MDM2 SWIB and p53 motifs of the last common ancestor of reptiles and mammals, which lived about 320 million years ago, and the ancestor of fishes and tetrapods, which lived roughly 390 million years ago.

We measured affinity for the interaction between MDM2 and p53 for present-day species using a combination of stopped-flow spectroscopy and FP. The advantage of using stopped flow is that we were able to obtain the intrinsic kinetic constants for the association and dissociation of the complex (k_1 and k_{-1} from equation 1), which provide additional layer of information about the underlying parameters of the interaction and allows a more accurate measurement of K_D. However, the benefit of kinetic constant resolution comes at the cost of high protein consumption, low throughput, and a lengthy optimization process. Of the living species, birds and mammals showed the tightest affinity in the range of 0.1 μM, with resurrected ancestors for this node binding with similar affinity. Affinity was 20-fold lower for the extant fish *D. rerio* and even further reduced for species evolutionarily distant from the tetrapods (Figure 19). This could be inferred given the wide divergence of p53 TAD but is however surprising considering the central role of the interaction in tetrapods. Interestingly, the predecessor of fish and tetrapods interacted with an affinity comparable to that of present-day tetrapods. Moreover, the bilateral p53 node bound to MDM2 with tetrapod-like affinity. Importantly, the interaction for the bilateral node was measured with human MDM2 since the resurrection of this node could not be performed due to large degree of sequence divergence. Therefore, conclusions about early evolutionary events should be drawn with caution. Nevertheless, the most plausible explanation is that the sub micromolar affinity between MDM2 SWIB and p53 TAD emerged early in evolutionary history and was then either lost or increased as in birds and mammals.

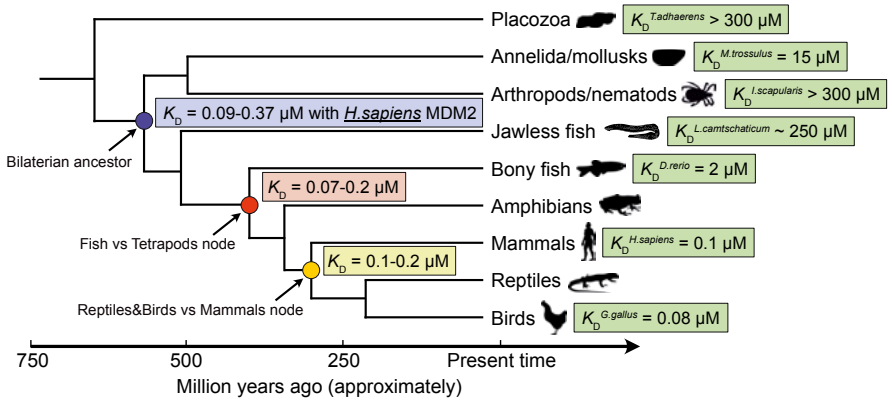


Figure 19: Evolutionary trajectory of the affinity between MDM2 and p53. (Figure adapted from Figure 4 in paper V)

Affinity of MDM2-p53 interaction is continuously calibrated in fish

While analysing the results from paper V, we became aware of the curious case of the evolution of the MDM2-p53 interaction in fish. As mentioned earlier, the affinity of the interaction for *D. rerio* was 2 μM, and the decrease in affinity came predominantly from the insertion of asparagine in position 26 of the canonical motif. We collected and analysed the p53 sequences from 51 ray-finned fish. Multiple sequence alignment revealed that the first and second specificity-determining positions of the motif (F_{19} and W_{23}) remained conserved across the fish lineage. However, we observed an interesting pattern of subtle variation around position 26 of the recognition motif, where L_{26} was frequently substituted with valine, or an asparagine was inserted increasing the gap between W_{23} and L_{26}. In addition, we noted that there were two paralogs of p53 in the sequence databases, which can be explained by the whole genome duplication at the base of the teleost fish lineage[175,176]. This prompted us to perform a similar ancestral reconstruction as described in paper V, but this time include only fish species (described in paper VI). As before, we observed that the ancestral complexes had affinities in the sub micromolar range that changed as the species diversified. These changes regulating affinity were found primarily in the C-terminal part of the canonical motif surrounding or affecting L_{26}. Overall, contemporary fish species have predominantly lost the high-affinity interaction between MDM2 and the canonical p53 TAD motif

To confirm that the affinities we measured for the MDM2-p53 interaction are also relevant in the context of full-length complexes (or at least a better approximation thereof), we measured the affinity between *D. rerio* MDM2 and a longer p53 TAD construct spanning the entire disordered region. We found that the full-length *D. rerio* p53 TAD binds to *D. rerio* MDM2 SWIB with an affinity of 0.2 μM, which is comparable to the affinity for the complex of other tetrapods. This was surprising as previous reports describing the human interaction established that only the canonical motif determines the affinity of the interaction[177]. In the case of *D. rerio*, it appeared that there were additional residues outside of the canonical motif that increased affinity. To address this issue, we tiled the entire *D. rerio* p53 TAD with shorter peptides (Figure 20). Surprisingly, all short peptides bound with a K_D of 2.2 μM or higher. Coincidentally, we noticed that the canonical sequence in Uniprot lacks the serine in position 29 (Uniprot entry P79734). This serin is however present in the non-canonical sequence (Uniprot entry G1K2L5), which was also used for sequence alignment and reconstruction because it better aligns with proteins from other fish species. When we tested a variant of the peptide that contained the serine at position 29, the high binding affinity was recapitulated. Based on these results we were able to propose a model for tight binding of full-length *D.rerio* p53 TAD, in which the secondary motif $W_{23}...L_{27}..I_{30}$ is recognised and increases the overall binding affinity (Figure 20B). The

result could have an alternative explanation, with the S_{29} and residues surrounding it forming additional specific interactions. This example provides an excellent illustration of the remarkable plasticity of SLiM-mediated interactions and underscores the importance of correct protein sequence annotation in established reference databases such as Uniprot.

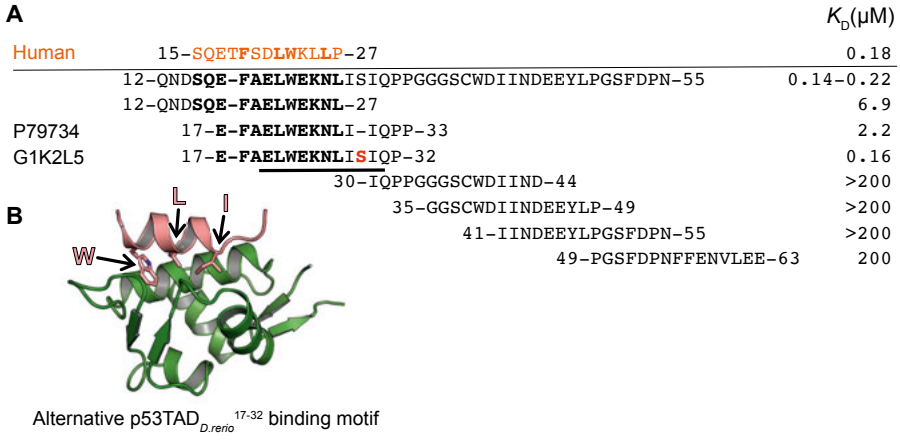


Figure 20: Alternative binding of full-length *D.rerio* p53. (A) Alignment of all peptides for which affinities were measured. The canonical binding motif is shown in bold. Remarkably, the peptide with the S_{29} insertion restored binding. The serine in question is shown in red and bold. **(B)** A ColabFold prediction for the interaction between the MDM2 SWIB domain and the alternative W_{23}...L_{27}..I_{30} binding motif. (Figure adapted from Figure 4 in paper VI)

Conclusions and future perspectives

Since the discovery of the cell in the 17th century by Robert Hooke[178], people have tried to decipher and understand the central patterns that allow for the emergence of life. In order to be able to manipulate fundamental cellular processes, a complete understanding of all participating components is necessary. This has proven to be a monumental challenge as the complexity of the system grows exponentially with the number of its constituent parts. Despite the great advances over the past 300 years that have led to the emergence of modern medicine and extended life expectancy by decades while producing the greatest human prosperity in history, detailed understanding of the most fundamental component of our body is still drastically incomplete.

In the work presented here, we have attempted to fill the knowledge gap regarding SLiM-mediated interactions. Inadequate screening techniques have largely overlooked this important group of protein interaction modules, and better approaches for systematic discovery and analysis need to be developed. We have designed and tested a novel approach based on ProP-PD to screen for SLiMs in humans and viruses. In our experiments, we screened only between 34 and 139 domains for interactions with SLiMs, but the experimental design allows for rapid scaling of the number of globular domains tested. As the domain purification was performed at low throughput in our experiments, this step proved to be the bottleneck. Nonetheless, since the publication of paper I, hundreds of additional domains have been investigated, considerably expanding our motif dataset and will be described in future publications. Nothing stands in the way of the complete annotation of all SLiMs in the human proteome, except the number of PhD students willing to work on this endeavour.

Regarding viral hijacking of SLiM-mediated interactions, we have presented the first non-computational technique that systematically screens the entire viral disorderomes for motif-based interactions. Based on this approach, we expanded the number of known ESCRT pathway hijacking motifs and showed that interference with clathrin by viral proteins disrupts cellular trafficking. In addition, we established the C-terminal domain of PABC as a novel example of viral SLiM mimicry. This work will be further expanded in the future as additional domains are investigated. Moreover, we have developed a second phage display library for the "viral disorderome" that also contains disordered regions of DNA viruses and thus represents the entire viral

disorderome (of viruses that infect humans). One of the key challenges ahead is to integrate these large datasets to gain new insights into the overarching principles that govern SLiM-mediated interactions. We are currently optimising a systems biology approach to address this issue.

The rapid technological development of our time means that various techniques and approaches are becoming obsolete at an unprecedented rate. Since the start of this project five years ago and the development of the first phage libraries, new advances are already demonstrating the need for an updated design. As an unintended consequence, the introduction of AlphaFold has produced the most accurate disorder prediction tool to date. New disorderome libraries based on AlphaFold and potential future improvements will be the next step in creating more reliable and accurate data at higher resolution. Another improvement would be to tile the disordered regions even more tightly, for example with a sliding window of two. Apart from improving the technique, the biggest challenge will be to validate all proposed novel SLiM-mediated interaction to uncover new biologically relevant functions and processes. As there is an increasing need for novel drug targets in the pharmaceutical industry and as the field moves away from small, easily druggable enzyme pockets towards targeting more complex interaction interfaces, SLiM-mediated interactions offer a new avenue for future drug design.

Finally, our exploration of the evolutionary development of a single motif mediated interaction between MDM2 and p53 highlights how complex even the simplest models can be. In addition, it showcased incredible plasticity and rapid evolution of SLiM-mediated interactions. It has been suggested that motifs may be one of the critical drivers of the evolution of interaction networks, as new motifs readily emerge *ex nihilo* and exhibit incredible versatility[32]. These and similar questions remain to be answered in this ever-growing field of interactions mediated by short linear motifs.

And so it will continue, with each question answered five new questions arise.

Acknowledgements

In the end I would like to thank all the people who have supported and guided me through the past five years.

First and foremost, I would like to express my deepest gratitude to my supervisor *Per Jemth*, whose doors were always open for any smart (or stupid) question. Thank you for accepting me into your group, teaching me about protein interactions, giving me the opportunity to pursue my scientific interests, and teaching me how to write scientific literature. I thoroughly enjoyed our discussion and journal clubs, where conversation often devolved from purely scientific to debates of more controversial topics. The amount of confidence and trust you placed in me never ceased to amaze me (I would never trust myself as much…). I could not ask for a more supportive supervisor. Hopefully I see you at the start of next Stockholm marathon. You are the best!

I was extremely fortunate to have *Ylva Ivarsson* as my co-supervisor, who partially adopted me into the B7:3 corridor. You were the motivation and driving force behind most of the projects I was involved in, and none of it would have happened without you. Every once in a while, I would pass your office to start a random chat and realize 25 minutes later that I was supposed to be running an experiment. Thank you for taking me under your wing and introducing me to all your colleagues at Seefeld, it was a great experience.

A special thanks to *Norman*, *Öla*, *Evangelia*, *Anna*, and *Doreen* for the guidance and expertise needed to steer such a large set of projects in the right direction. Your contribution was immensely appreciated.

Ravi, thank you for teaching me how to love cell work. Your positive attitude and big smile were a huge inspiration to me and helped me get through some tough times. I will deeply miss running into you in the corridors, playing ping pong with you, and discussing all topics under the sun with you.

Eva, who kept the lab on its feet, tirelessly fixed the äkta and the mass spec, and basically ran my projects for me towards the end of my PhD. For all your help, you should get at least half of a PhD!

Elin, with whom I shared the office throughout my entire PhD. I loved having lunch and fika with you, and you were a great lab companion. Your dedication and hard work are deeply inspiring.

NMR superstar *Chi*, who`s magic hands made the NMR machine do the things I had no idea it could (and still do not understand). Thank you also for your positive attitude and humour in the lab, it was great to have you around.

Louise and *Weihua* for all the interesting discussions and laughter we have shared over the years.

To my former students *Hanna*, *Niels*, and *Kit* who worked hard in the lab, I wish you all the best in your future endeavours.

Thank you*, Leandro,* for making the best figures*, Caroline,* for performing a million phage display selections*, Marie,* for battling with PLA to understand all of the ambiguous results, *Eszter*, for incorporating mass spec and SPOT arrays in our projects*, Dilip,* who patiently taught me the basics of crystallography, *Girolamo,* who performed bioinformatic analyses ten paygrades above my level, and *Richard*, who helped us with viruses and extended our hypotheses to cellular systems*,* your technical skills always amazed me.

Vladimir, my friend, you are one of the most knowledgeable people I know and the main reason I decided to pursue a PhD. I could always count on your sarcastic answer to any dumb question I had.

My personal Dutch support group *Jeroen* and *Lianne*, who put a roof over my head when I came back to Uppsala. Evenings with tea and the Nationals are some of the best memories I have of my time here.

Ali, I will miss losing to you at ping pong and squash. I hope one day we can take a road trip to the North Cape.

Viktor, Isa, Tifaine, Petra and *Chris*, for all the good times we shared at the IPhA meetings and after works.

Johanna and *Malin* for all the fikas

To all others who have contributed to this work, I am deeply grateful.

Finally, I want to express my deepest appreciation for *Janina,* who has stood by my side over the past year, designed the most beautifulest cover, and helped with suggestions and proofreading of this work. Without you, every “the” in this thesis would be wrong.

And lastly, I want to thank my family: mom *Suzana* for all the love you have shown me over the past 30 years and for always knowing what is best for everyone else except you (koprive)*,* dad *Rafko* for always begrudgingly help me study math one night before the exams*,* and sister *Zala* for being so exquisitely annoying as only younger sisters can be, you have grown into an amazing young woman, I am proud of you. Thank you all for supporting me and believing in me when the future seemed less promising. Knowing you were just a phone call away was a great comfort on those cold, dark winter evenings in Uppsala. I love you all deeply.

References

1 Stumpf, M. P. *et al.* Estimating the size of the human interactome. *Proc Natl Acad Sci U S A* **105**, 6959-6964 (2008).

2 Akey, C. W. *et al.* Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* **185**, 361-378.e325 (2022).

3 Davey, N. E., Travé, G. & Gibson, T. J. in *Trends in Biochemical Sciences* Vol. 36 159-169 (Trends Biochem Sci, 2011).

4 Davey, N. E. *et al.* Attributes of short linear motifs. *Molecular BioSystems* **8**, 268-281 (2012).

5 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).

6 Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* **49**, D437-D451 (2021).

7 Fersht, A. R. in *Nature Reviews Molecular Cell Biology* Vol. 9 650-654 (Nature Publishing Group, 2008).

8 Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**, 321-331 (1999).

9 Uversky, V. N. Recent Developments in the Field of Intrinsically Disordered Proteins: Intrinsic Disorder-Based Emergence in Cellular Biology in Light of the Physiological and Pathological Liquid-Liquid Phase Transitions. *Annu Rev Biophys* **50**, 135-156 (2021).

10 Dyson, H. J. & Wright, P. E. in *Nature Reviews Molecular Cell Biology* Vol. 6 197-208 (2005).

11 Gianni, S. & Jemth, P. Affinity versus specificity in coupled binding and folding reactions. *Protein Eng Des Sel* **32**, 355-357 (2019).

12 Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife* **8** (2019).

13 Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **37**, 509-516 (2012).

14 Wright, P. E. & Dyson, H. J. in *Nature Reviews Molecular Cell Biology* Vol. 16 18-29 (Nature Publishing Group, 2015).

15 van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 6589-6631 (2014).

16 Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. *J Biol Chem* **276**, 13505-13508 (2001).

17 Karamouzis, M. V., Konstantinopoulos, P. A. & Papavassiliou, A. G. Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis. *Cell Res* **17**, 324-332 (2007).

18 Goodman, R. H. & Smolik, S. in *Genes and Development* Vol. 14 1553-1577 (Cold Spring Harbor Laboratory Press, 2000).
19 Contreras-Martos, S. *et al.* Linking functions: An additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Scientific Reports* **7**, 1-13 (2017).
20 Chan, H. M. & La Thangue, N. B. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *Journal of Cell Science* **114**, 2363 LP-2373 (2001).
21 Oates, M. E. *et al.* D2P2: Database of disordered protein predictions. *Nucleic Acids Research* **41** (2013).
22 Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-645 (2004).
23 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
24 Fisher, C. K. & Stultz, C. M. Protein structure along the order - Disorder continuum. *Journal of the American Chemical Society* **133**, 10022-10025 (2011).
25 Yang, J., Gao, M., Xiong, J., Su, Z. & Huang, Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Science* **28**, 1952-1965 (2019).
26 Uversky, V. N. in *Protein Science* Vol. 22 693-724 (Wiley-Blackwell, 2013).
27 Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Science* **11**, 739-756 (2002).
28 Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. in *Current Opinion in Structural Biology* Vol. 18 756-764 (Elsevier Current Trends, 2008).
29 Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol Cell* **55**, 161-169 (2014).
30 Kumar, M. *et al.* The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res* **50**, D497-D508 (2022).
31 Boehr, D. D., Nussinov, R. & Wright, P. E. in *Nature Chemical Biology* Vol. 5 789-796 (Nature Publishing Group, 2009).
32 Davey, N. E., Cyert, M. S. & Moses, A. M. Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun Signal* **13**, 43 (2015).
33 Davey, N. E., Shields, D. C. & Edwards, R. J. SLiMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res* **34**, 3546-3554 (2006).
34 Kelil, A., Dubreuil, B., Levy, E. D. & Michnick, S. W. Fast and accurate discovery of degenerate linear motifs in protein sequences. *PLoS One* **9**, e106081 (2014).
35 Kozlov, G., Ménade, M., Rosenauer, A., Nguyen, L. & Gehring, K. Molecular determinants of PAM2 recognition by the MLLE domain of poly(A)-binding protein. *J Mol Biol* **397**, 397-407 (2010).
36 Kozlov, G., Safaee, N., Rosenauer, A. & Gehring, K. Structural basis of binding of P-body-associated proteins GW182 and ataxin-2 by the Mlle domain of poly(A)-binding protein. *J Biol Chem* **285**, 13599-13606 (2010).
37 Ruan, L. *et al.* Quantitative characterization of tob interactions provides the thermodynamic basis for translation termination-coupled deadenylase regulation. *Journal of Biological Chemistry* **285**, 27624-27631 (2010).

38 Kozlov, G. *et al.* Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase. *EMBO Journal* **23**, 272-281 (2004).
39 Johansen, T. & Lamark, T. in *Journal of Molecular Biology* Vol. 432 80-103 (Academic Press, 2020).
40 Birgisdottir, Å. B. *et al.* Members of the autophagy class III phosphatidylinositol 3-kinase complex I interact with GABARAP and GABARAPL1 via LIR motifs. *Autophagy* **15**, 1333-1355 (2019).
41 Jones, J. E., Le Sage, V. & Lakdawala, S. S. Viral and host heterogeneity and their effects on the viral life cycle. *Nat Rev Microbiol* **19**, 272-282 (2021).
42 Masters, P. S. The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292 (2006).
43 Finkel, Y. *et al.* The coding capacity of SARS-CoV-2. *Nature* **589**, 125-130 (2021).
44 Adhikari, S. *et al.* A high-stringency blueprint of the human proteome. *Nat Commun* **11**, 5301 (2020).
45 Faust, T. B., Binning, J. M., Gross, J. D. & Frankel, A. D. Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol* **4**, 241-260 (2017).
46 Freire, J. M., Santos, N. C., Veiga, A. S., Da Poian, A. T. & Castanho, M. A. Rethinking the capsid proteins of enveloped viruses: multifunctionality from genome packaging to genome transfection. *FEBS J* **282**, 2267-2278 (2015).
47 Wen, X. *et al.* Multifunctionality of structural proteins in the enterovirus life cycle. *Future Microbiol* **14**, 1147-1157 (2019). https://doi.org:10.2217/fmb-2019-0127
48 Copley, S. D. An evolutionary perspective on protein moonlighting. *Biochemical Society Transactions* **42**, 1684-1691 (2014).
49 Bornholdt, Z. A. *et al.* Structural rearrangement of ebola virus vp40 begets multiple functions in the virus life cycle. *Cell* **154**, 763-774 (2013).
50 Greber, U. F. & Fornerod, M. in *Current Topics in Microbiology and Immunology* Vol. 285 109-138 (Curr Top Microbiol Immunol, 2004).
51 Cohen, S., Au, S. & Panté, N. in *Biochimica et Biophysica Acta - Molecular Cell Research* Vol. 1813 1634-1645 (Biochim Biophys Acta, 2011).
52 Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem* **284**, 478-485 (2009).
53 Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509 (1984).
54 Adam, S. A., Lobl, T. J., Mitchell, M. A. & Gerace, L. Identification of specific binding proteins for a nuclear location sequence. *Nature* **337**, 276-279 (1989).
55 Lu, J. *et al.* Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Commun Signal* **19**, 60 (2021).
56 Shaw, M. L., Cardenas, W. B., Zamarin, D., Palese, P. & Basler, C. F. Nuclear Localization of the Nipah Virus W Protein Allows for Inhibition of both Virus- and Toll-Like Receptor 3-Triggered Signaling Pathways. *Journal of Virology* **79**, 6078-6088 (2005).
57 Abaitua, F., Hollinshead, M., Bolstad, M., Crump, C. M. & O'Hare, P. A Nuclear Localization Signal in Herpesvirus Protein VP1-2 Is Essential for

Infection via Capsid Routing to the Nuclear Pore. *Journal of Virology* **86**, 8998-9014 (2012).
58 Plafker, S. M. & Gibson, W. Cytomegalovirus Assembly Protein Precursor and Proteinase Precursor Contain Two Nuclear Localization Signals That Mediate Their Own Nuclear Translocation and That of the Major Capsid Protein. *Journal of Virology* **72**, 7722-7732 (1998).
59 Lyons, R. H., Ferguson, B. Q. & Rosenberg, M. Pentapeptide nuclear localization signal in adenovirus E1a. *Molecular and Cellular Biology* **7**, 2451-2456 (1987).
60 Fontes, M. R. M. *et al.* Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-α. *Biochemical Journal* **375**, 339-349 (2003).
61 Henderson, B. R. & Percipalle, P. Interactions between HIV Rev and nuclear import and export factors: The Rev nuclear localisation signal mediates specific binding to human importin-β. *Journal of Molecular Biology* **274**, 693-707 (1997).
62 Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. & Lührmann, R. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475-483 (1995).
63 Gilmore, T. D. in *Oncogene* Vol. 25 6680-6684 (Nature Publishing Group, 2006).
64 Sen, R. & Baltimore, D. Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. *Cell* **47**, 921-928 (1986).
65 Hiscott, J., Nguyen, T. L. A., Arguello, M., Nakhaei, P. & Paz, S. in *Oncogene* Vol. 25 6844-6867 (Nature Publishing Group, 2006).
66 Hiscott, J., Kwon, H. & Génin, P. in *Journal of Clinical Investigation* Vol. 107 143-151 (The American Society for Clinical Investigation, 2001).
67 Brinkmann, M. M. & Schulz, T. F. in *Journal of General Virology* Vol. 87 1047-1074 (Microbiology Society, 2006).
68 Gires, O. *et al.* Latent membrane protein 1 of Epstein-Barr virus mimics a constitutively active receptor molecule. *EMBO Journal* **16**, 6131-6140 (1997).
69 Thorley-Lawson, D. A. in *Nature Reviews Immunology* Vol. 1 75-82 (European Association for Cardio-Thoracic Surgery, 2001).
70 Izumi, K. M. & Kieff, E. D. The Epstein-Barr virus oncogene product latent membrane protein 1 engages the tumor necrosis factor receptor-associated death domain protein to mediate B lymphocyte growth transformation and activate NF-kappaB. *Proc Natl Acad Sci U S A* **94**, 12592-12597 (1997).
71 Ye, H., Park, Y. C., Kreishman, M., Kieff, E. & Wu, H. The structural basis for the recognition of diverse receptor sequences by TRAF2. *Mol Cell* **4**, 321-330 (1999).
72 Devergne, O. *et al.* Association of TRAF1, TRAF2, and TRAF3 with an Epstein-Barr virus LMP1 domain important for B-lymphocyte transformation: role in NF-kappaB activation. *Mol Cell Biol* **16**, 7098-7108 (1996).
73 Babst, M., Katzmann, D. J., Snyder, W. B., Wendland, B. & Emr, S. D. Endosome-Associated Complex, ESCRT-II, Recruits Transport Machinery for Protein Sorting at the Multivesicular Body. *Developmental Cell* **3**, 283-289 (2002).

74 Katzmann, D. J., Babst, M. & Emr, S. D. Ubiquitin-Dependent Sorting into the Multivesicular Body Pathway Requires the Function of a Conserved Endosomal Protein Sorting Complex, ESCRT-I. *Cell* **106**, 145-155 (2001).
75 Schöneberg, J., Lee, I. H., Iwasa, J. H. & Hurley, J. H. in *Nature Reviews Molecular Cell Biology* Vol. 18 5-17 (Nature Publishing Group, 2016).
76 Hurley, J. H. ESCRT s are everywhere. *The EMBO Journal* **34**, 2398-2407 (2015).
77 Henne, W. M., Buchkovich, N. J. & Emr, S. D. in *Developmental Cell* Vol. 21 77-91 (Dev Cell, 2011).
78 Votteler, J. & Sundquist, W. I. in *Cell Host and Microbe* Vol. 14 232-241 (Cell Press, 2013).
79 Bieniasz, P. D. in *Virology* Vol. 344 55-63 (Virology, 2006).
80 Garrus, J. E. *et al.* Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55-65 (2001).
81 Gottlinger, H. G., Dorfman, T., Sodroski, J. G. & Haseltine, W. A. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 3195-3199 (1991).
82 Strack, B., Calistri, A., Craig, S., Popova, E. & Göttlinger, H. G. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**, 689-699 (2003).
83 Puffer, B. A., Parent, L. J., Wills, J. W. & Montelaro, R. C. Equine infectious anemia virus utilizes a YXXL motif within the late assembly domain of the Gag p9 protein. *Journal of Virology* **71** (1997).
84 Zhai, Q. *et al.* Structural and functional studies of ALIX interactions with YPX nL late domains of HIV-1 and EIAV. *Nature Structural and Molecular Biology* **15**, 43-49 (2008).
85 Martin-Serrano, J., Eastman, S. W., Chung, W. & Bieniasz, P. D. HECT ubiquitin ligases link viral and cellular PPXY motifs to the vacuolar protein-sorting pathway. *Journal of Cell Biology* **168**, 89-101 (2005).
86 Xiang, Y., Cameron, C. E., Wills, J. W. & Leis, J. Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. *Journal of virology* **70**, 5695-5700 (1996).
87 Nagashima, S. *et al.* A PSAP motif in the ORF3 protein of hepatitis E virus is necessary for virion release from infected cells. *J Gen Virol* **92**, 269-278 (2011).
88 Licata, J. M. *et al.* Overlapping motifs (PTAP and PPEY) within the Ebola virus VP40 protein function independently as late budding domains: involvement of host proteins TSG101 and VPS-4. *J Virol* **77**, 1812-1819 (2003).
89 Perez, M., Craven, R. C. & de la Torre, J. C. The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. *Proc Natl Acad Sci U S A* **100**, 12978-12983 (2003). https://doi.org:10.1073/pnas.2133782100
90 Irie, T. & Harty, R. N. L-domain flanking sequences are important for host interactions and efficient budding of vesicular stomatitis virus recombinants. *J Virol* **79**, 12617-12622 (2005).
91 Griffiths, J. A brief history of mass spectrometry. *Anal Chem* **80**, 5678-5683 (2008).

92 Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
93 Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2**, 140-150 (2003).
94 Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**, 223-226
95 Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology* **21**, 255-261 (2003).
96 Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol Cell* **49**, 583-590 (2013).
97 Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712-723 (2015).
98 Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022-3040.e3028 (2021).
99 Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).
100 Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791-803 (1993).
101 Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E. & Boeke, J. D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc Natl Acad Sci U S A* **93**, 10315-10320 (1996).
102 Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* **10**, 2763-2788 (2009).
103 Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212-1226 (2014).
104 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402-408 (2020).
105 Smith, G. P. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317 (1985).
106 Smith, G. P. & Petrenko, V. A. Phage display. *Chemical Reviews* **97**, 391-410 (1997).
107 Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386-390 (1990).
108 Sundell, G. N. & Ivarsson, Y. Interaction analysis through proteomic phage display. *BioMed research international* **2014**, 176172-176172 (2014).
109 Parmley, S. F. & Smith, G. P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305-318 (1988).
110 Thie, H., Voedisch, B., Dübel, S., Hust, M. & Schirrmann, T. Affinity maturation by phage display. *Methods Mol Biol* **525**, 309-322, xv (2009).
111 Sharma, S. C., Memic, A., Rupasinghe, C. N., Duc, A. C. & Spaller, M. R. T7 phage display as a method of peptide ligand discovery for PDZ domain proteins. *Biopolymers* **92**, 183-193 (2009).
112 Sternberg, N. & Hoess, R. H. Display of peptides and proteins on the surface of bacteriophage lambda. *Proc Natl Acad Sci U S A* **92**, 1609-1613 (1995).
113 Krumpe, L. R. H. *et al.* T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. *Proteomics* **6**, 4210-4222 (2006).

114 Crissman, J. W. & Smith, G. P. Gene-III protein of filamentous phages: Evidence for a carboxyl-terminal domain with a role in morphogenesis. *Virology* **132**, 445-455 (1984).
115 Nasir, S. F. & Jaworski, J. Assessing the stability of assembled filamentous phage coat protein P8. *Supramolecular Chemistry* **26**, 329-337 (2014).
116 Kehoe, J. W. & Kay, B. K. in *Chemical Reviews* Vol. 105 4056-4072 (American Chemical Society, 2005).
117 Marks, J. D. *et al.* By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *Journal of Molecular Biology* **222**, 581-597 (1991).
118 McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554 (1990).
119 Barbas, C. F., Kang, A. S., Lerner, R. A. & Benkovic, S. J. Assembly of combinatorial antibody libraries on phage surfaces: The gene III site. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7978-7982 (1991).
120 Sifniotis, V., Cruz, E., Eroglu, B. & Kayser, V. Current Advancements in Addressing Key Challenges of Therapeutic Antibody Design, Manufacture, and Formulation. *Antibodies (Basel)* **8** (2019).
121 Vaughan, T. J. *et al.* Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat Biotechnol* **14**, 309-314 (1996).
122 Ali, M., Simonetti, L. & Ivarsson, Y. Screening Intrinsically Disordered Regions for Short Linear Binding Motifs. *Methods Mol Biol* **2141**, 529-552 (2020).
123 Alfaleh, M. A. *et al.* Phage Display Derived Monoclonal Antibodies: From Bench to Bedside. *Front Immunol* **11**, 1986 (2020).
124 Frenzel, A., Schirrmann, T. & Hust, M. in *mAbs* Vol. 8 1177-1194 (Taylor and Francis Inc., 2016).
125 Orchard, S. *et al.* The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358-363 (2014).
126 Ideker, T. A systems approach to discovering signaling and regulatory pathways--or, how to digest large interaction networks into relevant pieces. *Adv Exp Med Biol* **547**, 21-30 (2004).
127 Jordán, F., Nguyen, T. P. & Liu, W. C. Studying protein-protein interaction networks: a systems view on diseases. *Brief Funct Genomics* **11**, 497-504 (2012).
128 Jameson, D. M. & Ross, J. A. Fluorescence polarization/anisotropy in diagnostics and imaging. *Chemical Reviews* **110**, 2685-2708 (2010).
129 Perrin, F. Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'etat excité. *J. Phys. Radium* **7**, 390-401 (1926).
130 Lakowicz, J. R. *Principles of fluorescence spectroscopy*. Third edn, (Springer, 2006).
131 Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols* **6**, 365-387 (2011).

132 French, T. *et al.* Two-photon fluorescence lifetime imaging microscopy of macrophage-mediated antigen processing. *Journal of Microscopy* **185**, 339-353 (1997).
133 Duckworth, B. P. & Aldrich, C. C. Development of a high-throughput fluorescence polarization assay for the discovery of phosphopantetheinyl transferase inhibitors. *Analytical Biochemistry* **403**, 13-19 (2010).
134 Zhang, R. *et al.* Fluorescence polarization assay and inhibitor design for MDM2/p53 interaction. *Analytical Biochemistry* **331**, 138-146 (2004).
135 August, E. M. *et al.* Development of a high-throughput assay to measure histidine decarboxylase activity. *Journal of Biomolecular Screening* **11**, 816-821 (2006).
136 Nikolovska-Coleska, Z. *et al.* Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **332**, 261-273 (2004).
137 Blikstad, C. & Ivarsson, Y. High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Commun Signal* **13**, 38 (2015).
138 Davey, N. E. *et al.* Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J* **284**, 485-498 (2017).
139 Bernal, F., Tyler, A. F., Korsmeyer, S. J., Walensky, L. D. & Verdine, G. L. Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide. *J Am Chem Soc* **129**, 2456-2457 (2007).
140 Kussie, P. H. *et al.* Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948-953 (1996).
141 Luo, Z. *et al.* RNF115-mediated ubiquitination of p53 regulates lung adenocarcinoma proliferation. *Biochem Biophys Res Commun* **530**, 425-431 (2020).
142 Sczaniecka, M. *et al.* MDM2 protein-mediated ubiquitination of numb protein: identification of a second physiological substrate of MDM2 that employs a dual-site docking mechanism. *J Biol Chem* **287**, 14052-14068 (2012).
143 Fros, J. J. *et al.* Chikungunya virus nsP3 blocks stress granule assembly by recruitment of G3BP into cytoplasmic foci. *J Virol* **86**, 10873-10879 (2012).
144 Gorchakov, R., Garmashova, N., Frolova, E. & Frolov, I. Different types of nsP3-containing protein complexes in Sindbis virus-infected cells. *J Virol* **82**, 10088-10101 (2008).
145 Panas, M. D., Ahola, T. & McInerney, G. M. The C-terminal repeat domains of nsP3 from the Old World alphaviruses bind directly to G3BP. *Journal of virology* **88**, 5888-5893 (2014).
146 Panas, M. D. *et al.* Viral and cellular proteins containing FGDF motifs bind G3BP to block stress granule formation. *PLoS Pathog* **11**, e1004659 (2015).
147 McInerney, G. M. FGDF motif regulation of stress granule formation. *DNA Cell Biol* **34**, 557-560 (2015).
148 Wirblich, C., Bhattacharya, B. & Roy, P. Nonstructural protein 3 of bluetongue virus assists virus release by recruiting ESCRT-I protein Tsg101. *J Virol* **80**, 460-473 (2006).
149 Mattijssen, S., Kozlov, G., Fonseca, B. D., Gehring, K. & Maraia, R. J. LARP1 and LARP4: up close with PABP for mRNA 3' poly(A) protection and stabilization. *RNA Biol* **18**, 259-274 (2021).

150 Craig, A. W., Haghighat, A., Yu, A. T. & Sonenberg, N. Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* **392**, 520-523 (1998).
151 Khaleghpour, K. *et al.* Translational repression by a novel partner of human poly(A) binding protein, Paip2. *Mol Cell* **7**, 205-216 (2001).
152 Bessa, L. M. *et al.* The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci Adv* **8**, eabm4034 (2022).
153 Batra, J. *et al.* Non-canonical proline-tyrosine interactions with multiple host proteins regulate Ebola virus infection. *EMBO J* **40**, e105658 (2021).
154 Kirchdoerfer, R. N., Moyer, C. L., Abelson, D. M. & Saphire, E. O. The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis. *PLoS Pathog* **12**, e1005937 (2016).
155 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
156 Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459-468 (2020).
157 Thoms, M. *et al.* Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255 (2020).
158 Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc Natl Acad Sci U S A* **119**, e2204539119 (2022).
159 Ricciardi, S. *et al.* The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606**, 761-768 (2022).
160 Wang, Q. *et al.* Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell* **182**, 417-428.e413 (2020).
161 Park, G. J. *et al.* The mechanism of RNA capping by SARS-CoV-2. *Nature* (2022).
162 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
163 Miknis, Z. J. *et al.* Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol* **83**, 3007-3018 (2009).
164 Sutton, G. *et al.* The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* **12**, 341-353 (2004).
165 Littler, D. R., Gully, B. S., Colson, R. N. & Rossjohn, J. Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. *iScience* **23**, 101258 (2020).
166 Buchko, G. W., Zhou, M., Craig, J. K., Van Voorhis, W. C. & Myler, P. J. Backbone chemical shift assignments for the SARS-CoV-2 non-structural protein Nsp9: intermediate (ms - μs) dynamics in the C-terminal helix at the dimer interface. *Biomol NMR Assign* **15**, 107-116 (2021).
167 F Dudás, E. *et al.* Backbone chemical shift spectral assignments of SARS coronavirus-2 non-structural protein nsp9. *Biomol NMR Assign* **15**, 235-241 (2021).
168 Chen, Y. & Guo, D. Molecular mechanisms of coronavirus RNA capping and methylation. *Virol Sin* **31**, 3-11 (2016).
169 Rosas-Lemus, M. *et al.* High-resolution structures of the SARS-CoV-2 2'-*O*-methyltransferase reveal strategies for structure-based inhibitor design. *Sci Signal* **13** (2020).

170 Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237-1245 (1992).
171 Brooks, C. L. & Gu, W. p53 ubiquitination: Mdm2 and beyond. *Mol Cell* **21**, 307-315 (2006).
172 Nag, S., Qin, J., Srivenugopal, K. S., Wang, M. & Zhang, R. The MDM2-p53 pathway revisited. *J Biomed Res* **27**, 254-271 (2013).
173 Pauling, L. C., Zuckerkandl, E., Henriksen, T. & Lövstad, R. Chemical Paleogenetics. Molecular "Restoration Studies" of Extinct Forms of Life. *Acta Chemica Scandinavica*, 9-16 (1963).
174 Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* **345**, 86-89 (1990).
175 Glasauer, S. M. & Neuhauss, S. C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics* **289**, 1045-1060 (2014).
176 Taylor, J. S., Van de Peer, Y., Braasch, I. & Meyer, A. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* **356**, 1661-1679 (2001).
177 Åberg, E., Karlsson, O. A., Andersson, E. & Jemth, P. Binding Kinetics of the Intrinsically Disordered p53 Family Transactivation Domains and MDM2. *J Phys Chem B* **122**, 6899-6905 (2018).
178 Hooke, R. *Micrographia, or, Some physiological descriptions of minute bodies made by magnifying glasses : with observations and inquiries thereupon*. (London : Printed for James Allestry ... and are to be sold at his shop ..., MDCLXV11 [1667], 1667).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1877

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-486764

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022