UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2267

Diving into short linear motifs

Large-scale identification of endogenous and host-pathogen protein-protein interactions and further characterized by deep mutational scanning

CAROLINE BENZ

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023

ISSN 1651-6214
ISBN 978-91-513-1804-2
URN urn:nbn:se:uu:diva-498312

Dissertation presented at Uppsala University to be publicly examined in Room B42, BMC, Husargatan 3, Uppsala, Thursday, 8 June 2023 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Birthe B. Kragelund (University of Copenhagen).

Abstract
Benz, C. 2023. Diving into short linear motifs. Large-scale identification of endogenous and host-pathogen protein-protein interactions and further characterized by deep mutational scanning. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2267. 68 pp. Uppsala: Acta Universitatis Upsaliensis.
ISBN 978-91-513-1804-2.

Short linear motifs (SLiMs) are protein-protein interaction sites that play an essential role in distinct cellular processes. Those interactions are challenging to capture by common high-throughput methods. Therefore, we established an improved version of Proteomic Peptide Phage Display (ProP-PD) as a dedicated method to identify SLiM-based interactions. ProP-PD libraries were created for the discovery of endogenous and host-pathogen protein-protein interactions. The M13 bacteriophage libraries present 16 amino acid long peptides from the intrinsically disordered regions (IDRs) of the human (HD2) proteome or the proteomes of RNA viruses (RiboVD). Through benchmarking of the approach using 35 well-known SLiMs binding domains and the HD2 library, we defined parameters for assigning confidence levels to the results. The selections against the HD2 library revealed >2000 SLiMs-based interaction pairs. Regarding host-pathogen interactions, we focused on interactions mediated by coronavirus proteins, exploring how human proteins bind to viral peptides and how viral proteins bind to human SLiMs. By screening more than 130 human bait proteins against the RiboVD, we revealed several host proteins potentially being targeted by SARS-CoV-2 proteins. Viral hijacking of human G3BP1/2 by the N-protein from SARS-CoV-2 impacted stress granule formation, and inhibition of the interaction was found to have an antiviral effect. Using SARS-CoV-2 proteins in selections with our HD2 library, we found that viral proteins may bind host SLiMs. Selected interactions were validated via affinity measurements revealing a wide range of affinities. Finally, we uncovered that a peptide binding to the NSP9 has an antiviral effect. It is not always possible to establish binding determinants directly from ProP-PD derived peptides. Therefore, we developed a deep mutational scanning (DMS) by phage display protocol. To test the approach, we designed libraries in which all amino acid positions of binding peptides were individually mutated, and the effect on binding was investigated through peptide phage selection. The approach was validated against well-studied interactions and applied to SLiM-based interactions between human proteins and SARS-CoV-2 proteins. Based on the DMS by phage display data we could create a higher affinity binder for NSP9 with increased antiviral effects. The research presented in this thesis has established a platform for large-scale interaction screening through phage display. The results contribute to a deeper understanding of the SLiMs binding and function and also pinpoint novel potential targets for the development of antiviral agents.

Keywords: protein-protein interactions, short linear motif, endogenous interactions, host-pathogen, viral hijacking, deep mutational scanning

Caroline Benz, Department of Chemistry - BMC, Biochemistry, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1804-2
URN urn:nbn:se:uu:diva-498312 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-498312)

To my loved ones

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., Davey, N. E., & Ivarsson, Y. (2022). Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Molecular systems biology*, *18*(1), e10584.

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., Moliner Morro, A., Mund, A., Andersson, E., McInerney, G., Mann, M., Jemth, P., Davey, N. E., Överby, A. K., Ivarsson, Y. (2021). Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nature communications*, *12*(1), 6761.

III. Mihalic, F., **Benz, C.**, Kassa, E., Lindqvist, R., Simonetti, L., Inturi, R., Aronsson, H., Andersson, E., Chi, C. N., Davey, N. E., 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. *under revision*

IV. **Benz, C.**, Maasen L., Simonetti, L., Mihalic, F., Lindqvist, R., Tsitsa, I., Jemth, P., Överby, A. K., Davey, N. E., Ivarsson, Y. Parallel exploration of short linear motif-based interactions using deep mutational scanning by phage display. *manuscript*

The following publications are not included in this thesis:

I. Mihalic, F., Simonetti, L., Giudice, G., Sander, M. R., Lindqvist, R., Berit, M., Peters, A., **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., Davey, N. E., Ivarsson, Y. Large-scale phage-based screening reveals extensive pan-viral mimicry of host short linear motifs; *under revision*,

II. Lindqvist, R., **Benz, C.**, Sereikaite, V., Maassen, L., Laursen, L., Jemth, P., Strømgaard, K., Ivarsson, Y., Överby, AK. (2022) A Syntenin Inhibitor Blocks Endosomal Entry of SARS-CoV-2 and a Panel of RNA Viruses. *Viruses*.; 14(10):2202.

III. Wadie, B., Kleshchevnikov, V., Sandaltzopoulou, E., **Benz, C**., & Petsalaki, E. (2022). Use of viral motif mimicry improves the proteome-wide discovery of human linear motifs. *Cell reports*, *39*(5), 110764.

Contents

Abbreviations

aa	Amino acids
AI	Artificial intelligence
AP-MS	Affinity purifications-Mass spectrometry
ARM	Armadillo repeat
AUC	Area under the ROC curve
BM	Deep mutational scanning Benchmark (phage library)
CoV	Deep mutational scanning SARS-CoV-2 (phage library)
ds	double stranded
DMS	Deep mutational scanning
E.coli	Escherichia coli
ELISA	Enzyme-linked Immunosorbent Assay
ELM	Eukaryotic Linear Motif
ep-PCR	Error-prone polymerase chain reaction
FP	Fluorescence polarization
FITC	Fluorescein
GFP	Green fluorescent protein
HD2	Human disorderome 2 (phage library)
IC_{50}	half maximal inhibitory concentration
IDR	Intrinsically disordered region
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
mP	millipolarisation unit
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
NSP	Nonstructural proteins
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein data bank
PPI	Protein-protein interaction
ProP-PD	Proteomic Peptide Phage Display
PSSM	Position-specific scoring matrix
RiboVD	RNA virus disorderome (phage library)
ROC	Receiver operating characteristic
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SLiM	Short linear motif

SPOT	Synthetic peptide arrays on membrane support technique
SPR	Surface plasmon resonance
ss	single stranded
wt	Wildtype
Y2H	Yeast two-hybrid

Amino acid	**Three letter code**	**One letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Introduction

Imagine a coffee shop. A customer comes in, approaches the counter and orders a black coffee to go. After receiving the hot beverage, the customer runs into another person with a dog and spills the coffee. In this short scenario, the customer engages and interacts with two humans – the barista and the dog owner. Human-human interactions are in a delicate balance. If homeostasis is brought out of balance it leads to conflict and dysfunction in society. This is easily translatable to the interactions occurring in the human cell. To understand those in detail, let´s take a magnifying glass and dive into the incredible world of interactions in the cell. The players in cells are biomolecules such as nucleic acids (DNA and RNA), lipids, and proteins. The molecules have different roles in the network of interactions. The most diverse group of listed molecules are proteins. Interactions between two proteins are called protein-protein interactions, in short PPIs. One kind of PPI is mediated by short linear motifs (SLiMs) whereby a small flexible stretch of a protein interacts with a folded entity, a globular domain. The SLiMs are the objective of this thesis and the following points are the main questions that will be answered in this thesis:

- How can we identify SLiMs in large-scale?
- How do viruses such as severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) exploit SLiM-based interactions?
- Can we use information on SLiM-based interactions to inhibit viral infection?
- Can we use deep mutational scanning (DMS) by peptide-phage display to gain information on affinity determinants of SLiM-based interactions?

Don´t worry at this point if you find the questions just a combination of words. In the following sections, I will introduce the topic and give definitions.

Protein structure and intrinsically disordered regions

Before having a closer look at the interactions mediated by proteins, we have to understand their nature. Proteins are built of smaller units, called amino acids (aas), of which there are 20 commonly occurring (see abbreviations).

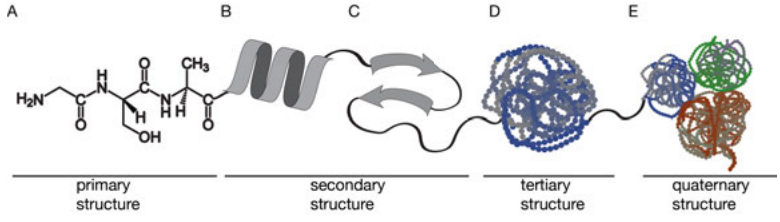

Figure 1 Protein structure at different levels. A) Primary structure (polypeptide chain). B) and C) Secondary structure ((B) α-helix; C) β-sheet). D) Tertiary structure (a folded domain). E) A quaternary structure built of three distinct proteins. The black line between folded entities indicates intrinsically disordered regions.

The proteinogenic aas are linked to each other through a peptide bond from the carboxyl group of one aa to the amino group of the following aa (Figure 1 A). The backbone of a polypeptide chain can form secondary structures such as an α-helix (Figure 1 B) and a β-sheet (Figure 1 C). The secondary structures are formed through the polypeptide backbone and stabilized through hydrogen bonds of amide groups. The tertiary structure describes the overall three-dimensional shape of a single polypeptide chain and is mainly built up by the side chains of the aas (Figure 1 D). Quaternary structure describes the structure formed by multiple protein subunits (Figure 1 E). The structure of a protein can be resolved by methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy.

So far, I described different levels of structured regions of proteins, but there are also unstructured, flexible regions, called intrinsically disordered regions (IDRs). In fact, the human proteome is between 30-40% intrinsically disordered which is a large proportion of the whole proteome, and for a long time, those regions have not been given the attention which they deserve due to the structure equals function dogma [1,2]. IDRs have no defined three-dimensional structure and may serve as connections between domains (linker or loops), or be flexible regions flanking a well-folded structure. Even whole proteins exist with no defined three-dimensional structures [3]. Amino acids which promote IDRs are proline, serin, glutamic acid, glutamate, lysin, alanine, and glycine. Less abundant aas in the flexible regions are the other aas, of which tryptophan and cysteines occur least frequently [4]. Experimentally IDRs can be indirectly determined with several techniques, such as X-ray crystallography, NMR and Circular dichroism-spectroscopy [5]. Unstructured regions lead in X-ray crystallography to missing electron density; this can

however have other explanations such as a wobbly domain (domain with several conformations) or technical problems [5]. Another approach to identifying structures and IDRs in proteins is through NMR. The protein is placed in a magnetic field and the chemical shift of atoms is detected. A sample consisting of only an IDR can be characterized by a very narrow range of the ^{1}HN chemical shift. Typically, those signals are located in the 7.5-8.5 ppm range as compared to a folded protein, which has a widely dispersed spectrum range (6-11 ppm) in the ^{1}H dimension. It results in the lack of dispersion of the proton signals as well as a high percentage of overlapping clusters of signals. To get a better resolution of the overlapping signals in IDRs NMR methods based on spin relaxation can be used [5,6] and taking the intramolecular motion of IDRs to help. The intra-molecular motions of IDRs have a slower relaxation rate, which results in narrower lines. Therefore, the spectra of IDRs can be clearly acquired.

Bioinformatic prediction tools have been developed to identify IDRs such as IUPRED, where the amino acid composition is analyzed and intrinsically disordered properties can be calculated [7]. Recently AlphaFold was launched, which is an AI (artificial intelligence) -based system. This tool predicts based on its primary sequence the fold of the protein with high accuracy compared to experimental data. As AlphaFold predicts structures it also predicts IDRs [8,9].

Protein-protein interactions

Protein-protein interactions (PPIs) describe the association of two or more protein entities and can be manifested on different levels (Figure 2). We can describe PPIs on the level of residues involved in the interactions and the instances estimated in the human cell (Figure 2 D). The formation of large multi protein complexes involve many aas and large surfaces. The number of distinct complexes in a human cell has been estimated to in the order of 600-1,000 [10,11] (Figure 2 A, D). Most of these complexes are stable and might not dissociated within the life-time of the individual proteins [12]. The second type of PPIs is the interactions of two globular domains with each other, which involves fewer amino acids than multi-protein complexes, and the number of instances estimated is higher, approximately 35,000 [13] (Figure 2 B). A third kind of PPIs are the SLiM-based interactions (Figure 2 C). These motif-based interactions involve usually only three to ten amino acids found in IDRs that bind to a folded domain. Motif-based interactions are characterized by low to medium affinity, transient behavior and it has been estimated that the human proteome contains in the order of 100.000 SLiMs [13]. The Eukaryotic Linear Motif (ELM) database, which is most comprehensive database of SLiMs, lists only around 4,000 instances [14,15]. Motif-based interactions are important for multiple cellular processes such as signal transductions, activation, and

localization [13] through enzyme docking, localization signal, degradation motif, or binding site [16]. Motif-based interactions between MDM2 and the proto-oncogene TP53 is mediated through three key specificity-determining residues (Figure 2 C in red - FxxxWxxL) [17].

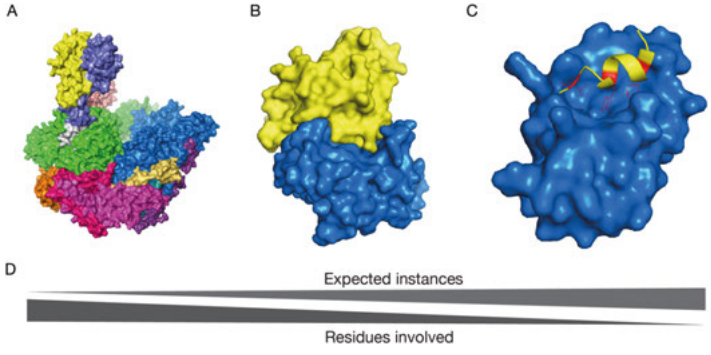

Figure 2 Examples of different PPIs. A) The RNA polymerase complex (PDB 6DRD), with each subunit colored in distinct colors. B) Heterodimeic interaction between RhoA GDP and RhoGAP (PDB 1OW3). C) SLiM-based interaction between the SWIB domain of MDM2 (blue) and the TP53 degradation motif that takes an alpha-helical structure upon binding (yellow). The key residues of the motif (FxxxWxxL) are shown in red (PDB 4HFZ). D) Graphical representation of expected number of instances and residues involved.

How-to-read regular expressions

A capital letter indicates the amino acid in that position in the single-letter code (see abbreviations). x indicates that any amino acid can be placed in that position. A squared bracket with multiple aas indicates that any of the aas can be found in that position. A circumflex (^) indicates that the aas indicated are excluded from that position. ϕ indicates the group of hydrophobic aas [18].

At the start of this work, draft maps of the human interactome had been reported based on data generated through affinity purification coupled to mass spectrometry (AP-MS) [19–21] and yeast two-hybrid (Y2H) [22,23].

Due to the properties of the interactions and the limitations of those two techniques, there is a bias towards identification of stable protein complexes and interactions between globular domains. Consequently, the low affinity transient SLiM-based interactions remain disproportionally underrepresented. The need to develop and standardizes a high throughput method for discovering motif-based interactions such ProP-PD (see later) was clear and is strengthened by the discrepancy between the estimated number of SLiM-mediated interactions and the ELM instances.

So far I only described endogenous PPIs, which means that the interacting proteins are from the same organism. Protein interactions can also occur between proteins from two different organisms such as viruses and humans, which I describe in the next section.

Host-virus interactions

Viruses are obligate intracellular parasites that takes over the cell by interactions with host proteins. The Covid-19 pandemic has shown the importance of understanding the ways viruses hijack and perturb host cell function to understand the molecular basis of viral infection, and to identify novel targets for therapeutic intervention. This can be done through host-virus systems biology where the goal is to systematically explore interactions between host and virus proteins, such as systematic AP-MS experiments [24]. However, as mentioned above, these experiments often fail to capture SLiM-based interactions, which are often exploited by different viruses.

It is advantageous for viruses to hijack the host using SLiMs, in part because they are involved in many processes that are critical for cell function, and in part because a SLiM, due to the small number of involved residues, can be quickly mutationally generated de novo [25]. In the literature, several examples of viral hijacking and viral proteins containing multiple motifs for interactions with different host proteins can be can be found [26–28]. Well-studied examples include the human Papillomavirus oncoproteins E6 and E7. The E6 contains at least two SLiMs, of which one is C-terminal motif and bind to PDZ domains in host proteins involved in establishing and maintaining cell polarity and adhesion. E6 has also LxxLL motif that binds to the ubiquitin ligase E6AP [26], the interferon regulator factor 3 (IRF3) [29] and NOTCH co-activator MAML1 [30]. Targeting the ubiquitin system gives the viruses the advantage to enhance virus replication [31]. IRF3 plays a critical role in immune response and targeting IRF3 can circumvent normal antiviral response [29]. These are not the only processes viruses are interfering with through SLiM-based interactions. In fact, viruses have been found to use hijack or perturb various cellular processes such as signaling, trafficking, protein degradation and transcriptional regulation using SLiMs [16,25] (Figure 3). In addition, viruses may also have folded proteins or protein domains that can bind to host SLiMs [32–35].

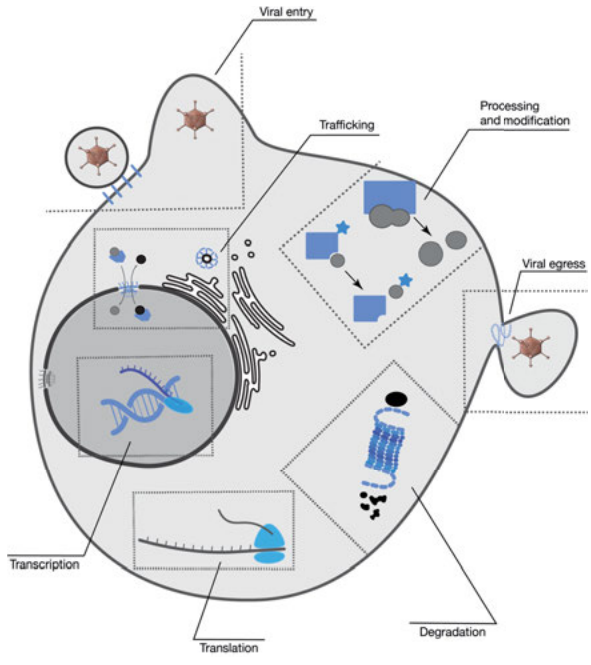

Figure 3 Schematic representation of pathways and processes viruses targeting in the eukaryotic host by mimicking endogenous motifs based on ELM annotations [15] (Figure inspired by Simonetti *et al*. [36]).

At the start of this thesis, the knowledge of SLiM-based viral hijacking of the host cell was based on numerous studies over many years [15,36], which motivated us to attempt to systematically explore these interactions in a higher throughput. We decided to explore the topic both from the side of viral SLiMs binding to human proteins, and from the side of viral domains binding to host motifs [32], in both cases taking a peptide phage display approach.

Proteomic peptide phage display (ProP-PD)

George P. Smith was awarded the Nobel prize in Chemistry in 2018 for the development of peptide display on the filamentous M13 phage. Phage, or bacteriophage, stands for a virus that infects and replicates in bacteria. The filamentous M13 phage has a single-stranded (ss) genome surrounded by a protein coat. There are around 2,700 copies of the major coat protein P8, which covers the long sides of the phage particle [37,38]. It is capped by around five copies of minor coat proteins, of which P3 is important for infecting the host. The coat proteins can be used to present proteins, peptides or antibodies, by genetically fusing the gene that codes for a coat protein with a gene for the polypeptide to be display. Since phage display was developed in the 1980s, numerous groups in the world have applied and further developed the

approach as a tool to explore several questions including the development of high affinity antibodies [39,40].

For experiments aiming for high affinity binders, such as in the case of interactions of antibodies, the display is usually on the minor coat protein P3. In the case of low affinity ligands, such as SLiMs binding to target proteins, the display is typically on the major coat protein P8, as the multivalent display increases the affinity through avidity effects. In this thesis, the research questions are related to the capture of SLiM-based interactions, therefore we generated phage libraries that display peptides from target proteomes. The peptides were from the IDRs of the human proteome or from RNA viruses. The approach is called ProP-PD, which stands for **Pro**teomic **P**eptide-**P**hage **D**isplay and as described later. The method was first established 10 years ago [41] and has later been developed by the group and applied to various bait proteins [42–44]. In Figure 3 the ProP-PD workflow is schematically shown.

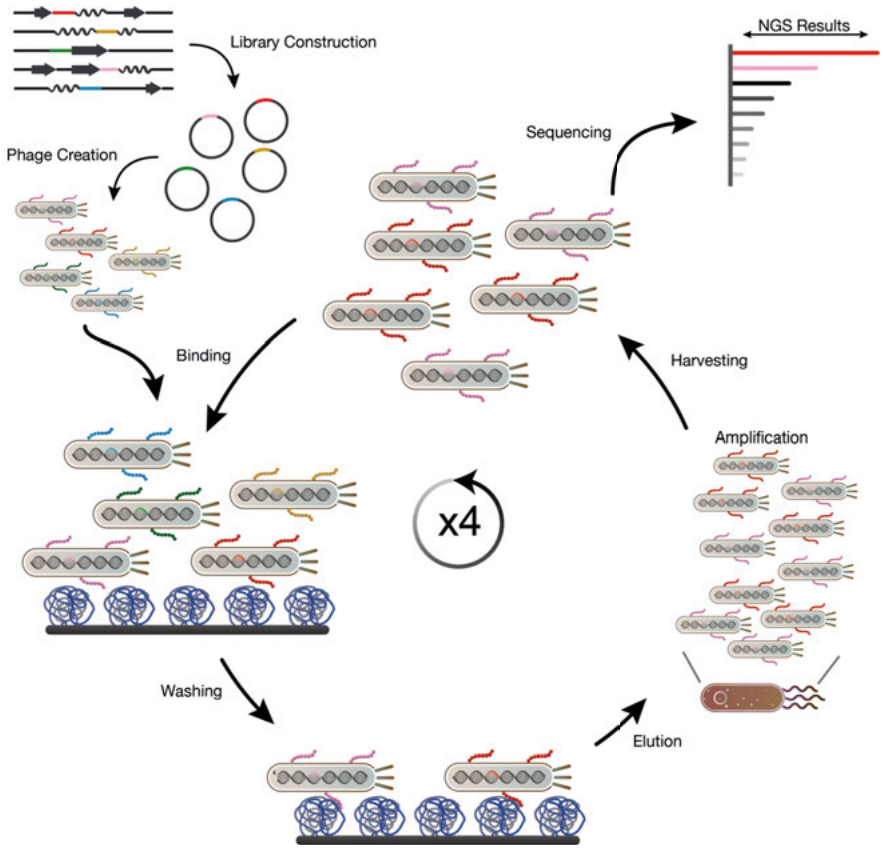

Figure 4 Overview of ProP-PD workflow. Library construction (left upper corner) involves bioinformatical determination of IDRs in the proteome and translation into oligonucleotides. The oligonucleotide library is obtained from a commercial provider and used to create a phagemid library using oligonucleotide-directed mutagenesis as described by Kunkel *et al.* [45]. The phagemid library is electroporated into *E.coli* which produces the phage library. The phage library is used in selection against a bait protein (blue) immobilized in a microtiter plate. Washing steps remove unbound phages followed by elution of bound phages with *E.coli*. Amplified phages are harvested and used in the next round of selection, usually 3-4 rounds of selections are performed. The genotype of the peptide coding region of the enriched phages are determined through next-generation sequencing (NGS). (Figure created by Elias Tjärnhage)

In the first step, a proteome of interest is bioinformatically scanned for IDRs. The identified regions are titled by 16 aa long overlapping peptides. In the next step, synthetic oligonucleotides encoding for the peptides are obtained from a commercial provider. The oligonucleotides are amplified by polymerase chain reaction (PCR) and then fused to the gene for the coat protein (P3 or P8) in a phagemid vector via the classical oligonucleotide-directed mutagenesis method (Kunkle reactions) [45]. The phagemid library is then electroporated into *E.coli* that are already infected with a helper phage M13KO7 to provide all genetic information for phage particle assembly. The bacteria then produce

phage particles that express peptides on their surface. Through those steps, a diverse phage library has been generated. The coverage of the experimentally generated phage library versus the library design can be determined with next-generation sequencing (NGS) [41,46,47]. Before starting the selection process, recombinant proteins/protein domains (baits) are expressed and purified. Bait proteins are immobilized non-specifically in wells of a microtiter plate which is surface treated to bind hydrophilic and hydrophobic proteins. The phage library is applied to the protein and allowed to bind. In the next step, unbound phages are washed away and bound phages are eluted with actively growing *E.coli*. The bacteria amplify phages and the enriched and amplified phages are then used as the starting point for the selection in the next round. After an enrichment process of several days depending on the library size, the binding enriched pools are analyzed via NGS [39,42].

For this work several M13 phage libraries are relevant, each of them presenting 16 aa long peptides: The HD2 library presents the IDRs of the human proteome and the HD is a previous version of that library. The RiboVD presents peptides of the IDRs from RNA virus's proteome.

Next-generation sequencing and data analysis

Simply expressed, NGS is Sanger sequencing providing deeper information [48]. Sanger sequencing determines the sequence of DNA via random incorporation of chain-terminating nucleotides followed by electrophoresis. Therefore, Sanger sequencing can only determine which order of nucleotides are in one oligonucleotide chain [49] but with NGS the content of a mixture can be determined quantitively and qualitatively. How is this achieved? In the following, I will describe one of the NGS platforms – Illumina MiSeq – due to its relevance for this work. The to-be-analyzed oligonucleotides are the region of the phagemid encoding for the peptide (Figure 5 A). Those regions are amplified via PCR from the enriched binding phage pools. For NGS analysis, the to-be-analyzed sequences have to be fused with oligonucleotide adaptors on the 5′ and 3′ via PCR (Figure 5 B). The adaptor regions are complementary to oligonucleotide adaptors in the flow cell used for sequencing. In the same step, barcodes are added to the sequences which are unique for each experiment. Through the barcodes, samples from multiple experiments can be pooled, and after the NGS sequences can be sorted (Figure 5 B green).

The sample mixture is applied to a flow cell and each oligonucleotide is binding to one adaptor. Through PCR the complementary strand is synthesized. The mature strand is denatured and washed away. In the next step, there is bridging of the strand to the second adaptor. Again, the double strand is synthesized via DNA polymerase. After denaturation of the strands, this is resulting in two attached opposite-orientation strands on adaptors. In the following amplification step, several rounds of bridging and synthesizing of

complementary strands are performed. Through that, a concentrated density of the same oligonucleotide is gained in the flow cell. The process is called solid-phase bridge amplification (Figure 5 C). Oligonucleotides in one orientation are digested to end up with uniform oligonucleotide clusters [48,50,51]. In the final step, cyclic reversible termination with fluorophore color-labeled nucleotides are performed, similar to Sanger sequencing. Through the amplification of the analyte, the camera is able to detect the color signal and convert it to the sequence.

For our studies, the NGS facility is providing a single FASTQ file containing the sequencing data. Oligonucleotide sequences are separated back to their original selection experiments by a demultiplexing step using an in-house Python script based on the barcodes. All sequences are translated into peptides and those matching the library design are annotated using PepTools [52], which is a bioinformatic tool developed by our collaborator. PepTools has several features such as annotations with evolutionary, functional, structural, genomic, and proteomic pieces of information [52]. PepTools uses the SLiMFinder pipeline on confident set of peptides and is finding enriched motif. In the next step. The motif is used to align peptides and generate a PSSM (position-specific scoring matrix) from aligned peptides. In the final step all peptides are scored with PSSM to find similar peptides that don´t contain the consensus motif [53] (Figure 5 D).

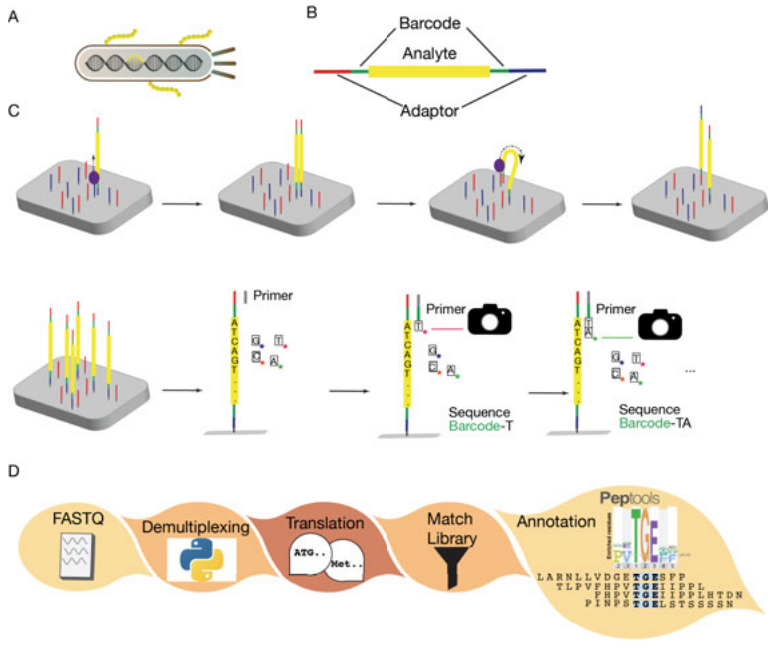

Figure 5 NGS scheme and data analysis. A) M13 phage representation genotype presented as P8 phenotype. B) Sample scheme of an oligonucleotide to analysis, red and blue adaptor, yellow analyte-peptide presented on the phage, green barcodes. C) Illumina solid-phase bridge amplification and sequencing step with labeled nucleotides, in purple DNA polymerase. One oligonucleotide is represented for simplicity. The to analyzed pool contains samples from multiple experiments and thousands of sequences. D) Scheme of NGS processing from data delivery in FASTQ format followed by demultiplexing using an in-house Python script including translation and matching sequencing to library design. Annotation and alignment of peptides and PSSM generation with PepTools.

Deep mutational scanning (DMS)

ProP-PD experiments ideally provide information on binding peptides from proteins in a given proteome, which can be used to establish consensus motifs. However, sometimes the experiments only enrich a couple of peptides from which a motif cannot be derived. Motif and key residues can still be derived by systematically evaluating the importance of each aa position through mutational analysis. However, such studies typically only evaluate the effect of single points mutations (e.g. alanine scanning). These experiments can be scaled up by approaches such as peptide arrays [54], where the effect of substitutions on binding can be tested in a cost-effective and easy way [55].

However other mutations than alanine may be more informative. Therefore, mutational studies investigating the change of each amino acid in a

sequence to all possible variants may reveal important information about the functional relevance of aa positions and give insight into protein folding and stability [54]. Deep mutational scanning (DMS) approaches have been developed to systematically study non- [56] and proteinogenic amino acid substitution at every position [54]. DMS studies have been performed to reveal the potential of optimizing enzymes [57], coevolution of domains to optimize binding [58,59], and sequence-activity relationships [60]. On the peptide level, DMS studies have been performed to design therapeutics such as antimicrobial inhibitors [56,61] but also to explore peptide binding to TRAF6 MATH domain [62]. In these studies, peptide sequences and activity/binding relationships are generated for numerous variants and in a single experiment. This allows the extraction of information for every position in each peptide sequence [61]. Many studies that use DMS in their analysis employ error-prone polymerase chain reaction (ep-PCR) to generate random mutations in the DNA library members that encode for the protein/peptide being investigated [63]. Those libraries may suffer from limited coverage of the complete space of possible options in the given context. To investigate the effect of mutations, the libraries are often used in display methods such as phage [64], yeast [65], and bacterial [62] -display to gain genotype-phenotype information.

Biophysical characterization and validation of protein-peptide interaction

After the primary identification of binding peptides through ProP-PD or DMS by peptide phage display, validations are needed to ensure that the selections, and the subsequent data processing, identified biophysically binding peptides. Moreover, we aim to understand not only that we found binding peptides, but also to understand the strength of the found interactions.

There is a large number of methods for determining the affinities of protein-peptide interactions, such methods can provide information about dissociation constants (K_D), the kinetics of the binding, and thermodynamics. For the calculation of K_D, complex formation (Reaction 1) is used. In the reaction, P stands for protein and L for ligand and PL for the formed complex. The calculation of the K_D is based on equations 1 and uses the concentrations of the reaction partners. A high affinity is manifested as a small K_D-value.

$$P + L \rightleftharpoons PL \qquad \text{(Reaction 1)}$$

$$K_D = \frac{[P]\,[L]}{[PL]} \qquad \text{(Equation 1)}$$

ITC (Isothermal titration calorimetry) is a gold standard method for affinity determinations [66]. In ITC, solution of protein of interest is located in a closed

chamber. A second chamber called the reference cell is filled with buffer. Both chambers are placed in an adiabatic jacket. The binding peptide is titrated stepwise into the sample chamber. The change of heat while titrating peptide into the chamber compared to the reference cell is detected and directly counteracted by a heating/cooling source in the sample chamber. The power to counteract is plotted versus the time [66,67]. In this plot, a curve is fitted in and the K_D is calculated. ITC gives information on the thermodynamic parameters of binding from which the K_D value can be derived. A disadvantage of ITC is the consumption of high quantities of peptide and protein compared to other methods, and that it has a low throughput.

Another method for affinity determination is SPR (surface plasmon resonance) [68]. In this technique, the protein is immobilized on a surface and the binding peptide will flow over the protein (or vice versa). Binding changes the angle of the light beam which is directed on the binding surface. That change can be detected and from there, affinity determinants are calculated. Additionally, the kinetic parameters (k_{on} and k_{off}) can be determined. In SPR the surface can be regenerated to analyze several peptides binding to the same protein, and the sample consumption is thus low. One major thing to consider is, that the protein has to be in optimal conditions to be constrained on a surface [69].

One alternative scalable method which has been developed over the last 10-15 years is the hold-up assay [70,71]. It is a chromatography retention assay to determine affinities. In the hold-up assay a resin is saturated with peptide and is incubated with a protein until the system reaches equilibrium. The protein solution can be a recombinantly expressed purified protein, or a cell lysate. This is followed by a filtering step where the resin is held up, and the flow-through contains the unbound protein fraction. The protein concentration, in the flow-through is analyzed by SDS-PAGE, by fluorescence, or by MS, and compared to a control [72], and referenced to interactions with known affinities. The hold-up assay could be an orthogonal approach to validate the ProP-PD identified peptides and at the same time give an estimate about the affinities. A drawback is a need for the synthetic peptides to be coupled to the beads, and the detection method of choice, which may limit the scalability of the approach.

For this thesis, we used FP (fluorescence polarization) based affinity measurements. The technique is described in the next paragraph.

Fluorescence polarization

FP-based affinity measurements are based on the Brownian molecular rotation. This method utilizes that fluorescence-labeled peptide depolarizes polarized light depending on molecular weight. In saturation experiment, a small labeled peptide is binding to an in size much bigger protein [73,74] (Figure 6 B, C). In different words, upon binding to a bigger protein the tumbling of the

labeled peptide decreases, which leads to a smaller proportion of light being depolarized leading to an increased signal (Figure 6 A, B). FITC (fluorescein) can be used as a fluorescent probe but also other fluorescent labels are available, and the choice depends on the fluorometer and filters available. The advantage of FP is the performance in microtiter plates with small volumes and in a high throughput approach. Measurements are taken at equilibrium. This setup can also be used in a displacement assay, where a premade complex of protein and labeled peptide is displaced with an unlabeled peptide in increasing amounts (Figure 6 D). This approach has the advantage of determining K_D of unlabeled proteins with unlabeled peptides. Limiting factors for FP are the ligand size and change in molecular weight upon binding.

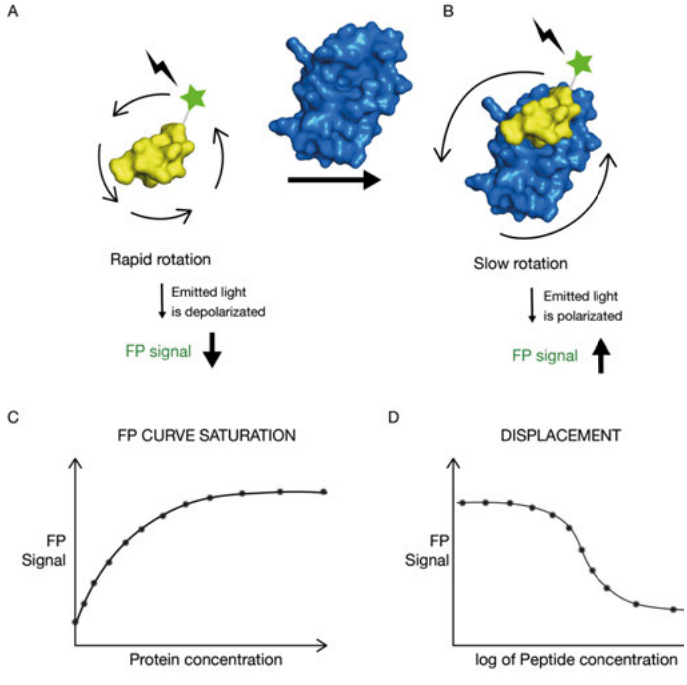

Figure 6 FP Scheme. A) An unbound labeled peptide is rotating quickly which leads to depolarization of the light which consequently results in a low FP signal. B) Upon binding of an interaction partner with a higher molecular weight the rotation slows down and the light is more polarized which leads to an increased FP signal. C) Example curve of saturation experiment with an increasing amount of binding protein. D) Schematic representation of the result of a displacement experiment, x-axes is representing the log of unlabeled peptide concentration.

Present investigations

The common theme in this thesis is the detection and investigation of SLiM-based interactions with globular domains or small proteins. Let´s dive into those interactions, as I state in the thesis title. Phage display was used as the tool to identify SLiM-based interaction. I generated several M13 phage libraries and purified numerous, almost uncountable various proteins and domains for the four publications following in the next chapter. With this research, we wanted to establish and showcase how to use ProP-PD to identify endogenous and host-pathogen interactions. This is highly relevant to the scientific community and society. The discrepancy between the number of reported SLiMs in ELM, which is considered the most comprehensive database, and the estimation of 100,000 instances is huge [13–15]. We have to close the gap in order to gain a better insight into the interactome and to potentially identify novel targets for drug development. Therefore, we identified endogenous as well as host-virus in SLiM-based interactions. After the identification, we validated selected interactions through affinity measurements. The function of the SliMs has been shown in the example of the nuclear importin system. Furthermore, some identified peptides showed an antiviral effect.

For further investigation and to get a deeper understanding of some SliMs and the context of SLiM-based interaction, we developed a new peptide phage display-based DMS approach. The information can be used to predict the effect of mutations before they are observed in the biological context during viral evolution. Additionally, with the DMS approach amino acids which lead to a greater affinity can be determined and can be the starting point for inhibitor development within various areas of application.

Contribution to manuscript

All studies included in this thesis involved collaborations with different groups. My contributions to these studies are as follows:

Paper I **Proteome-scale mapping of binding sites in the unstructured regions of the human proteome**
Planning and carrying out experiments. Library generation. Protein purification. ProP-PD selections. NGS analysis. Affinity measurements. Data analysis. Participated in writing and editing manuscript.

Paper II **Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities**
Planning and performing ProP-PD selections and affinity measurements. NGS preparation. Data analysis. Participated in writing and editing manuscript.

Paper III **Identification of motif-based interactions between SARS-CoV-2 protein domains and human peptide ligands pinpoint antiviral targets**
Planning and performing experiments. Purification of proteins. ProP-PD selection. NGS preparation. Data analysis. Participated in editing manuscript.

Paper IV **Parallel exploration of short linear motif-based interactions using deep mutational scanning by phage display**
Design and planning of the experiments. Carried out and supervised experiments. Purification of proteins. ProP-PD selections. Affinity measurements. NGS preparation, Data analysis. Participated in writing the manuscript.

In the description of the results in the next section, I focus on the aspects and results of the studies that I was mostly involved in.

Paper I Proteome-scale mapping of binding sites in the unstructured regions of the human proteome

In Paper I we developed an improved second-generation ProP-PD library for the identification of human SLiM-based interaction. The library, called HD2 for human disorderome version 2, contains peptides from the IDRs of the human proteome. The first-generation HD library had been developed a couple of years earlier and had been used to successfully identify endogenous interactions [41,43,44,75]. However, the HD library had some quality issues. In the years between the design of the HD and HD2 libraries, knowledge had been gained that made it possible to improve the library design in terms of how IDRs were defined, and in terms of how densely the regions were tiled by overlapping peptides. A bioinformatic tool was developed (PepTools) for peptide analysis and annotation. We benchmarked the HD2 against 34 well-studied protein domains with ligands curated in the ELM database [14,15]. Based on the results we established confidence scores. Furthermore, we determined the K_D values of a representative set of interactions and showed the functionality of the motif-based interaction based on the nuclear import system. In the following section, we have a closer look at the results.

The HD2 library design

The design of the novel HD2 library is different in some aspects in comparison to the previous HD library. The design parameters and coverage of the constructed libraries are shown in Figure 7. Both libraries display 16 aa long peptides, but the overlap between flanking peptides was changed from 7 aa in HD to 12 aa in HD2. Therefore, a SLiM can be presented several times in different peptides and we can have the discovery of overlapping peptides as one of the quality measures for assigning confidence levels. The HD2 library contains almost one million distinct peptides and the coverage of the phage library as determined by NGS analysis is more than 95 %, a great improvement in comparison to the previous version.

HD Library		HD2 Library
0.3	Cut off IUPred:	0.4
16 aa	Peptide length:	16 aa
7 aa	Overlap between peptides:	12 aa
479 846	# Peptides:	961 847
18 682	# Proteins:	18 332
63 %	Coverage:	91 %
P8	Coat protein:	P8

Figure 7 Comparison between HD and HD2 library designs and coverage. In the bottom: a graphical representation of the distinct library designs. Indicated in blue is the overlap.

Standardizing the data analysis and benchmarking of results

Phage selections were performed in triplicates for each protein/domain against various libraries. In the end, peptide-coding regions of enriched phage pools were amplified and barcoded for NGS analysis. The NGS analysis provided a wealth of information on peptides for each target, and the challenge was to understand which of them represented background noise and which were real biophysical binders of the bait proteins used and to develop guidelines for the standardization of the ProP-PD approach.

We evaluated the selection quality metrics available with receiver operating characteristic (ROC) curve analysis. The metrics used were a reproducible discovery of a peptide between replicate selections, identification of overlapping peptides, the proportion of NGS counts, and the presence of a consensus binding motif. The analysis revealed that all metrics had some predictive power, and we finally combined all metrics to assign confidence levels. We categorized the peptides obtained from the selections into four confidence levels: high, medium, low, and filtered, based on the number of quality metrics they fulfilled. Through this binning process, we can enrich the datasets for biophysical binders (true positives) in a standardized manner. As with any technique also ProP-PD has disadvantages or detection limits. The HD2 library does not contain any PTM or free N- or C-terminal residues which leads to a lack of opportunities to detect interactions that rely on these features. But we used this fact, as a negative control in your evaluation process.

Next we benchmarked our results against other large-scale PPI datasets generated through AP-MS or Y2H, using the recall of known binders. Through the HD2 selections, we should be able to find 337 previously validated motif-based interactions in total, of which we were able to find 65 in our results in the high/medium confidence set. This is a recall of 18.5% which is greater compared to other high datasets generated by Y2H (HuRI [76]) and AP-MS (Bioplex [77]) methods.

Comparing multivalent P8 display with monovalent P3 display

In addition to the P8 displayed library, we used the oligonucleotide library to generate a library displaying the 16 aa long peptides on the minor coat protein P3. Monovalent P3 display is often used to generate information on high-affinity binders such as antibodies [37,78,79]. We compared the outcome of selections against the P3 displayed and the P8 displayed HD2 library. The screening results of all 34 baits with P3 showed that the recall of known motif instances was lower (9 versus 65). The lower recall from the P3 display is likely caused by the failure to capture low-to-medium affinity ligands without the avidity effect of provided by the multivalent presentation of peptides on the P8 coat protein. There was one interesting exception, the N-terminal domain of peroxisomal membrane protein PEX14. The selection with the monovalent

P3 display library returned 3 out of the 8 known PEX14 ligands for the protein, compared to the selection against the P8 displayed library which did not return any previously known interaction partners.

The future observations are based on the P8 selection data.

Validation of identified peptides in ProP-PD selections

To validate the results from the selections we performed FP-based affinity determinations for several peptide-domain pairs, here showcased by the affinity measurements of TLN1 PTB and KEAP1 Kelch (Figure 8A). KEAP1 is an adaptor subunit of a Cullin 3-based E3 ubiquitin ligase and binds with high affinity to its known substrate NFE2L1 ($NFE2L1_{228-243}$, K_D=160 nM), and with lower affinity to the novel ligands identified ($PLEC_{4459-4474}$, $PAK1_{87-102}$, $KLHL24_{335-350}$; 1.5 µM-19 µM). Talin 1 (TLN1) is a focal adhesion-related protein that link integrins to the cytoskeleton. The tested domain is a phosphotyrosine binding domain (PTB), and based on our results we found it to bind to a WxYP motif, which is different from the reported TLN1 PTB binding NyxP motif [80]. Using a FITC-labeled probe peptide containing the WxYP motif we found that the peptide bound with high affinity (Figure 8 B). We tested through FP-based competition three unlabeled peptides found in selections ($PIP5K1C_{640-655}$ $TPTE2_{112-128}$, $NYAP1_{542-557}$) and a previously reported TLN1 binding NyxP containing $ITB3_{763-788}$ peptide [80]. The peptides found in the selection bound with affinities in a range from 0.07 µM to 20 µM, while the $ITB3_{763-788}$ peptide was not identified in selections bound with low affinity (500 µM) (Figure 8 C). Appreciably is the wide range of affinities that can be recovered by ProP-PD but low-affinity binder such as $ITB3_{763-788}$ peptide could get out-competed. Consequently, low-affinity binders are mostly likely identified, if there are no high-affinity binders available.

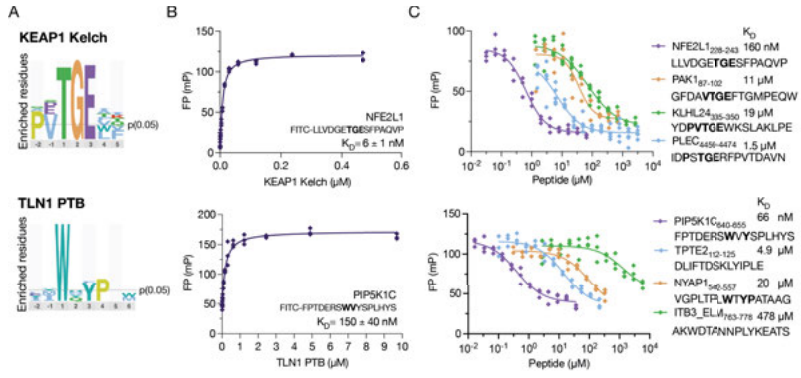

Figure 8 PSSMs and affinity determinations of the two domains KEAP1 Kelch and TLN1 PTB. A) Enriched PSSM based on peptides from selections against the HD2 P8 library; Affinity measurements using FP as determined using B) Direct binding to a FITC-labeled probe peptide, and through displacement of the probe peptide in C) with unlabeled peptides. Bold letters in the peptide sequencing indicated sequences matching enriched PSSM. (Figure adapted from Benz *et al.* [52])

As mentioned in the introduction SliMs can serve as localization signals (NLSs). KPNA4 (importin alpha-3) binds to proteins with NLS, and shuttles the proteins through the nuclear pore into the nucleus together with KPNB1 (importin beta). The key residues of NLSs are KR but additional amino acids are playing a role so that NLSs bind to KPNA4. KPNA4 is composed of 10 armadillo repeats (ARM). The KPNA4 ARM repeat has two peptide binding pockets, named the minor and major pockets (Figure 9 A). KPNA4 binding peptides can have specificity towards one of the pockets (Figure 9 B top and bottom) or bind both pockets with similar affinities (Figure 9 B middle). The ProP-PD selections identified ligands for both pockets. The functionality of the Identified localization signals was evaluated by a cell-based assay. KPNA4 binding peptides were fused to three repeats of GFP (green fluorescent protein) and the localization of the fluorescent signal was determined. The result revealed that 12 out of 16 tested KPNA4 binding peptides were functional as NLSs (Figure 9 C loc). The two lowest affinity ligands did not confer shuttling of GFP into the nucleus, which thus correlates affinity with functionality. Through the experiments we showed that the ProP-PD selections identified binders for both the minor and major site that bind with a range of affinities and serve as NLSs.

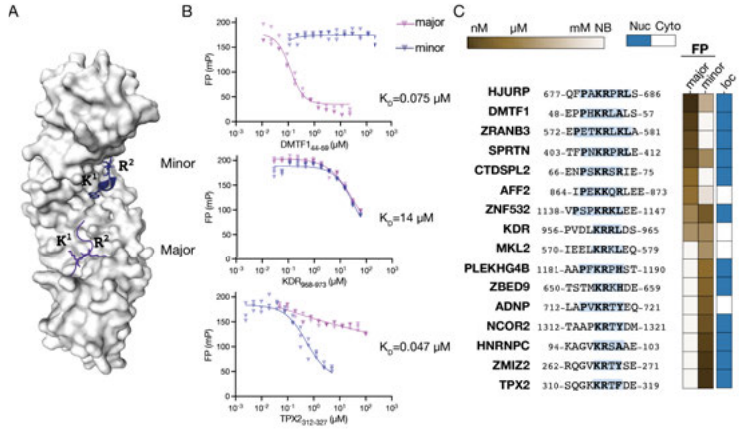

Figure 9 Evaluation of the binding and function of KPNA4 binding peptides. A) Crystal structure of KPNA2 (PDB: 1PJN) with a minor groove binding peptide (PBD: 3ZIP) in blue and major groove binding peptide in purple. B) FP-based displacement experiment showing displacement of probe-peptides binding to the major (purple) and the minor (blue) pockets. Top: DMTF1$_{44-59}$, middle: KDR$_{958-973}$, bottom: TPX2$_{312-327}$. C) Summary of the measured affinities and the outcome of the evaluation of their function as NLSs. "loc" indicates that the localization of the GFP-tagged peptides as evaluated by microscopy. (Figure adapted from Benz *et al.* [52])

Prediction of disease mutations and the effect of PTMs

Next, we focused on using the amino acid resolution binding site information gained from the selection data to predict the effect of disease mutations and PTMs on binding. Found peptides were annotated with information on disease mutation information and phosphorylation sites. The information for two of the domains, KEAP1 Kelch and KPNA4 ARM, are represented in PPI networks (Figure 10 A, C). We found more than 300 mutations in the binding peptides, that could potentially affect binding, and we experimentally tested the effect of four mutations. Two of the mutations E79K and T80K, with high impact on the binding affinity, were found in the KEAP1 Kelch binding NFE2L2 peptide, and these mutations are linked to a disease: early onset of multisystem disorder. In contrast, for the R4466C mutation in the KEAP1 Kelch binding PLEC peptide, we did not observe any effect on binding, and the mutation is annotated with "uncertain significance on pathogenicity" (Figure 10 B). Importantly, while the mutations of the NFE2L2 peptide overlap with key residues of the motif, the R4466C mutation in PLEC is found in a position outside of the motif, which explains the results.

For KPNA4 we found that the mutation of K194R in NKX2-5 decreased the affinity 10-fold, and the mutation is known to lead to an atrial septal defect. The effect on the affinity reflects the relevance of the aa, which is one key residue of the specificity determinants.

We further tested the effect of phosphorylation on KPNA4 binding exploring three phosphorylation sites in NCOR1 ($_{1240}$-EISL**KR**SYESVEGNIK-$_{1256}$). The strongest effect was observed for pS1246, which may serve to regulate the nuclear localization of the protein (Figure 10 D).

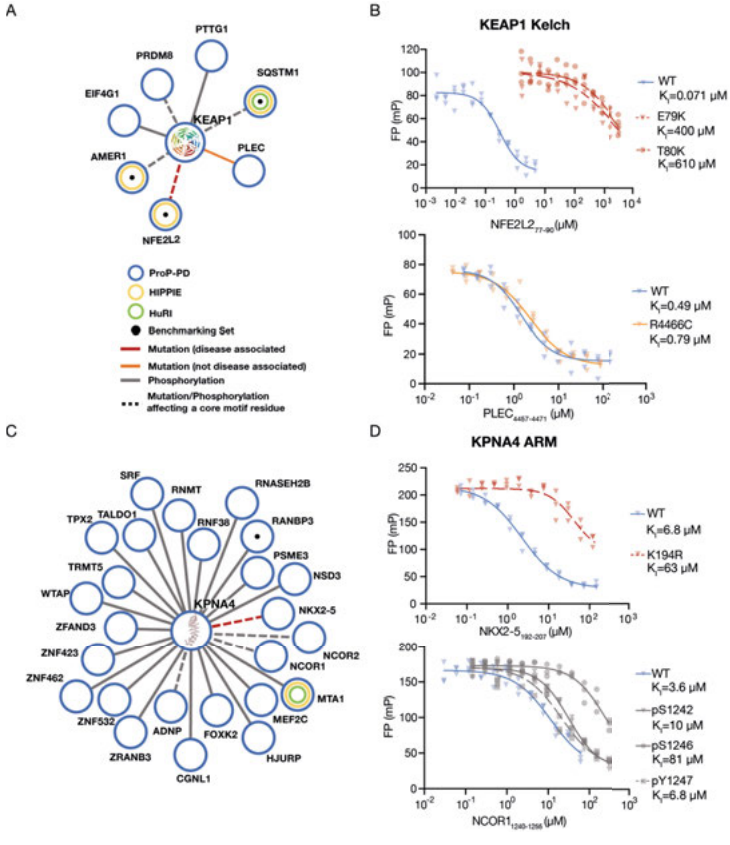

Figure 10 Effects of disease mutations and PTMs on binding. A) Partial ProP-PD PPI network of KEAP1 Kelch with information on mutations (red, disease-associated; orange if not disease-associated and phosphosites (grey)). Dashed edges indicate that mutations or phosphorylation sites overlap with motif residues. B) FP competition experiments of KEAP1 Kelch binding to wt (blue) and disease mutant peptides (red) (technical replicates n=3) C) Partial PPI network of KPNA4, legend the same as in A. D) FP competitions experiment of KPNA4 ARM binding to wt (blue), disease mutant (red) and 32hosphor-peptides (grey). (Figure adapted from Benz *et al.* [52])

To sum up, by using the detailed binding site information provided by the ProP-PD data it is possible to pinpoint mutations and PTMs that may affect the affinities of the interactions.

Conclusion and outlook

In Paper I we describe the development and benchmarking of our novel ProP-PD library HD2. In comparison to other high-throughput PPI detection methods, ProP-PD performs better in terms of identifying SLiM-based interactions. The affinities of the interactions identified by ProP-PD selection cover a wide range.

The functionality of found SliMs was showcased for NLSs binding to KPNA4. We further showed how the binding site information can be used to predict the consequences on binding.

With the new resource established and a framework for data analysis and processing, we are now in a position to scale up to many more proteins and domains. The large-scale PPI data provide much more information than we can explore within one group. Hopefully, groups all over the world will take our data and further validate the interactions identified by ProP-PD to provide a deeper insight into each of the motif-based interactions.

Paper II Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities

The recent pandemic was a drastic experience for all of us but one positive effect was the joint efforts and synergy in the scientific community. In the shortest time, collaborations were formed and complementary expertise were combined to learn more about the virus. We wanted to contribute and use our expertise in the identification and validation of SLiM-based interactions, with the ambition to find novel host-virus interactions that could be targeted by antiviral agents. Therefore, we designed a new library containing intrinsically disordered regions of single-stranded (ss) and double-stranded (ds) RNA viruses. In particular, SARS-CoV-2 and all the related coronavirus strains such as SARS-CoV, MERS, and the common cold-causing human coronavirus 229E [81] were included in this so-called RiboVD library design. Coronaviruses are the largest RNA viruses with around 30 proteins [82]. The intrinsically disordered parts are below 8% for most coronaviruses. By using more than 100 human proteins and domains in selections against the novel RiboVD library, we provided novel insights into the coronavirus host-virus interactome. Further validation brought to light vulnerabilities which can be targeted for treatment.

Let´s have a closer look at our findings, with a focus on the parts that I was directly involved in.

ProP-PD revealed common targets between the different corona strains

In Paper II we describe the design and construction of a new M13 phage library to find motifs in IDRs of ss and ds RNA viruses (RiboVD). At the time of the design, there were 23 corona strains listed in UniProt including the endemic SARS-CoV-2, but also, the less threatening variants such as the common cold-causing virus 229E. The library design contains around 19,500 unique 16 aa long peptides, which were displayed on the P8 protein. After ensuring the quality of the constructed ProP-PD resource [83], we screened 139 human proteins/domains for binding to viral peptides. The bait domain collection was compiled to contain domains from previously reported interactors of SARS-CoV-2 proteins [84] together with well-established peptide-binding domains [14,15,52]. The bait proteins were used in three rounds of selection against the RiboVD, which is one less round of selection than previously used for the HD2 library selections (Paper I) due to the smaller library size. NGS analysis of the enriched binding phage pools generated information on interactions

involving proteins from a broad panel of viruses. Here, we focused on our findings for coronaviruses.

The results were analyzed to identify commonalities and differences between the SLiM-based interactions mediated by coronavirus proteins. We explored the data to find human baits that were targeted by several different corona strains, as well as human proteins that were only targeted by some viruses, and SARS-CoV-2 in particular. A network of those interactions was generated (Figure 11 A). Interestingly, among the common targets we found the MATH domain of ubiquitin carboxyl-terminal hydrolase 7 (USP7), a protein that deubiquitinates various host proteins. In our screen we found it to be bound by peptides from several coronavirus proteins such as NSP4 and NSP1. The FERM domains from Moesin, Radixin, and Ezrin were found to interact with peptides from E protein from the SARS-CoV-2, SARS-CoV, and MERS. In contrast, the interaction of ATG8 from MAP1LC3 A/B/C with NSP3 was only found for SARS-CoV and CoV-2. Similarly, we found that the NTF2 domains of the homologs G3BP1 and G3BP2 interact with a peptide from N-protein from SARS-CoV and CoV-2.

After identifying common and distinct targeting points, we decided to systematically determine the affinities of the motif-based interactions between SARS-CoV-2, SARS-CoV, and MERS to the human proteins. The affinities of the peptides were measured via FP-competition assays, as exemplified in Figure 11 C. In total, we determined the K_D values for 27 interactions, and an overview of the affinities is shown in the heatmap in Figure 11 B. As expected from the phage data, AP2M1 bound $NSP14_{6384\text{-}6397*}$ from SARS-CoV, CoV-2 and MERS with similar affinities (K_D=1.6 and 3.0 µM). Similarly, the G3BP2 NTF2 domain bound to the peptide from the N proteins of SARS-CoV and SARS-CoV-2 with similar affinities (K_D=1.7 µM and 2.8 µM) whereas a MERS $N_{12\text{-}26*}$ peptide, that was not found in the phage selection, was binding 10-fold weaker (K_D=51 µM). For the ATG8 domain of MAP1LC3C, we found that the $NSP3_{2019\text{-}2038*}$ peptide of SARS-CoV was binding almost 10-fold tighter than the same peptide from SARS-CoV-2 (K_D=210 µM and 28 µM).

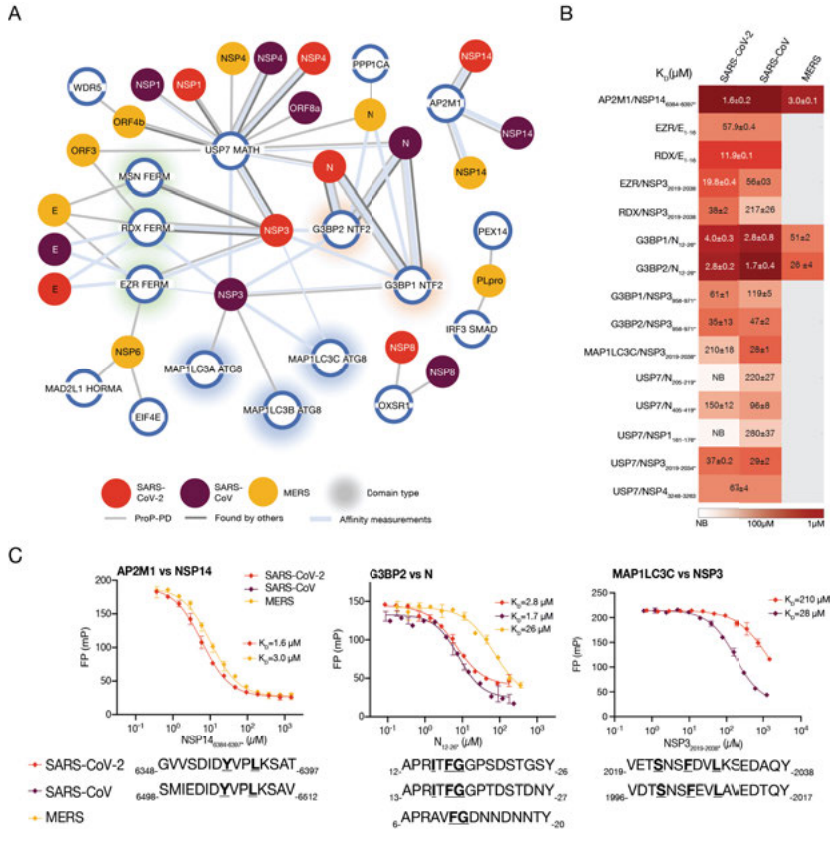

Figure 11 Highlights of the RiboVD selection results and affinity measurements. A) Network of host-virus PPIs for MERS, SARS CoV and CoV-2. Blue circles: human bait domains; red circles: SARS-CoV-2 proteins; purple circles: SARS-CoV proteins; orange circles: MERS proteins; light grey line indicates ProP-PD selection results and dark grey indicates interactions also found in other studies. Blue line indicates interactions subjected to affinity measurement. The thickness of the blue line shows the relative affinity. B) Heatmap of the 27 measured affinities, first column for peptides from SARS-CoV-2 followed by SARS-CoV and MERS. Dark red color indicates high affinity (low K_D value). C) Representative FP displacement curves for AP2M1 vs NSP14$_{6384\text{-}6397*}$, G3BP2 vs N$_{12\text{-}26*}$ protein, MAP1LC3C vs NSP3$_{2019\text{-}2038*}$ color-coded as in A, peptide sequences are indicated below and the specificity determining residues shown in bold and underlined. (Figure adapted from Kruse *et al.* [85])

Antiviral effect of SARS-CoV-2 peptides

To explore if it would be a valid antiviral strategy to inhibit the found interactions, we engaged in a collaboration with the Överby lab in Umeå to test the antiviral effects of the found peptides. The SARS-CoV-2 peptides found in the selections and validated as ligands of human proteins through affinity measurement were presented intracellularly as GFP-fusions in VeroE6 cells using a lentiviral delivery system. In this experiment, three peptides were

genetically fused to a carrier protein (GFP), and the effect of the peptides on virus infection was tested through a viral focus-forming assay (Figure 12). The peptides from NSP3 and E protein did not confer any significant inhibitory effect on viral replication, but for the N peptide, the virus titer of SARS-CoV-2 was 3.4-fold less than the mutated control peptide. It thus seemed like the G3BP binding peptide could serve as an inhibitor of viral replication. Building on this result, we found in the literature a peptide of the Semliki Forest virus, which is known to bind G3BP1/2 with high affinity [86]. The peptide contains a double FG motif and is in the future called G3Bpi wt. The effect of the GFP-tagged G3Bpi was even greater than the N peptide on reducing viral formation (Figure 12).

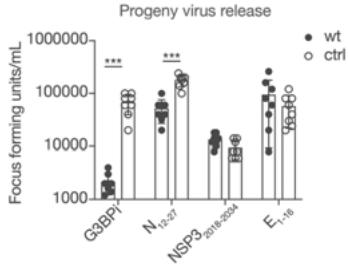

Figure 12 Screen for peptides with SARS-CoV-2 antiviral activity. The amount of SARS-CoV-2 virus release was determined 16 h post-infection by focus forming assay (independent n=8), ctrl peptide motif mutated. (Figure adapted from Kruse *et al.* [85])

The G3Bpi wt peptide was found to bind 10-fold stronger than the SARS-CoV-2 N_{12-26} (K_D=4 µM vs K_D=0.3 µM) (Figure 13 A). The higher affinity likely explains the greater effect of the G3BP1 on viral release.

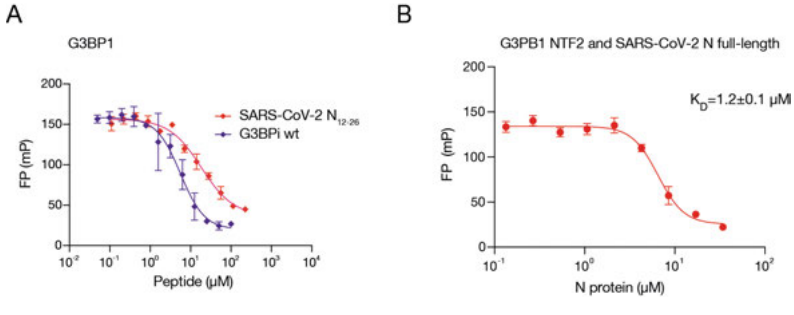

Figure 13 Affinity measurements by FP competition assay. A) Semliki Forest virus peptide G3Bpi in purple and SARS-CoV-2 N peptide in red. B) Full-length N protein displacing the N probe on G3BP1 NTF2 domain. (Figure adapted from Kruse *et al.* [85])

The further efforts were focused on the interaction between G3BP1 and G3BP2 and the SARS-CoV-2 N protein.

Relevant knowledge G3BP1/2

G3BP1 and G3BP2 stand for Ras GTPase-activating protein-binding protein 1 and 2. Their sequence identity is 58 % (determined with NCBI blast). G3BP1/2 binds to peptides with an ϕxFG motif with the NTF2 domain. The proteins play an essential role in the formation of stress granules and viral infections [87,88]. What are stress granules? They are aggregations in the cytosol including proteins and RNAs, which are induced when the cell is under stress [89,90]. Recent publications report that SARS-CoV-2 N induces stress granule disassembly but the mechanism is unclear [91,92].

We determined the affinity of the full-length N protein to the NTF2 domain of G3BP1 and found it to bind with a K_D value of 1.2 µM, which is a 4-fold higher affinity than the 15 amino acids long peptide used before (Figure 13 B).

The effect of N protein on stress granule formation was investigated in the presence and absence of ϕxFG (FG to AA) in the N full-length protein. The result has shown that the motif is relevant to disrupt stress granule formation.

Endogenous interactions of G3BP1/2

In the next step, we investigated the endogenous interactions of G3BP1/2 (Figure 14 B, D). We used G3BP1/2 as baits in selections against our HD2 library described in Paper I and identified endogenous interactions, with which the SARS-CoV-2 N protein is competing. The results revealed several proteins associated with stress granules and nuclear pore association such as USP10, TRIM25, and NUP214 (Figure 14 A). We determined the affinities for several of the peptides through FP displacement experiments, which revealed that of the peptides tested only $USP10_{5-19}$ bound with higher affinity than the SARS-CoV-2 N peptide (K_D=0.52 µM) (Figure 14 B). Since this paper was published the crystal structure of G3BP1 binding to the N peptide was solved [93]. We can superimpose this crystal structure with the structures of the G3BP1 NTF2 domain bound to G3Bpi [94] and a human peptide [95]. All three peptides are bound to the same pocket (Figure 14 E).

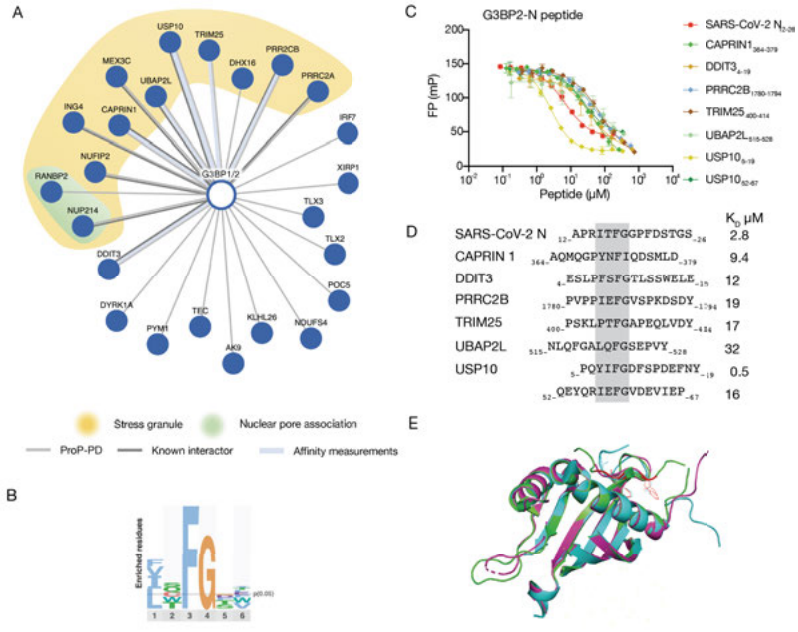

Figure 14 Human interaction partners of the G3BP1/2 NTF domains. A) Network of G3BP1/2 NTF2 ligands in light grey. Dark lines indicate previously known interactors and light blue lines indicate affinity measurements. The yellow background indicates proteins that are associated with stress granules and the green indicates nuclear pore association. B) PSSM generated based on high/medium confidence peptides of endogenous interactions. C) FP-based affinity determinations of human G3BP binding peptides, in comparison to the G3BP-SARS-CoV-2 N peptide interaction. D) Sequence alignment of peptides in C with determined K_D values. E) Superimposition of G3BP1 NTF2 structures with N peptide from SARS-CoV-2 in green (PDB 7SUO); G3Bpi in cyan (PDB 5DRV), Nucleoporin repeat peptide in purple (PDB 4FCM). (Figure adapted from Kruse *et al.* [85])

Conclusion and outlook

We developed a resource to identify SLiM-based host-coronavirus interactions. The selections identified more than 250 motif-based interactions with proteins from 18 coronavirus strains. We determined affinities of 27 interactions between human baits and coronavirus peptides and showed similarities and differences between different strains.

G3BP1/2 were identified as a potential target for viral inhibition, and the double FG containing G3Bpi peptide was found to have the strongest inhibitory effect on virus infection replicates.

The SARS-CoV-2 N peptide was found to bind G3BP1/2 with higher affinity than most endogenous ligands. Further studies are necessary to understand the importance of the interaction between N and G3BP1 and stress granule formation, and its relevance for viral infection.

Taken together, we found that ProP-PD is a great tool for investigating host-pathogen interaction and that identified interactions can be the starting point for inhibitor development. The approach can be easily adapted to other viruses.

Paper III Identification of motif-based interactions between SARS-CoV-2 protein domains and human peptide ligands pinpoints antiviral targets

Being a PhD student working on a host-pathogen project was exciting at the beginning of 2020, despite the worldwide fear and anxiety. In Paper III we applied our platform for finding human SLiM-based to the folded domains of the SARS-CoV-2. The topic of viral proteins binding to human SliMs is underexplored and less discussed in the literature [32,33,96,97].

In Paper III we recombinant expressed and purified the folded protein domains from SARS-CoV-2 and used them as baits in ProP-PD selections. We divided the genome into 31 expression constructs based on known and predicated modular domains (Figure 15). The structural proteins envelope (E), spike protein (S), and membrane glycoprotein (M) are extracellular proteins and are involved in the docking of the virus to the human cell. Therefore, those proteins were not investigated for intracellular SLiM-based interactions.

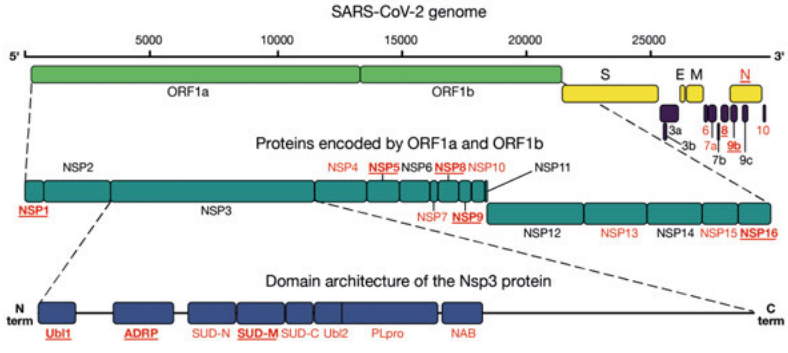

Figure 15 SARS-CoV-2 genome. Divided in two ORFs (open reading frames), which are further divided into 16 nonstructural proteins (NSP), in red all proteins/domains which have been successfully expressed for the Paper III, underlined the ones with enriched phage pool ELISA. (Figure adapted from Mihalic *et al.* [98])

Selections results

Twenty-six domains from SARS-CoV-2 proteins were expressed and purified in sufficient amounts for phage selections. After performing triplicate selections with the HD2 library 11 domains showed binding enriched phage pools. The peptide-coding regions of the phage pools were analysed by NGS, and processed using our established pipeline. The selection results are summarized in Figure 16. The peptides found were assigned different confidence levels following the system developed in Paper I (overlap of peptides, found in

several replicates, NGS score, and/or motif present). In total 281 high/medium confidence peptides for nine viral proteins (NSP1, NSP3 (Ubl1, ADRP and SUD-M), NSP5, NSP8, NSP9, NSP16, ORF8, ORF9b and N NTD) were identified. Enriched peptides with the highest scores were found for the catalytic inactive main protease (NSP5 Mpro C3408A) and for NSP9. NSP9 also had the most identified interactions with 118 peptides found followed by NSP1 (47) and then the main protease NSP5, also called Mpro (32). The enriched consensus motif for Mpro was found to be [FLM]x[AS]. The known proteolytic motif of the protease (LQ↓[GAS]) is similar to the determined recognition motif with position one having a leucine and three alanine or serine (numbering based on the logo in Figure 16 B) [99]. The motif for NSP9 was found to be G[F/L]xL.

Validations on the domain-peptide interaction were performed through FP affinity measurements (eight of the domains) and alanine scanning peptide SPOT arrays for five of the domains. The affinities were found to range from low micromolar to more than 300 µM.

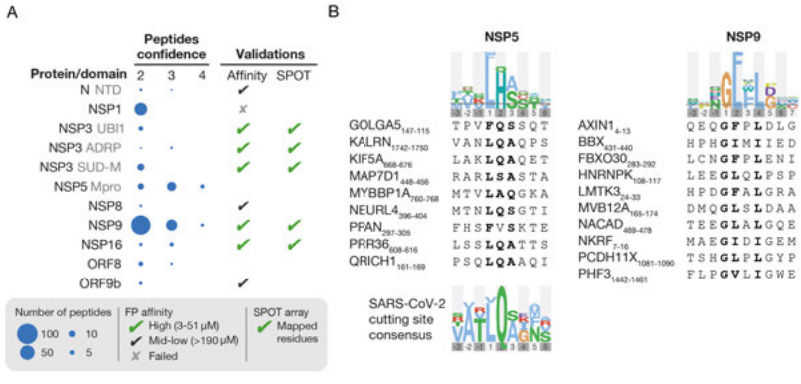

Figure 16 HD2 selections results with SARS-CoV-2 baits. A) Summary of the results for the 11 SARS-CoV-2 domains, confidence levels are indicated by numbers 2 and 3 (medium), 4 (high). The area of the blue circles is proportional to the number of peptides found. Further validation by affinity measurements (FP) and peptide SPOT arrays are indicated. B) Consensus motifs and alignments of catalytic inactive NSP5 Mpro and NSP9, in the bottom logo of proteolytic consensus for the protease based on available literature [99]. (Figure adapted from Mihalic *et al.* [98])

Furthermore, we investigated the structural determinants of the detected interactions using AlphaFold2. The key residues of interaction were confirmed via alanine scanning peptide SPOT array.

In addition, we screened peptides for antiviral activities. Two peptides targeting NSP9 and NSP16 were found to have antiviral effects. Those peptides can be starting point for antiviral drug development.

Conclusion and discussion

We showed that the HD2 ProP-PD resources can be easily used to investigate viral-human motif-based interaction. In a large-scale fashion, we can expand the knowledge about virus-host interactions. Additionally, for each new virus or viral protein the technique could be applied in a fast way.

Peptides revealed in selections can be starting point for antiviral drug development.

Paper IV Parallel exploration of short linear motif-based interactions using deep mutational scanning by phage display

The manuscripts I have presented so far focused on the identification and elucidation of SLiM-based interactions based on IDRs of human and viral proteomes. To define a motif several peptides with the same features are required. Sometimes the ProP-PD results are not enriched in enough many distinct peptides to build a PSSM and identify the key amino acids. Deep mutational scanning (DMS), where every single position in a sequence is exchanged to all possible amino acids, may be used to obtain a better understanding of the binding determinants in a given sequence. In the case of the motif-based interaction and the 16 aa long peptides, also the flanking regions of the specificity determinants can be investigated because the context of SLiM matters and may modulate the affinity [100]. In this manuscript, we aimed to establish a DMS by peptide-phage display protocol. We used highly defined DMS libraries to identify key residues and amino acid alternations that fine-tune the binding affinities.

Library design for evaluating the DMS approach

For the benchmarking library (BM) we chosen 23 wt peptides for 11 well-known peptide binding domains (Figure 17 B) from Paper I. For each wt peptide a by two aa shifted peptide was included. The peptides were systematically mutated and the peptide encoding oligonucleotide sequences were obtained from a commercial provider. An M13 bacteriophage library with P8 display was generated (Figure 17 C). The coverage of the generated BM library was determined to 99.6% through NGS (Figure 17 D). The BM library was used against the 11 bait proteins in four rounds of selections. Binding enriched phages pools were analyzed by NGS (Figure 17 A).

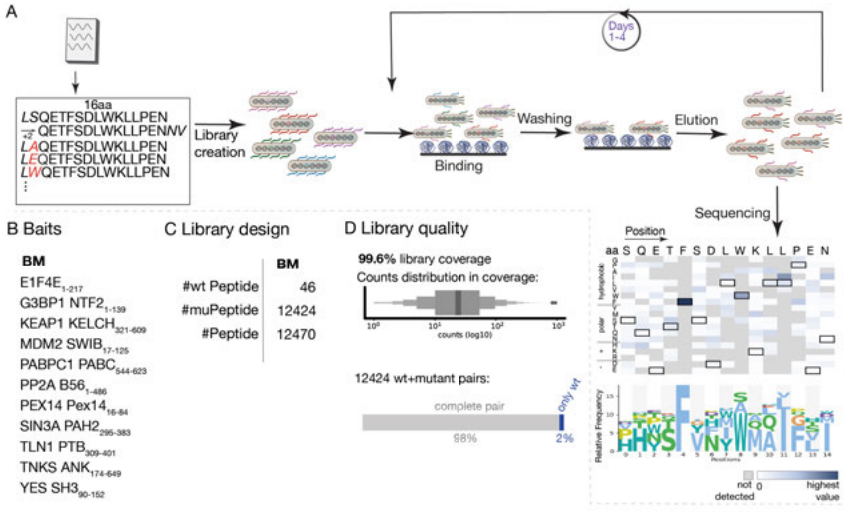

Figure 17 Overview of the DMS by phage display and analysis, together with BM library characteristics. A) Scheme of DMS by phage experimental work flow. DMS libraries were designed based on previous protein-peptide interactions literature. Overlapping peptides, shifted by two aas, were used as parental peptides. All possible amino acid substitutions were introduced to each position (indicated in red). The designed peptides were displayed on the major coat protein P8 of the M13 phage. Bait proteins (in blue) were immobilized, phages were allowed to bind, unbound phages were washed away and on the bound phages were eluted, amplified and used in the next day of selection. After four days of selection, enriched pools for day 1 to 4 were analysed by NGS. For each bait protein and peptide, the proportions of associated NGS counts results were visualized in heatmaps (in dark blue highest value in matrix, white indicates 0, grey indicates not detected) and PSSM were generated. B) Bait proteins for the BM library with domain type, in index the peptide chain borders. C) BM library design, indicating the number of wt peptides, mutant peptides and total number of peptides. D) BM library coverage and count distribution. (Figure adapted from Benz *et al.* [101])

Benchmarking and selections results

DMS by phage selection functions as a dropout assay, which means in each round we are losing more and more low-affinity peptides and enriching for higher-affinity ligands. In the evaluation of the DMS by phage protocol we identified which selection round/combination of selection rounds that were resulting in the most relevant PSSMs in comparison to the PSSMs built on annotated motif instances listed in the ELM database. A rank and a p-value that represent the similarity to the expected ELM was calculated. With those results, ROC curves with scores of the PSSMs of all combinations of days (single and combined) were plotted. The greatest AUC (area under the ROC curve) was achieved for a combination of results from selection days 2 and 3 (0.87) (Figure 18 A). Based on these results all further representations and

observations were based on data from the combined day 2 and day 3 of selection.

The PSSMs for the overlapping peptides were significantly different from the peptides designed for the same bait and also between baits (Figure 18 B). The biological background of G3BP1 was described in Paper II and is reported in ELM to bind [FYLIMV]xFG[DES]F. The observed specificity determinates based on the DMS experiment for the two peptides USP10 and CAPRIN are different (Figure 18 C) whereby the USP10 was matching to the ELM reported. Literature suggests that the two peptides are oriented in opposite directions in the peptide binding pocket [102]. The DMS of the two peptides from AXIN1 and CDCA2 that bind to the protein phosphatase 2A (PP2A) adaptor subunit B56 resulted in the previously determined motif [LFM]xx[ILV]xE.

The DMS analysis of two TLN1 PTB domain binding peptides from PIP4K1C and TPTE2 resulted in similar PSSMs, converging on a [FW]xxSxL motif. Interestingly, for the TPTE2 peptide the DMS analysis revealed additional positions contributing to binding (FxxSxLxYxP), and further suggested that a tryptophan would be favored over phenylalanine on the first position of the motif (Figure 18 D). Affinity determinations confirmed that an F95W mutation in $TPTE2_{92\text{-}107}$ conferred a major improvement of the affinity (Figure 18 E).

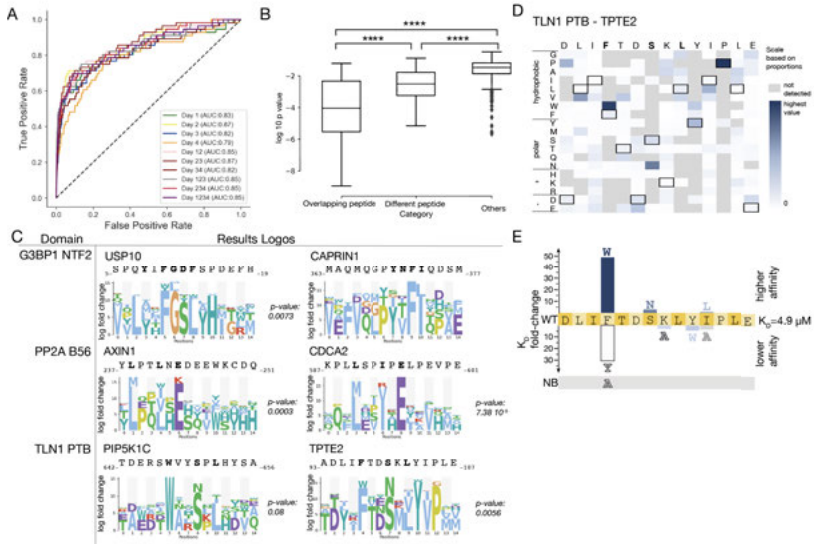

Figure 18 Benchmarking and selections result for the DMS BM library. A) ROC curve of PSSMs generated based on DMS results for different selections round and combinations of rounds as compare to the reported instances in ELM. B) Comparison of p-values for PSSMs identified for overlapping peptides, peptides for the same bait and peptides between different baits. **** indicates p<0.0001. C) PSSMs for selected cases from the BM set. G3BP1 NTF2 with USP10 and CAPRIN1, PP2A B56 with AXIN1 and CDCA2, TLN1 PTB with PIP5K1C and TPTE2. D) Heatmap of DMS results for the TLN1 PTB binding TPTE2 peptide. The parental peptide sequence is indicated on top, the highest value is indicated in dark blue, white indicates 0, grey indicates that there was no information. E) Fold change of affinity (K_D) of single mutation peptides of TPET2 in FP binding assay with TLN1 PTB. (Figure adapted from Benz *et al.* [101])

More challenging cases in the context of SARS-CoV-2

After successfully establishing the DMS-approach with the BM library we wanted to apply the method to gain information on SLiM-based interactions between human proteins and SARS-CoV-2 proteins. Therefore, we generated a new DMS library designed to display peptides from SARS-CoV-2 binding to human bait proteins, as well as human peptides binding to viral proteins (Figure 19 A). Binary peptides were previously identified with ProP-PD in Paper II and Paper III [85,98]. We used the same design parameters as above, displaying 16 aa long peptides on P8 and for each parental peptide including two sequences shifted by two amino acids. The coverage was validated through NGS and reached 96.5% (Figure 19 B). 2% of the pairs were identified as absent.

In benchmarking the approach using the BM library we could use the ELM information as references to evaluate the quality of the results generated. For the DMS-CoV we did not have that information. In terms of a characterizing

value, we instead looked at the sparsity of the data. Sparsity is defined by the lack of data or values. Given that the DMS by phage analysis is functioning as a dropout assay it is expected that the sparsity increases with every round of selection. However, if the data is too sparse (e.g. the selection being dominated by a very limited set of peptides) then it is not possible to generate PSSMs. We, therefore, used sparsity as a measurement of selection quality (Figure 19 C). The overall sparsity of the DMS CoV data (average 0.27) was lower as compared to the DMS BM data (average 0.69). For further analyses, we considered selection results with a sparsity below 0.5 as being of low quality.

Selections results of the challenging interactions

Representative PSSMs generated based on DMS results with sparsity indication for human bait proteins are shown in Figure 19 D. The EZR FERM motif was previously described to be YxΦ (where Φ is a hydrophobic amino acid). The DMS analysis correctly identified the expected binding motif and suggests a preference for L over F in position four of the E protein. Similarly, the DMS analysis correctly identified the expected G3BP1 NTF2 consensus motifs in N and NSP3 from SARS-CoV-2. The results further suggested a set of mutations that would be tolerated in the flanking regions of the motif, which we evaluated by FP-based affinity measurements (Figure 19 E). Here we can see that the selection results don´t correlate as great with the heatmaps suggested as they did for the TLN1 PTB-TPTE2, possibly because the affinity of the starting interaction is fairly high.

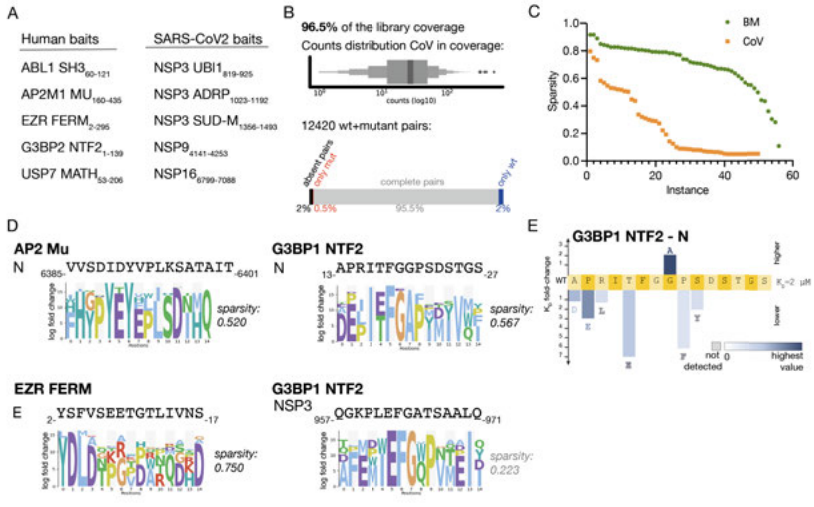

Figure 19 Library quality and DMS selection results using the CoV library. A) Protein domains used for the DMS-CoV analysis. B) Library coverage, mean NGS count of sequences, and the % of complete wild-type/mutant pairs. C) Sparsity of DMS-CoV results in comparison to the DMS-BM results. Green indicates BM and in orange CoV. D) PSSMs generated by DMS for AP2 Mu and $N_{6385-6401}$, G3BP1 NTF2 and N_{13-27}, EZR FERM and E_{2-17}. The sparsity value is indicated if above 0.5, and in grey if less than 0.5. E) Fold changes of K_D values of $N_{6385-6401}$ binding to G3BP1 NTF2 domain upon mutations. The color code relates to the associated proportion of NGS counts in the DMS NGS data. (Figure adapted from Benz *et al.* [101])

DMS results for peptides binding to viral bait protein domains

Previously, alanine scanning SPOT arrays were performed for the set of parental peptides binding to the viral bait domains (Paper III) [98]. The residues suggested to be important for binding through alanine scanning are shown in the sequences in Figure 20 (underlined residues). We could confirm the importance of most of the key residues suggested by the SPOT array analysis with the DMS-CoV results, and the analysis further suggested mutations that could improve the affinity of the interactions. In each case, mutations indicated to be favorable for binding based on the DMS results were tested through FP-based affinity determinations (Figure 20 D, E, F). The analysis confirmed mutations that improved the affinities for NSP3 Ubl1 and NSP3 SUD-M. We further generated double mutants of the $NCOA2_{1073-1088}$ peptide that binds to NSP3 Ubl1 and the $PRDM14_{196-211}$ peptide that binds to NSP3 SUD-M. The binding affinities were determined, which in both cases confirmed improved affinities in comparison to the respective wt and single mutants (Figure 20 G,H).

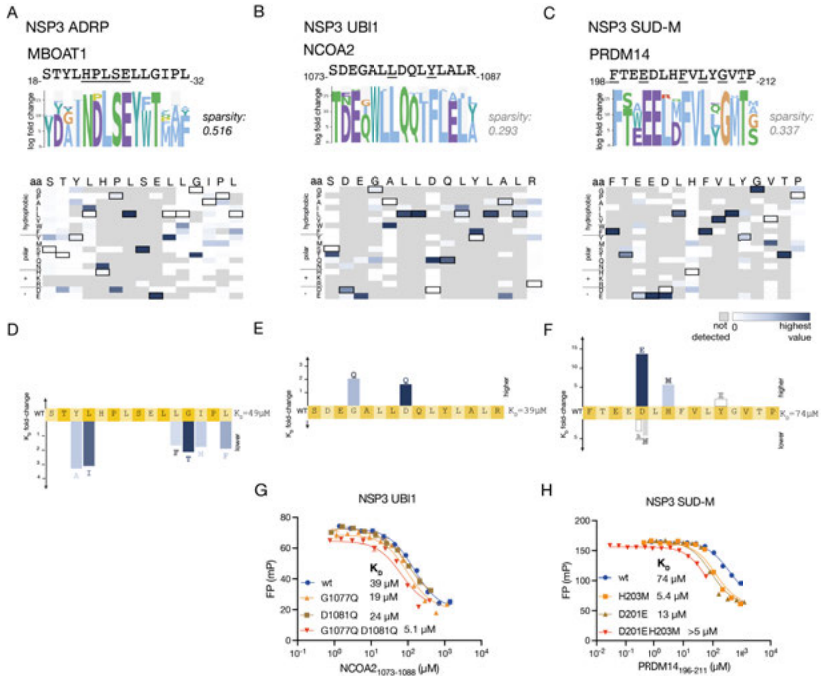

Figure 20 DMS results for viral bait proteins and human peptide ligands. A-C) Parental peptide sequences with key residues identified via alanine scanning peptide SPOT arrays underlined together with PSSMs and heatmaps generated based on the DMS data for A) NSP3 ADRP, B) NSP3 Ubl1, and C) NSP3 SUD–M. D-F) Fold change of K_D values upon mutations of indicated peptides binding to D) NSP3 ADRP, E) NSP3 Ubl1, and F) NSP3 SUD-M. G) FP displacement curves of FITC-labeled probe peptide bound to NSP3 Ubl1 displaced by wt NCOA2$_{1073\text{-}1088}$ (blue), G1077Q (orange), D1081Q (brown) and combined double mutant (red). I) FP displacement curves of FITC-labelled probe peptide bound to NSP3 SUD-M outcompeted with wt PRDM14$_{196\text{-}211}$ (blue), H203M (orange), D201E (brown) and double mutant (red). (Figure adapted from Benz *et al.* [101])

Optimization of NSP9 binding peptide for a more potent viral inhibitor

For NSP9 we explored the binding determinants using three distinct peptides. Based on the results, PSSMs were successfully generated, and revealed similar but distinct PSSMs depending on the parental peptide used (Figure 21A). Here, we focus on the NOTCH4 peptide for which the analysis resulted in a GxWLG motif. The heatmaps (Figure 21 B) indicated several amino acid substitutions in the motif-flanking regions that could potentially be introduced to improve the affinity of the interaction (Figure 21 C). Affinity measurements with a set of mutant peptides confirmed that the DMS data correctly identified affinity-enhancing point mutation. By combing two mutations (A1608V/P1613V) we improved the affinity of the initial NSP9-NOTCH4

interaction 12-fold. As previously described, the NOTCH4 peptide has an antiviral effect against SARS-CoV-2 infection [98]. To explore if the enhanced affinity would translate into a more potent inhibitor we turned the double mutant peptide into a cell-permeable variant (Tat-tagged peptide) and tested its effect on SARS-CoV-2 infection (Figure 21 E). We found that the mutations conferred a minor improvement of the IC50 value (from 211 µM to 170 µM) without any impact on cell viability (Figure 21 F).

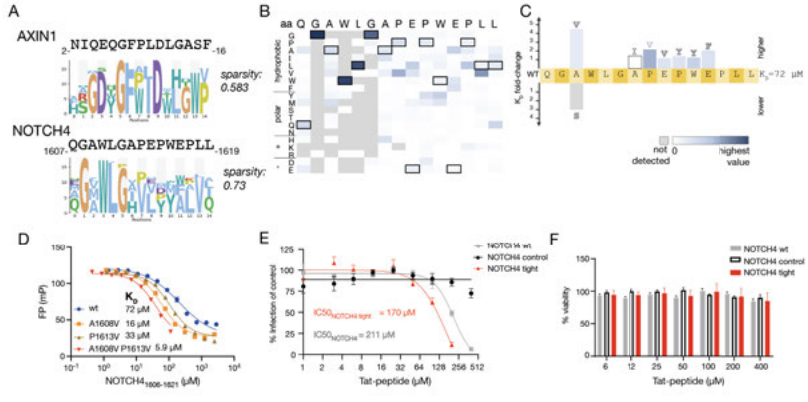

Figure 21 DMS results for NSP9 binding peptides. A) PSSMs generated based on DMS results for the AXIN1 and NOTCH4 peptides. B) Heatmap of DMS results for the NOTCH4 peptide. C) Fold-change of affinities upon point mutations suggested by the DMS results. D) FP displacement curves of FITC-labeled probe peptide bound to NSP9 and unlabeled NOTCH4 wt (blue), single mutant (A1608V (orange) and P1613V (brown)) and double mutant peptides (red). E) Inhibition study of viral replication upon addition of cell-permeable peptides. Grey indicates the wt peptide, black the negative control peptide (with mutated motif), ´red is the NOTCH4 tight (A1608V P1613V) peptide. F) Cell viability assay with increasing concentrations of Tat-tagged peptides. The colors are as in E). (Figure adapted from Benz *et al.* [101])

Conclusion and discussion

The motif-based interactions found in Paper I-III were further investigated via DMS. For this, two novel phage libraries were designed and generated, the first based on well-established peptide-bait protein interactions from Paper I, and the second based on more unexplored host-virus interactions from Paper II and Paper III. We were able to establish and benchmark DMS by peptide phage display. The result led to an optimized NSP9 peptide with an improved antiviral effect.

Concluding remarks and future perspective

Diving into SLiM-based interactions has been exciting but we reached the point of concluding remarks, where I want to summarize and point out, where all this could lead to and the importance for the research field.

In this thesis, I have described the development and improvement of ProP-PD to investigate human SLiM-based interactions as well as host-virus PPIs. The number of endogenous motif-based interactions is estimated to be in the order of 100,000 [13] and at the start of this PhD, the most comprehensive database (ELM) contained roughly 4,000 annotated instances [14,15,103]. Now after a little bit more than five years later, we contributed with more than 2,000 SLiM-mediated interaction pairs (with Paper I). This is a small step closer to the predicted 100,000, but a great step considering that we only screened 35 bait proteins.

With Paper II we were able to identify SLiMs in the IDRs of SARS-CoV-2 binding to human domains. Through this study, we identified 269 peptide-based interactions. The results contributed to a deeper understanding of the SARS-CoV-2 virus responsible for the Covid-19 pandemic. We pinpointed that the interaction between human G3BP1/2 and the viral N protein is involved in stress granule formation.

We also contributed to understand that SARS-CoV-2 proteins can bind peptides from human protein (Paper III). We were able to identify 281 high/medium confidence human peptides binding to domains from SARS-CoV-2.

Through FP-based affinity measurements, we validated several motif-based interaction pairs and conclude at this point that the identified pairs show a span of affinities in the range from nanomolar to millimolar.

As a part of method development, we further showed that we can use DMS by phage to determine the specificity determinant and the relevance of flanking regions, as outlined in Paper IV. The information can be further used to optimize the affinity of a peptide ligand. The results provided a foundation for optimizing a cell-permeable antiviral peptide inhibitor against SARS-CoV-2.

Several different phage libraries were developed in this thesis. We can categorize the application of different libraries for different research questions.

- The HD2 and RiboVD libraries are utilized for the identification of potential biologically relevant interaction partners.
- The DMS libraries answer specific research questions such as what is the binding motif in a given sequence, and how mutations in the motif and flanking regions affect binding for the given SLiM interactions.

My research has shown that the HD2 and RiboVD ProP-PD libraries are potent interaction discovery tools, which can be combined with DMS libraries to gain a deeper understanding of the molecular determinants of specific interactions.

The resources, tools, and results of this thesis contribute to a deeper understanding of motif-based protein-protein interaction in health and disease. In the future perspective, efforts should be continued to close the gap between the estimated number of 100,000 SLiM-based interaction. This will help to understand and elucidate motif-based interactions in human and viral contexts. The gained knowledge can be used to build protein networks. Using a systems biological approach rewiring of the network caused by disease and infection may be identified. The foundation is laid to tackle this tremendous task of identifying the missing SLiM-based interactions. The peptides which have been identified to have antiviral effects could be further characterized and used as starting points for drug development [36].

There are not only RNA viruses out there, which are pathogens to the human species. Also, the ss and ds DNA viruses capable to infect humans should be explored further. On top of this, bacteria and their intracellular effector proteins are a harm to the human system [104]. Much thus remains to be done in the field in terms of understanding how pathogens exploit SLiM-based interactions to take over the host cell.

As a scientific community, let's *dive into* motifs and elucidate the world with a magnifying glass.

Populärwissenschaftliche Zusammenfassung

Das ist einer meiner Lieblingsabschnitte in dieser Thesis. Ich kann mit einfachen Worten, und hoffentlich auch für alle meine Liebsten verständlich erklären, was ich die letzten 5 Jahre in Schweden gemacht habe.

Wo fange ich am besten an?- Wir Menschen bestehen aus einer Vielzahl von Zellen. In diesen Zellen sind viele Mitspieler zur Bewältigung all der komplexen Aufgaben. Proteine sind eine Gruppe dieser Mitspieler. Sie sind aus 20 verschiedenen Bausteinen, den sogenannten Aminosäuren, aufgebaut, und hängen in einer Kette zusammen. Die Kette allein kann Aufgaben bewältigen, aber wie wir auch im Alltag sehen, im Team geht vieles besser.

Wie erkennt ein Protein nun seine Mitspieler? Dafür gibt es unterschiedliche Möglichkeiten, eine davon ist die „motiv-basierte Interaktion". Bevor ich diese näher beschreibe, möchte ich beschreiben, wie man diese Interaktion findet.

Diese Technik heißt „*Proteomic Peptide Phagen Display*" (kurz ProP-PD). Phagen sind Viren, die nur Bakterien infizieren können. Sie sind so aufgebaut, dass man Informationen einfügen kann, welche dann auf die Phagen aufgesetzt werden. Man könnte es mit einer Krone auf den Viren vergleichen. Diese „Kronen", fachlich richtig nennt man sie Peptide, bestehen aus einer kurzen Kette von Aminosäuren. Das Wort „Proteomic" beschreibt den Ursprung dieser Peptide. Ich habe für meine Arbeit Peptide sowohl vom Menschen als auch von Viren verwendet.

Vereinfacht könnte man die Technik so erklären: Man nimmt ganz viele Phagen und bringt sie auf engsten Raum mit einem Protein. Die Phagen, die gerne Zeit mit dem Protein verbringen, weil sie gut zusammenpassen, binden aneinander und die, die nicht binden werden weggewaschen. Nun kann man die Phagen, die gebunden haben, analysieren und ihre Kronen, also die Peptide bestimmen. Alle Peptide für ein Protein werden dann darauf untersucht, ob sie ähnliche Charakter haben. Das bedeutet, dass man beim Aufbau der Peptide, die ja aus einer Kette von Aminosäuren bestehen, ein Muster erkennen kann. Dieses wiederkehrende Muster nennt man auch Motiv. Diese Motive bestehen normalerweise aus 3 bis 10 Aminosäuren in einer Kette.

Dieses Binden auf Grund des passenden Motivs nennt man motiv-basierte Interaktion wie bereit zuvor schon einmal kurz erwähnt. Diese Interaktionen finden in den Zellen statt, im gesunden, wie auch im kranken Zustand.

Wie wir in den letzten Jahren zur Genüge gesehen haben, können Viren Krankheiten auslösen. Ein Virus ist nicht lebendig, deshalb braucht es einen Wirt, zum Beispiel den Menschen, um sich zu vermehren. Das Virus benutzt nun dasselbe Motiv, um die menschliche Zelle zu missbrauchen und sie dazu zu bringen, das Virus zu vermehren. Das Virus nimmt die Zelle sozusagen als Geisel.

In meiner Doktorarbeit habe ich „Proteomic Peptide Phage Display" benutzt, um rasch viele dieser Motive sowohl im Menschen als auch im Zusammenhang mit Infektionen, die von Viren ausgelöst werden, zu entdecken. Wir konnten die Technik verbessern und in einem nächsten Schritt bestimmen, wie fest die Bestandteile, Protein und Motiv zusammenhalten. Diese Motive können auch als Transportsignal dienen. So konnten wir im Paper I zeigen, dass Proteine aufgrund eines bestimmten Motivs vom Zellplasma in den Zellkern gebracht werden.

In der zweiten Publikation (Paper II) konnten wir im Falle einer Coronavirusinfektion zeigen, wie Viren Motive verwenden, um Zellen als Geiseln zu nehmen. Und wie die Peptide, die das entsprechende Motiv beinhalten, dazu verwendet werden können, die Vermehrung des Virus zu verlangsamen. Dies kann als Ausgangspunkt für die Entwicklung von Medikamenten dienen.

Im Paper III benutzten wir Proteine von SARS-CoV-2 und testeten, welche menschlichen Proteine eine Interaktion damit eingehen. Das ist nicht gefährlich, da wir nur einen Teil des Virus für diese Tests genommen haben. Durch die Versuche konnten wir ein tieferes Verständnis über das Coronavirus erlangen. Auch hier konnten wir vielversprechende Ausgangspunkte für eine Medikamentenentwicklung finden.

In der vierten Publikation (Paper IV) gingen wir noch einen Schritt weiter. Wir verwendeten die „Peptide Phage Display"- Technik, um ein besseres Verständnis für die motiv-basierende Interaktion von Peptiden zu Proteinen zu gewinnen. Peptide bestehen ja aus einer Kette von Aminosäuren. Wir haben in dieser Versuchsreihe nun jede Aminosäure durch jede mögliche andere Aminosäure ausgetauscht und anschließend getestet, ob sich die Bindung dadurch verbesserte oder schwächer geworden ist. Dadurch können wir schneller und genauer voraussagen, wenn sich ein Virus verändert oder Krankheiten auftreten.

Zusammenfassend kann ich feststellen, dass ich mit meiner Doktorarbeit dazu beigetragen habe die Technik ProP-PD zu etablieren, auch konnte ich zeigen, dass die Technik geeignet ist, Wissenslücken in Bereich der motiv-basierten Interaktion zu schließen. Diese Interaktionen sind sowohl in

gesundem Körper als auch wenn es um Erkrankungen geht, von großer Bedeutung und sollten darum weiter erforscht werden.

Populärvetenskaplig sammanfattning

Människan består av många olika celler. I dessa celler finns många olika spelare som utför olika komplexa uppgifter. Proteiner är en av dessa spelare. De består av 20 olika byggstenar, så kallade aminosyror, som länkas ihop till en kedja. Proteiner kan själva utföra vissa uppgifter, men som vi ser i vardagen, många saker funkar bättre som ett team.

Så hur känner proteiner igen sina medspelare? Det finns olika sätt att göra detta, ett av de sätten är ”motivbaserade interaktioner”. Innan jag beskriver detta i mer detalj vill jag förklara hur man kan hitta dessa interaktioner. Tekniken jag har använt kallas ”proteomisk peptid fagdisplay” eller förkortat (ProP-PD). Fager, eller bakteriofager, är virus som bara kan infektera bakterier.

De är utformade på ett sätt som gör att man relativt enkelt kan tillföra information, som sedan presenteras på ytan av fagerna. Man kan jämföra det med de proteiner som gör att virus ser ut att ha spiror som in en krona på sig. Spirorna på fagen består av peptider som är korta aminosyrakedjor. Ordet proteomisk vittnar om var dessa peptider kommer ifrån nämligen naturliga proteiner från olika källor, jag har använt delar från både mänskliga proteiner, och proteiner från virus, för att skapa peptiderna som används i mitt arbete. När man har skapat dessa fager med peptider på utsidan används de i fagdisplay. De tekniken kan beskrivas på följande sätt: man tar många fager, med olika peptider på sig, och lägger dem i ett litet utrymme med ett större protein. Fagerna som gillar att hänga med proteinet för att de passar bra ihop, binder till varandra. Fager som inte matchar och därför inte binder tvättas bort. Där efter kan man analysera fagerna för att se vilka spiror, eller peptider, som bundit. Alla olika peptider analyseras för att de om an kan hitta ett mönster i sekvensen av aminosyrakedjorna. Hittar man ett mönster av liknande sekvenser kan man beskriva dem som et motiv. Dessa motiv som är gemensamma för peptiderna som binder brukar bestå av kedjor mellan 3 och 10 aminosyror, och interaktioner mellan proteiner och dessa motiv kallas motivbaserade interaktioner, vilket är ett ord som jag använde innan. Interaktionerna jag har beskrivit händer inuti i celler, detta sker i celler som både kan vara friska och sjuka.

Som vi har lärt oss de senaste åren så kan virus orsaka sjukdomar. Men virus till skillnad från bakterier är inte levande så därför behöver de ett värddjur som

till exempel en människa för att fortplanta sig. Ett sätt den gör det på är att använda samma motiv för att parasitera på värdproteiner och där med ta celler som gisslan.

I min avhandling har jag använt proteomisk peptid-fagdisplay för att på ett snabbt vis hitta många av dessa motivbaserade interaktioner i mänskliga protein men också i virusprotein som vid en virusinfektion. Detta är en vidareutveckling och förbättring av tekniken från tidigare studier. I ett nästa steg kunde jag visa hur starkt proteinet och peptiden med motivet binder samman. I den fösta artikeln i avhandlingen (Paper I) visar jag hur dessa motiv kan fungera som transportsignal i celler, där ett protein genom att binda ett särskilt motiv kan bli transporterat från cytoplasman till cellens kärna.

I den andra artikeln (Paper II) visar vi hur coronavirus använder motivbaserade interaktioner för att inkräkta på cellen vid en virusinfektion. Vi kunde också visa att om man tillförde peptider med samma motiv som viruset använde kunde man sakta ner skapandet av nya virus i cellen. Det kan vara en bra startpunkt för att skapa antivirala läkemedel.

I den tredje artikeln (Paper III) använde vi oss av proteiner från SARS-CoV-2 och testade vilka motiv från mänskliga protein som band till dem. Eftersom vi bara använde oss av proteiner och inte det faktiska viruset är det ingen fara för infektion. Denna studie gav oss nya djupare insikter om coronaviruset. Även i den här studien hittade vi intressanta mål för vidare läkemedelsutveckling.

I den fjärde artikeln (Paper IV) tog vi det hela ett steg längre genom att använda peptider vi vet binder ett protein och skapa ett bibliotek av peptider en efter en av aminosyrorna byts ut till de andra 19 möjliga och på det sättet kan man kolla om interaktionen blir starkare eller svagare på grund av utbytet. Resultaten vi fick från det kan vi senare använda för att förutspå om förändringar av ett virusmotiv genom mutationer (det som händer när vi får en ny variant av viruset) kan skapa farligare virus.

Sammanfattningsvis, med min avhandling har jag bidragit med att etablera metoden med proteomisk peptid-fagdisplay och jag kunde också visa att metoden är användbar för att hitta interaktioner viktiga motivbaserade interaktioner. De interaktionerna är viktiga både I den friska kroppen och vid sjukdom så som infektioner, detta är något som det är viktigt att forska vidare på.

Acknowledgments

The acknowledgment is probably the most read section in any thesis. Hopefully I don´t disappoint. A small disclaimer at the beginning. If you don´t find yourself here and you think you should be stated here. You were and are essential for my thesis and to me, but in the craziness of the last days -mistakes happen.

I would like to start to thank **Ylva**, for coming up with such an amazing project, who could have thought that it perfectly fitted into world affairs. Probably you... Trusting me with the project, hiring amazing people in the network, the endless scientific advice- I think now I´m capable to spot motifs also by eye in a peptide list- a achievement. But mostly the interhuman relationship and conversation. I think you can agree that 90% of those conversations were better not heard by others' ears.

Norman Davey one of my three Co-supervisors, none of my projects would have been anyhow existing and presentable without your input or bioinformatical genius mind. **Helena Danielson** in 2015 you trusted me with a small student project and without that, I would have never ended up in the Department of Chemistry-BMC and a lot of things would look differently today. During my PhD, you have been always willing to help out and give on so many levels beneficial advice. Number three in my Co-supervisor round **Ola Söderberg**, glad to have your advice for my thesis and progress.

I want to thank the whole **SSF team**, I think what we had and have during this project is crazy, would not have liked to do it without you. In particular **Leandro**- I remember the first Zoom meeting my first week in BMC and you in Argentina. You arrived in Uppsala and small city tour … but most importantly your character, basically always calm when I had my freakout moment, the massive will to have leg day and to explain the 100s game rules to me, because I didn´t know the game. **Eszter** partner in crime to find SLiMs, in the dark days but even more on the sunny days, in BMC but more preferably outside of BMC. We will keep doing pep talks and having all the sushi in the world. They always say a group of three females is not a good combination. I think this is completely wrong with **Marie** we formed an amazing crowd. Unforgettable trips but even more important, lunches to ventilate and reenergized for going back to work. **Filip** thank you for every heated discussion,

challenging my scientific thoughts, and giving me space to grow. **Ahmed** time was short but you taught me that in the field words can be used differently and is important to choose your words wisely. In my PhD time, I have seen and had the pleasure to work with amazing people in the **Ivarsson group**. **Johanna** where should I start: Thank you for gym sessions at 6 o´clock in the morning, accepting that a straight-shooted comment which never meant to hurt just "Caroline phrasing". **Susanne** - I was so glad to have a female in the lab when I started. The last years you showed me so many times that the medal has always two sides, to see the other perspective. Will never forget the Marlies presentation discussion (I think you know which conversation I talk about)… **Ali** every beginning is tough but I think the development was great. I learned so much from you- the king of cloning. **Gustav** – "Can I have a stupid question?" I don´t know how many times I asked you that, and not a single time you actually told me that it is a stupid question. My first office mate. I´m super glad that still know that I can send you the "stupid" questions. **Emily,** I enjoyed seeing you growing in the lab, open to hearing what the last generation has to say and the nice lunch we had in Elma, always ready to repeat that. **Priyanka** your positive energy filled the whole office and pushed the german grumpiness several times away. **Max,** keep up the good work.

While teaching I had once a student which said, "It sucks to be successful" and I promised him to have that as a statement in my thesis. Now I kept my promise. I had the pleasure to supervise several **students** (**Elias**, **Emma**, **Hanna**, **Kalle**, **Lara**, **Lars**, **Margot**, **Nikoleta**, **Shady** and **Victor**) during my PhD. The progress we achieved would have been alone not possible.

Elias without your passion for illustrator, I would still fight with InkScape. The Pamela Reif discussion with you, **Lara,** was essential during that time. The DMS project would not exist without your workforce **Lars**. **Margot** and **Nikoleta** I´m so glad that the experience with my supervision didn´t lead to leaving science and wish you amazing research results.

The whole **Biochemistry sections** including former and recent co-workers, **Eldar**, **Maxim**, **Nadine**, **Giulia**, **Joana** and **Ted**. Thank you for making the workspace a place to enjoy good company. Dear **Organic chemist, Scott, Stefan** and **Fabio** you have been essential for realizing a PhD isn´t just working, there is life outside of BMC. **Kate** your hugs kept my soul even in the coldest Swedish days warm- never change that. You are my walking organic chemistry encyclopedia.

Thank you also to the **Analytic chemistry section**, your sparkling water kept me hydrated and the company of your staff is always a pleasure. **Catiá** amazing to meet you, and for sharing a similar past with me, it created a special bond.

I´m glad that the Department of Chemistry has such amazing permanent staff, **Johanna**, **Gunnar**, **Farshid**, **Eva, Francoise** and **Hanna.** I don´t know

how many times you saved me time or an experiment, helped me out or you Hanna, took the “smelly garbage” out.

A special thanks is going to all my **friends** outside of Department, from Konstanz, Black Forest, Toronto or Uppsala. **Valentina** without Marzipian-milchreis and hot chocolate in SciLife I probably would have quit in the first year. **Helen** we know each other longer, than this PhD was going on and fikas and beers at super random times were life-saving. I hope we keep doing this. **Sofia** and **Gloria** we should get a special spot in the Bierhuis for life time to enjoy after works. **Eva** glad that our ways have crossed. **Lisa** danke, dass du in der alten Heimat immer da bist, freue mich immer wieder zurück zu kommen. Last but not least **my family**, **Jean Paul** and **Emil** I love our bubble, let´s keep that one safe together. Bizu nini cyane. **Oma** und **Lisbeth**, dass ihr immer an mich geglaubt haben, für die moralische Unterstützung und für euer Interesse an dem was ich tue. Wir sind nicht eine Familie der großen Worte aber ohne euch **Mama**, **Papa, Florian** mit **Susanne** und **Sophia** mit **Lukas** wäre das nicht möglich gewesen. Danke fürs anrufen, hinterherreisen und die alte Heimat immer zuhause fühlen lassen.

References

1. Fukuchi, S., Hosoda, K., Homma, K., Gojobori, T. & Nishikawa, K. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct Biol* **11**, (2011).
2. Basile, W., Salvatore, M., Bassot, C. & Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput Biol* **15**, e1007186 (2019).
3. Deiana, A., Forcelloni, S., Porrello, A. & Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* **14**, e0217889-(2019).
4. Theillet, F. X. *et al.* The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24360 (2013).
5. Dunker, A. K. *et al.* Intrinsically disordered protein. *J Mol Graph Model* **19**, 26–59 (2001).
6. Bracken, C. NMR spin relaxation methods for characterization of disorder and folding in proteins. *J Mol Graph Model* **19**, 3–12 (2001).
7. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
8. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
9. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
10. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
11. Perica, T. *et al.* The emergence of protein complexes: Quaternary structure, dynamics and allostery. *Biochem Soc Trans* **40**, 475–491 (2012).
12. Wells, J. N., Bergendahl, L. T. & Marsh, J. A. Co-translational assembly of protein complexes. *Biochem Soc Trans* **43**, 1221–1226 (2015).
13. 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).
14. Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* **48**, D296–D306 (2020).

15. Kumar, M. *et al.* The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res* **50**, D497–D508 (2022).
16. Van Roey, K. *et al.* Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* **114**, 6733–6778 (2014).
17. Kussie, P. H. *et al.* Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science (1979)* **274**, 948–953 (1996).
18. Aasland, R. *et al.* Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett* **513**, 141–144 (2002).
19. Drew, K. *et al.* Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol Syst Biol* **13**, 932 (2017).
20. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
21. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
22. 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).
23. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
24. Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan, N. J. A systems approach to infectious disease. *Nat Rev Genet* **21**, 339–354 (2020).
25. Davey, N. E., Cyert, M. S. & Moses, A. M. Short linear motifs - Ex nihilo evolution of protein regulation Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Communication and Signaling* **13**, (2015).
26. Zanier, K. *et al.* Structural Basis for Hijacking of Cellular LxxLL Motifs by Papillomavirus E6 Oncoproteins. *Science (1979)* **339**, 694–698 (2013).
27. Kruse, T. *et al.* The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30. *Mol Cell* **69**, 136-145.e6 (2018).
28. Hagai, T., Azia, A., Babu, M. M. & Andino, R. Use of Host-like Peptide Motifs in Viral Proteins Is a Prevalent Strategy in Host-Virus Interactions. *Cell Rep* **7**, 1729–1739 (2014).
29. Ronco, L. V, Karpova, A. Y., Vidal, M. & Howley, P. M. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes Dev* 2061–2072 (1998).

30. Rozenblatt-Rosen, O. *et al.* Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* **487**, 491–495 (2012).
31. Valerdi, K. M., Hage, A., van Tol, S., Rajsbaum, R. & Giraldo, M. I. The role of the host ubiquitin system in promoting replication of emergent viruses. *Viruses* **13**, (2021).
32. Davey, N. E., Trave, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem Sci* **36**, 159–169 (2011).
33. Mamigonian Bessa, L. *et al.* The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci. Adv* **8**, 4034 (2022).
34. Batra, J. *et al.* Non-canonical proline-tyrosine interactions with multiple host proteins regulate Ebola virus infection. *EMBO J* **40**, (2021).
35. 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**, (2016).
36. Simonetti, L., Nilsson, J., McInerney, G., Ivarsson, Y. & Davey, N. E. SLiM-binding pockets: an attractive target for broad-spectrum antivirals. *Trends Biochem Sci* (2023) doi:10.1016/j.tibs.2022.12.004.
37. Sidhu, S. S., Weiss, G. A. & Wells, J. A. High Copy Display of Large Proteins on Phage for Functional Selections. *J. Mol. Biol.* **296**, 487–495 (2000).
38. Sidhu, S. S. Engineering M13 for phage display. *Biomol Eng* **18**, 57–63 (2001).
39. Smith G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (1979)* **228**, 1315–1316 (1984).
40. Nelson, F. K., Friedman, S. M. & Smith, G. P. Filamentous Phage DNA Cloning Vectors: A Noninfective Mutant with a Nonpolar Deletion in Gene III. *Virology* **108**, 338–350 (1981).
41. 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).
42. Hertz, E. P. T. *et al.* A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase. *Mol Cell* **63**, 686–695 (2016).
43. Ueki, Y. *et al.* A Consensus Binding Motif for the PP4 Protein Phosphatase. *Mol Cell* **76**, 953-964.e6 (2019).
44. Wu, C. G. *et al.* PP2A-B′ holoenzyme substrate recognition, regulation and role in cytokinesis. *Cell Discov* **3**, (2017).
45. T. A. Kunkel, J. D. Roberts & R. A. Zakour. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Methods Enzymol* **154**, (1987).

46. Davey, N. E., Simonetti, L. & Ivarsson, Y. ProP-PD for proteome-wide motif-mediated interaction discovery. *Trends Biochem Sci* **47**, 547–548 (2022).
47. Derda, R. *et al.* Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* **16**, 1776–1803 (2011).
48. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016).
49. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Biochemistry* **74**, 5463–5467 (1977).
50. Adessi, C. *et al.* Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res* **28**, 87 (2000).
51. Mercier, J.-F., Slater, G. W. & Mayer, P. *Solid Phase DNA Amplification: A Simple Monte Carlo Lattice Model*. *Biophysical Journal* vol. 85 (2003).
52. Benz, C. *et al.* Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol Syst Biol* **18**, (2022).
53. Davey, N. E., Haslam, N. J., Shields, D. C. & Edwards, R. J. SLiMFinder: A web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Res* **38**, (2010).
54. Gray, V. E., Hause, R. J. & Fowler, D. M. Analysis of large-scale mutagenesis data to assess the impact of single amino acid substitutions. *Genetics* **207**, 53–61 (2017).
55. Dreier, J. E. *et al.* A context-dependent and disordered ubiquitin-binding motif. *Cellular and Molecular Life Sciences* **79**, (2022).
56. Rogers, J. M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc Natl Acad Sci U S A* **115**, 10959–10964 (2018).
57. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* **42**, (2014).
58. Ernst, A. *et al.* Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst* **6**, 1782–1790 (2010).
59. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat Methods* **7**, 741–746 (2010).
60. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
61. Koch, P. *et al.* Optimization of the antimicrobial peptide Bac7 by deep mutational scanning. *BMC Biol* **20**, (2022).
62. Foight, G. W. & Keating, A. E. Comparison of the peptide binding preferences of three closely related TRAF paralogs: TRAF2, TRAF3, and TRAF5. *Protein Science* 1273–1289 (2016) doi:10.1002/pro.2881.

63. Zinkus-Boltz, J., Devalk, C. & Dickinson, B. C. A Phage-Assisted Continuous Selection Approach for Deep Mutational Scanning of Protein-Protein Interactions. *ACS Chem Biol* **14**, 2757–2767 (2019).
64. Goodyear, C. S. & Silverman, G. J. Phage-display methodology for the study of protein-protein interactions: Overview. *Cold Spring Harb Protoc* **3**, (2008).
65. Younger, D., Berger, S., Baker, D. & Klavins, E. High-throughput characterization of protein–protein interactions by reprogramming yeast mating. *Proc Natl Acad Sci U S A* **114**, 12166–12171 (2017).
66. Ramirez, J. & Nominé, Y. High-Quality Data of Protein/Peptide Interaction by Isothermal Titration Calorimetry. *Methods in Molecular Biology* **1964**, 99–117 (2019).
67. Freire, E., Mayorga, O. L. & Straume, M. Isothermal Titration. *Anal Chem* **62**, 950–959 (1990).
68. Stroth, N. A surface plasmon resonance-based method for monitoring interactions between G protein-coupled receptors and interacting proteins. *J Biol Methods* **3**, e36 (2016).
69. Yang, J. *et al.* Macrocyclic Peptides Uncover a Novel Binding Mode for Reversible Inhibitors of LSD1. *ACS Omega* **5**, 3979–3995 (2020).
70. Gogl, G. *et al.* Quantitative fragmentomics allow affinity mapping of interactomes. *Nat Commun* **13**, (2022).
71. Charbonnier, S., Zanier, K., Masson, M. & Travé, G. Capturing protein-protein complexes at equilibrium: The holdup comparative chromatographic retention assay. *Protein Expr Purif* **50**, 89–101 (2006).
72. Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat Methods* **12**, 787–793 (2015).
73. Moerke, N. J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Curr Protoc Chem Biol* **1**, 1–15 (2009).
74. Jameson, D. M. & Seifried, S. E. Quantification of Protein-Protein Interactions Using Fluorescence Polarization. *Methods* **19**, 222–233 (1999).
75. Wigington, C. P. *et al.* Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling. *Mol Cell* **79**, 342-358.e12 (2020).
76. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
77. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022-3040.e28 (2021).
78. Huovinen, T. *et al.* The selection performance of an antibody library displayed on filamentous phage coat proteins p9, p3 and truncated p3. *BMC Res Notes* **7**, (2014).
79. Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous Bacteriophage: Biology, Phage Display and Nanotechnology Applications. *Current Issues Mol. Biol.* **13**, 51–76 (2011).

80. Eigenthaler, M., Höfferer, L., Shattil, S. J. & Ginsberg, M. H. A conserved sequence motif in the integrin β3 cytoplasmic domain is required for its specific interaction with β3-endonexin. *Journal of Biological Chemistry* **272**, 7693–7698 (1997).
81. Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P. & Templeton, K. E. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Microbiol* **48**, 2940–2947 (2010).
82. Grellet, E., L'Hôte, I., Goulet, A. & Imbert, I. Replication of the coronavirus genome: A paradox among positive-strand RNA viruses. *Journal of Biological Chemistry* **298**, 101923 (2022).
83. Mihalic, F. *et al.* Large-scale phage-based screening reveals extensive pan-viral mimicry of host short linear motifs. doi:10.1101/2022.06.19.496705.
84. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
85. Kruse, T. *et al.* Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nat Commun* **12**, (2021).
86. Schulte, T. *et al.* Combined structural, biochemical and cellular evidence demonstrates that both fgdf motifs in alphavirus nsp3 are required for efficient replication. *Open Biol* **6**, (2016).
87. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345.e28 (2020).
88. Sanders, D. W. *et al.* Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324.e28 (2020).
89. Kim, S. S. Y., Sze, L. & Lam, K. P. The stress granule protein G3BP1 binds viral dsRNA and RIG-I to enhance interferon-β response. *Journal of Biological Chemistry* **294**, 6430–6438 (2019).
90. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol* **26**, 668–679 (2016).
91. Luo, L. *et al.* SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemble stress granules and facilitate viral production. *Sci Bull (Beijing)* **66**, 1194–1204 (2021).
92. Wang, J., Shi, C., Xu, Q. & Yin, H. SARS-CoV-2 nucleocapsid protein undergoes liquid–liquid phase separation into stress granules through its N-terminal intrinsically disordered region. *Cell Discov* **7**, (2021).
93. Biswal, M., Lu, J. & Song, J. SARS-CoV-2 Nucleocapsid Protein Targets a Conserved Surface Groove of the NTF2-like Domain of G3BP1. *J Mol Biol* **434**, (2022).
94. Kristensen, O. Crystal structure of the G3BP2 NTF2-like domain in complex with a canonical FGDF motif peptide. *Biochem Biophys Res Commun* **467**, 53–57 (2015).

95. Vognsen, T., Møller, I. R. & Kristensen, O. Crystal structures of the human G3BP1 NTF2-like domain visualize FxFG Nup repeat specificity. *PLoS One* **8**, e80947–e80947 (2013).
96. 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**, (2016).
97. Batra, J. *et al.* Non-canonical proline-tyrosine interactions with multiple host proteins regulate Ebola virus infection. *EMBO J* **40**, (2021).
98. Mihalic, F. *et al.* Identification of motif-based interactions between SARS-CoV-2 protein domains and human peptide ligands pinpoint antiviral targets. doi:10.1101/2022.10.07.511324.
99. Scott, B. M., Lacasse, V., Blom, D. G., Tonner, P. D. & Blom, N. S. Predicted coronavirus Nsp5 protease cleavage sites in the human proteome. *BMC Genom Data* **23**, (2022).
100. Bugge, K. *et al.* Interactions by Disorder – A Matter of Context. *Front Mol Biosci* **7**, (2020).
101. Benz, C. *et al. Parallel exploration of short linear motif-based interactions using deep mutational scanning by phage display*.
102. Schulte, T. *et al.* Caprin-1 binding to the critical stress granule protein G3BP1 is regulated by pH. doi:10.1101/2021.02.05.429362.
103. Gouw, M. *et al.* The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res* **46**, D428-d434 (2018).
104. Sámano-Sánchez, H. & Gibson, T. J. Mimicry of Short Linear Motifs by Bacterial Pathogens: A Drugging Opportunity. *Trends Biochem Sci* **45**, 526–544 (2020).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2267

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, 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 Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-498312

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023