Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 2289

Exploring conditional motif-based protein interactions in health and disease

JOHANNA KLICHE

ACTA UNIVERSITATIS
UPSALIENSIS
2023

ISSN 1651-6214
ISBN 978-91-513-1859-2
urn:nbn:se:uu:diva-508635

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in BMC A1:107a, Husargatan 3, Uppsala, Friday, 22 September 2023 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Ulrich Stelzl (Institute of Pharmaceutical Sciences, Pharmaceutical Chemistry, University of Graz, Austria; BioTechMed-Graz, Austria; Field of Excellence BioHealth - University of Graz, Austria.).

Abstract

Kliche, J. 2023. Exploring conditional motif-based protein interactions in health and disease. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2289. 70 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1859-2.

Protein-protein interactions (PPIs) orchestrate a variety of cellular events, ranging from signal transduction, scaffolding to subcellular localisation. A subclass of PPIs is mediated by short linear motifs, which are short amino acid stretches found in the intrinsically disordered regions of the proteome. Regulation of these interactions, which determines which proteins interact, as well as when and where interactions occur, is vital for performing cellular tasks. Phosphorylation can act as cue for this regulation, creating, breaking or fine-tuning a given interaction site. Disease-associated mutations may, in turn, deregulate motif-based PPIs. The Ivarsson lab established proteomic peptide-phage display (ProP-PD) for the discovery of binding peptides and motifs of protein domains. I extended the approach to kinase domains to assess their peptide binding properties and uncovered potential docking interactions of CASK and MAPK8.I further investigated the modulation of motif-based PPIs by phosphorylation guided either by bioinformatic predictions or by phosphomimetic ProP-PD. This led to unravelling of phospho-modulated binding motifs in the cytoplasmic tails of coronavirus host receptors. In addition, I developed an improved phosphomimetic phage library combining the intrinsically disordered regions of the human proteome with functionally prioritised phosphosites. Screening protein domains against the phosphomimetic library suggested novel interaction partners and their phospho-modulation. I demonstrated the dependency of clathrin binding on S839 HURP phosphorylation, which was further found to be required for the mitotic function of HURP. Lastly, I assessed, with the genetic variation phage library, whether mutational ProP-PD is suited to capture changes in motif-based PPIs as a consequence of disease-associated mutation. The method can with high confidence identify PPI-disruptive mutations, such as the P348L SQSTM1 mutation that diminishes binding to KEAP1 and a R157C CDC45 mutation that disrupts the nuclear localisation of CDC45 and its interaction with KPNA7. Together, I have investigated motif-based PPIs in health and disease and probed their identification by (mutational) ProP-PD. Mutational ProP-PD offers the advantage to identify conditional interaction partners, which might be overlooked in conventional ProP-PD experiments.

Keywords: ProP-PD, protein-protein interaction, short linear motif, phosphorylation, disease mutation

Johanna Kliche, Department of Chemistry - BMC, Biochemistry, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1859-2
URN urn:nbn:se:uu:diva-508635 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-508635)

To Papa who inspired me to become a scientist.

List of Papers

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

I. **Kliche, J.**, Simonetti L., Kraemer A., Benz C., Rask E., Knapp S., Ivarsson, Y. Exploring peptide binding of kinase domains by proteomic peptide-phage display. *Manuscript (Letter)*.

II. **Kliche J**., Kuss H., Ali M., Ivarsson Y. Cytoplasmic short linear motifs in ACE2 and integrin β3 link SARS-CoV-2 host cell receptors to mediators of endocytosis and autophagy. *Science Signaling*. 2021 Jan 12;14(665):eabf1117.

III. **Kliche J.,** Garvanska DH., Simonetti L., Badgujar D., Dobritzsch D., Nilsson J., Davey NE., Ivarsson Y. Large-scale phosphomimetic screening identifies phospho-modulated motif-based protein interactions. *Molecular Systems Biology.* 2023 Jul 11;19(7):e11164.

IV. **Kliche J.,** Simonetti L., Krystkowiak I., Kuss H., Diallo M., Rask E., Nilsson J., Davey NE., Ivarsson Y. Mutational ProP-PD for elucidating binding effects of disease mutations in the intrinsically disordered regions of the proteome. *Manuscript*.

The supplementary tables for Paper I (Table S1-S3) and Paper IV (Table S1-S10), as well as Figure S4, S7-10 for Paper IV can be accessed here: https://uppsala.box.com/s/yqqkcln59v7qgy4u005777rg0g92c3l7.

Contribution to the Papers

The work in all papers is a joint effort of the co-authors. Below I am specifying my contributions as first author.

I. Phage selections against the HD2 library with kinase domains from collaborators; fluorescence polarisation experiments and kinase assay including protein expression and purification; data analysis; data visualisation; writing of the original draft of the letter, editing and reviewing of the letter.

II. Protein expression and purification; fluorescence polarisation experiments; data analysis; data visualisation; writing of the original draft of the paper, editing and reviewing of the paper.

III. Generated the PM_HD2 library; protein expression and purification for the display, affinity measurements and crystallisation, phage display selections, affinity measurements (fluorescence polarisation and isothermal titration calorimetry); co-immunoprecipitation, RNA interference experiments and live cell microscopy under the guidance of the Nilsson lab; data analysis; data visualisation; writing of the original draft of the paper, editing and reviewing of the paper.

IV. Protein expression and purification; phage display selections; affinity measurements (fluorescence polarisation and isothermal titration calorimetry); co-immunoprecipitation and confocal microscopy partially under the guidance of the Nilsson lab; data analysis; data visualisation, writing of the original draft of the paper, editing and reviewing of the paper. Credit for the generation of the GenVar_HD2 library goes to my master student Hanna Kuss under guidance of Ylva Ivarsson.

The following papers are not part of this thesis.

I. **Kliche, J.**, Ivarsson Y. (2022) Orchestrating serine/threonine phosphorylation and elucidating downstream effects by short linear motifs. *Biochem J.* 479(1):1-22.

II. Benz C., Ali M., Krystkowiak I., Simonetti L., Sayadi A., Mihalic F., **Kliche, J.**, Andersson E., Jemth P., Davey NE., Ivarsson Y. (2022) Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol Syst Biol*, 18(1):e10584.

III. Kruse, T., Benz C., Hristoforova Garvanska, D., 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 NE., Överby AK., Nilsson J., Ivarsson Y. (2021) Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nat Commun.*, 12(1):6761.

Contents

Abbreviations

ATG8	autophagy-related protein 8
CDK	cyclin-dependent kinase
DNA	deoxyribonucleic acid
ELISA	enzyme-linked immunosorbent assay
ELM	eukaryotic linear motif database
FITC	fluorescein isothiocyanate
GenVar_HD2	genetic variation HD2 library
GFP/YFP	green/yellow fluorescent protein
HD2	second generation disorderome library
K_A/K_D	equilibrium association/ dissociation constant
K_I	equilibrium inhibition constant
KPNA	karyopherin alpha or importin alpha
LIR	LC3-interacting region
MAPK	mitogen-activated protein kinase
NGS	next generation sequencing
NTD	N-terminal domain
PPI	protein-protein interaction
PM_HD2	phosphomimetic HD2 library
PDZ	Postsynaptic density 95; Discs large; Zonula occludens-1
PRISMA	Protein Interaction Screen on Peptide Matrix
ProP-PD	proteomic peptide-phage display
PTM	post-translational modification
SARS-CoV-2	severe acute respiratory coronavirus 2
SH2/3	SRC homolog 2/3
SLiM	short linear motif
SNV	single nucleotide variance

In the interest of conciseness, abbreviations for gene and protein names are not listed.

Introduction

Essential to the function of life, intriguing in their shape and diversity, proteins represent the molecular toolbox performing the work of the cell and the body. Renowned for their role as enzymes, proteins have various tasks beyond that capacity, including their involvement in signalling networks, transport and scaffolding (1). Importantly, proteins rarely act as isolated entities but form networks among themselves or with other macromolecules, allowing them to communicate, interact and coordinate their endowed cellular tasks. Mapping those protein networks is pivotal for understanding the context in which proteins meet and interact, as well as predicting their function. These networks are mediated by protein-protein interactions (PPIs) and by now, databases report on more than 750,000 different PPIs found in the human (IntAct, 07/2023; (2)). This pinpoints the enormous complexity and dynamics of protein networks, which challenges and motivates to explore them in depth.

Motif-based PPIs

PPIs can be described at the different levels of protein modularity, covering protein complexes, globular domains, as well as globular domains and binding motifs (3). Globular domains are building blocks within a protein, both structurally and functionally. They often fold independently and folds are conserved among domain families (4). Domain-domain interactions have been proposed to rely on extensive surface recognition of the domains (5). Alternatively, globular domains recognise binding motifs, also called short linear motifs (SLiMs), in their interaction partners. SLiMs are short stretches of amino acids, typically 3-10 amino acids long and often found in the intrinsically disordered regions of the human proteome (**Figure 1**; (3,6)).

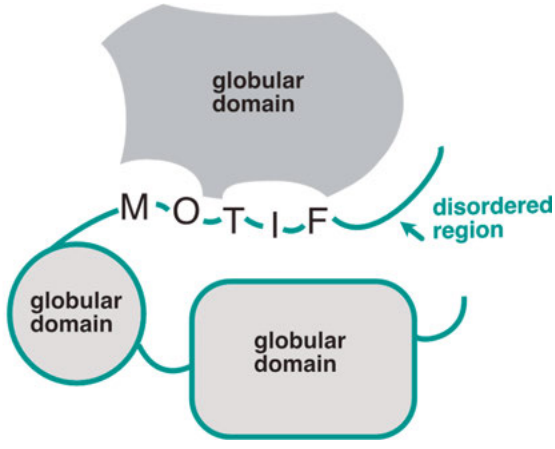

Figure 1: Motif-based PPIs. A globular domain (all grey) binds to a motif found in the disordered regions (green) of another protein, which consists of disordered regions (green) and globular domains (grey with green lining).

Intrinsic disorder itself can be defined as the absence of a well-defined three-dimensional structure in a given protein region (7). It can be estimated by metrics such as IUPred, which calculates the probability of a given amino acid to engage in interactions (8,9), and assessed by protein structure determination, using for example X-ray protein crystallography or nuclear magnetic resonance spectroscopy (10,11), as well as structure prediction such as AlphaFold (12). Intrinsically disordered regions are commonly enriched in polar and charged amino acids and depleted in hydrophobic residues (7,13). They represent accessible interaction platforms on protein surfaces and it has been predicted that more than 50% of the eukaryotic proteins contain long regions of disorder (14,15).

SLiM-based PPIs play a role in various cellular processes, such as transport, signalling, stress granule formation and degradation (16). In line with their function, these motifs confer often low- to mid-micromolar affinity upon binding and their interactions are transient (6). Specificity of the interactions is achieved by core residues within the motif, interspersed residues, residues in the flanking regions but also time and space of the interaction occurring, as well as local concentrations of the interaction partners within the cell (16,17). As of now 186 ligand classes have been defined in the eukaryotic linear motif (ELM) database (ELM 07/2023; (18,19)), but the human proteome has been predicted to harbour over 100,000 binding motifs (3).

A variety of globular domains have been reported to bind to binding motifs, and the most abundant and well-studied domain families include, among others, PDZ (Postsynaptic density 95; Discs large; Zonula occludens-1), SH3 (SRC homology 3) and WW domains (20–22). Nonetheless, peptide recognition goes beyond these classical recognition domains. Well-known representatives are proteins involved in trafficking and transport, such as the clathrin N-terminal domain (NTD) recognising LΦxΦ[DE]-motifs (with x representing any and Φ representing hydrophobic amino acids; [] : any of the amino acids in the brackets is allowed) in its cargo proteins or the μ2 subunit of the clathrin adaptor AP2 (AP2M1) binding to Yxx[LMVIF]-motifs, as well as the nuclear localisation signals (KR-motifs) recognised by the importin alphas (KPNAs) (23–25). Equally, recruitment of autophagy receptors to the autophagosome has been shown to rely on the recognition of the LC3-interacting region (LIR) motif (simplified: [WFY]xxΦ) by the autophagy-related protein 8 (ATG8) family of proteins (26,27).

Moreover, substrate targeting of enzymes such as phosphatases and kinases has in several cases been described to be motif-guided, either by auxiliary domains or by the catalytic subunit itself (28). These interactions are important for substrate selectivity and targeting catalytic activity and hence the downstream effects of, for example, phosphatases and kinases (28). Together, this pinpoints the plethora of SLiMs and peptide-recognition domains yet to be discovered, characterised and to be contextualised within protein networks.

Modulation of motif-based PPIs by phosphorylation

PPIs are often conditional and their restriction in time and space adds an additional level of regulation. Post-translational modifications (PTMs), including methylation, sumoylation, ubiquitination and phosphorylation (29), can act as cues orchestrating when and where a given interaction occurs. One key aspect of this thesis evolves around protein phosphorylation, which is the most abundant PTM and important for cell cycle regulation, trafficking events, gene transcription and signalling cascades (28,30). While kinases are responsible for the addition, phosphatases catalyse the removal of the phosphate moiety. On proteins, phosphorylation has most extensively been described for serine, threonine and tyrosine residues, but may also occur on basic amino acids such as arginine, lysine and histidine (30–33).

As for its effect on motif-based PPIs, phosphorylation can act as binary binding switch, either creating or breaking a binding site (**Figure 2**). However, it can also act to fine-tune interactions when the phosphorylation site is within the flanking sites or interspersed within the motif. The phosphate moiety is in this context not strictly required for the binding event but may enhance or diminish the binding affinity (28).

Figure 2: Phospho-modulated motif-based PPIs. Phosphorylation of residues within a motif creates (red domain) or breaks (grey domain) a binding site.

Impact of genetic variation on motif-based PPIs

A vast amount of non-synonymous single nucleotide variances (SNVs) has been identified by genome and exome sequencing, which contributes to the genetic variation in humans. However, pinpointing their functional and biological relevance on the protein level and for disease progression remains a major challenge at hand (34–36). When SNVs are not silent, they lead to a change in the sequence of the encoded protein by one amino acid, which may in turn affect protein functionality. Perturbed protein functionality manifests for instance in altered activity, stability, conformation, cellular localisation and also interaction with their binding partners (37–40).

Whereas a majority of the disease-associated mutations reportedly cluster to the protein core, affecting protein stability (41), mutations found on protein surfaces populate protein interaction interfaces (42–44). There are several studies suggesting that perturbation of PPI networks is a widespread phenomenon caused by disease-associated mutations (45–47). It can hence be proposed that mutations can affect PPIs and their networks by creating or breaking binding motifs within the intrinsically disordered regions of proteins (**Figure 3**).

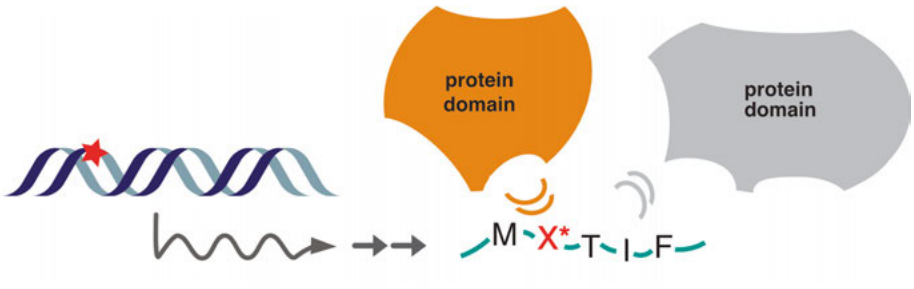

Figure 3: Impact of genetic variation on motif-based PPIs. Genetic variation which results in amino acid changes and occurs within a binding motif can impact motif-based PPIs, either creating (orange protein domain) or breaking (grey protein domain) a given binding site.

Proteomic peptide-phage display (ProP-PD)

Various methods are deployed to investigate motif-based PPIs, which are notoriously difficult to assess due to their low affinity and transient nature. These methods include yeast two-hybrid approaches (48–51), peptide arrays (52–55), peptide arrays coupled to mass spectrometry (PRISMA: Protein Interaction Screen on Peptide Matrix; (56–58)), as well as different display techniques such as yeast (59,60), bacterial (61,62) and proteomic peptide-phage display (ProP-PD) (63–65) (**Table 1**, **Figure 4**).

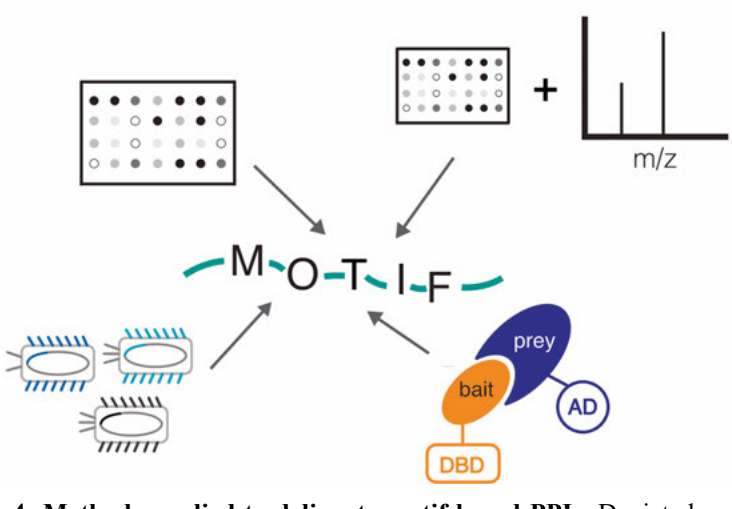

Figure 4: Methods applied to delineate motif-based PPIs. Depicted as examples are the yeast two-hybrid assay (right bottom), peptide arrays (left top), PRISMA (right top) and peptide phage display (left bottom).

Table 1: Methods applied to delineate motif-based PPIs including their advantages, disadvantages and selected references.

Approach	Advantages	Disadvantages	References
Yeast two-hybrid assay	Binary, direct PPIs of full-length proteins	- Yeast as host system, e.g. lack of necessary cofactors/chaperones - Steric constraints of the fusion - Interaction only detected if it occurs nuclear	(48–51)
Peptide arrays	- Can include PTMs - Information with amino acid resolution	- Limited number of peptides - Interaction not in biological context	(52–55)
PRISMA	- Full-length proteins from whole cell lysates - Advantages of a peptide array	- Not necessarily direct interactions captured - Difficulties in detection of low abundance-proteins	(56–58)
ProP-PD	- Information with amino acid resolution - Peptides from human proteome	- Interaction not in context of full-length proteins - Immobilisation strategies - Amount of purified protein (domains)	(63–65)

Phage display has been pioneered by George P. Smith and Sir Gregory P. Winter (66,67), which was awarded with the Nobel Prize in 2018, and is based on the display of proteins or peptides by bacteriophages and subsequent selection for ligands. In peptide phage display, thousands to billions of different peptides are encoded in bacteriophages and presented on the coat of the phage. By different panning strategies, bait protein (domains) can be screened for their binding to the displayed peptides (68). Phage libraries often contain randomised peptides and the retrieved binding motifs are not necessarily biologically relevant. ProP-PD differs in that all peptides are derived from the human proteome, more specifically from the intrinsically disordered regions, which have a higher propensity to harbour binding motifs. ProP-PD has proven to be a suitable technique to determine the binding preferences of bait protein domains to SLiMs (64,65). In comparison to its orthogonal screening methods for motif-based PPIs, ProP-PD offers the advantage of uncovering binding peptides from the human proteome and binding motifs

with amino acid resolution while maintaining a high throughput (**Table 1**, (65)).

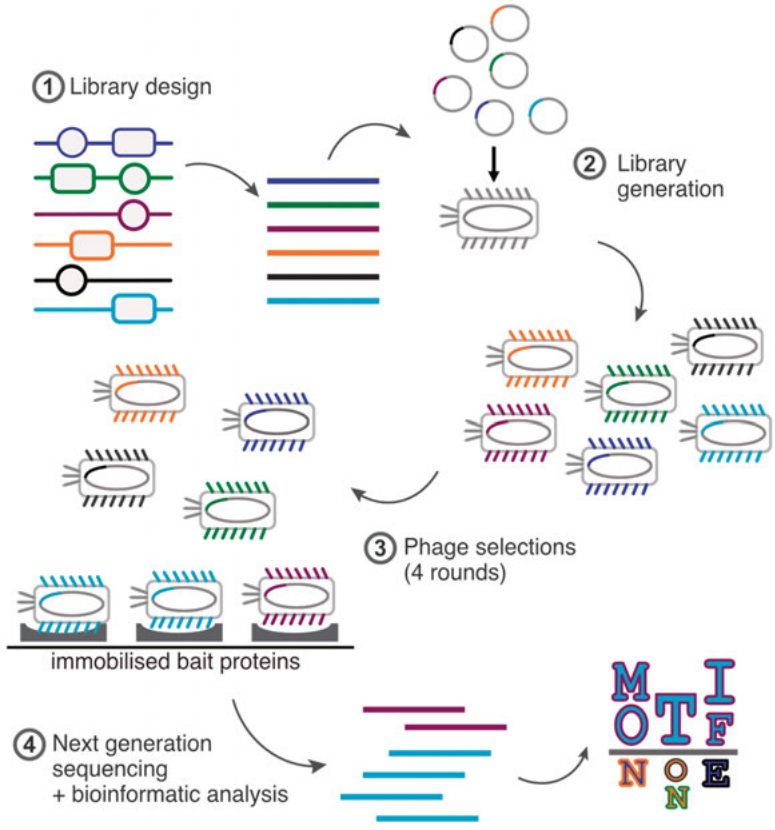

Figure 5: Principle of ProP-PD. The library design of the human disorderome 2 (HD2) library is based on the intrinsically disordered regions of the human proteome. They are tiled in 16 amino acid long peptides with a 12 amino acid overlap. Naïve phagemids are transformed to contain the peptides (1 million different peptides) in fusion with either of the p3 or p8 coat protein of the M13 bacteriophages. In phage selections, bait protein domains are immobilised and challenged with the phage library. Unbound phages are washed off, whereas phages displaying peptides binding to the bait protein domain are amplified in *Escherichia coli*. Enriched phage pools identified by phage pool ELISA are sent for NGS. The bioinformatical pipeline and hit analysis identifies potential binding peptides, binding motifs and interactors.

To briefly describe ProP-PD, M13 bacteriophages are modified to display 16 amino acid long stretches from the human proteome in fusion with either of the coat proteins p8 or p3. The regions are further tiled by overlapping peptides (12 amino acids overlap), which allows for a potential redundancy of the entailed binding motifs. During the phage display experiment, immobilised bait protein domains are challenged to

bind to the presented peptides. Multiple selection rounds (commonly four) allow to enrich for phages displaying peptides specifically binding to the bait protein domain. As a first evaluation, phage pool enzyme-linked immunosorbent assay (ELISA) allows to identify enriched phage pools. Next-generation sequencing (NGS) is subsequently applied to delineate the peptides encoded by the enriched pools. Analysis of the binding peptides provides information on potential and endogenous ligands, as well as binding motifs (**Figure 5**; (65,69)).

Mutational ProP-PD

Conditional PPIs are challenging to analyse in depth and in context of the human interactome due to their fluctuating nature. I focus here on phosphorylation and genetic variation by SNVs as modulators/ regulators of conditional motif-based PPIs.

In order to assess conditional interactions in large-scale, a plethora of different approaches tailored to specific cellular states have been taken, and general strategies include conditional proteomics approaches (46,70) and the cellular thermal shift assay (71). More specific for the investigation of phospho-modulated interactions, peptide arrays and the hold-up assay including phosphorylated amino acids (55,72,73) or conditional yeast two-hybrid screens (74–76) have for example contributed to our current understanding of the phosphorylation code. Similarly, proteomics-based approaches (77,78) including PRISMA (56), computational efforts (45,46), deep-mutational scanning studies (79) and again yeast two-hybrid screens (45,47,80) have been applied to elucidate the effect of disease-associated mutations on PPIs in large-scale.

This thesis promotes the application and contributes to the continued development of mutational ProP-PD to assess conditional motif-based PPIs. The technique is a variation of the conventional ProP-PD approach. The phage libraries in mutational ProP-PD are designed to contain two versions of a given peptide, a wild-type and a mutant version. The approach relies on challenging a given bait protein domain simultaneously with tens of thousands of those wild-type and mutant peptide pairs (**Figure 6**). Binding preferences for either wild-type or mutant peptide are thus extracted in the same experiment and quantified by NGS. Mutational phage display offers the advantage of both suitability for high-throughput and concomitant specificity for delineating conditional SLiM-based PPIs. It was piloted in our lab to decipher the

phospho-modulated binding of PDZ domains. The domains were screened in a phage display using the C-termini of the human proteome, in which all potential serine/threonine phosphosites have been mutated to glutamate to approximate the phosphorylated status of those amino acids (81). I here apply mutational phage display to elucidate both the phospho-modulation of motif-based PPIs (phosphomimetics as mutant peptides) and their disruption by genetic variation (non-synonymous SNVs as basis for the mutant peptides) in large scale.

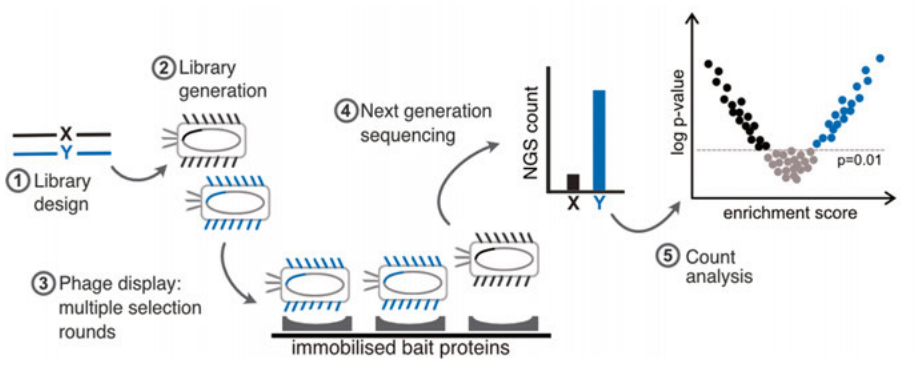

Figure 6: Principle of mutational ProP-PD experiment. The library design is based on the intrinsically disordered regions of the human proteome, together with information on either functionally prioritised phosphosites or disease-associated mutations. In both cases, peptides are represented as wild-type and mutant peptide, so that the immobilised bait protein domains are challenged with both versions of the peptides. NGS of the enriched phage pools identifies the preference of the bait protein domains to either of the versions. This preference is quantified by the enrichment score and statistical power is attributed to NGS count differences between wild-type and mutant peptide by the Mann-Whitney test.

Validation of motif-based PPIs

While ProP-PD is an excellent tool for screening for putative binding partners in high-throughput, it is necessary to validate these interactions both on the peptide level and in context of the full-length proteins. On the peptide level, **affinity measurements**, such as surface plasmon resonance, microscale thermophoresis, isothermal titration calorimetry, stopped flow spectrophotometry and fluorescence polarisation, represent convenient tools to confirm the interaction between the putative motif-containing peptide and the bait protein domain. The two techniques applied in this thesis will be described briefly in the following.

Isothermal titration calorimetry is based on measuring the temperature difference to a reference cell when sequentially adding ligand to protein (or vice versa) to the measurement cell. An interaction feasible to be sampled by this method should result in a heat release or heat consumption upon binding (82–84) (**Figure 7**; here: heat release). The heat corresponds, under constant pressure, to the enthalpy ΔH of the binding event. The instrument applies power to adjust the temperature of the reference cell to the measurement cell, and the resulting power peaks are then integrated and normalised to the amount of added ligand. By fitting the data to an equation suitable for the binding event, the equilibrium association constant (K_A) can be retrieved (82–84).

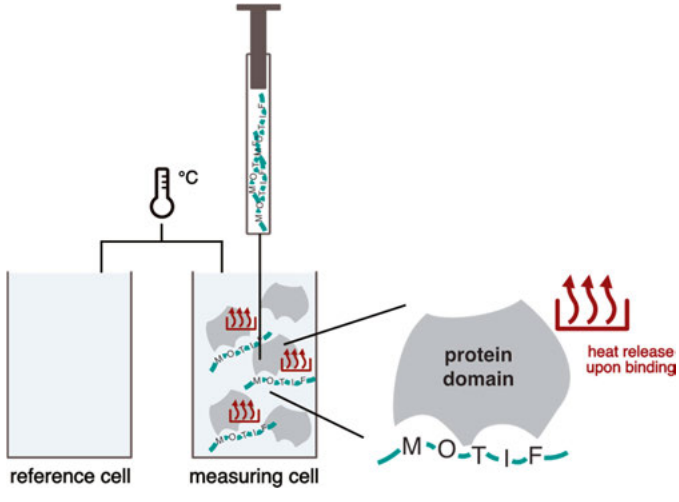

Figure 7: Principle of isothermal titration calorimetry. A peptide ligand is sequentially titrated to the protein domain in the measuring cell. The change in temperature caused by the binding event is registered in comparison to the reference cell. From this, thermodynamic parameters and the K_D-value can be derived.

From this, the thermodynamic parameters of the binding event and the equilibrium dissociation constant (K_D) can be calculated according to the following equation (82,84).

$$\Delta G = -RT * \ln K_A = \Delta H - T\Delta S \qquad \text{Equation 1}$$

$$\text{with}: K_A = \frac{1}{K_D}$$

Isothermal titration calorimetry offers the advantage of label-free measurement of protein binding affinities, thermodynamic parameters and the stoichiometry of binding in free solution.

Fluorescence polarisation is based on measuring the polarisation of light and utilises the changes in the tumbling behaviour of a small fluorescently labelled molecule (here: peptide) upon binding of a larger interaction partner (here: protein domain). The free fluorescently labelled peptide tumbles faster resulting in the depolarisation of light (**Figure 8A**), whereas its rotation is slowed down when bound by the interacting protein domain, maintaining the light polarisation (**Figure 8B**, (85–87)).

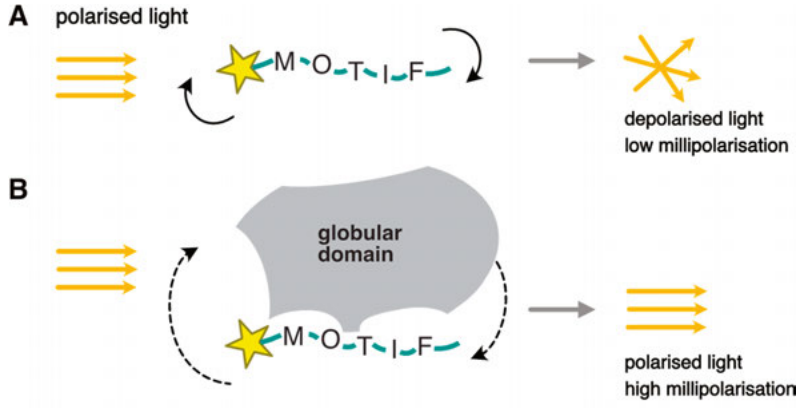

Figure 8: Principle of fluorescence polarisation. Exciting a small fluorescently labelled molecule (here: peptide) with polarised light results in the depolarisation of the emitted light due to the fast tumbling of the molecule. Tumbling of the peptide is slowed down by the binding of the protein domain, so that the polarisation of the light is maintained.

The fluorescence polarisation signal in millipolarisation is calculated by measuring the emitted light of the fluorophore in two angles and according to the following equation:

$$\text{Fluorescence polarisation signal} = F_{\parallel} - F_{\perp} / F_{\perp} + F_{\parallel} \qquad \text{Equation 2}$$

with:
$F_{\parallel}$ = fluorescence intensity parallel to excitation plane
$F_{\perp}$ = fluorescence intensity perpendicular to excitation plane

Both direct (saturation curves) and indirect (displacement curves) binding events can be measured. For saturation curves, the protein domain is titrated against a constant fluorescently labelled peptide. The label used throughout the experiments in this thesis is a fluorescein isothiocyanate (FITC)-label. The K_D-value can be obtained by plotting the fluorescence polarisation signal against the different protein domain

concentration and fitting the data using the quadratic equation for binding (88). For displacement curves, first a complex between protein domain and labelled peptide is formed, and then unlabelled peptides are titrated against this complex. Fitting the data against a sigmoidal dose-response equation allows to extrapolate the IC50-values of the displacement experiments, which can be used to calculate the equilibrium inhibition constant (K_I), which corresponds by extension to the K_D-value of the displacing peptide (89). With this, it is possible to describe the affinities independent of the fluorescent label on the reporting peptide (85). Fluorescence polarisation allows hence the determination of K_D-values in high-throughput while consuming low amounts of protein given that the fluorescently labelled peptide is a high affinity ligand.

Additionally on the peptide level, the interaction can be structurally assessed for instance by X-ray protein crystallography or nuclear magnetic resonance spectroscopy by analysing the complex of protein domain and peptide (90–93). This allows to identify key residues of the interaction on both the protein domain and the peptide.

After confirming the interaction on the peptide level, its validation in context of the full-length proteins is considered the next step in exploring the biological relevance of the probed interaction. **Co-immunoprecipitation** of bait and prey protein in mammalian cells is commonly used for this purpose. Tagging the bait proteins with e.g YFP or GFP (yellow/green fluorescent protein) and expressing them either transiently or stably allows to affinity-purify them from the lysed cells. The co-immunoprecipitation of the prey protein can be analysed by western blotting detecting either tagged- or endogenous prey proteins as well as the successful pull-down of the bait protein (94). An alternative detection method is by mass spectrometry, which is then known as affinity purification mass spectrometry (95). In order to pinpoint the biological role of the interaction, it requires functional read-outs tailored to the respective biological process. Examples explored in this thesis include tracking cells through mitosis in a time-lapse microscopy experiment and determining the nuclear localisation behaviour of a protein by mutating its putative nuclear localisation signal.

Aim of the thesis

Focus of this thesis is on the discovery and validation of SLiMs mediating PPIs, and more specifically on their regulation by phosphorylation and deregulation by genetic variation. As delineated above, motif-based PPIs are challenging to assess due to their commonly moderate affinity and transient nature, and conditional interactions even more so due to their restricted spatial and temporal occurrence.

I aimed with my research to assess motifs involved in orchestrating the phosphorylation code (**Paper I-III**) and investigate how genetic variation impacts motif-based PPIs (**Paper IV**). Initial experiments were based on ProP-PD screens or bioinformatic predictions providing information on putative binding peptides, motifs and their (de-)regulation (**Figure 9**). They form the basis for follow-up experiments. Motif discovery by conventional ProP-PD has been established in the field (65) and I extended its application to kinase domains. I further focussed on the development of mutational ProP-PD as resource to assess conditional motif-based PPIs, regulated by phosphorylation or deregulated by genetic variation.

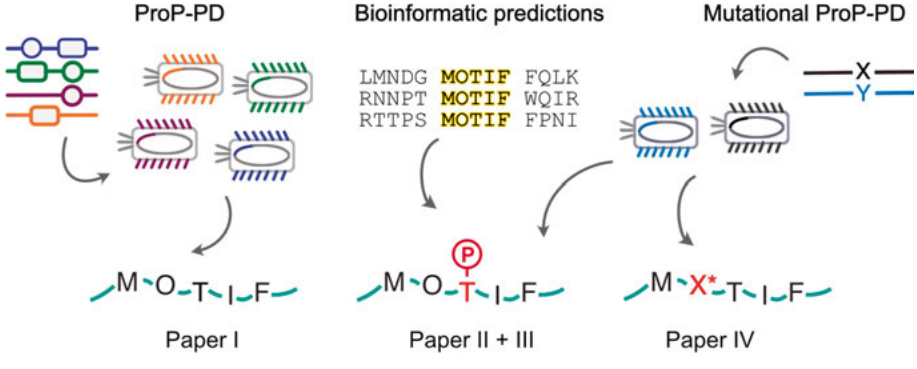

Figure 9: Illustration of the aim of this thesis, which encompasses finding binding motifs of kinase domains by conventional ProP-PD (**Paper I**), as well as assessing phospho-modulated motif-based PPIs (**Paper II+III**) or their deregulation by disease-associated mutations (**Paper IV**).

Summary of findings

The work presented here contributes to the exploration of conditional motif-based PPIs in health and disease. At the core is the phospho-modulation of PPIs, which is addressed from different angles in **Paper I-III**. First, in **Paper I (Letter)**, substrate targeting of kinase domains guided by SLiMs is explored by screening kinase domains in ProP-PD experiments. Secondly, binding motifs found in the cytoplasmic tail of host receptors of the severe acute respiratory corona virus 2 (SARS-CoV-2) are, together with their modulation by phosphorylation, assessed in **Paper II**. Since it has been proven a daunting task to link a given phosphorylation event to its function, we further expanded phosphomimetic ProP-PD as a screening platform to allow for the large-scale identification of phospho-modulated motif-based PPIs (**Paper III**). Phosphomimetic ProP-PD served as a proof of concept-study for the feasibility of mutational ProP-PD, which we extended in the following in **Paper IV** for the assessment of interactome changes caused by genetic variation. This thesis is hence centred on the modulation of motif-based interactions by serine/threonine phosphorylation, and expands to the development of mutational ProP-PD to assess conditional interactome changes.

Paper I (Letter): Exploring peptide binding of kinase domains by proteomic peptide-phage display

Protein phosphorylation is governed by the concerted action of kinases and phosphatases, responsible for the addition or removal of the phosphate moiety on e.g. serine/threonine/tyrosine residues of the substrate protein. The functional consequences of phosphorylation are manifold ranging from protein activation, conformational changes or changes in interaction partners, and phosphorylation orchestrates numerous cellular processes (30,96).

This motivates the tremendous interest in elucidating kinase substrates and strategies evolved by kinases to target them in the cellular pool of possible substrate candidates. Serine/threonine/tyrosine kinases can be subdivided based on their preferred phosphorylation site, which is determined by the central phosphorylated residue and the flanking residues. Recently, a systematic peptide array-based study elucidated phosphorylation site-motifs for more than 300 serine/threonine kinases (55). Beyond the phosphorylation site, substrate specificity can be increased by the kinase engaging in interactions outside the catalytic cleft. These additional interactions are mediated by either distal sites on the kinase domain, distinct domains of the kinase (e.g. SRC Homology 2 (SH2) domains of tyrosine kinases) or separate protein moieties such as the cyclins of the cyclin-dependent kinases (CDKs) (**Figure 10**; (28)).

Figure 10: Substrate interactions of kinases. This includes the phosphorylation site and potentially interactions outside of the catalytical cleft, either on the kinase domain itself or with the help of auxiliary modular domains. Figure modified from (28).

Taking the complexity of the phosphorylation event into account with numerous regulators contributing to its precision, we here set out in a broad fashion to sample kinase domains for their ability to bind peptides. To this end, we screened 87 kinase domains from different kinase groups through ProP-PD selections against the second generation disorderome (HD2) library (**Figure 11A**). The library tiles the disordered regions of the human proteome in 16 amino acid stretches and displays them on the major p8 coat protein of the M13 bacteriophage. The peptide-coding regions of successfully enriched phage pools were barcoded and sent for NGS (67 kinase domains). The NGS results were filtered according to previously established criteria including (i) reoccurrence in replicate selection, (ii) overlapping peptides, (iii) motif matching and (iv) NGS count numbers (65). This ranks the found binding peptides in four confidence levels, among which medium and high confidence peptides (confidence 2-4) are considered in the subsequent analysis. We further increased the specificity filtering (≥ 0.4) and filtered out peptides occurring in only one selection replicate. For the remaining 65

kinase domains in the analysis, we found 1,148 binding peptides corresponding to 1,046 interacting prey proteins (**Figure 11B**).

One main limitation is that the display does not discriminate between the types of interaction the kinase domains can engage in. The binding peptides might hence represent phosphorylation sites, docking sites or even inhibitory peptides. Comparison with the phospho-proteomic data assembled in the Ochoa *et al*. study (97) revealed that for the 457 comparable prey proteins, we found a match for 26 between the annotated kinases in the Ochoa study and our bait kinase. In accordance with the bias of the proteomics data towards CDKs and mitogen-activated kinases (MAPKs), almost all matches were from these kinase families.

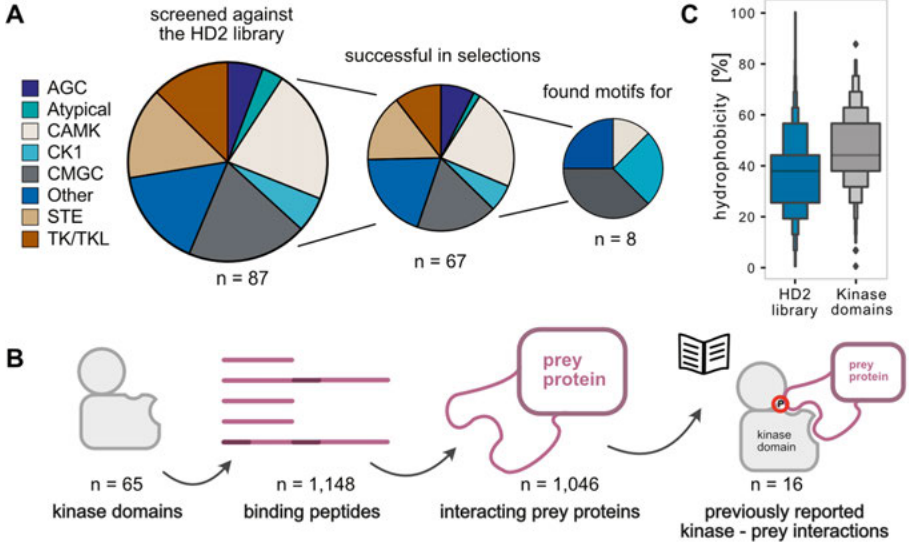

Figure 11: ProP-PD selections with kinase domains. **A:** Overview of the numbers of kinase groups used in and successful in selections against the HD2 library. **B:** Overview of the medium/high-confidence binding peptides from selections with 65 kinase domains. **C:** Comparison of the hydrophobicity in % between the binding peptides from selections with the kinase domains and the naïve HD2 phage library.

The found binding peptides were in average more hydrophobic (47%) than the naïve library (37%) (**Figure 11C**), which is also reflected by the binding motifs established for eight kinase domains, among which five are, to our knowledge, novel. We validated binding peptides and motifs for six kinase domains by fluorescence polarisation experiments, which returned, with a few exceptions, affinities in the high micromolar range. Among the interactions tested, I will highlight in the following the binding of the peripheral plasma membrane protein CASK (CASK) and MAPK8 kinase domain.

There is scarce information on CASK kinase activity and its major function is, as scaffolding protein, to recruit other proteins to the plasma membrane (98,99). Through the ProP-PD selections, we identified the CASK kinase domain to bind to a [ILPV]W-motif (**Figure 12B**), which matches binding sites from previous reports and the motif reported in the ELM database (18,19,100,101). Among the prey proteins were several with shared gene ontology terms, such as Period circadian protein homolog 2 (PER2), as well as reported binders, such as Caskin-1 (CASKIN1) (**Figure 12C**).

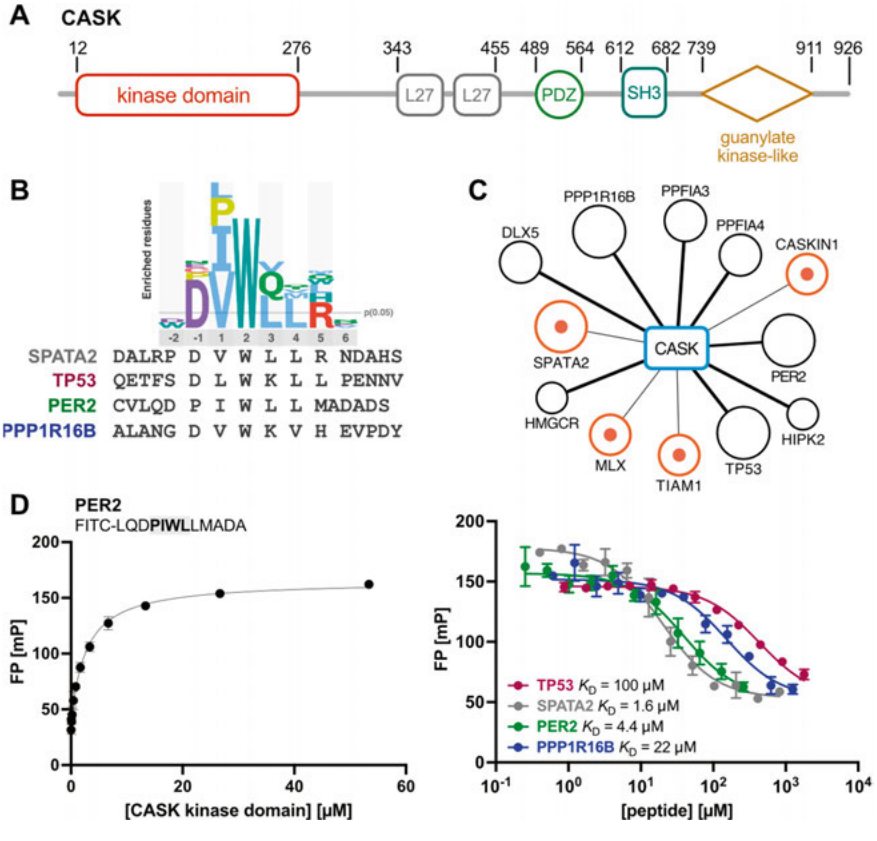

Figure 12: The CASK kinase domain acts as peptide-binding domain. A: Domain structure of CASK. **B:** Consensus motif established by the ProP-PD selections together with aligned peptides used in fluorescence polarisation experiments. **C:** Network of prey proteins with shared gene ontology terms with CASK. Orange with a dot indicates previously reported interactors. Straight lines indicate that the consensus motif was found and the confidence of the binding peptide (2-4) is encoded by the circle size. **D:** Fluorescence polarisation experiments. Left: Saturation experiments with the FITC-labelled $PER2_{1132-1143}$ peptide. Right: Displacement experiments with the $SPATA2_{351-366}$, $TP53_{16-31}$, $PER2_{1130-1145}$ and $PPP1R16B_{438-453}$ peptides.

We measured the affinities for a selected set of peptides and found them to bind within an affinity range of 1.6-100 µM, with the SPATA2$_{351-366}$ peptide (Spermatogenesis-associated protein 2) as the highest affinity binder (**Figure 12B, D**). Our findings support hence the notion that the CASK kinase domain has evolved as dedicated peptide-binding domain contributing to the scaffolding function of CASK. Of note, I performed co-immunoprecipitation experiments between SPATA2 and CASK full-length proteins, which looked promising but the results were not clean enough in comparison to the negative control to draw definite conclusions.

MAPK8 is activated in response to environmental stimuli and cytokines and acts as an important regulator of signal transduction (102,103). Of all kinase domains, docking interactions of the MAPK kinase domains are characterised in most detail and described as D-motifs in the literature (28). The ProP-PD selections established a matching RPxLx[LI]-motif for the MAPK8 kinase domain (**Figure 13B**). We found known interactors, such as WD repeat-containing protein 62 (WDR62), and also prey proteins annotated as MAPK substrates in the phospho-proteomics dataset assembled by Ochoa *et al.* (97), e.g. Eukaryotic translation initiation factor 2-alpha kinase 3 (EIF2AK3). For the affinity measurements, we selected peptides from WDR62 (WDR62$_{1282-1297}$) and EIF2AK3 (EIF2AK3$_{850-868}$) for the displacement experiments and found them to bind with affinities in the low µM range, < 0.2 µM and 3.2 µM respectively (**Figure 13D**). In contrast, mutation of the motif-critical residues in the EIF2AK3$_{850-868}$ peptide reduced the affinity to > 100 µM (EIF2AK3_dockmut; **Figure 13D**).

We next asked whether the EIF2AK3$_{850-868}$ peptide presents a substrate of MAPK8, and if so whether S856 is the site being phosphorylated. For assessing MAPK8 kinase activity, we used a commercial kinase assay relying on the coupled activity of a phosphatase on the released ADP and concomitant detection of the free phosphate. We used different versions of the EIF2AK3$_{850-868}$ peptide, including the S856A mutant (EIF2AK3_S856A) and the docking mutant (EIF2AK3_dockmut), as substrates at 0.9 mM concentration and found for all of them weak phosphorylation activity (~26 pmol/min/µg, **Figure 13E**). This was surprising for the EIF2AK3_dockmut peptide due to the weak affinity found in the fluorescence polarisation experiment, but we hypothesise that there is sufficient binding at 0.9 mM peptide concentration for the phosphorylation reaction to occur. We also concluded that the S856 is not the relevant or exclusive phosphosite in the EIF2AK3$_{850-868}$

peptide, since we detected no difference in phosphorylation activity for the EIF2AK3_S856A peptide.

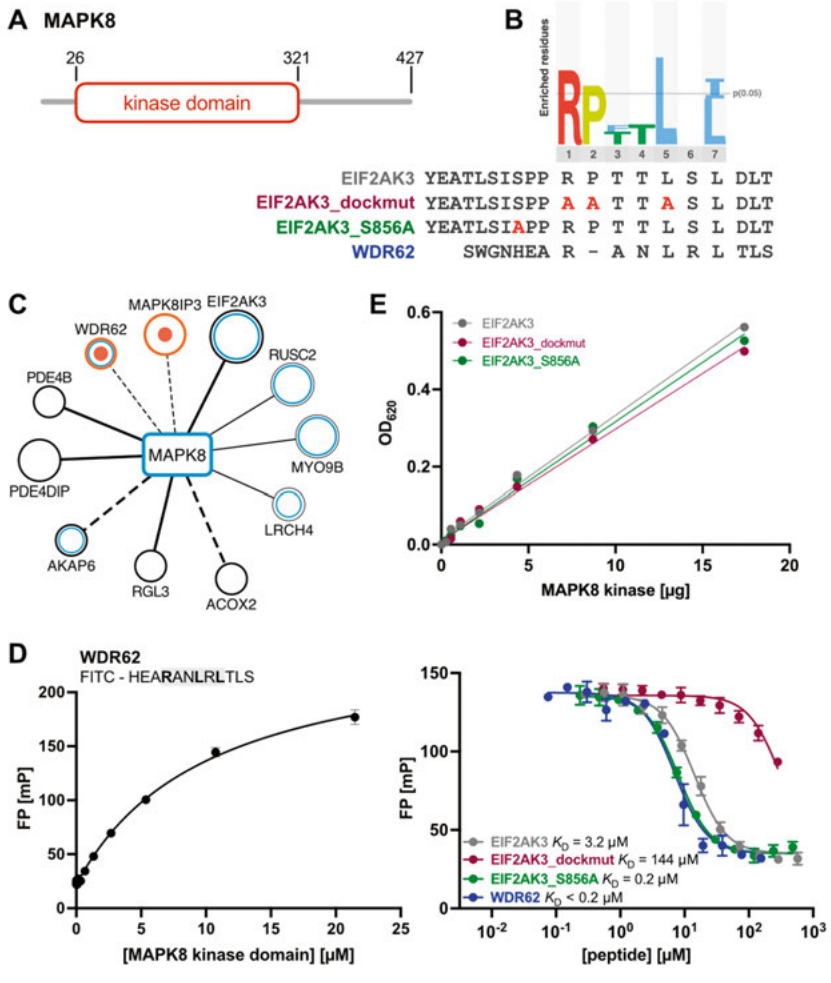

Figure 13: ProP-PD retrieves a classical D-motif for MAPK8. A: Domain structure of MAPK8. **B:** Consensus motif established by the ProP-PD selections and aligned peptides used in fluorescence polarisation experiments. **C:** Network of prey proteins with shared gene ontology terms with MAPK8. Orange with a dot indicates previously reported interactors and a blue circle indicates that the prey protein is a MAPK-substrate in the Ochoa data set (97). Straight lines indicate that the consensus motif was found and the confidence (2-4) is encoded by the circle size. **D:** Fluorescence polarisation experiments. Left: Saturation experiments with the FITC-labelled $WDR62_{1286-1297}$ peptide. Right: Displacement experiments with the $WDR62_{1282-1297}$ and $EIF2AK3_{850-868}$ peptides. **E:** MAPK8 phosphorylation activity for the different $EIF2AK3_{850-868}$ peptides. Read-out was at 620 nm measuring the malachite-phosphate complex formation, formed with free phosphate released from the coupled phosphatase activity on ADP.

In conclusion, we here sampled the peptide-binding properties of 87 kinase domains through a ProP-PD experiment and judge the result as heterogenous among the kinase domains. The increased hydrophobicity of the binding peptides for the kinase domains evokes caution since it might be associated with non-specific binding. Nevertheless, we demonstrate the validity of the hydrophobic interactions of a selected set of kinase domains by affinity measurements. We hypothesise that the ProP-PD selections return in large potential docking interactions due to the hydrophobicity and the high micromolar affinity range obtained. It can be envisioned that spot arrays could be deployed to sample the binding of the kinase domains to our binding peptides more systematically. Besides, it is pivotal to validate the interactions and the importance of the motif for it in the context of the full-length proteins.

Of note, this letter on kinase domains will be part of a larger study, titled "Towards the human motif map", which samples the peptide binding properties of various domain families. This will include WW or SH3 domains, which have evolved as dedicated peptide binders. In comparison, the ProP-PD results here suggest that while some kinase domains clearly contribute to substrate/ interactor targeting, such as the CASK and MAPK8 kinase domains, other kinase may in turn rely on alternative strategies.

Paper II: Phospho-modulation of the binding motifs harboured by the cytoplasmic tails of SARS-CoV-2 host receptors

SARS-CoV-2 has, as the biological agent of the coronavirus disease 19, been extensively studied within the recent years. Much effort has, up to now, been invested in understanding its biology and infection mechanism. Mészaros *et al.* set out in their bioinformatic study to investigate potential SLiMs harboured by the C-termini of the host cell receptor, the angiotensin converting enzyme 2 (ACE2), and the (putative) receptor integrin β3 (ITGB3). The aim was to shed light on binding/signalling events downstream of the receptor binding in physiological conditions and potentially during SARS-CoV-2 infection (104). Our contribution to this effort, as represented by **Paper II**, was to experimentally test the binding of the bioinformatically predicted motifs in the C-termini of ACE2 and ITGB3 to their respective binding domains.

To this end, we purified protein domains and measured their affinity for short synthetic peptides derived from ACE2 or ITGB3 by

fluorescence polarisation measurements. Particular focus was on the phospho-modulation of binding. Among the motifs tested was an overlapping motif for both AP2M1 and SH2 domains in ACE2, which was proposed to be regulated by Y781 phosphorylation (**Figure 14A**, in green), as well as a phospho-modulated LIR motif for the ATG8 proteins in ITGB3 (**Figure 14B,** in blue).

Figure 14: C-termini of ACE2 (A) and ITGB3 (B) with the tested peptides harbouring the different putative binding motifs highlighted by colour. **Green**: overlapping motif for AP2M1, SH2 domains and ATG8 proteins. **Purple**: FERM (4.1 (F) protein, ezrin, radixin and moesin) domain binding motif. **Orange**: C-terminal class I PDZ binding motif. **Blue**: LIR motif for ATG8 proteins. Figure modified from (105).

Both direct binding (saturation) and displacement experiments were performed throughout the study. First, for the unphosphorylated $ACE2_{775–790}$ C-terminus, we could demonstrate the binding of AP2M1 with a K_D-value of 100 µM, which is within the range of previous reports (106,107). In accordance with its cognate binding motif (YxxΦ), phosphorylation of Y781 in the $ACE2_{775–790}$ peptide abolished the binding of the domain (**Figure 15A**). Since, we could not detect any interaction of the NCK Adaptor Protein 1 (NCK1) SH2 domain with the $ACE2_{775–790}$ peptide tested (**Figure 15A**), we next sampled the binding of the Tyrosine-protein kinase Fyn (FYN) SH2 domain to the FITC-labelled versions of the $ACE2_{775–790}$ C-terminus (**Figure 15B**). The FYN SH2 domain bound to the pY781 $ACE2_{775–790}$ C-terminus with 7 µM affinity, but showed no binding to the unphosphorylated or pS783 peptide (**Figure 15B**). Due to the low solubility of AP2M1, we could not obtain binding curves of good quality for the FITC-labelled $ACE2_{775–790}$ peptides. Nonetheless, it shows the preference of the domain for the unphosphorylated $ACE2_{775–790}$ peptide over the pY781 ACE2 peptide (**Figure 15B**). Taken together, we confirm here an overlapping binding site in the $ACE2_{775–790}$ C-terminus for AP2M1 and the FYN SH2 domain, which is regulated by Y781 phosphorylation.

As part of the clathrin adaptor AP2 (108), AP2M1 could act to recruit the endocytic machinery upon virus infection by binding to the $ACE2_{775–790}$ C-terminus, whereas FYN, as tyrosine kinase involved in cytoskeletal rearrangements (109), can be envisioned to facilitate viral entry by remodelling the actin cytoskeleton.

Secondly, we explored the phospho-modulation of the putative LIR motif in the C-terminus of ITGB3 (**Figure 15C**). To this end, we tested the binding of the ATG8 proteins MAP1LC3A, -B, -C and GABARAPL1 and L2 to peptides from ITGB3 ($ITGB3_{774-787}$) sampling phosphorylation of several residues, including T777, S787, T779 and Y785. The ATG8 proteins did not bind the unphosphorylated $ITGB3_{774-787}$ peptide (**Figure 15C**) but the affinity was considerably enhanced by phosphorylation on Y785 or S778. The double phosphorylation at T779 and Y785 had the most pronounced effect on binding and MAP1LC3A and -B bound the pT779/pY785 $ITGB3_{774-787}$ peptide with 8.5 μM and 15 μM affinity respectively. We could hence conclusively show that phosphorylation of the $ITGB3_{774-787}$ peptide enhances ATG8 protein binding and elucidate on the synergistic effect of the T779 and Y785 phosphosites. By validating the phospho-dependent LIR motif in the ITGB3 tail, we demonstrate a potential direct link between ITGB3, integrin-mediated signalling and the autophagic pathway.

The study sheds light on the phospho-modulation of binding which governs the interaction of the ACE2 and ITGB3 cytoplasmic tails with possible downstream binders. Several layers of evidence, such as the interactions in the cellular setting and activation of the different downstream pathways, are however required to show the biological relevance of the binding events sampled in our study as well as their contribution to viral infection. For instance, a following study showed in an animal model and in cell culture that SARS-CoV-2 infection down-regulates ACE2 by endocytosis dependent on clathrin and AP2 (110). This supports our finding of AP2M1 binding to the ACE2 C-terminus. Another follow-up study focused on the PDZ-binding motif in ACE2, which we found with high affinity to bind to sorting nexin 27 (SNX27), and the SNX27-retromer was shown to route the ACE2/SARS-CoV-2 complex to the early endosome preventing virus cell entry (111,112). In contrast, it has also been reported that SARS-CoV-2 entry does not essentially rely on the ACE2 C-terminus in HEK293T cells (113). I came to believe that SARS-CoV-2 research has been moving tremendously fast in the past years and it will consequently take considerable time to put all the

puzzle pieces together and determine the relevant mechanistic details of its infection route.

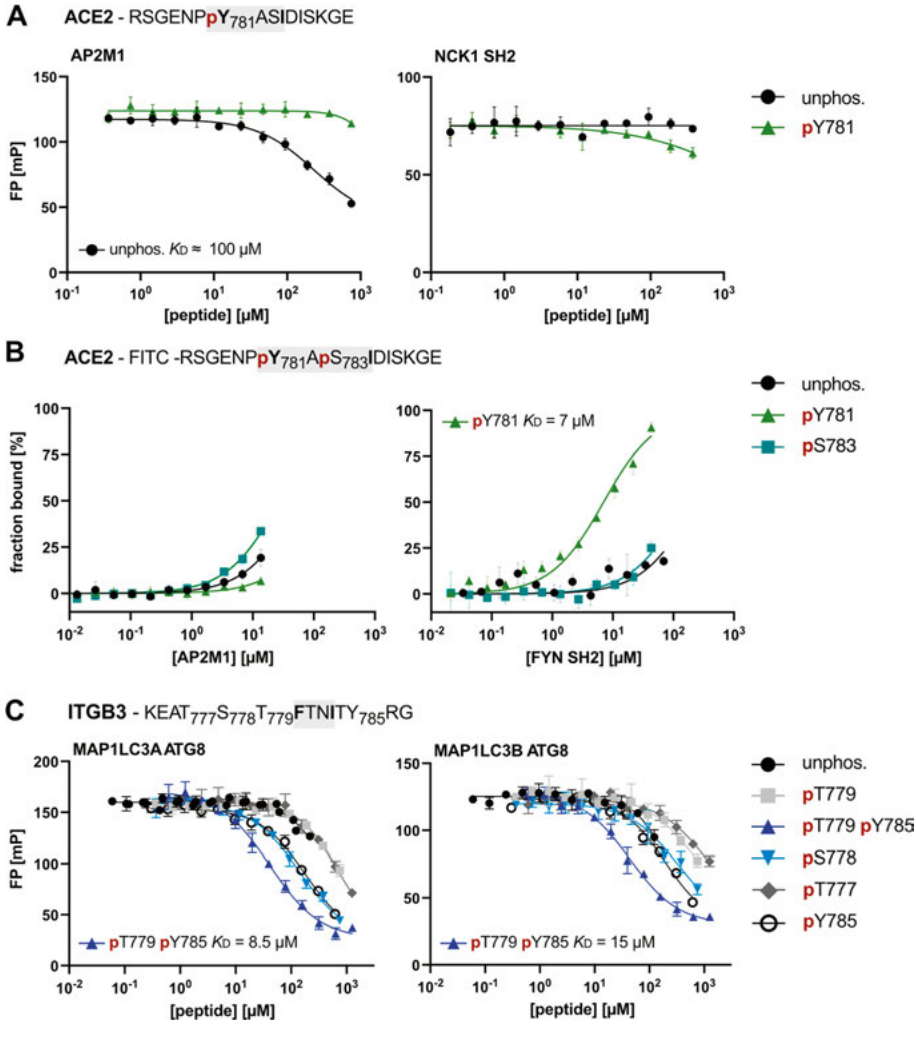

Figure 15: Fluorescence polarisation curves assessing the binding of protein domains to the C-termini of ACE2 and ITGB3. A: Displacements curves of AP2M1 and NCK1 SH2 domain with unphosphorylated and pY781 $ACE2_{775–790}$ peptides. **B:** Saturation curves of AP2M1 and FYN SH2 domain with FITC-labelled pY781, pS783 and unphosphorylated $ACE2_{775–790}$ peptides. **C:** Displacements curves of MAP1LC3A and -B ATG8 proteins with unphosphorylated, pT779, pT779pY785, pS778, pT777 and pY785 $ITGB3_{774-787}$ peptides. Figure modified from (105).

Paper III: Large-scale phosphomimetic screening identifies phospho-modulated motif-based protein interactions

Phosphorylation can, as elaborated above, impact motif-based PPIs by acting as binding switch or by fine-tuning affinities between interaction partners (28). Mass spectrometry-based approaches have identified hundreds of thousands of phosphosites (114) and the challenge at hand is to attribute functionality to them, since a majority of the follow-up methods are low-throughput. In order to elucidate the phospho-modulation of motif-based PPIs, the Ivarsson lab has previously developed phosphomimetic ProP-PD (81). The approach is based on the approximation of phosphorylated serine and threonine by glutamate for a given phosphosite, so that every peptide in the phage library exists as wild-type and phosphomimetic version. From the NGS results, it is then possible to establish binding preferences of the bait protein domains for either the wild-type or phosphomimetic peptide. This pinpoints the putative phospho-modulation of the interaction and simultaneously possible interactors. In their initial screen, they used a small C-terminal library and investigated the phospho-modulated binding of PDZ domains (81).

Here, we present a novel design of the phosphomimetic phage library to identify putative phosphosites impacting interactions of a variety of different bait protein domains in large-scale. For the design, we combined the functionally prioritised phosphosites, defined by Ochoa *et al.,* with the unstructured regions of the human proteome, as defined by the HD2 library (**Figure 16A**; (65,97)). The phage peptidome is displayed in fusion with the major coat protein p8 of the M13 bacteriophage. The PM_HD2 library covers 13,539 phosphosites from 6,246 prey proteins, of which 65% and 0.5% are associated with a kinase or phosphatase, respectively (**Figure 16B**). In comparison to the previous C-terminal phosphomimetic phage library, the PM_HD2 library offers the possibility of screening a variety of different protein domains. Among the domains tested were obligate phospho-binders, such as the 14-3-3 domains, known binders of the HD2 library and domains for which the phospho-modulation of binding has previously been reported (**Figure 16C**). Of the in total 71 bait protein domains, selections of 54 were successful as judged by phage pool ELISA. The phage pools were PCR-barcoded and the amplicons sent for NGS. The obligate phospho-binders notably failed in the selections (six out of ten failed), showing either no enrichment of the phage pools or returning, when sent for NGS, low-

confidence peptides. This is in line with the strict requirement of those domains for the phosphate moiety, and prompted us to focus on the bait protein domains, the binding of which may be modulated by but is not strictly dependent on phosphorylation. The NGS results were analysed for in total 54 bait protein domains (**Figure 16C**) and peptides were filtered for medium- and high-confidence peptides according to the prioritisation criteria established for the HD2 library (65) and described in **Paper I**. In total, we found, for the 54 bait protein domains, 895 interacting peptides, which corresponded to 537 interactors and 884 PPIs, of which 144 were previously reported (71 by selections with the HD2 library).

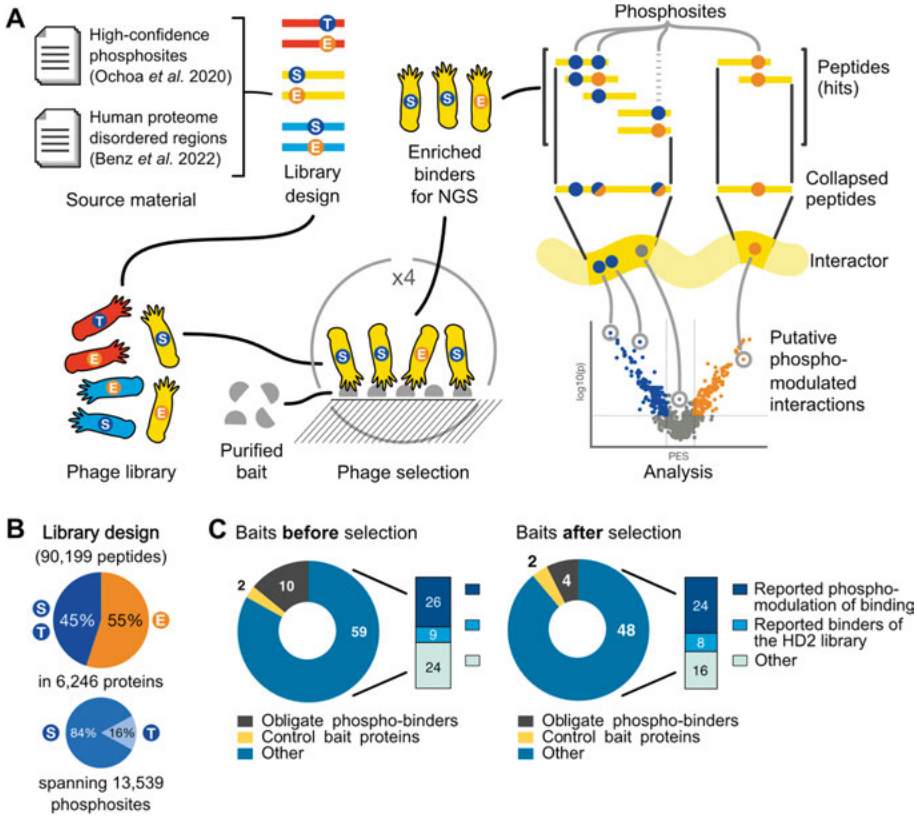

Figure 16: Phage selections with the PM_HD2 library. A: Outline of the design of the library, its generation, use in selection and analysis of the NGS results, which encompassed mapping of potential interactors and evaluating the effect of the phosphomimetic mutation. **B:** Parameters of the PM_HD2 library. **C:** Bait protein domain collection screened against the PM_HD2 library (left) and sent for NGS (right). Figure modified from (115).

We further improved the NGS analysis pipeline of the phosphomimetic phage display. Overlapping peptides reporting on the same phosphosite were, to this end, collapsed, so that the mean effect of the phosphomimetic mutation at a given position is considered (**Figure 16A**). In order to evaluate the effect of the phosphomimetic mutation for the collapsed domain-phosphosite pairs, we calculated the (phosphomimetic) enrichment score, as indicated by the following equation, which is based on the NGS counts of wild-type and phosphomimetic peptides.

$$\text{Enrichment score} = \sum_{i=0}^{n} \frac{nc^{i}{}_{pm}}{nc^{i}{}_{wt} + nc^{i}{}_{pm}} - \sum_{i=0}^{n} \frac{nc^{i}{}_{wt}}{nc^{i}{}_{wt} + nc^{i}{}_{pm}} \quad \text{Equation 3}$$

with:
nc^{i} = normalised sequencing count of wild-type
nc^{i} = normalised sequencing count of phosphomimetic

In addition, a Mann-Whitney test was performed to attribute statistical power to the differences in the NGS counts between wild-type and mutant peptide pairs. Together, those evaluation metrics give rise to a V-shaped plot, in which the left arm (blue) represents cases in which the phosphomimetic mutation is disruptive, and the right arm (orange) cases in which it enhances the interaction. For defining a high confidence dataset of phosphosites impacting binding, we chose an absolute phosphomimetic enrichment score cut-off ≥ 2, as well as a p-value ≤ 0.01 or ≤ 0.001 for the Mann-Whitney test. We found 248 domain-phosphosite pairs (for $p \leq 0.01$; for $p \leq 0.001$: 86 interactions) to be significantly modulated by the phosphomimetic mutation, with an equal contribution of enabling and disabling effects (**Figure 17A**).

Next, we sought to validate the binding preferences established in the display by affinity measurements using fluorescence polarisation. We sampled in total 32 interactions of 11 protein domains, covering a phosphomimetic enrichment score range from -15 to 14 and measuring the binding of the respective wild-type, phosphomimetic and phosphorylated peptides in displacement experiments. When correlating the fold-change in affinity of the wild-type/phosphomimetic or wild-type/phosphorylated peptides, we observed a near two-fold difference indicating that the phosphorylated peptide confers a more pronounced effect on binding in comparison to the phosphomimetic peptide (**Figure 17B**). This was true for interactions enabled or disabled by phosphorylation, and might represent the differences in net charge or molecular geometry between glutamate and the phosphate group. This is exemplified by the interaction between MAP1LC3A ATG8 and the TPD52L$_{91\text{-}106}$ peptides,

where the S96 peptide bound with 20 µM, the phosphomimetic with 78 µM and the wild-type with > 200 µM affinity (**Figure 17C**). One exception to this observation is on the other hand the binding of SNX27 PDZ domain to the $DAXX_{606-616}$ peptide, for which the phosphomimetic mutation promoted the affinity more than the phosphorylated peptide (**Figure 17C**).

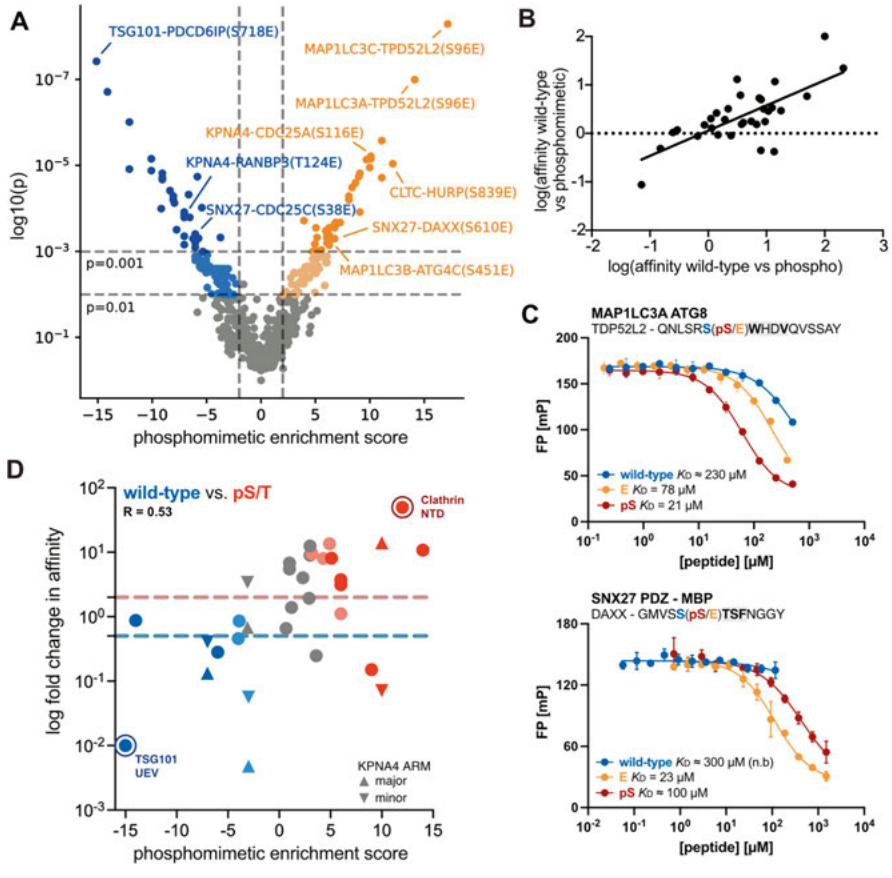

Figure 17: Phosphomimetic phage display results and validation at the peptide level. A: V-shaped plot of the domain-phosphosite pairs with their phosphomimetic enrichment score and plotted against the p-value of the Mann-Whitney test. Blue indicates that the bait protein domain prefers the wild-type and orange that it prefers the phosphomimetic peptide in the phage display. **B:** Correlation of the log-fold change in affinity between wild-type/phosphomimetic and wild-type/phospho-peptides as measured by fluorescence polarisation. **C:** Examples of displacement experiments, in which phosphorylation had a more pronounced effect (MAP1LC3A ATG8) or less pronounced effect (SNX27 PDZ) on binding than the phosphomimetic mutation. **D:** Phosphomimetic enrichment score of the domain-phosphosite pairs plotted against the log-fold change in affinity between wild-type and phospho-peptides (red: interactions indicated as enabled in the display and blue: disrupted interactions; darker shade: p-value ≤ 0.001, lighter shade: p-value ≤ 0.01). Figure modified from (115).

For evaluating the performance of the phosphomimetic phage display in indicating the impact of phosphorylation on binding, we correlated the interactions with an absolute phosphomimetic enrichment score cut-off ≥ 2, as well as a p-value ≤ 0.01 or ≤ 0.001, to the fold-change in affinity of the wild-type/phosphorylated peptides (**Figure 17D**). For 14 out of the 18 interactions (78%) within the chosen cut-off values (p-value ≤ 0.01; for p-value ≤ 0.001: 9 out of 11 interactions), we found the phage display to correctly reflect preferences for either wild-type or phosphorylated peptide, as judged by a, at least, two-fold difference in affinity (**Figure 17D**). I consider a 78% positive rate as fair given the approximation of phosphorylation by phosphomimetics in the phage library.

In the following, we focused on the binding of the clathrin NTD to a C-terminal peptide of $HURP_{832-846}$ (Hepatoma-upregulated protein), which was indicated by the display to be enhanced by the phosphomimetic mutation (and phosphorylation) at S839. We confirmed the binding preference by fluorescence polarisation and isothermal titration calorimetry, which found the clathrin NTD to bind the pS839 $HURP_{832-846}$ peptide with ~ 5 μM K_D, whereas the wild-type (unphosphorylated) peptide bound with > 100 μM K_D (**Figure 18A, B**). Co-crystallisation of the domain with pS839 $HURP_{826-846}$ peptide elucidated the preference structurally, as the phosphate moiety is clamped between two basic residues (R64 and K96) in the binding pocket 1 on the clathrin NTD (**Figure 18C**).

Since the predicted kinase for this interaction is CDK1 and both clathrin and HURP have reported roles in mitosis (116,117), we turned to assess the interaction of the full-length proteins in asynchronous and mitotic HeLa cells. For this purpose, cell lines inducibly expressing YFP-tagged wild-type and C-terminal truncated (Δmut) HURP were generated and the interaction with endogenous clathrin probed in a GFP-trap experiment. First, we could demonstrate that the interaction between clathrin and wild-type HURP occurs only in mitotic and not in asynchronous HeLa cells (**Figure 18D**). Secondly, truncation of the HURP C-terminus, which harbours the clathrin binding motif, abolishes the interaction (**Figure 18D**).

In order to show the dependency of clathrin binding on HURP phosphorylation, we generated the S839A HURP mutant and probed again for co-immunoprecipitation of clathrin in mitosis. Similar to Δmut HURP, S839A HURP did not pull-down clathrin pinpointing the importance of the phosphosite for the interaction (**Figure 18E**).

Moreover, we assessed the behaviour of the different HURP constructs during mitotic progression by an RNA interference experiment (**Figure 18F**). It has been reported that HURP knockdown results in mitotic delay (117), i.e. in prolonged time from nuclear envelope breakdown to anaphase, which we could reproduce in our experiment (median time control: 45 min and median time knockdown: 69 min). Re-expression of wild-type HURP in a background of endogenous HURP knockdown rescued the phenotype (median time: 55 min), whereas re-expression of Δmut or S839A HURP maintained the delay of the knockdown (median time Δmut: 71 min and S839A: 64 min). We hence conclusively demonstrated that clathrin binding and S839 HURP phosphorylation are required for the mitotic function of HURP (**Figure 18G**).

Overall, we demonstrated that the binding preference indicated by the phosphomimetic ProP-PD screen were in good agreement with the affinity measurements on the peptide level and, as showcased with the clathrin-HURP interaction, in context of the full-length proteins. The PM_HD2 phage library presents hence a convenient resource to screen for the putative phospho-modulation of binding from various bait protein domains against thousands of different phosphosites. As with all phosphomimetic experiments, follow-up experiments including actual phosphorylation (here: phosphorylated peptide) are required for validation. Nonetheless, phosphomimetic ProP-PD offers the advantage of a sensible library design and feasibility for high-throughput, which allows to identify phosphosites impacting motif-based PPIs and facilitates to contextualise phosphosite functionality.

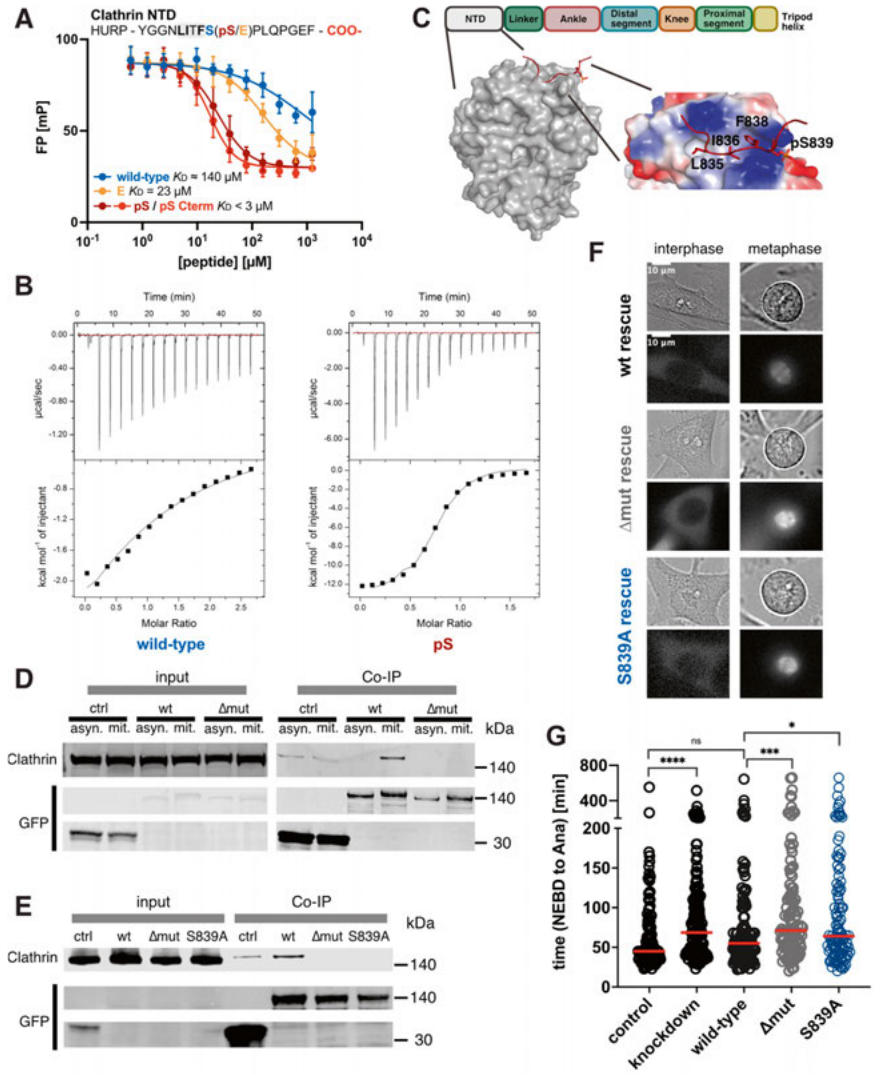

Figure 18: Importance of S839 HURP phosphorylation for its interaction with clathrin and mitotic function. A: Displacement experiments of clathrin NTD with wild-type, phosphomimetic, phosphorylated and C-terminal phosphorylated $HURP_{832-846}$ peptides. **B:** Isothermal titration curves of clathrin NTD with wild-type and phosphorylated $HURP_{832-846}$ peptides. **C:** Domain structure of clathrin and crystal structure of clathrin NTD with the phosphorylated $HURP_{826-846}$ peptide bound to the binding pocket 1. Close view of the charge representation of binding pocket 1. **D:** Co-immunoprecipitation of YFP only, YFP-tagged wild-type or Δmut HURP with endogenous clathrin in asynchronous and mitotic HeLa cells. **E:** Co-immunoprecipitation of YFP only, YFP-tagged wild-type, Δmut or S839A HURP with endogenous clathrin in mitotic HeLa cells. **F:** Representative images of the time-lapse microscopy of the RNA interference experiment knocking down HURP and re-expressing wild-type, Δmut or S839A HURP. **G:** Single cell representation of the time required to progress through mitosis for control, knockdown, wild-type, Δmut or S839A HURP (n= 2). NEBD: nuclear envelope breakdown, Ana: anaphase. Figure modified from (115).

Paper IV: Mutational proteomic peptide-phage display elucidates genetic variation impacting protein binding

High-throughput sequencing approaches have identified a plethora of missense mutations in the human genome and bridging the gap between genomic information and consequences on the protein level presents as bottleneck for the prioritisation of the mutation data (34,36). An emerging theme among the functional consequences is the perturbation of PPIs and the networks they form (42–45,47). In line and in context of motif-based PPIs, approximately 22% of disease-associated mutations map to intrinsically disordered regions and occur within SLiMs more frequently than neutral mutations (118,119).

We here set out to explore the impact of genetic variation as represented by non-synonymous SNVs on motif-based PPIs by using our mutational ProP-PD approach (**Figure 19A**). The genetic variation (GenVar_) HD2 library combines the unstructured regions of the human proteome, as defined by the HD2 library (65), with missense mutations reported across different databases (e.g. COSMIC and ExAC). The phage peptidome is displayed in fusion with the major coat protein p8 of the M13 bacteriophage. The total 12,301 mutations from 1,915 prey proteins are represented as 36,479 pairs of wild-type and mutant peptide pairs in the GenVar_HD2, where multiple pairs cover individual mutations (**Figure 19B**). We categorised the diseases associated with the mutations in eleven disease categories, such as neurological, metabolic and musculoskeletal diseases, as well as cancer. Of note, a considerable part of the diseases was classified into more than one category (mixed category, e.g. metabolic & cancer) (**Figure 19C)**.

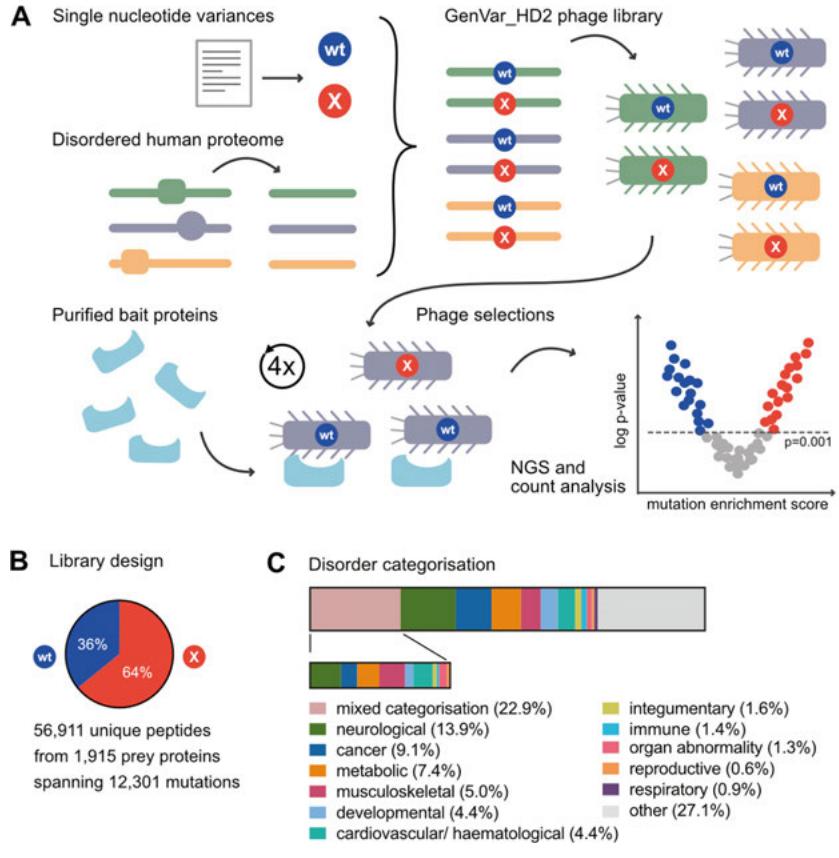

Figure 19: Design and phage selections with the GenVar_HD2 library. **A:** Outline of the design of the library, its generation, use in selection and analysis of the NGS results. **B:** Parameters of the GenVar_HD2 selections. **C:** Categorisation of the diseases underlaying the different mutations. In the mixed category, mutations belong to more than one disease category and their distribution is detailed in the lower panel.

After initial phage selection rounds with 91 bait protein domains, we selected 80 domains based on phage pool ELISA results for selection in 5-6 replica against the GenVar_HD2 library to gain statistical confidence in the following NGS analysis. Binding peptides were ranked according to the established criteria described above (medium/high-confidence ligands; confidence 2-4; **Paper I**) and we found for the 80 bait protein domains (+ 2 control proteins) 2,223 domain-peptide interactions, which map to 1,229 PPIs, of which 106 have been previously reported. The binding peptides were in the following collapsed on the mutation site, according to the analysis pipeline of the PM_HD2 selections (**Paper III**) and in order to infer the average effect of a given mutation on binding across the different wild-type/mutant peptide pairs. Further, the mutation enrichment score was calculated in the same

fashion as the phosphomimetic enrichment score (see Equation 3) and significant differences in the NGS counts of wild-type/mutant peptide pairs were derived based on the Mann-Whitney test. For our bait protein domain collection, this gave rise to 1,787 domain-mutation pairs of which we considered 367 to be significantly modulated by the mutation ($p \leq 0.001$ of the Mann-Whitney test), with half indicated as enhanced and half as diminished by the mutation. This is visualised by plotting the mutation enrichment scores of the domain-mutation pairs against the p-value of the Mann-Whitney test, which creates a V-shaped plot (**Figure 20A**). The left arm in blue represents the interaction pairs where the mutation diminishes/breaks the interaction, and the right arm in red the cases in which the mutation enhances or creates the interaction.

In order to determine whether the binding preferences indicated by phage selection with the GenVar_HD2 ($p \leq 0.001$ of the Mann-Whitney test) translate to differences in affinity, we tested the binding of 24 wild-type and mutant peptide pairs to their respective bait protein domain (15 domains in total) by fluorescence polarisation experiments. We found 19 out of the 24 cases (79%) to be in accordance with preferences of the phage selections, as judged by at least two-fold change in affinity (**Figure 20B**). This compares well with the results of the PM_HD2 library and confirms mutational ProP-PD as appropriate screening method to identify binding difference on the peptide level.

We further noticed that contextualising the mutation to the binding motif of the bait protein domain facilitates our data prioritisation. We found that 13 out of 15 of the domain-mutation pairs (87%), where the mutation occurs within key residues and which were tested in affinity measurements, were positively correlated with the preferences of the phage display. Based on this observation, we decided to map the motif instances of the bait protein domains to the binding peptides of the display and subsequently mapped the mutation position in relation to the motif. This resulted in four mutation categories for the cases, in which a motif instance was found: mutations occurring in key residues, in wild-card positions or in flanking regions, as well as mutations creating the motif instance. Based on this, we defined a high-confidence set of mutation-modulated interaction pairs with the mutation occurring in key residues of the motif instance or creating the motif instance (99 out of 367 domain-mutation pairs).

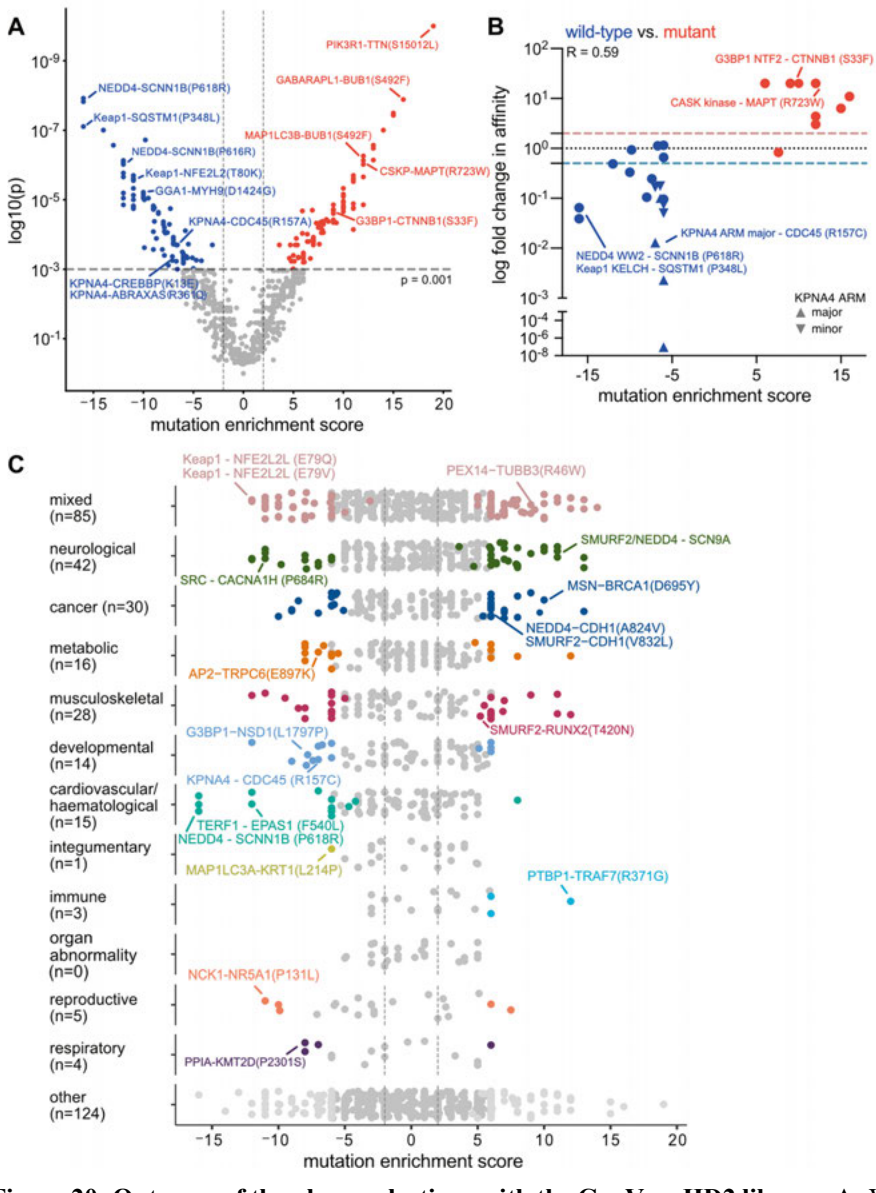

Figure 20: Outcome of the phage selections with the GenVar_HD2 library. A: V-shaped plot of the domain-mutation pairs with their mutation enrichment score plotted against the p-value of the Mann-Whitney test performed on the NGS counts of collapsed wild-type and mutant peptide pairs. Red indicates that the mutation significantly enhances and blue that it significantly diminishes the interaction. **B:** Correlation of affinity measurements (fluorescence polarisation) with the preferences of the phage selections for wild-type and mutant peptides as indicated by the mutation enrichment score and the associated p-value of the Mann-Whitney test. **C:** Domain-mutation pairs with their mutation enrichment score with the mutations mapped back to the different disease categories. Colour-coding is by the p-value of the Mann-Whitney test and disease category (grey: $p > 0.001$; colour of disease category: $p \leq 0.001$).

Finally, we mapped back the diseases associated with the different mutations, which were found to modulate interactions with our bait protein domains. For the domain-mutation pairs modulated by the mutation and within our disease categorisation, we found 42 pairs linked to neurological diseases, 30 linked to cancer, 28 linked to musculoskeletal diseases and 84 within the mixed category (**Figure 20C**).

Due to the heterogeneity of the linked diseases, it is challenging to draw overarching conclusions of the GenVar_HD2 selections and I will instead highlight in the following observations and findings. For the mutations enhancing the binding of the bait protein domains in the phage selections, we found several linking to autophagy, scaffolding and trafficking. One example includes the R723W MAPT (Microtubule-associated protein tau) mutation, associated with frontotemporal dementia and Parkinson's disease, which creates a partial binding site for the CASK kinase domain. While there is good accordance between the phage display preferences and the affinity measurements on the peptide level, validation of the binding effect of the enabling mutations in context of the full-length proteins proved challenging. We hypothesise that the de novo generated binding motifs tested may not be competitive enough to prevail in the pool of endogenous ligands, particularly when sampled in stringent co-immunoprecipitation experiments. Their biological relevance might also only be revealed in a cellular setting resembling the disease state more closely, for example within the context of multiple cancer-associated mutations in the same protein or other proteins.

Of note, we found that 20% of the domain-mutation pairs indicated as disrupted by the mutation map to previously reported PPIs. Among those we note several PPIs involving E3 ligases, the function of which is to transfer ubiquitin to their substrate proteins, targeting them for example for proteasomal degradation (120). For instance, the E3 ubiquitin protein ligase NEDD4 (NEDD4) and/or NEDD4-like (NEDD4L) reportedly interact with and regulate the turnover of the Amiloride-sensitive sodium channel subunit beta (SCNN1B) (121,122). Mutation of SCNN1B is associated with Liddle's syndrome, a cardiovascular disorder characterised by hypertension due to elevated Na^{2+} transport by SCNN1B (123). On the molecular level, it has been proposed that the loss of NEDD4 binding on account of SCNNB1 mutation within the PPxY WW domain binding-motif results in increased cell surface expression of SCNN1B (124). In line and extending on the mutations affecting WW domain-binding, we found in our phage selection that the P616S/R, P617S, P618R and Y620H SCNNB1 mutations diminish the

binding of NEDD4 WW2 domain. We confirmed the disrupting effect of the P616R and P618R mutation in affinity measurements using fluorescence polarisation (P616R: two-fold decrease, P618R: 15-fold decrease; **Figure 21A**).

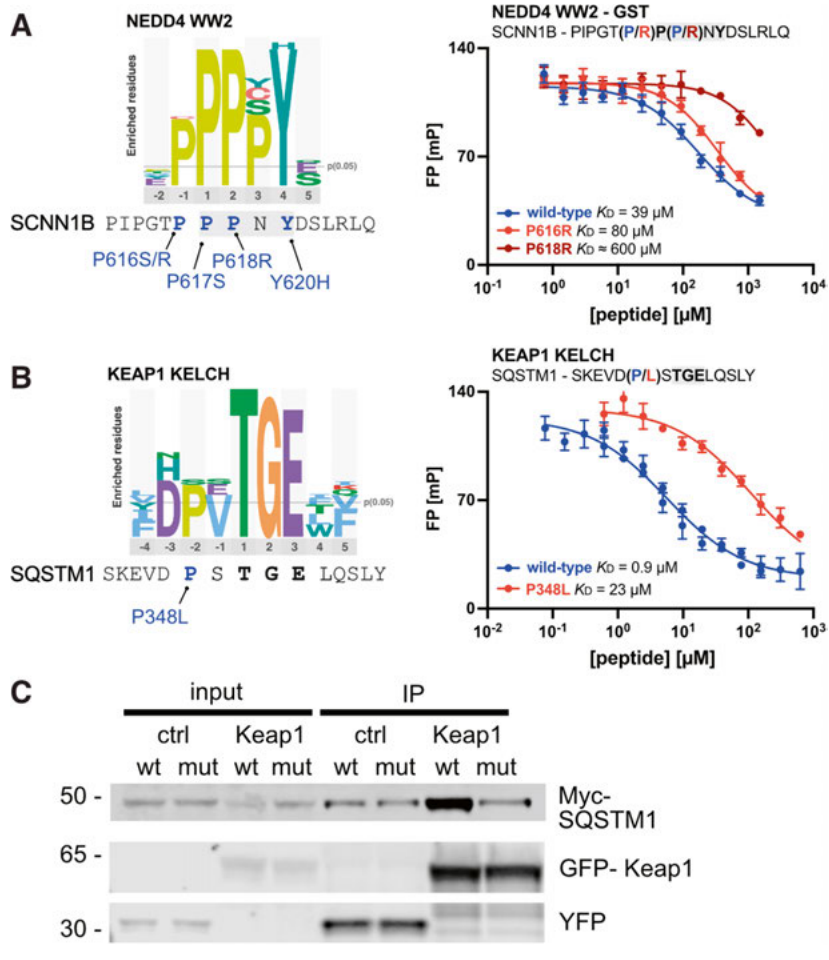

Figure 21: Exploring PPIs disrupted by disease-associated mutations. A: NEDD4 WW2 domain binding motif as established by selections with the HD2 library, the aligned SCNN1B$_{611\text{-}626}$ peptide with the mutations found to disable binding in the GenVar_HD2 selections and displacement curves of fluorescence polarisation experiments with wild-type, P616R and P618R SCNN1B$_{611\text{-}626}$ peptide. **B:** KEAP1 KELCH domain binding motif as established by selections with the HD2 library, the aligned SQSTM1$_{343\text{-}356}$ peptide with the mutation found to disable binding in the GenVar_HD2 selections and displacement curves of fluorescence polarisation experiments with wild-type and P348L SQSTM1$_{343\text{-}356}$ peptide. **C:** Co-immunoprecipitation of GFP-tagged KEAP1 with wild-type and P348L Myc-tagged SQSTM1 in HeLa cells.

Within our phage selections we found several mutations in NFE2L2 (substrate nuclear factor erythroid-derived 2-like 2), mapping to the TGE-motif, which abolish binding to the Kelch-like ECH-associated protein 1 (KEAP1) KELCH domain. NFE2L2 is an important transcriptional regulator in response to oxidative stress and its levels are regulated by the interaction with KEAP1. Loss of the interaction results in increased NFE2L2 levels and resistance of cancer cells to reactive oxygen species (125). In line, we find for instance the E79Q and E79L mutation, associated with cancer, at the p-1 position of the TGE-motif indicated to diminish the interaction with the KEAP1 KELCH domain. Selections with the GenVar_HD2 library identified moreover the P348L SQSTM1 mutation as disruptive for KEAP1 KELCH binding, which was proposed in an earlier study (126) and which we confirmed by affinity measurements and in co-immunoprecipitations (**Figure 21B, C**). The P348L SQSTM1 mutation is linked to frontotemporal dementia and/or amyotrophic lateral sclerosis (126,127) and the loss of KEAP1 binding was proposed to result in augmented turnover of NFE2L1 by KEAP1 and hence reduced transcription of anti-oxidative genes (126). Overall, we can demonstrate that selections with the GenVar_HD2 identify and confirm deregulation of E3 ligase interactions as a consequence of disease mutation. The associated altered turnover of the target proteins may contribute to various diseases.

We also explored the interactions of the KPNA4 ARM domain, modulated by disease-associated mutations, which may impact the nuclear localisation and hence functionality of the mutant prey proteins. Among those, I will highlight the R157C mutation in CDC45 indicated in our display to disrupt the binding to the domain. The mutation is found at the p+2 position of the KR-motif of the KPNA4 ARM domain and is linked to the Meier-Gorlin syndrome, a developmental disorder characterised by short stature. We confirmed in affinity measurements with wild-type and R157C $CDC45_{152\text{-}166}$ peptides that the mutation confers a 70-fold decrease in affinity for the KPNA4 ARM domain (**Figure 22A**). The co-immunoprecipitation experiments of wild-type CDC45 with KPNA4 were however unsuccessful, which we rationalise with the importin alpha isoforms binding to similar motifs but nuclear translocation in cases to be importin alpha-specific (128–130). Hence, we moved on to testing possible interactions with five other importin alpha isoforms, among which we find KPNA7 to co-immunoprecipitate with wild-type CDC45 and the binding to be abolished by the R157C mutation (**Figure 22B**).

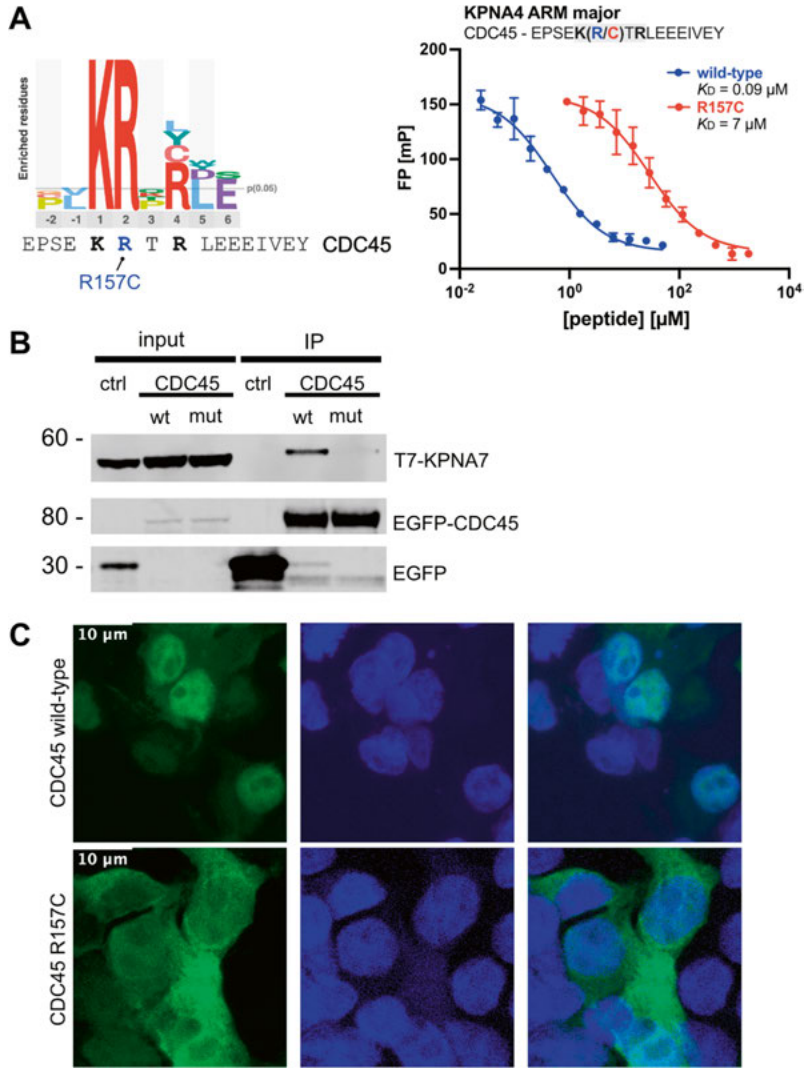

Figure 22: R157C CDC45 mutation impairs nuclear import. A: KPNA4 ARM domain binding motif as established by selections with the HD2 library, the aligned $CDC45_{152\text{-}166}$ peptide with the mutation found to disable binding in the GenVar_HD2 selections and displacement curves of fluorescence polarisation experiments with wild-type and R157C $CDC45_{152\text{-}166}$ peptide. **B:** Co-immunoprecipitation of EGFP-tagged wild-type and R157C CDC45 with T7-tagged KPNA7 in HEK293T cells. **C:** Representative confocal images of EGFP-tagged wild-type and R157C CDC45. Green: EGFP signal, blue: Hoechst staining used to indicate the nuclei.

In addition, assessment of the cellular localisation by confocal microscopy revealed that wild-type CDC45 localises to the nucleus whereas R157C CDC45 shows a more cytoplasmic localisation (**Figure 22D**). Of note, KPNA7 is proposed to play a role in embryogenesis and

development (131–133). This, together with the R157C CDC45 mutation being linked to a developmental disease, might hint at the importance of the KPNA7-CDC45 interaction during developmental stages. We could hence identify, through selections with the GenVar_HD2 library, a novel interaction between KPNA7 and CDC45, which is perturbed by the disease-associated R157C CDC45 mutation.

To summarise, I extended in this study the approach of mutational ProP-PD to sample the impact of genetic variation, as represented by non-synonymous SNVs, on protein binding. Selection with the GenVar_HD2 library indicates mutations affecting binding on the peptide level, which is true for both enhancing/creating and diminishing/disrupting mutations. We could further validate the binding effect of selected disruptive mutations in context of the full-length proteins. In comparison to low-throughput methods dissecting individual mutations, selection with the GenVar_HD2 library offers the advantage to identify mutations resulting in altered protein binding in large-scale by challenging a bait protein domain simultaneously with thousands of wild-type/mutant peptide pairs. Since the sampled mutations are linked to the diseases, it can be envisioned to use the information from the GenVar_HD2 selection as starting point to investigate the molecular mechanism of a given mutation and its contribution to disease progression.

Conclusion

My thesis work has been dedicated to the understanding of motif-based PPIs with particular focus on conditional interactions modulated by phosphorylation or impacted by genetic variation. I investigated motif-based PPIs orchestrating the phosphorylation code of binding, involving kinases, phosphatases and phospho-binding domains, from different angles. This included screening various kinase domains in a conventional ProP-PD experiment to find binding peptides contributing to substrate targeting of the kinase domains. We hypothesise based on the hydrophobicity of the binding peptides that the selections mainly retrieve docking interactions, which are interactions distal of the catalytical cleft. This can, in combination with phosphorylation-site motifs determined by other methods, shed light on kinase-substrate relationships.

From the perspective of the interpreter of the phosphorylation code, we investigated the phospho-modulated binding of protein domains on the basis of either bioinformatic predictions or phosphomimetic ProP-PD. While phosphorylation can act as switch, either creating (e.g. FYN SH2 binding to pY781 ACE2; **Paper II**) or breaking (e.g. TSG101 UEV binding to pT718 PDCD6IP; **Paper III**) binding sites, it often fine-tunes affinities by impacting residues interspersed or within the flanking regions of the core motif. I demonstrated that phosphomimetic ProP-PD can serve as resource to screen for putative phospho-modulation of binding in high-throughput, even though the phosphomimetics do not allow to capture the binding of obligate phospho-binders. This is convincingly shown by the dependency of clathrin binding on S839 HURP phosphorylation, suggested by the phosphomimetic ProP-PD screen and recapitulated on the level of the phospho-peptide and of the full-length proteins.

Expanding on the feasibility of mutational ProP-PD to derive binding preference for wild-type or mutant peptides in a single experiment, I used the GenVar_HD2 library to elucidate the impact of genetic variation on motif-based PPIs. While there is good accordance on the peptide level for both disruptive and enabling mutations, only the disruptive

mutations could be recapitulated on the level of the full-length proteins. Of note, previous studies report mainly on the relevance of disruptive mutations in disease (45,47). This could hence caution the interpretation of the enabling mutations, but it can also be envisioned that orthogonal methods, such as BioID (proximity-dependent biotin identification), could capture the binding preference more sensitively. We could outline for a selected set of cases, the relevant information provided by the screen with the GenVar_HD2 library on lost and found interactions associated with neurological and metabolic diseases, as well as in cancer. This is demonstrated convincingly for the P348L SQSTM1 mutation, associated with frontotemporal dementia and diminishing the interaction with KEAP1, as well as the R157C CDC45 mutation, associated with Meier-Gorlin syndrome and disrupting the interaction with KPNA7. Together, we show the potential of mutational ProP-PD to report on genetic variation perturbing motif-based PPIs.

One aspect, which has to be considered for mutational ProP-PD is that binding preferences are established by sequencing counts. The analysis assumes hence that the phage particles display the same amount of wild-type or mutant peptide on their surface. Discrepancy between the phage display and affinity measurements could be explained by avidity effects elicited by a differential display of wild-type and mutant peptide on the phage. Nonetheless, mutational ProP-PD convinces by screening thousands of phospho- and mutation sites in parallel and a positive correlation between the display and affinity measurements of ca. 80% for both the phosphomimetic and the genetic variation project.

Future perspectives

Despite enormous efforts and a variety of techniques applied, we are still far away from a holistic understanding of PPIs and the networks they form, and will assumably never reach it. We are often limited by the systems we choose to work with and the techniques used to investigate them. ProP-PD offers the advantage of covering the intrinsically disordered region proteome-wide, tiling the potential motif-based interaction surfaces of the human proteome in peptide stretches. At the same time, the method is limited by the length of the peptides displayed and the number and types of protein domains screened against the libraries. The latter will be addressed in the umbrella project "Towards the human motif map", in which the screening of the kinase domains will be incorporated. Assessment of binding motifs and specificities of various protein domains and families will improve our global understanding of motif-based PPIs. It can moreover be envisioned to go beyond the length of 16 amino acids displayed on the p8 coat protein with the downside that it will affect phage stability. Nonetheless, this may allow to capture longer binding motifs, such as the bipartite nuclear localisation signal, and provide additional information on flanking regions.

Direct biological translation of ProP-PD is furthermore inherently limited by the in vitro approach of the selections sampling peptide-protein domain interactions. Interactor prioritisation has been benchmarked as outlined above and also takes co-localisation and shared biological processes into account. Nonetheless, it can be suggested that orthogonal interaction screens on the basis of the full-length proteins, such as bimolecular fluorescence complementation analysis, could be implemented routinely to increase confidence in the uncovered interactors.

Conditional motif-based interactions pose a unique challenge since their assessment might require prior knowledge of when and where the interaction occurs. As demonstrated in this thesis, mutational ProP-PD can be applied to screen for phospho-modulated PPIs and disease-relevant mutations impacting PPIs in large-scale. This allows to identify phosphosites and mutations in vicinity of the core binding motif

modulating the binding behaviour of a protein domain. With regard to the phosphomimetic mutation, improvement of the screen can be achieved by either encoding phosphorylated amino acids as amber stop codons or by phosphorylating the phage libraries. By incorporating actual phosphorylation in the phage library, it could be possible to capture the binding of obligate phospho-binders in the phage selections. Mapping of the phosphorylation-site motifs established by Johnson *et al.* (55) on the interactions found to be modulated by the phosphomimetic mutation may further facilitate to contextualise them and render them more amenable for follow-up studies.

Mutational ProP-PD applied to elucidate interactome changes caused by genetic variation is mainly limited by its transferability from peptide to full-length protein levels. As noted above, it would be advantageous to implement an interaction screen of the full-length proteins with reasonable throughput. Co-immunoprecipitations are valuable assays, but they are of low throughput and low affinity interactions are often lost in the washing process. Of note, I tried to assess some of the interactions by proximity ligation assay. However, the method was in my hands too sensitive to the expression levels of the transiently transfected proteins, so that anything abundant enough will appear as co-localised. Proving the biological relevance of the peptide-domain interaction changes established by the display remains hence a major challenge at hand. Building on this, exploring the functional consequences of identified interaction changes and how they contribute to genetic disorders will require dedicated studies, tailored towards the proteins involved in the interaction.

To summarise and assumably the most common statement of any PhD thesis or study: much more is to be done in order to comprehend conditional motif-based PPIs in health and disease. I hope I have convinced the reader that I have contributed with my work to their exploration.

Populärwissenschaftliche Zusammenfassung

In diesem Teil möchte ich gerne eine Kurzversion meiner Arbeit wiedergeben, die eventuell ein wenig verständlicher ist. Wenn man sich so eine Nährwertangabe auf unseren Nahrungsmitteln anschaut, dann findet man darauf Kohlenhydrate, Fett und auch Proteine. Proteine sind die Moleküle, die für einen Muskelprotz wichtig sind, damit er stark bleibt. Aber es sind auch die Moleküle, mit welchen ich mich in meiner Doktorarbeit beschäftigt habe. Proteine sind nämlich nicht nur für die Ernährung wichtig, sondern auch für die verschiedenen Arbeitsabläufe im Körper und in den Zellen, aus denen unsere Gewebe bestehen. Wenn wir also so eine Zelle näher betrachten, so finden wir dort eine Menge verschiedener Prozesse, welche zeitlich und räumlich koordiniert werden müssen. Man kann sich das wie einen Flughafen oder eine Lagerhalle vorstellen: wenn nicht jeder an seinem designierten Platz genau das tut was von ihm erwartet wird, so kommt es zur großen Katastrophe. Proteine kommunizieren miteinander, um ihre Funktion zu erfüllen, in dem sie miteinander interagieren. Das nennt man Protein-Protein Interaktion.

Protein-Protein Interaktionen können auf unterschiedliche Art und Weise geschehen, und in meiner Arbeit konzentriere ich mich auf Interaktionen, in denen das eine Protein nur einen kleinen Teil des anderen Proteins erkennt und auch nur diesen braucht, um zu binden. Diesen kleinen Teil nennt man Peptid und oftmals enthält dieses Peptid eine Sequenz aus charakteristischen Aminosäuren, die spezifisch für das interagierende Protein ist. Diese Sequenz nennt man Motif. In unserem Labor wurde eine Methode (ProP-PD) weiterentwickelt, um all die verschiedenen Motive, die von unterschiedlichen Proteinen erkannt werden, zu identifizieren. Dazu benutzt man Phagen, um Peptide und damit potenzielle Motive auf der Hülle der Phagen anzuzeigen. Die Peptide sind als Nukleinsäuren in der DNA der Phagen enkodiert. Man fragt dann die Proteine in den Selektionsrunden welche Peptide, die von den Phagen angezeigt werden, sie lieber binden und damit eine Interaktion eingehen. Durch Sequenzieren der Phagen-DNA findet man dann heraus welche Peptide das genau sind und kann dann im besten Fall ein

Motiv herausbekommen. Dazu habe ich auch in meinem ersten Projekt beigetragen und mir angeschaut ob Kinase Proteindomänen auch solche Motive binden. Das hat für eine kleine Anzahl der Domänen auch gut geklappt (**Paper I**).

Der Hauptteil meiner Arbeit beschäftigt sich allerdings mit einer Variation von ProP-PD, in welcher die angezeigten Peptide immer in zwei verschiedenen Varianten vorhanden sind (Mutations-ProP-PD). Damit kann man herausfinden, ob ein Protein eine der zwei Varianten mit großer Vorliebe bindet. Ich habe das einmal verwendet, um zu schauen, ob das Hinzufügen einer kleinen Molekülgruppe, der sogenannten Phosphatgruppe, einen Einfluss darauf hat, ob die Proteine an die Peptide binden (**Paper III**). Die Phosphatgruppe ist eine wichtige Modifikation, die viele unterschiedlich Prozesse in der Zelle reguliert. Deshalb ist es wichtig herauszufinden, welche Interaktionen durch sie hervorgerufen und verstärkt, oder aber gebrochen und abgeschwächt werden. In **Paper II** habe ich mir das auch für verschiedene Motive angeschaut, allerdings basierend auf bioinformatischen Voraussagungen von Kollaborateuren und nicht dem Mutations-ProP-PD. In meinem letzten Project (**Paper IV**) verwende ich auch den Mutations-ProP-PD mit dem Fokus auf Mutationen, hauptsächlich diejenigen die in Krankheiten auftreten. Die Hypothese ist, dass Mutationen, die zu einer Veränderung im Peptide führen, Interaktionen zwischen den Proteinen hervorrufen oder brechen können. Das hat auch auf der Ebene der Peptide gut geklappt. Es scheint allerdings als wäre es nicht ganz trivial die biologische Relevanz dieser Interaktionen in der Zelle nachzuweisen. Das muss dann noch optimiert werden.

Insgesamt habe ich mir in meiner Doktorarbeit mir Interaktionen, basierend auf Motiven, mit ProP-PD angeschaut. Viele der Interaktionen sind konditionell. Das heißt die Proteine interagieren nur unter bestimmten räumlichen und zeitlichen Umständen miteinander. Der Mutations-ProP-PD hat dabei geholfen diese konditionellen Interaktionen zu analysieren. Es bleibt jedoch noch viel zu optimieren und zu lernen, um ein vollständiges Bild der Interaktionen zu generieren.

Populärvetenskaplig sammanfattning

I den här delen ville jag presentera en kort version av mitt arbete, som kanske är lite mer förståeligt. Om du tittar på näringsinformation för den mat vi äter så hittar du kolhydrater, fett och även proteiner. Proteiner är de molekyler som är nödvändiga för en muskelbyggare att hålla sig stark. Men det är också de molekyler som jag har studerat i min doktorsavhandling. Proteiner är inte bara viktiga för näringsämnen som bygger upp på våra muskler, utan de är också viktiga för de olika arbetsprocesserna i kroppen och i cellerna av vilka våra vävnader består. Så om vi tittar närmare på en sådan cell hittar vi ett antal olika processer där som måste koordineras rumsligt och tidsmässigt. Du kan föreställa dig det som en flygplats eller ett lager: om alla inte gör exakt vad som förväntas av dem på sin angiven plats, då blir det en stor katastrof. Proteiner kommunicerar och interagerar med varandra för att fylla sina funktioner. Detta kallas för en protein-protein interaktion.

Protein-protein interaktioner kan ske på olika sätt, och i mitt arbete fokuserar jag på interaktioner där ett protein känner igen och bara behöver en liten del av det andra proteinet för att binda. Denna lilla del kallas en peptid och ofta innehåller denna peptid en sekvens av karakteristiska aminosyror som är specifik för det interagerande proteinet. Denna sekvens kallas för ett motiv. En metod (ProP-PD) utvecklades vidare i vårt laboratorium för att identifiera vilka motiv som binds av olika proteiner. För att göra detta används fager för att presentera peptider med potentiella motiv på faghöljet. Peptiderna kodas som nukleinsyror i fagens DNA. Proteinerna går igenom selektionsrundor där de frågas sedan vilka peptider, som visas av fagen, de föredrar att binda och interagera med. Genom att sekvensera fag-DNA:t kan man sedan ta reda på vilka peptider det är exakt och i bästa fall kan man ta reda på ett motiv. Jag bidrog även till detta i mitt första projekt och tittade på om kinas domäner också binder sådana motiv. Detta fungerade bra för ett litet antal av domäner (**Paper I**).

Huvuddelen av mitt arbete handlade dock om en variant av ProP-PD, där de visade peptiderna alltid finns i två olika varianter (mutations-

ProP-PD). Detta gör det möjligt att ta reda på om ett protein binder till någon av de två varianterna med stor preferens. Jag använde detta å ena sidan för att se om att lägga till en liten molekylgrupp, som kallas fosfatgrupp, skulle påverka om proteinerna binder till peptiderna (**Paper III**). Fosfatgruppen är en viktig modifiering som reglerar många olika processer i cellen. Därför är det viktigt att ta reda på vilka interaktioner som framkallas och stärks av dem, eller bryts och försvagas. I **Paper II** tittade jag också på fosfatgruppens betydelse för bindning för olika motiv, som i den studien var baserade på bioinformatiska förutsägelser från samarbetspartners. I mitt senaste projekt (**Paper IV**) använder jag även den mutations ProP-PD med fokus på mutationer, främst de som uppstår i sjukdomar. Hypotesen är att mutationer som leder till en förändring i peptiden kan skapa eller bryta interaktioner mellan proteinerna. Detta fungerade också bra på peptidnivå. Det verkar dock inte vara helt trivialt att bevisa den biologiska relevansen av dessa interaktioner i cellen. Det måste fortfarande optimeras.

Sammantaget tittade jag i min doktorsavhandling på interaktioner baserade på motiv med ProP-PD. Många av interaktionerna är villkorade. Detta innebär att proteinerna endast interagerar med varandra under vissa rumsliga och tidsmässiga konditioner. Mutations-ProP-PD hjälpte till att analysera dessa villkorade interaktioner. Men mycket återstår av att finjustera och lära sig för att skapa en komplett bild av protein-protein-interaktionerna i cellen.

Acknowledgments

The last part of the thesis and undoubtedly the one paid most attention to is the one, in which I would like to thank everyone who has accompanied and supported me in my PhD adventures.

My first thank you goes to my supervisor **Ylva**, who has guided me through the past 4.5 years. In the beginning very much on top of my projects, you soon let go which gave me the liberty and responsibility to handle them. I always knew with confidence that you will safely support me towards PhD. Thank you for the past years as mentor.

I also want to thank my co-supervisor, **Micke**. You are the very kind face of the corridor and someone every PhD knows can turn to. Thank you for the feedback on my thesis.

Obviously always there and just shifting in shape and people around is the Ivarsson lab. **Caroline**, I believe you have accompanied me the longest in my PhD through all the highs and all the lows. Even if you are a bit rough on the outside, I could always rely on being a stronghold to me when I was feeling small. **Susanne**, I am happy that I told you that we need to get to know each other and be friends. You have supported me through all times and guided me when I was confused. I deeply appreciate our friendship. I also want to say my thanks to **Emily**, who came a little later in my PhD life and I grew very fond of. I am happy you show me your loud side and I appreciate you listening to me about all the things that worry me. And there we have **Eszter**, always there for a möh and to hang out on your couch once being fed. I have very fond memories of Fikas and dinners, with you, **Marie** and Caroline who integrated me from time to time in the SSF crew. **Ali**, as my senior PhD you have helped me a lot with the advices you gave. You suck because you left before my defence though and I miss the friendly squabbles we were having. I am also happy to share the office with **Priyanka**, who always makes me welcome in the postdoc crew, is always there to have a chat about food and is a very loyal 6:30 am in the morning running companion. **Leandro**, you are the sweet soul of the lab. Thank you for doing all the bioinformatic stuff, for nice chats about this

and that and never losing your patience with me asking you for all the things all the time. **Max**, I wish you the best for the rest of your PhD. And I want to thank my students, **Hanna, Emma and Alice**, who have contributed to my projects.
I also want to express my appreciation for the enormous help which you, **Norman**, provided with all the mutation tables and the critical contributions to ALL our projects. Thanks goes also to **Doreen** and **Dilip** for their help with crystallising clathrin.

The biochemistry corridor harbours many more faces, which I grew fond of over the years. **Maxim**, always there for a beer and a chat. Even though you and Leonidas did not start the fire, I will always remember the ski jumper and will miss having you around. Da hat man schon Pipi in den Augen. **Nadine and Petra**, I am happy to have shared some time with you in the corridor and obviously even more outside with Fika. And thank you Nadine for the lovely cover of my thesis. I admire the creativity. I also very much appreciated the postdoc crew, including **Mostafa, Marija and Candice**, who adopted me for lunch every now and then. Always shameful to bring my cheese bread to that crowd, such foodies. I also want to thank the people **Giulia, Erika and Ted** who have already left the corridor and accompanied me in my journey nonetheless. I enjoyed my time with the analytical chemists at football games with **Johan** (you definitely sing better than Leonidas) or at dinners with **Catía** and **Ioanna,** and obviously at cornhole. Who would want to miss that? **Filip and Malin,** I wanted to say my thanks as well for being part of my PhD life at work and outside work.

Very much integral to my Uppsala life is the Swedish Fika crew with **Elin, Hanna and Mia**. You made Uppsala and Sweden home to me with numerous walks, Fikas, dinners, activities and the support you provide at all times. Even though you were only for a short time part of the Ivarsson lab, **Christin**, you stuck with me from Uppsala to Copenhagen, where you always looked out for me. And I know we will keep on sticking together.

A big shout out goes also to the Nilsson lab, who integrated me twice during my PhD time and made me feel so very welcome. **Jakob**, you inspire with your energy, particularly early in the morning. **Dia**, I am very appreciative of all the help provided, and **Marcel**, I do appreciate you but maybe not the German memes.

Far away but very close to my heart is my family. Ich vermisse euch alle und bin sehr froh, dass es euch gibt. **Mom**, deine Begeisterung für das Leben ist eine Inspiration für alle und ich bin unglaublich dankbar für die Wärme, die du und **Daddy** mir und der ganzen Familie spendet. Ich möchte mich auch bei meinen Cousins, Tanten und Onkels, insbesondere **Luki, Tine, Sani, Steffi und Armin**, bedanken, die mich in meinem Leben viel begleitet haben. Und da gibt es noch die Kröten, also meine Geschwister **Nini** und **Tom**, die ich so dolle liebhabe. Ich freue mich immer, wenn wir mal Zeit für einander finden. Wie war das mit den Krücken? **Mama** und **Papa**, so viel Unterstützung und so viel Liebe wie man sich nur wünschen kann. Ohne euch hätte ich gar nichts geschafft, wäre vielleicht ein bisschen mehr Chill am See gewesen aber brauch man gar nicht. Und nein, Papa, ich weiß leider immer noch nicht die Länge einer C-C Bindung.

Leonidaki – I just wanted to say: Dom känslor du får mig att känna; Dom minnen du skapar i mig; Trots torskar och tårar är kärleken störst; Leonidaki - jag står här för dig! Always - patooties stick together. Words cannot fathom the support and love you provide.

References

1. Braun P, Gingras AC. History of protein-protein interactions: From egg-white to complex networks. Proteomics. 2012;12(10):1478–98.
2. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014 Jan;42(Database issue):D358-63.
3. Tompa P, Davey NE, Gibson TJ, Babu MM. A Million peptide motifs for the molecular biologist. Mol Cell. 2014;55(2):161–9.
4. Sun PD, Foster CE, Boyington JC. Overview of protein structural and functional folds. Curr Protoc protein Sci. 2004 May;Chapter 17(1):Unit 17.1.
5. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev. 2008 Apr;108(4):1225–44.
6. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. Mol Biosyst. 2012;8(1):268–81.
7. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014 Jul;114(13):6589–631.
8. Dosztányi 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. 2005 Aug;21(16):3433–4.
9. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021 Jul;49(W1):W297–303.
10. Papageorgiou AC, Poudel N, Mattsson J. Protein Structure Analysis and Validation with X-Ray Crystallography. Methods Mol Biol. 2021;2178:377–404.
11. Bax A, Clore GM. Protein NMR: Boundless opportunities. J Magn Reson. 2019 Sep;306:187–91.
12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
13. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins. 2000 Nov;41(3):415–27.
14. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn. 2012;30(2):137–49.
15. Uversky VN. Protein intrinsic disorder and structure-function continuum. Prog Mol Biol Transl Sci. 2019;166:1–17.
16. Ivarsson Y, Jemth P. Affinity and specificity of motif-based protein–protein interactions. Curr Opin Struct Biol. 2019;54:26–33.

17. Bugge K, Brakti I, Fernandes CB, Dreier JE, Lundsgaard JE, Olsen JG, et al. Interactions by Disorder - A Matter of Context. Front Mol Biosci. 2020;7:110.
18. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, et al. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020;48(D1):D296–306.
19. Kumar M, Michael S, Alvarado-Valverde J, Mészáros B, Sámano-Sánchez H, Zeke A, et al. The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res. 2022 Jan;50(D1):D497–508.
20. Ren R, Mayer BJ, Cicchetti P, Baltimore D. Identification of a ten-amino acid proline-rich SH3 binding site. Science. 1993 Feb;259(5098):1157–61.
21. Saras J, Heldin CH. PDZ domains bind carboxy-terminal sequences of target proteins. Trends Biochem Sci. 1996 Dec;21(12):455–8.
22. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P. Characterization of a novel protein-binding module--the WW domain. FEBS Lett. 1995 Aug;369(1):67–71.
23. ter Haar E, Harrison SC, Kirchhausen T. Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. Proc Natl Acad Sci U S A. 2000 Feb;97(3):1096–100.
24. Sorkin A, Mazzotti M, Sorkina T, Scotto L, Beguinot L. Epidermal growth factor receptor interaction with clathrin adaptors is mediated by the Tyr974-containing internalization motif. J Biol Chem. 1996 Jun;271(23):13377–84.
25. Lu J, Wu T, Zhang B, Liu S, Song W, Qiao J, et al. Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Commun Signal. 2021 May;19(1):60.
26. Ichimura Y, Kumanomidou T, Sou Y, Mizushima T, Ezaki J, Ueno T, et al. Structural basis for sorting mechanism of p62 in selective autophagy. J Biol Chem. 2008 Aug;283(33):22847–57.
27. Johansen T, Lamark T. Selective Autophagy: ATG8 Family Proteins, LIR Motifs and Cargo Receptors. J Mol Biol. 2020;432(1):80–103.
28. Kliche J, Ivarsson Y. Orchestrating serine/threonine phosphorylation and elucidating downstream effects by short linear motifs. Biochem J. 2022 Jan;479(1):1–22.
29. Duan G, Walther D. The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol. 2015 Feb;11(2):e1004049.
30. Ardito F, Giuliani M, Perrone D, Troiano G, Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med. 2017 Aug;40(2):271–80.
31. Huang B, Zhao Z, Zhao Y, Huang S. Protein arginine phosphorylation in organisms. Int J Biol Macromol. 2021 Feb;171:414–22.
32. Cieśla J, Frączyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochim Pol. 2011;58(2):137–48.
33. Puttick J, Baker EN, Delbaere LTJ. Histidine phosphorylation in biological systems. Biochim Biophys Acta. 2008 Jan;1784(1):100–5.
34. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011;12(9):628–40.
35. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12(11):745–55.
36. Kunowska N, Stelzl U. Decoding the cellular effects of genetic variation through interaction proteomics. Curr Opin Chem Biol. 2022 Feb;66:102100.

37. Soto C, Pritzkow S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. Nat Neurosci. 2018 Oct;21(10):1332–40.
38. Kim M, Park J, Bouhaddou M, Kim K, Rojc A, Modak M, et al. A protein interaction landscape of breast cancer. Science. 2021 Oct;374(6563):eabf3066.
39. Zheng F, Kelly MR, Ramms DJ, Heintschel ML, Tao K, Tutuncuoglu B, et al. Interpretation of cancer mutations using a multiscale map of protein systems. Science. 2021 Oct;374(6563):eabf3067.
40. Hung M-C, Link W. Protein localization in disease and therapy. J Cell Sci. 2011 Oct;124(Pt 20):3381–92.
41. Wang Z, Moult J. SNPs, protein structure, and disease. Hum Mutat. 2001 Apr;17(4):263–70.
42. David A, Razali R, Wass MN, Sternberg MJE. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum Mutat. 2012 Feb;33(2):359–63.
43. David A, Sternberg MJE. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. J Mol Biol. 2015 Aug;427(17):2886–98.
44. Wong ETC, So V, Guron M, Kuechler ER, Malhis N, Bui JM, et al. Protein-Protein Interactions Mediated by Intrinsically Disordered Protein Regions Are Enriched in Missense Mutations. Biomolecules. 2020 Jul;10(8):1097.
45. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, et al. Widespread macromolecular interaction perturbations in human genetic disorders. Cell. 2015 Apr;161(3):647–60.
46. Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, et al. Comprehensive characterization of protein-protein interactions perturbed by disease mutations. Nat Genet. 2021 Mar;53(3):342–53.
47. Fragoza R, Das J, Wierbowski SD, Liang J, Tran TN, Liang S, et al. Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. Nat Commun. 2019 Sep;10(1):4141.
48. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989 Jul;340(6230):245–6.
49. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005 Oct;437(7062):1173–8.
50. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005 Sep;122(6):957–68.
51. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. Nature. 2020 Apr;580(7803):402–8.
52. Frank R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications. J Immunol Methods. 2002 Sep;267(1):13–26.
53. Rodriguez M, Li SS-C, Harper JW, Songyang Z. An oriented peptide array library (OPAL) strategy to study protein-protein interactions. J Biol Chem. 2004 Mar;279(10):8802–7.
54. Volkmer R, Tapia V, Landgraf C. Synthetic peptide arrays for investigating protein interaction domains. FEBS Lett. 2012 Aug;586(17):2780–6.

55. Johnson JL, Yaron TM, Huntsman EM, Kerelsky A, Song J, Regev A, et al. An atlas of substrate specificities for the human serine/threonine kinome. Nature. 2023 Jan;613(7945):759–66.
56. Meyer K, Kirchner M, Uyar B, Cheng J-Y, Russo G, Hernandez-Miranda LR, et al. Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. Cell. 2018 Sep;175(1):239-253.e17.
57. Dittmar G, Hernandez DP, Kowenz-Leutz E, Kirchner M, Kahlert G, Wesolowski R, et al. PRISMA: Protein Interaction Screen on Peptide Matrix Reveals Interaction Footprints and Modifications- Dependent Interactome of Intrinsically Disordered C/EBPβ. iScience. 2019 Mar;13:351–70.
58. Meyer K, Selbach M. Peptide-based Interaction Proteomics. Mol Cell Proteomics. 2020 Jul;19(7):1070–5.
59. 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. 2017 Nov;114(46):12166–71.
60. Huisman BD, Grace BE, Holec P V, Birnbaum ME. Yeast Display for the Identification of Peptide-MHC Ligands of Immune Receptors. Methods Mol Biol. 2022;2491:263–91.
61. Hwang T, Parker SS, Hill SM, Grant RA, Ilunga MW, Sivaraman V, et al. Native proline-rich motifs exploit sequence context to target actin-remodeling Ena/VASP protein ENAH. Elife. 2022 Jan;11:e70680.
62. Halpin JC, Whitney D, Rigoldi F, Sivaraman V, Singer A, Keating AE. Molecular determinants of TRAF6 binding specificity suggest that native interaction partners are not optimized for affinity. Protein Sci. 2022 Nov;31(11):e4429.
63. Seo M-H, Nim S, Jeon J, Kim PM. Large-Scale Interaction Profiling of Protein Domains Through Proteomic Peptide-Phage Display Using Custom Peptidomes. Methods Mol Biol. 2017;1518:213–26.
64. Ivarsson Y, Arnold R, McLaughlin M, Nim S, Joshi R, Ray D, et al. Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. Proc Natl Acad Sci U S A. 2014;111(7):2542–7.
65. Benz C, Ali M, Krystkowiak I, Simonetti L, Sayadi A, Mihalic F, et al. Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. Mol Syst Biol. 2022 Jan;18(1):e10584.
66. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science. 1985 Jun;228(4705):1315–7.
67. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. Nature. 1990 Dec;348(6301):552–4.
68. Sundell GN, Ivarsson Y. Interaction Analysis through Proteomic Phage Display. Biomed Res Int. 2014;2014:176172.
69. Ali M, Simonetti L, Ivarsson Y. Screening Intrinsically Disordered Regions for Short Linear Binding Motifs. Methods Mol Biol. 2020;2141:529–52.
70. Schopp IM, Amaya Ramirez CC, Debeljak J, Kreibich E, Skribbe M, Wild K, et al. Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. Nat Commun. 2017 Jun;8:15690.
71. Dai L, Zhao T, Bisteau X, Sun W, Prabhu N, Lim YT, et al. Modulation of Protein-Interaction States through the Cell Cycle. Cell. 2018 May;173(6):1481-1494.e13.

72. Liu BA, Jablonowski K, Shah EE, Engelmann BW, Jones RB, Nash PD. SH2 domains recognize contextual peptide sequence information to determine selectivity. Mol Cell Proteomics. 2010 Nov;9(11):2391–404.
73. Gogl G, Jane P, Caillet-Saguy C, Kostmann C, Bich G, Cousido-Siah A, et al. Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study. Structure. 2020 Jul;28(7):747-759.e3.
74. Grossmann A, Benlasfer N, Birth P, Hegele A, Wachsmuth F, Apelt L, et al. Phospho-tyrosine dependent protein-protein interaction network. Mol Syst Biol. 2015 Mar;11(3):794.
75. Erce MA, Low JKK, Hart-Smith G, Wilkins MR. A conditional two-hybrid (C2H) system for the detection of protein-protein interactions that are mediated by post-translational modification. Proteomics. 2013 Apr;13(7):1059–64.
76. Jehle S, Kunowska N, Benlasfer N, Woodsmith J, Weber G, Wahl MC, et al. A human kinase yeast array for the identification of kinases modulating phosphorylation-dependent protein-protein interactions. Mol Syst Biol. 2022 Mar;18(3):e10820.
77. Roumeliotis TI, Williams SP, Gonçalves E, Alsinet C, Del Castillo Velasco-Herrera M, Aben N, et al. Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. Cell Rep. 2017 Aug;20(9):2201–14.
78. Ryan CJ, Kennedy S, Bajrami I, Matallanas D, Lord CJ. A Compendium of Co-regulated Protein Complexes in Breast Cancer Reveals Collateral Loss Events. Cell Syst. 2017 Oct;5(4):399-409.e5.
79. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014 Aug;11(8):801–7.
80. Woodsmith J, Apelt L, Casado-Medrano V, Özkan Z, Timmermann B, Stelzl U. Protein interaction perturbation profiling at amino-acid resolution. Nat Methods. 2017 Dec;14(12):1213–21.
81. Sundell GN, Arnold R, Ali M, Naksukpaiboon P, Orts J, Güntert P, et al. Proteome-wide analysis of phospho-regulated PDZ domain interactions. Mol Syst Biol. 2018;14(8):e8129.
82. Johnson CM. Isothermal Titration Calorimetry. Methods Mol Biol. 2021;2263:135–59.
83. Biltonen RL, Langerman N. Microcalorimetry for biological chemistry: experimental design, data analysis, and interpretation. Methods Enzymol. 1979;61:287–318.
84. Freyer MW, Lewis EA. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods Cell Biol. 2008;84:79–113.
85. Du Y. Fluorescence polarization assay to quantify protein-protein interactions in an HTS format. Methods Mol Biol. 2015;1278:529–44.
86. Rossi AM, Taylor CW. Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc. 2011 Mar;6(3):365–87.
87. Checovich WJ, Bolger RE, Burke T. Fluorescence polarization--a new tool for cell and molecular biology. Nature. 1995 May;375(6528):254–6.
88. Gianni S, Engström Å, Larsson M, Calosci N, Malatesta F, Eklund L, et al. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. J Biol Chem. 2005;280(41):34805–12.
89. Nikolovska-Coleska Z, Wang R, Fang X, Pan H, Tomita Y, Li P, et al. Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. Anal Biochem. 2004;332(2):261–73.

90. Deller MC, Rupp B. Models of protein-ligand crystal structures: trust, but verify. J Comput Aided Mol Des. 2015 Sep;29(9):817–36.
91. Pozharski E, Deller MC, Rupp B. Validation of Protein-Ligand Crystal Structure Models: Small Molecule and Peptide Ligands. Methods Mol Biol. 2017;1607:611–25.
92. Clarkson J, Campbell ID. Studies of protein-ligand interactions by NMR. Biochem Soc Trans. 2003 Oct;31(Pt 5):1006–9.
93. Takeuchi K, Wagner G. NMR studies of protein interactions. Curr Opin Struct Biol. 2006 Feb;16(1):109–17.
94. Burckhardt CJ, Minna JD, Danuser G. Co-immunoprecipitation and semi-quantitative immunoblotting for the analysis of protein-protein interactions. STAR Protoc. 2021 Sep;2(3):100644.
95. Dunham WH, Mullin M, Gingras A-C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. Proteomics. 2012 May;12(10):1576–90.
96. Ubersax JA, Ferrell JE. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol. 2007;8(7):530–41.
97. Ochoa D, Jarnuczak AF, Viéitez C, Gehre M, Soucheray M, Mateus A, et al. The functional landscape of the human phosphoproteome. Nat Biotechnol. 2020;38(3):365–73.
98. Hsueh Y-P. Calcium/calmodulin-dependent serine protein kinase and mental retardation. Ann Neurol. 2009 Oct;66(4):438–43.
99. Mukherjee K, Sharma M, Urlaub H, Bourenkov GP, Jahn R, Südhof TC, et al. CASK Functions as a Mg2+-independent neurexin kinase. Cell. 2008 Apr;133(2):328–39.
100. Borg JP, Lõpez-Figueroa MO, de Taddèo-Borg M, Kroon DE, Turner RS, Watson SJ, et al. Molecular analysis of the X11-mLin-2/CASK complex in brain. J Neurosci Off J Soc Neurosci. 1999 Feb;19(4):1307–16.
101. Zhang Z, Li W, Yang G, Lu X, Qi X, Wang S, et al. CASK modulates the assembly and function of the Mint1/Munc18-1 complex to regulate insulin secretion. Cell Discov. 2020 Dec;6(1):92.
102. Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, Berman K, et al. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocr Rev. 2001 Apr;22(2):153–83.
103. Weston CR, Davis RJ. The JNK signal transduction pathway. Curr Opin Cell Biol. 2007 Apr;19(2):142–9.
104. Mészáros B, Sámano-Sánchez H, Alvarado-Valverde J, Čalyševa J, Martínez-Pérez E, Alves R, et al. Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. Sci Signal. 2021 Jan;14(665).
105. Kliche J, Kuss H, Ali M, Ivarsson Y. Cytoplasmic short linear motifs in ACE2 and integrin β(3) link SARS-CoV-2 host cell receptors to mediators of endocytosis and autophagy. Sci Signal. 2021 Jan;14(665):eabf1117.
106. Grass I, Thiel S, Höning S, Haucke V. Recognition of a basic AP-2 binding motif within the C2B domain of synaptotagmin is dependent on multimerization. J Biol Chem. 2004;279(52):54872–80.
107. Smith SM, Baker M, Halebian M, Smith CJ. Weak molecular interactions in clathrin-mediated endocytosis. Front Mol Biosci. 2017;4:72.
108. Owen DJ, Collins BM, Evans PR. Adaptors for clathrin coats: structure and function. Annu Rev Cell Dev Biol. 2004;20:153–91.
109. Parsons SJ, Parsons JT. Src family kinases, key regulators of signal transduction. Oncogene. 2004;23(48):7906–9.

110. Lu Y, Zhu Q, Fox DM, Gao C, Stanley SA, Luo K. SARS-CoV-2 down-regulates ACE2 through lysosomal degradation. Mol Biol Cell. 2022 Dec;33(14):ar147.
111. Yang B, Jia Y, Meng Y, Xue Y, Liu K, Li Y, et al. SNX27 suppresses SARS-CoV-2 infection by inhibiting viral lysosome/late endosome entry. Proc Natl Acad Sci U S A. 2022 Jan;119(4).
112. Lu Y, He P, Zhang Y, Ren Y, Zhang L. The emerging roles of retromer and sorting nexins in the life cycle of viruses. Virol Sin. 2022 Jun;37(3):321–30.
113. Karthika T, Joseph J, Das VRA, Nair N, Charulekha P, Roji MD, et al. SARS-CoV-2 Cellular Entry Is Independent of the ACE2 Cytoplasmic Domain Signaling. Cells. 2021 Jul;10(7).
114. Hornbeck P V, Kornhauser JM, Latham V, Murray B, Nandhikonda V, Nord A, et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. Nucleic Acids Res. 2019 Jan;47(D1):D433–41.
115. Kliche J, Garvanska DH, Simonetti L, Badgujar D, Dobritzsch D, Nilsson J, et al. Large-scale phosphomimetic screening identifies phospho-modulated motif-based protein interactions. Mol Syst Biol. 2023 Jul;19(7):e11164.
116. Rondelet A, Lin Y-C, Singh D, Porfetye AT, Thakur HC, Hecker A, et al. Clathrin's adaptor interaction sites are repurposed to stabilize microtubules during mitosis. J Cell Biol. 2020 Feb;219(2):e201907083.
117. Wong J, Fang G. HURP controls spindle dynamics to promote proper interkinetochore tension and efficient kinetochore capture. J Cell Biol. 2006 Jun;173(6):879–91.
118. Uyar B, Weatheritt RJ, Dinkel H, Davey NE, Gibson TJ. Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? Mol Biosyst. 2014 Oct;10(10):2626–42.
119. Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, Uversky VN, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS Comput Biol. 2012;8(10):e1002709.
120. Yang Q, Zhao J, Chen D, Wang Y. E3 ubiquitin ligases: styles, structures and functions. Mol Biomed. 2021 Jul;2(1):23.
121. Staub O, Abriel H, Plant P, Ishikawa T, Kanelis V, Saleki R, et al. Regulation of the epithelial Na+ channel by Nedd4 and ubiquitination. Kidney Int. 2000 Mar;57(3):809–15.
122. Rotin D, Staub O. Nedd4-2 and the regulation of epithelial sodium transport. Front Physiol. 2012;3:212.
123. Enslow BT, Stockand JD, Berman JM. Liddle's syndrome mechanisms, diagnosis and management. Integr Blood Press Control. 2019;12:13–22.
124. Lu C, Pribanic S, Debonneville A, Jiang C, Rotin D. The PY motif of ENaC, mutated in Liddle syndrome, regulates channel internalization, sorting and mobilization from subapical pool. Traffic. 2007 Sep;8(9):1246–64.
125. Taguchi K, Yamamoto M. The KEAP1-NRF2 System in Cancer. Front Oncol. 2017;7:85.
126. Goode A, Rea S, Sultana M, Shaw B, Searle MS, Layfield R. ALS-FTLD associated mutations of SQSTM1 impact on Keap1-Nrf2 signalling. Mol Cell Neurosci. 2016 Oct;76:52–8.
127. Rubino E, Rainero I, Chiò A, Rogaeva E, Galimberti D, Fenoglio P, et al. SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Neurology. 2012 Oct;79(15):1556–62.

128. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem. 2009 Jan;284(1):478–85.
129. Köhler M, Speck C, Christiansen M, Bischoff FR, Prehn S, Haller H, et al. Evidence for distinct substrate specificities of importin alpha family members in nuclear protein import. Mol Cell Biol. 1999 Nov;19(11):7782–91.
130. Pumroy RA, Cingolani G. Diversification of importin-α isoforms in cellular trafficking and disease states. Biochem J. 2015 Feb;466(1):13–28.
131. Wang X, Park K-E, Koser S, Liu S, Magnani L, Cabot RA. KPNA7, an oocyte- and embryo-specific karyopherin α subtype, is required for porcine embryo development. Reprod Fertil Dev. 2012;24(2):382–91.
132. Oostdyk LT, Wang Z, Zang C, Li H, McConnell MJ, Paschal BM. An epilepsy-associated mutation in the nuclear import receptor KPNA7 reduces nuclear localization signal binding. Sci Rep. 2020 Mar;10(1):4844.
133. Paciorkowski AR, Weisenberg J, Kelley JB, Spencer A, Tuttle E, Ghoneim D, et al. Autosomal recessive mutations in nuclear transport factor KPNA7 are associated with infantile spasms and cerebellar malformation. Eur J Hum Genet. 2014 May;22(5):587–93.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2289

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-508635

ACTA UNIVERSITATIS
UPSALIENSIS
2023