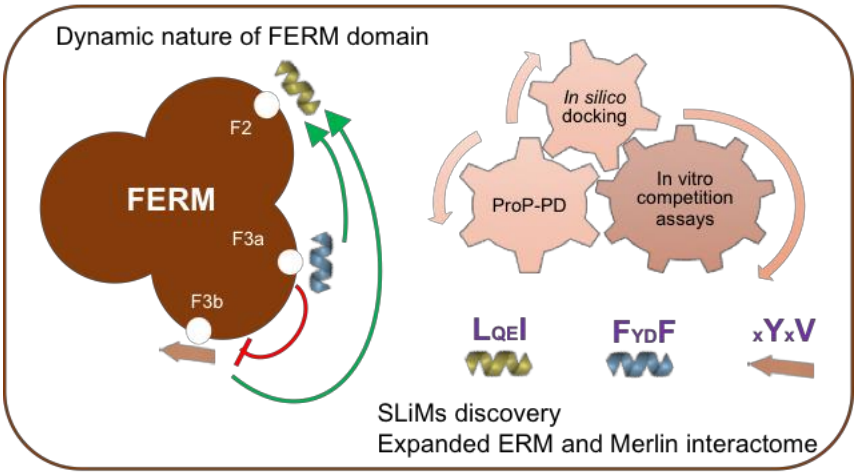
Defining binding motifs and dynamics of the multi-pocket FERM domain from ezrin, radixin, moesin and merlin

Muhammad Ali[1,*], Alisa Khramushin[2,*], Vikash K Yadav[1,3], Ora Schueler-Furman[2,#] & Ylva Ivarsson[1,#]

1. Department of Chemistry – BMC, Uppsala University, Husargatan 3, 751 34 Uppsala, Sweden
2. Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, the Hebrew University of Jerusalem, Jerusalem, Israel.
3. Current address: Department of Chemical engineering, Loughborough University, Loughborough, LE11 3TU, UK

* Shared first authors

\# Shared communicating authors



Abstract

The ERMs (ezrin, radixin and moesin) and the closely related merlin (NF2) participate in signaling events at the cell cortex through interactions mediated by their conserved FERM domain. We systematically investigated the FERM domain mediated interactions with short linear motifs (SLiMs) by screening the FERM domains againsts a phage peptidome representing intrinsically disordered regions of the human proteome. We uncovered a diverse set of interacting partners with similar but distinct binding motifs (FYDF, xYxV, FY(D/E)L and LQE(I/L) that bind to distinct binding pockets. We validated interactions between moesin and merlin FERM domains and full-length FAM83G, HIF1A, LATS1, NOP53, PAK6, RRBP1 and ZNF622 through pull-down experiments. Using biophysical binding assays, we determined affinities of, and uncovered allosteric interdependencies between, different binding partners, suggesting that the FERM domain acts as a switchable interaction hub. Using Rosetta FlexPepDock computational peptide docking protocols, we investigated the energy landscapes of identified interactions, which provide a detailed molecular understanding of the binding of the distinct binding motifs, as well as possible allosteric interconnections. This study demonstrates how experimental and computational approaches together can unravel a complex system of protein-peptide interactions that includes a family of proteins with multiple binding sites that interact with similar but distinct binding motifs.

Key words: FERM domain, ProP-PD, disordered regions, short linear motif, protein-protein interactions, allostery, Rosetta FlexPepDock

Highlights

- We screened the human disorderome for motif-containing partners of the FERM domains
- We expand the ERM and merlin interactomes of the ERMs and merlin
- We identify four distinct motif classes that bind the ERM and merlin FERM domains: FYDF, xYxV, FY(D/E)L and LQE(I/L)
- *In-vitro* and *in-silico* data suggest that the FYDF motif binds to the F3a site and that xYxV motif binds to the F3b site
- *In-silico* modelling sheds light on the underlying conformational changes responsible for ligand interdependencies

INTRODUCTION

The ERMs (ezrin, radixin and moesin) are membrane-associated proteins that provide linkage between membrane and actin cytoskeleton [1, 2]. The ERM proteins interact with filamentous (F)-actin, transmembrane or membrane-associated proteins, and phospholipids, and have important roles in controlling the localization of peripheral membrane proteins and the signaling from membrane receptors [1, 2]. The ERMs share an N-terminal FERM domain, followed by a region with alpha-helical propensity, and a C-terminal domain that binds to F-actin or to the FERM domain (**Fig. 1A, B**). The function of the ERM proteins is regulated by an autoinhibitory interaction of this same C-terminal region with the FERM domain [3-5]. The protein merlin, encoded by the NF2 gene is highly similar to the ERMs and is also regulated by an open-closed transition [6, 7], but lacks the C-terminal F-actin binding region and has different tissue localization and function [8]. Merlin is a well-known tumor suppressor protein and an upstream regulator of the Hippo pathway that regulates LATS1 and MAST kinases [9]. Here we focus on the FERM domains and its interactions with short linear motifs (SLiMs, 3-12 amino acid stretches) typically found in the intrinsically disordered regions of target proteins [10].

The FERM domain has a cloverleaf-like structure with three subdomains (F1, F2, and F3) [11]. The interdependent subdomains take distinct structures, with F1, F2 and F3 forming ubiquitin-like, acyl-CoA binding protein like, and phosphotyrosine binding (PTB) like folds, respectively. Several binding interfaces, here denoted as F2, F3a and F3b sites are distributed over the different subdomains (**Fig 1C**). The FERM domain binding sites engage in interactions with different types of peptides: a LATS1-derived peptide binds to the F2 site in a helical conformation [12], the C-terminal peptide of EBP50 engages in a helical conformation with the F3a site [13, 14], and the F3b site binds to a variety of ligands by ®-strand addition [15-18]. Despite the high amino acid identities between the ERMs and merlin, it has been found that their F3b sites have somewhat distinct specificities [17, 18].

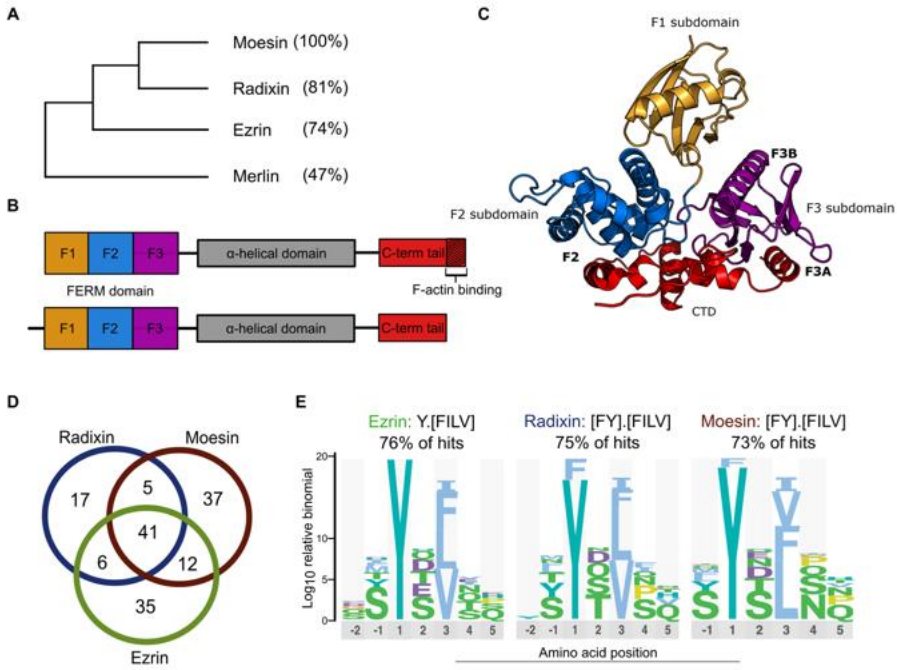


Figure 1. Overview of ERM domain family and merlin, together with results of ProP-PD selections against the ERMs. (A) Sequence identities between the full-length ERM family members and merlin (Clustal Omega). (**B**) Overview of the protein architecture of the ERM proteins and merlin (which lacks the F-actin binding site). (**C**) Structure of moesin with CTD in closed confirmation. The different subdomains and the binding sites are indicated (PDB: 1EF1 [11]). Colors are according to the domain architecture shown in (B). **(D)** Venn diagram of the overlap peptide hits identified for the ERM FERM domains through ProP-PD selection. **(E)** Consensus ERM binding motifs established from the ProP-PD identified peptides. Note that the apparently simple [FY]x[FILV] motif conceals additional levels of complexity as described in the main text.

The FERM domain shows a high versatility in terms of ligand binding and serves as a dynamic hub that binds to a large number of partners that may compete for interactions. The presence of several distinct peptide binding interfaces, and the versatility of ligand binding that is characteristic for the FERM domains makes it challenging to provide consensus motifs for the FERM domain family, as compared to other large peptide binding families with single binding pockets (e.g. PDZ domains, WW domains or SH3 domains) [19]. Here, we took on this challenge using a combination of proteomic peptide phage display (ProP-PD) and *in silico* modelling [20, 21] in order to expend the interactome and uncover the dynamic binding mode of FERM domains to its novel interactors. We screened the FERM domains of the ERMs and merlin using a phage library that tiles intrinsically disordered regions of the human proteome [20] and identified a large number of putative FERM domain ligands. We validated interactions of moesin and merlin for a select set of ligands through biophysical affinity measurements, and confirmed through cell-based experiments that they can occur in the context of the full-length proteins. Experimental data and computational analysis revealed that peptides with similar but distinct motifs (FYDF, xYxV and FY(D/E)L) interact with different binding sites of the FERM

domain. In addition, we found that binding of a ligand to the F3a site on one side of the F3 lobe confers conformational changes that reduces the affinity for binding to the F3b pocket the opposite side (but not the inverse), an allosteric effect that is likely operating during the transition from the open to the closed forms of the full-length proteins. The study provides an improved understanding of the SLiM-based molecular interactions that contribute to the common and distinct roles of the ERMs and merlin in regulating different signaling pathways from cytoplasm to the nucleus.

RESULTS & DISCUSSION

We first describe the results from the ProP-PD assay of ERM and merlin baits. We show that while most peptides selected for the ERMs adhere to a similar binding motif, the smaller merlin set is more diverse. We validate interactions through affinity measurements and pulldown experiments, and unveil a hidden complexity in the apparently simple binding motif of the ERMs. We next discuss in depth the features that govern ERM-binding specificity using moesin as an example, with emphasis on binding patterns associated with different binding sites on the FERM domain. We demonstrate allosteric effects upon ligand binding, and provide detailed structural models of these interactions, using our FlexPepDock protocol for the atom-resolution modeling of peptide-protein complexes.

ERM proteins share ligands and binding preferences

To capture potential interactors, we used the FERM domains of the ERMs as bait proteins in selections against our ProP-PD library that tiles intrinsically disordered regions of the human proteome [20]. Binding-enriched phage-pools were recovered after selections and analyzed by next-generation sequencing (NGS). The peptide coding regions were translated into peptides, and the data was filtered for peptides identified in replicate selections (obtaining at least 0.05% of the total NGS counts), which resulted in between 84 to 96 peptide hits for each of the FERM domains (**Table S1-S3**). As expected, there was a considerable overlap of peptide hits among the closely related ERM proteins (**Fig. 1D**). Consensus motifs were generated using the SLiMFinder algorithm [22], which suggested that the ProP-PD data of the ERM FERM domains is dominated by ligands with a shared, apparent, [FY].[FILV] motif (**Fig. 1E**). The motif is similar to the -T**Y**G**V**L- containing peptide of ICAM2 and the -M**Y**P**V**R- containing peptide of PSGL-1 that both have been co-crystallized bound to the F3b site in the PTB-like subdomain of radixin (Protein Data Bank, PDB [23]; code 1J19 and 2EMT [16, 24]), and the -T**Y**S**P**S- containing crumbs peptide that interacts with the moesin FERM domain using the same site (PDB code 4YL8) [17]. This pointed in the direction that the ProP-PD selection against the ERMs enriched for F3b site ligands. However, as we discuss below, our data suggest a more complex binding pattern.

We next enriched the data for ligands of potential biological relevance, by removing peptides found in proteins, or protein regions, associated with locations that would prevent interactions with the bait proteins in the cell (e.g. lumenal, extracellular or secreted localizations) and therefore less likely to be of physiological relevance. We further narrowed down the set to proteins that share gene ontology terms with the ERMs that are unlikely to occur by chance. (e.g. cell adhesion molecule binding, actin filament binding). These high confidence ERM data sets contain between 27 and 47 hits in between 35 to 66 different proteins. Only around 5% of the identified interactions of the ERMs involved previously reported ligands (**Table S1-3**; ezrin: ANK2 [25], ICAM3 [26, 27], MISP [28], SLC12A2 [29, 30]; radixin: MISP [28]; moesin; ICAM3 [26] and MISP [28]. Notably, we found that ezrin interacts with ANK2 through three different SLiM-binding sites ($_{2545}$-EVS**Y**E**V**TPKTTDVSTP-$_{2560}$, $_{2950}$-HTTSFHSSEV**Y**S**V**TIT-$_{2965}$ and $_{3761}$-PEESSLEYQQE**Y**F**V**TT-$_{3776}$). Oligomeric ezrin may thus exhibit avidity effects upon binding to ANK2. The result illustrates how ProP-PD can complement existing interaction data with insights at the level of the amino acids that are involved in the interaction.

Merlin FERM domain ProP-PD data is enriched in a smaller but more diverse set of ligands

The screening of merlin-binding peptides was performed through ProP PD selection assay in the same way using merlin FERM as a bait. The extraction of merlin-binding peptides from the phage display results was more challenging as only 7 peptides remained after applying the same filtering as for the ERM proteins. The main reason for this was that one dominant peptide (FAM83G$_{111-126}$) obtained close to 90% of all counts in the NGS analysis. We therefore relaxed the NGS data filtering criteria (replicate selections, cut-off $\geq 0.01\%$). To ensure that the relaxed filtering criterion was valid, we tested binding for a selected set of peptides through clonal phage ELISA. The tested set included peptides that passed the relaxed filtering, as well as peptides from the filtered-out cohort of peptides (e.g. low counts or no replicate). We found that all but one of the peptides from the first category were successfully confirmed as binders, while only three of the peptides from the second category bound to merlin (**Fig. 2A**). The assigned filtering parameters and cut-off value were thus efficient in separating true binders from non-binders.

The final set of merlin hits contained 30 peptides, matching 42 sequences in the proteome in 38 proteins (Table S4). However, in contrast to the ERM data there was no clear consensus motif. Inspection of the hits revealed that around half of the merlin ligands contain the ERM binding motif [FY].[FILV]. In the remaining set we noticed the already know merlin binder LATS1 ($_{73}$-THHKA**LQEI**RNSLLPF-$_{88}$) that binds merlin's F2 pocket to through a -

LQEI- stretch [18]. Similar sequence motifs were noticed in two novel merlin ligands, $_{1365-}$RAATR**LQEL**LKTTQEQ$_{-1380}$ from RRBP1 and $_{37-}$PQAASH**LQEL**LASTES$_{-52}$ from IRGC, and we would expect them to bind to the F2 pocket. The FAM83G (also called PAWS1) peptide ($_{111-}$AEPLPSLEYWPQKSDR$_{1-26}$) was the most enriched merlin ligand but does not conform to the general ERM consensus or the LATS1-like peptides, and may thus represent yet another type of motif (**Fig 2A-B**). To gain more information about the key determinants of the merlin binding peptides, we selected two peptides, FAM83G$_{111-126}$ and NOP53$_{192-207}$ for further characterization. NOP53 has previously been shown to interact with merlin using its 181-479 region (17). The determinants of the interactions were tested through a mutational analysis and the effects of the point mutations were evaluated by clonal phage ELISA, where radical changes (mainly a combination of lysine and alanine scanning) were introduced for a clear readout in the relatively insensitive binding assay (**Fig 2C**). For FAM83G$_{111-126}$ we found that the central -EYWP- stretch was necessary and sufficient for binding, as mutations of any of these residues abrogated binding and truncation of either the N-terminal region (AEPLP deletion) or the C-terminal region (KSDR deletion) only had minor effects. For NOP53$_{192-207}$ we found that the interaction required an extended FYDLW….PLD stretch, indicating that the interaction with this peptide required more than the apparent [FY].[FILV] motif. This was confirmed by the finding that a C-terminal truncation of the peptide (deletion of PLDRPLV) conferred loss of binding. The analysis thus established that the limited set of merlin ligands contained several distinct binding motifs. Since ten of the ProP-PD merlin peptide hits overlapped with the ERM ligands (**Fig. 2D**), we further explored the overlapping specificities of the ERMs and merlin through clonal phage ELISA. This revealed that the ERM FERM domains can bind to all tested merlin ligands, including the merlin specific peptides FAM83G$_{111-126}$ and RRBP1$_{1365-1380}$, although likely with varying affinities (**Fig 2E**).

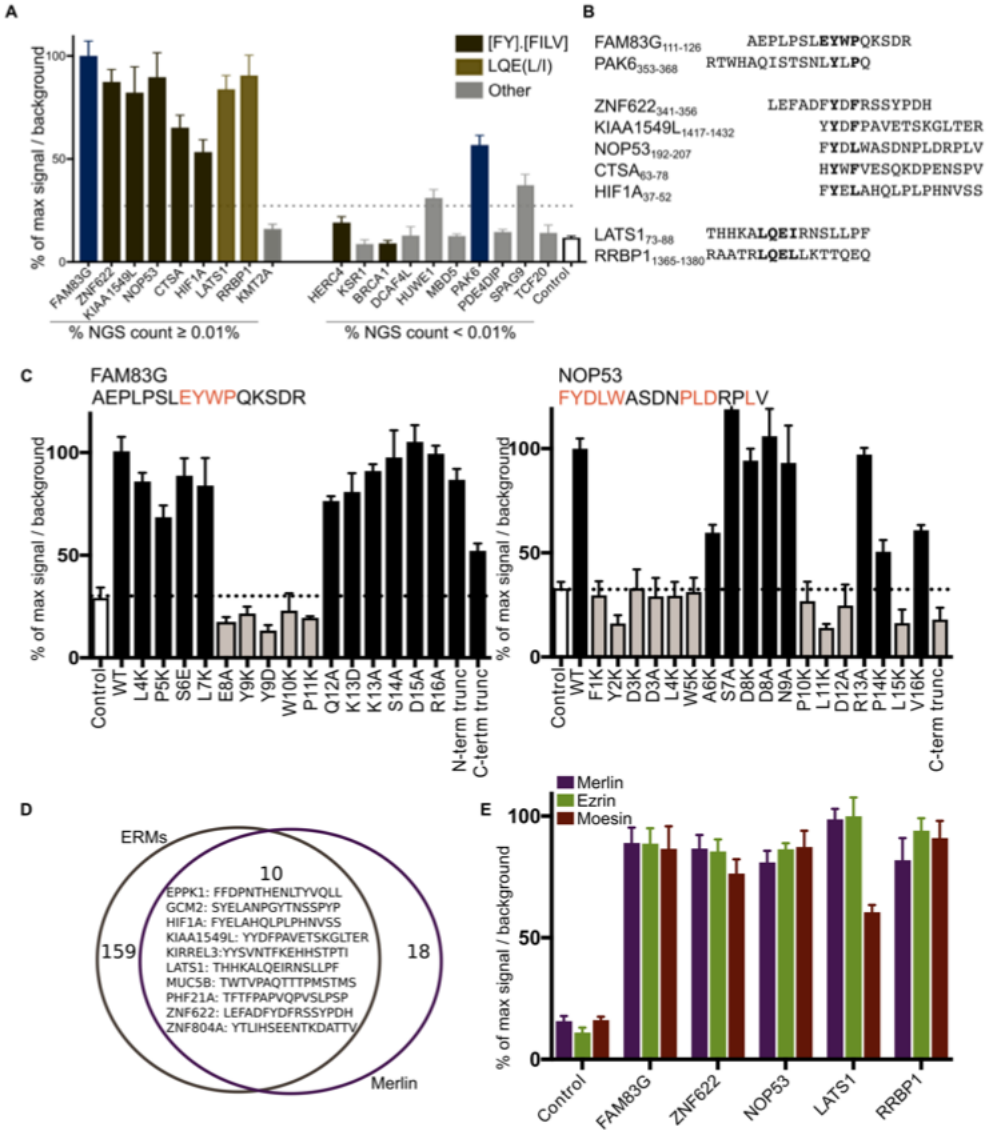


Figure 2. Analysis of peptide hits identified as merlin FERM domain ligands in ProP-PD selections. **(A)** Clonal phage ELISA using immobilized GST-tagged merlin FERM domain in ELISA against phage displaying indicated peptides (sequences in B). The y-axis represents the ratio between the A450 signal obtained for ELISA of GST-tagged merlin FERM in comparison to the background signal obtained for GST. A phage pool of wild-type phage (e.g. no peptide displayed) was used to establish the background signal (indicated as control). The analysis confirmed the interactions between merlin FERM and most peptides associated with relative NGS counts ≥0.01%. In contrast most peptides below the cut-off showed no or marginal binding. The binding signal is presented relative to FAM83G$_{111-126}$, and a cutoff of 25% relative activity defines substrates based on at least 2 x the signal of the negative control. **(B)** Sequence alignments of confirmed merlin FERM binding peptides. Ligands are grouped based on apparent consensus motifs. **(C)** Point mutations of key amino acids in the FAM83G$_{111-126}$ and NOP53$_{192-207}$ sequences affected binding of merlin FERM as detected by clonal phage ELISA. **(D)** Shared ERM and merlin ligands, as revealed by ProP-PD selection. **(E)** Clonal phage ELISA demonstrates binding of moesin and ezrin FERM domains to the same ligands as the merlin FERM domain, even if these were not identified in the ProP-PD selection (e.g., merlin ligands FAM83G$_{111-126}$ and RRBP1$_{1365-1380}$).

.

We enriched the merlin data set for interactors of potential biological relevance following the same procedure as the ERM data. This set of high confidence merlin ligands contains 12 peptides equally many proteins (although the EPPKI occurs multiple times in the protein, **Table S4**), of which three are previously reported interactors (LATS1 [12, 31], NOP53 [32], and HIF1A [33]). FAM83G$_{111\text{-}126}$ was further included in the high confidence set as it was consistently the dominating ligand in the merlin ProP-PD selection, and because it plays a role in regulation of the actin and focal adhesion dynamics, and cell migration [34]. About half of the GO term enriched set of merlin ligands were shared with the ERMs.

GST-pulldowns confirm that binding motifs are accessible to the FERM domains in the full-length proteins

The phage displayed peptides represent regions of human proteins and the selections identify thus putative interaction partners in the proteome. However, it is necessary to evaluate if the binding sites are relevant in the context of the full-length proteins, as the binding interfaces may not be accessible in the host proteins. We therefore validated interactions with 7 full-length proteins through GST-pulldown experiments using moesin and merlin FERM domains as baits (**Fig 3**). Through these experiments, we confirmed that the moesin and/or merlin FERM domain interact with the FAM83G, HIF1A, LATS1, NOP53, PAK6, RRBP1 and ZNF622. Through mutational analysis we further confirmed that the identified binding motifs are crucial for the interactions with the full-length proteins. We noticed that the target proteins have overlapping but distinct specificities for the FERM domains, as for example the HIF1A interaction was only confirmed for moesin, and the PAK6 and RRBP1 interactions were only confirmed for merlin (even though the clonal phage ELISA experiment with the peptide revealed interaction of RRBP1 with both, **Fig. 2E**). Among the confirmed interaction sites we further note an additional, previously not reported FERM binding site in LATS1.

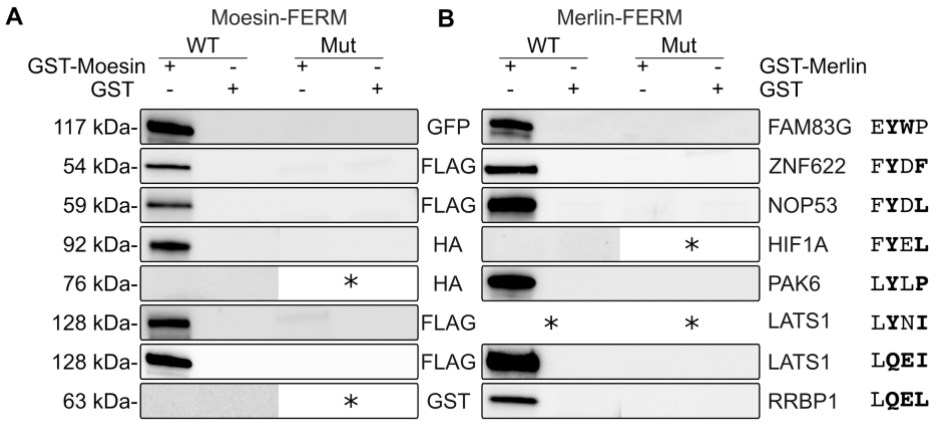


Figure 3. GST-tagged moesin FERM **(A)** or merlin FERM **(B)** were used to pull down ligands transiently expressed in HEK293 cells. Mutated consensus motif sequences are indicated to the right (bold residues

were mutated to A). Note that LATS1 has two moesin FERM domain binding regions, and both are crucial for the interaction with the full length protein. Results shown are representative of at least two replicated experiments. * indicates not tested.

ProP-PD derived ligands bind to moesin and merlin with micromolar affinities

To learn more about binding affinity and specificity in the FERM domain family, we determined the affinities of the moesin and merlin FERM domains for a set of ligands, namely KIRREL3$_{637-652}$, TBX4$_{428-443}$, ZNF622$_{341-356}$ and NOP53$_{192-207}$ as variants of the x[FY]x[FILV]x group, and FAM83G$_{111-126}$ as a unique case. The first two and latter two are ProP-PD derived moesin and merlin ligands, respectively, while ZNF622$_{341-356}$ was found in the datasets of both proteins. We also included the LATS1$_{73-88}$ peptide in the affinity measurements as a reference for the F2 pocket [18], and the EBP50$_{343-358}$ peptide as a reference for the F3a pocket [13, 14] (**Fig. 4A-B** and **Table 1**). We found the K_d values for interactions with moesin to be in the low micromolar range, except for KIRREL3$_{637-652}$ (10x higher affinity), and FAM83G$_{111-126}$ (30x weaker affinity, as expected, since it was not found among the moesin ligands in the phage selection experiment). For merlin, we found that ZNF622$_{341-356}$ was the best merlin ligand, followed by NOP53$_{192-207}$ and FAM83G$_{111-126}$. The affinity for FAM83G$_{111-126}$ was low (93 µM), which is somewhat surprising considering that it was the predominant ligand of merlin in the ProP-PD results. A plausible explanation for this may be the established bias for tryptophan containing peptides in phage display experiments [35]. Since the merlin FERM domain was rather challenging to use in affinity measurements, we decided to use moesin as a model protein for further analysis of FERM domain binding specificity.

Unveiling the complexity of the [FY]x[FILV] motif

The ERM FERM ProP-PD data are dominated by the shared consensus motif [FY]x[FILV] (**Fig. 1**). However, as the FERM domain has multiple binding pockets, it was expected to select ligands with distinct motifs that engaged in interactions with the different pockets. We therefore designed a set of competition experiments to obtain clues into whether different FERM ligands bind to the same or distinct pockets. In these experiments, we used FITC-labeled EBP50$_{343-358}$, ZNF622$_{341-356}$, KIRREL3$_{637-652}$ and NOP53$_{192-207}$ as probes and attempted to outcompete them with an excess of unlabeled peptides (**Fig. 4C-E**). This analysis revealed a complex competition pattern, where the EBP50$_{343-358}$ and ZNF622$_{341-356}$ peptides outcompeted all probe peptides with K_I values similar to their respective K_D values for direct binding. In contrast, KIRREL3$_{637-652}$ and NOP53$_{192-207}$ efficiently outcompeted only their own respective probe peptides, which suggested that these ligands bind to distinct moesin FERM pockets despite sharing a common [FY]x[FILV] motif. The results further suggested that EBP50$_{343-358}$ and the [FY]x[FILV] containing ZNF622$_{341-356}$, likely bind to the same pocket (F3a) and that they outcompete ligands binding to the other sites allosterically (**Fig. 4** and **Table 2**). Thus, the apparent [FY]x[FILV]

motif hides at least three different types of ligands. Along this line, we note that binding of the NOP53$_{192-207}$ peptide to merlin involves additional residues as shown by mutational analysis (**Fig. 2**). Testing the competition of an additional set of 11 peptides (**Table 2**) allowed us to define a consistent competition pattern, where **i)** FYDF containing ligands competed with all probes with probe-independent K_I values, **ii)** xYxV type peptides efficiently outcompeted FITC-labeled KIRREL3, and **iii)** FY(D/E)L containing peptides only efficiently outcompeted each other. Taken together, the analysis unveiled a hidden complexity of the [FY]x[FILV] motif.

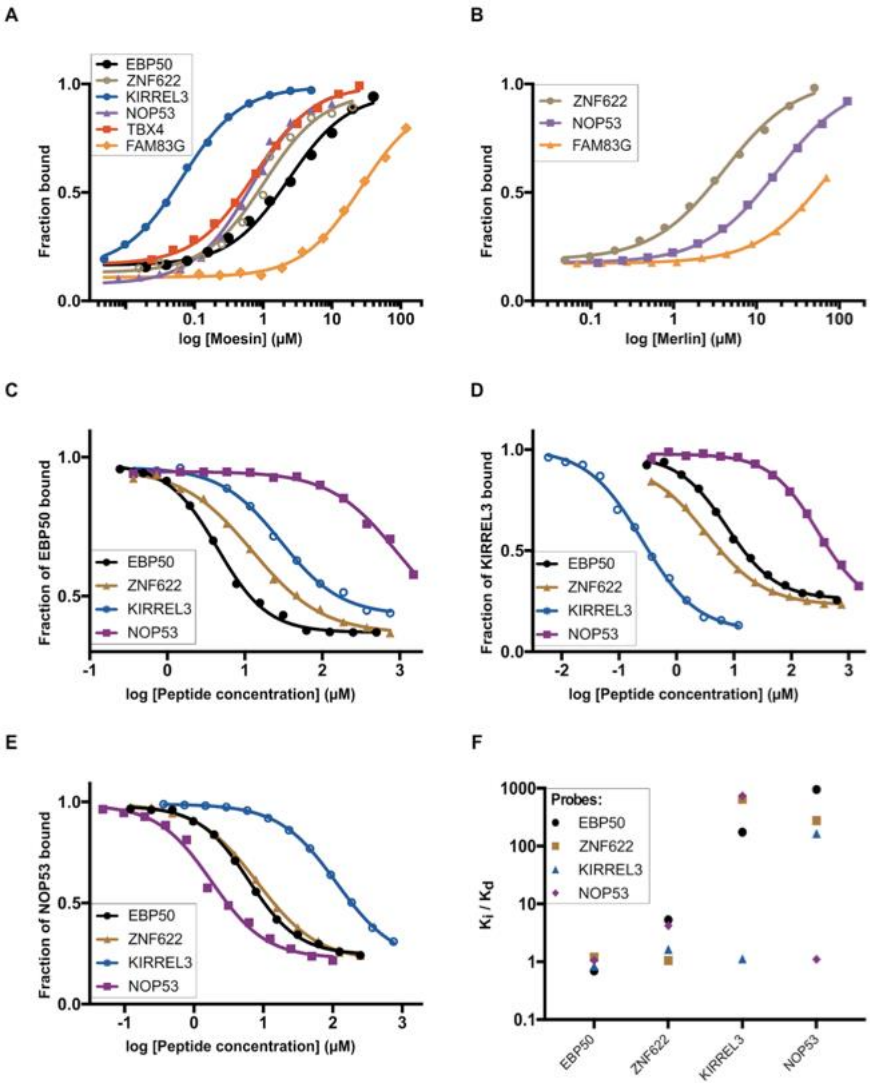


Figure 4. **Complex competition pattern for moesin ligands revealed by FP competition experiments.** Moesin **(A)** and merlin **(B)** bind to their ligands with nano or micromolar affinities (see also Table 1). **(C-E)** Unlabeled peptides were assayed for their ability to outcompete FITC-labeled EBP50$_{343-358}$ (**C**), KIRREL3$_{637-652}$ (**D**) or NOP53$_{192-207}$ (**E**). **(F)** Ratio between the apparent affinities (K_I, as determined by competitive FP) and direct binding (K_D, as determined by direct binding). While the K_I values of ZNF622$_{341-356}$ and the F3a binding EBP50$_{343-358}$ are probe independent, and similar to the K_D values in all cases, the KIRREL3$_{637-652}$ and the NOP53$_{192-207}$ peptides efficiently outcompete only themselves.

To gain additional information regarding the interplay between the distinct ligands we performed ProP-PD selections in the presence of saturating concentration of EBP50$_{343-358}$,

TBX4$_{428-443}$ or NOP53$_{192-207}$, each representing a different motif class. The results of this selection experiment clearly demonstrate that the presence of EBP50, and to less extent TBX4$_{428-443}$, reduces the enrichment of several peptides (**Fig. 5A)**. In contrast, NOP53$_{192-207}$ had less effect on the cohort of selected peptides, except for FYEL containing HIF1A$_{37-52}$, in line with the direct competition between NOP53$_{192-207}$ and HIF1A$_{37-52}$ revealed by the FP competition analysis (**Table 2**). Interestingly, the presence of either EBP50$_{343-358}$ or TBX4$_{428-443}$ led to a relative increase in the enrichment of phage particles displaying the F2 binding LATS1$_{73-88}$ peptide (**Fig. 5A**). This likely resulted from EBP50$_{343-358}$, and to less extent TBX4$_{428-443}$, outcompeting other ligands. However, since the effects of the two peptides on the enrichment of the LATS1$_{73-88}$ were striking, we hypothesized that the effect could be enhanced by a positive allosteric effect. To test this hypothesis, we measured the direct binding of LATS1$_{73-88}$ in the presence of saturating concentrations of EBP50$_{343-358}$, TBX4$_{428-443}$ or NOP53$_{192-207}$. Consistent with the phage display results, the presence of EBP50$_{343-358}$ or TBX4$_{428-443}$ lead to an increased apparent affinity of moesin for LATS1$_{73-88}$ (**Fig 5B**), suggesting that the ligands may prime the domain for LATS1$_{73-88}$ binding. Unlabeled LATS1$_{73-88}$ in turn was found to efficiently outcompete FITC-labeled EBP50$_{343-358}$, TBX4$_{428-443}$ and NOP53$_{192-207}$. A detailed analysis of K_D values determined as a function of the second ligands would be needed to clarify the mechanistic details of these potentially allosteric events.

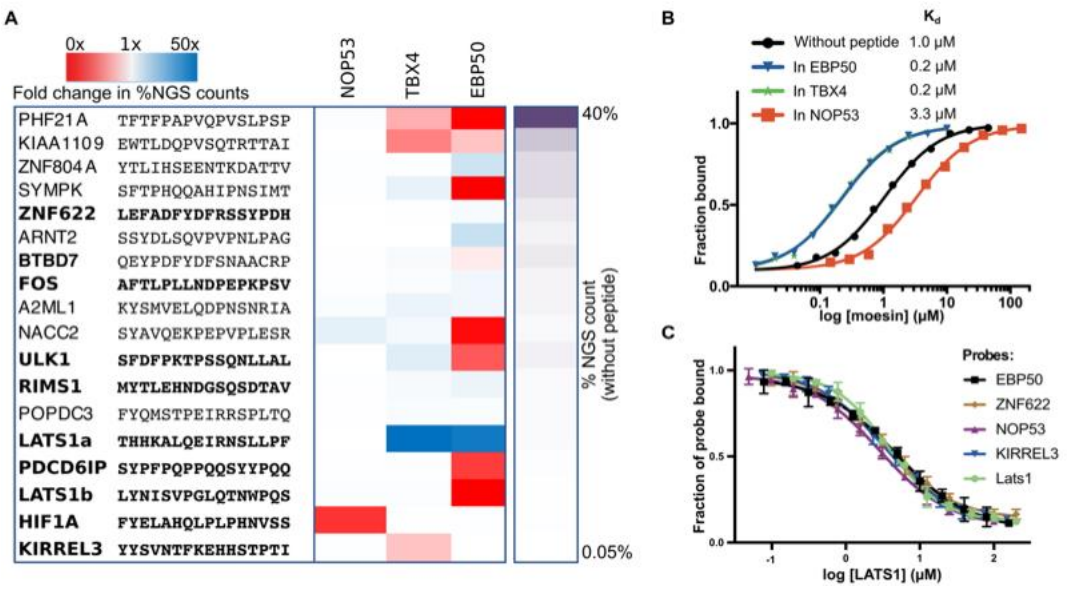


Figure 5. LATS1 binding is affected by other bound peptides, and outcompetes all. (A) Competitive ProP-PD selections reveal that the presence of NOP53$_{192-207}$, TBX4$_{428-443}$ or EBP50$_{343-358}$ have differential effects on the relative representation of the selected cohort of peptides. Bold peptide names and sequences indicate that affinity data is available in Table 2. The exceptional pattern observed for LATS1$_{73-88}$, whose selection is increased upon binding of the latter two ligands, is reconfirmed in (**B**), by FP affinity determination of moesin FERM for FITC-labeled LATS1$_{73-88}$ in the presence and absence of a saturating concentration of NOP53$_{192-207}$, TBX4$_{428-443}$ or EBP50$_{343-358}$. (**C**) Unlabeled LATS1$_{73-88}$ outcompetes FITC-labeled EBP50$_{343-358}$, KIRREL3$_{637-652}$, NOP53$_{192-207}$ and ZNF622$_{341-356}$ as determined by FP.

Establishing the structural basis of the different observed competition patterns

The phage display results and the binding assays revealed that the FERM domains bind four main classes of motifs: FYDF, xYxV, FY(D/E)L and LQE(I/L). Of these motifs, it was already known that the LQEL motif of $LATS1_{73-88}$ binds to the F2 pocket of merlin [18]. To better understand the binding of the other types of ligands, we applied peptide docking tools from the Rosetta FlexPepDock modeling suite [36] to model the structures of these interactions, using two main approaches: **i)** We applied FlexPepBind [37] to independently define the binding motif contained within the 16 amino acid long peptides. This protocol utilizes solved protein-peptide complexes to which different peptide sequences can be threaded in order to differentiate between binders and non-binders, and in the case of a binder, to identify the bound peptide conformation. **ii)** Following the suggested motif definition, we applied a blind, global docking protocol - PIPER FlexPepDock [21] to confirm the binding site and peptide conformation. In this approach a pre-defined set of fragments that represents the peptide conformational ensemble is rigid-body docked to the receptor protein to sample all possible orientations of the peptide relative to the receptor, followed by full-atom resolution of the top-scoring complexes.

For our simulations we used the available solved structures of moesin FERM domains, including the free moesin FERM domain (PDB ID 6TXQ [38]), the structure of moesin FERM domain with its own C-terminal domain (CTD) bound to the F3a and F2 sites (PDB ID: 1EF1 [11]), moesin with the crumbs CTD bound at the F3b site and in the cleft between the F3 and F1 lobes (PDB ID: 4YL8 [17]) and moesin with the CD44-derived peptide bound to structures the F3b site (PDB ID: 6TXS [39]).

The structural variability of F3a -binding FYDF containing peptides

The experimental results showed that the FYDF containing $ZNF622_{341-356}$ and $BTBD7_{940-950}$ peptides exhibit a similar competition pattern as the F3a binding $EBP50_{343-358}$ peptide [13]. We therefore attempted to thread these peptides onto the existing template of moesin bound to its CTD. However, no conclusive results could be obtained for the putative F3a binding peptides, although the two binding sites covered by moesin's CTD (F3a and F2) were identified

To further investigate where and how the peptides bind the FERM domain, we therefore performed a global docking simulation on the full FERM domain, as well as on the F3 subdomain. Docking onto the full FERM domain positions the 10 top scoring models of the $ZNF622_{341-356}$ peptide at the F3a site, as well as in the cleft between the F3 and F1 domains (**Fig. S1**). In turn, docking the peptides onto the isolated F3 domain located the peptide in the F3a and F3b binding sites (See **Fig. 9A** discussed below). Notably, in all global and local simulations the top-scoring peptide models are localized to F3a. These F3a bound peptide models do not converge to one single conformation, but rather adopt two distinct orientations: one similar to the part of the CTD bound at F3a, and another perpendicular to it (**Fig. 6A**), the

latter including the top scoring model. Notably, even though the backbone conformation of the peptide is different from the known binding mode at this site, the hydrophobic side-chains of the top-scoring model occupy the exact same sites as the CTD of moesin (**Fig. 6C**). The preference between these two possible orientations could be determined by the flanking regions, as for CTD and EBP50 [13].

Similar results were obtained for the BTBD7 derived peptide. Global docking simulation onto the F3 subdomain of the moesin template placed most of the top-scoring models onto the F3a site, however, no single defined binding conformation stood out. In addition, simulation on the full moesin template structure (after removal of the CTD) predicted the cleft between F3 and F1 binding lobes as the most probable binding site. Although the binding motif is very similar between the two peptides (BTBD7: EYP**DFYDF** vs. ZNF622: EFA**DFYDF**), PSIPRED [40] predicted the ZNF622 but not the BTBD7 peptide to form a helix, probably because of the presence of a proline upstream of the motif. Consequently, most of the starting fragments used for BTBD7 docking were in extended conformation. Since the quality of the initial fragments greatly affects the protocol performance, we ran another simulation, explicitly defining the BTBD7 fragment to be helical. This simulation gave very similar results to those obtained for ZNF622, with the top-scoring models concentrated at the F3a site, and the BTBD7 peptide adopting the same "non-canonical" conformation as the ZNF622 derived peptide (**Fig 6B**).

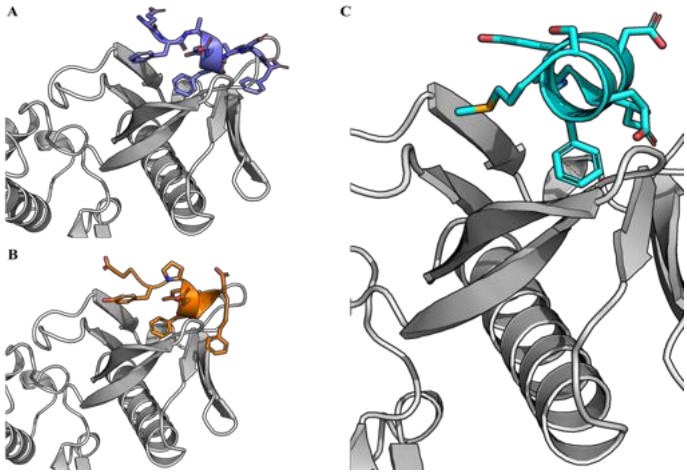


Figure 6. Structural modeling suggests that ZNF622 and BTBD7 bind the Moesin FERM domain at F3a using the same binding pocket as moesin CTD and EBP50. A-B. Models of the interactions of ZNF622 (**A**) and BTBD7 (**B**) that both suggest a perpendicular conformation, however using aromatic side-chains to fill the same hydrophobic pocket. As expected, this distinct conformation is only identified by a global docking simulation but not by threading. **C.** Fragment of moesin CTD at the same site (PDB code: 1EF1).

Structural modeling of the F3b binding xYxV motif

We next focused on the KIRREL3$_{637-652}$, TBX4$_{428-443}$ and MISP$_{595-610}$ group that shares a xYxV motif, and shows a common competition pattern (**Table 2** and **Fig. 4D,F**). A closer inspection

of the moesin binding $MISP_{595-610}$ revealed that its **-GSYSV-** sequence is similar to a **-GTYSP-** stretch of crumbs that is crucial for its binding to the F3b site of moesin [17], and the ICAM2 **-GTYGV-** peptide that has been co-crystallized in the F3b site of radixin (PDB 1J19 [16]). We therefore decided to apply the FlexPepBind approach using two existing F3b bound structures (PDB codes 4YL8 and 6TXS), as templates. We started by threading the $MISP_{595-610}$ peptide onto the peptides solved in these structures (the crumbs peptide and CD44, respectively), to independently identify the motif that could potentially bind at this site. The results converged for both simulations, identifying the IT<u>GSY</u>SVS sequence as the best motif for the given binding site (**Fig. 7A,D)**. To reaffirm this binding mode, and to make sure we did not miss another possible binding site, we used the identified motif in a global docking simulation with PIPER-FlexPepDock. The top-scoring structure identified in this simulation hit the same site as the one identified by threading in a very similar binding conformation (**Fig. 7A**).

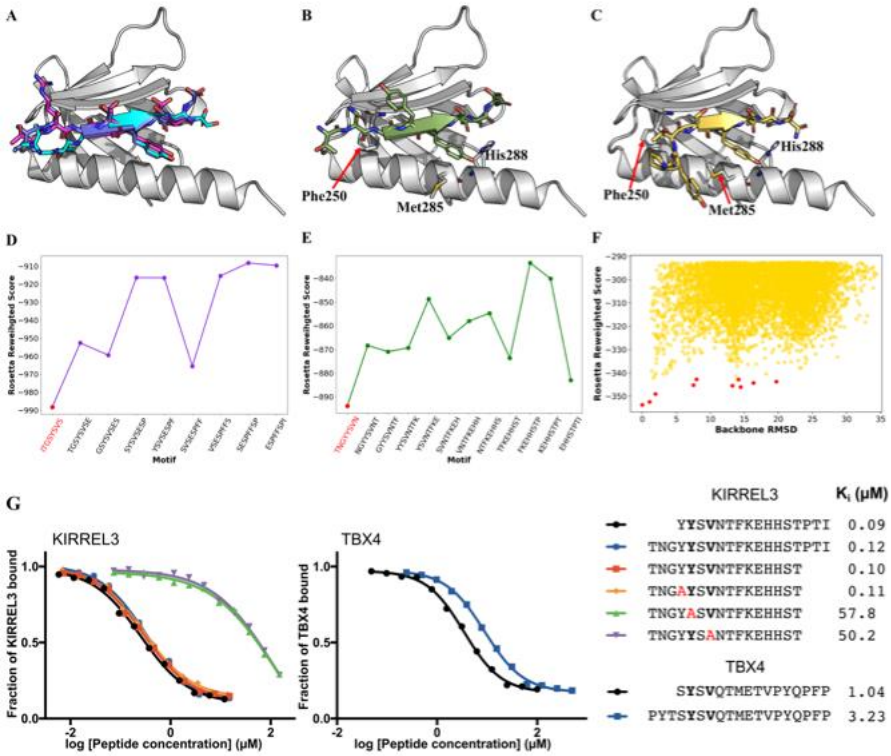


Figure 7. The MISP3, KIRREL3 and TBX4 top-scoring peptide models adopt similar conformations as the crumbs peptide when bound in the F3b binding pocket. **A**. Bound crumbs peptide (PDB id 4YL8) shown in magenta, MISP3 top-scoring model from FlexPepBind threading and PIPER-FlexPepDock global docking are shown in purple, and cyan, respectively. **B, C**. Top-scoring threading model of KIRREL3 and the global-docking model of TBX4 peptides are shown in green and yellow, respectively. The supposedly crucial residues that interact with the peptides are labeled. For each model, the corresponding energy landscapes are shown for threading and global docking simulations. **D-E.** Threading binding energy landscapes of MISP and KIRREL, respectively. The identified motif is highlighted in red. **F.** Global docking simulation binding energy landscape of TBX4, in which the lowest energy models converge towards the F3b site. **G.** Competition FP experiment using variants of the KIRREL3 peptide (left; sequences indicated to the right) or the TBX4 peptide (middle) for competition, and FITC-KIRREL3 as probe. Mutations of the core xYxV motif in KIRREL3 confer loss of affinity, while an N-terminal extension of the KIRREL3 peptide does not affect binding affinity. In contrast, we noted a minor decrease in affinity upon an N-terminal extension of the TBX4 peptide (middle).

The KIRREL3$_{637-652}$ and TBX4$_{428-443}$ peptides belong to the same group as MISP$_{595-610}$, with -S**Y**S**V**- and -Y**Y**S**V**- motifs in the peptide sequences of TBX4$_{428-443}$ and KIRREL3$_{637-652}$, respectively. However, they show distinct amino acids upstream of the motif. Both threading and global docking of the identified TBX4 and KIRREL3-derived peptides failed to yield conclusive results. We therefore extended each motif at the N-terminus to include the upstream residues. For KIRREL3$_{637-652}$, we added -TNG- to the motif -S**Y**S**V**-. The resulting peptide, TN**GSY**SV contains a -GS**Y**- motif that matches the crumbs sequence (featuring a glycine at position p-2). In contrast, the upstream residues of the TBX4 peptide are -PYT, and the extended peptide will feature threonine at p-2 instead, which could potentially interfere with binding. We used the same strategy as implemented for MISP$_{595-610}$ to model the extended peptides: FlexPepBind using the F3b bound template to identify the binding motifs, followed by global docking simulation of the best motif. For KIRREL3 the simulations converged in the same way as for MISP: the TNGY**Y**S**V**N motif identified by the FlexPepBind simulation was ranked best in the global simulation, adopting the same binding conformation as in threading **(Fig. 7B, E**). For the TBX4 peptide, threading failed to identify any specific motif. However, when an extended motif was used (PYTS**Y**S**V**S), global docking placed the peptide in a very similar position as the previous two peptides, with a difference noted at the N-terminal part of the peptide **(Fig. 7C, F**). As mentioned by Wei *et al.* [17] in their description of the moesin-crumbs complex, the glycine at the p-2 position allows tight packing with moesin residue F250 due to the lack of the side-chain (**Fig. 7B**). Since the TBX4 peptide features a threonine instead of a glycine at the p-2 position, the peptide has likely to be rearranged in this region, causing a change in the conformation of the N-terminal part of the peptide (**Fig. 7C**), which also explains why threading failed to identify the motif in this case.

To validate the extended motif, we experimentally tested how an N-terminal extension of the peptides with three amino acids would affect the affinity of moesin for KIRREL3 and TBX4 **(Fig. 7G**). Notably, we found that the N-terminal extension of the KIRREL3 peptide did not change the affinity (K_D values for KIRREL3$_{637-652}$ and KIRREL3$_{6334-652}$ being 0.09 μM and 0.12 μM, respectively). In contrast, we noted a slightly reduced affinity for TBX4, which likely is due to a failure of the bulkier threonine at p-2 to pack against moesin F250. Hence, although being necessary for docking, the upstream residues do not contribute to the affinity, but may cause minor steric clashes with the domain. Possibly, more rigorous sampling of the receptor rearrangement is needed in global docking when a bulky amino acid is present instead of glycine in order to model a more standard ®-addition peptide binding. We further confirmed the importance of the key residues of the xYxV motif through a mutational analysis of the KIRREL3 peptide, validated that the tyrosine at position p1 and valine at position p3, but not tyrosine at position p-1, are crucial for binding.

A diffuse binding pattern for the FY(D/E)L containing NOP53 and HIF1a-derived peptides

Docking simulations of the peptides in the third group, NOP53 and HIF1a, showed a diffuse distribution of binding modes that covered predominantly the F1-F3 cleft, and other interdomain sites (**Fig. S2**). Thus, while our structural models of peptide-FERM domain complexes can explain the observed binding and competition ability of the peptides in the two first classes, explaining the third class is more challenging, reflecting the difficulties to explain the experimental results: While these peptides are outcompeted by the peptides from the first group (F3a binders), they themselves are unable to compete, indicating that they bind in a different way, as also observed in our simulations.

To summarize, the models generated using the FlexPepDock protocols are in overall agreement with the experimental results and suggest full atomic structures of the complexes that can be used to further examine the binding mode and to manipulate the interactions.

Ligand binding induces dynamic conformational changes in the moesin FERM domain

The docking experiments identified the binding sites of the distinct classes of peptides. However, they did not explain the complex competition pattern observed in the experiments, with the F3a binding peptides outcompeting the F3b binding ligands, as well as the FYD(E/D)L containing ligands. To investigate a possible allosteric communication between the sites, we explored the dynamics of the FERM domain.

Comparison of structures co-crystallized with different peptides bound at different binding sites (moesin bound to its CTD at F3a and F2 (PDB code 1EF1*)*, moesin with the CD44 peptide bound at F3b (6TXS), and the free moesin structure (6TXQ)), revealed a significant conformational change localized to the F3 and F2 lobes, dependent on the peptide binding at F2, F3a and F3b sites. When the F3a site is occupied, the ®-sheet constituting the F3b site moves down closer to a helix, supposedly closing the F3b site (even more so than in the unbound conformation) (**Fig. 8A**). Alternatively, when the F3b is occupied by a ligand that provide an additional strand, the ®-sheet rises, opening the binding site at the grove between the ®-sheet and the helix (**Fig. 8B**). As expected, the free moesin structure resides in a conformation between the two states.

To recapitulate the ligand induced conformational changes, we started from the bound conformation, removed the binding partners, and relaxed the structures using the Rosetta FastRelax protocol allowing the structure to move without any constraints [41]. For these simulations we used the F3b bound structure 6TXS and the 1EF1 structure, bound at F3a and F2, and compared them to the unbound structure 6TXQ. In both cases, the relaxed structures approached the solved free conformation (**Fig. 8A,B,D,E**). The two regions directly involved

in binding at the F3b site, namely the beta edge strand and the helix, moved significantly (**Fig. 8D, E**) to close the binding pocket when relaxing an F3b-site bound conformation or to open it when relaxing a CTD bound structure.

More interestingly, structural alignment of the F3b bound peptide (derived from CD44) onto the free structure (6TXQ), followed by the relax protocol, enabled it to generate a bound-like conformation (**Fig. 8C**) that can be used in subsequent docking simulations, as described below. Overall, this data explained how one group of peptides (*i.e.,* those binding to F3a) can outcompete all the other binders including those binding at a distinct pocket *(i.e.,* F3b) by inducing a conformational change at the binding site.

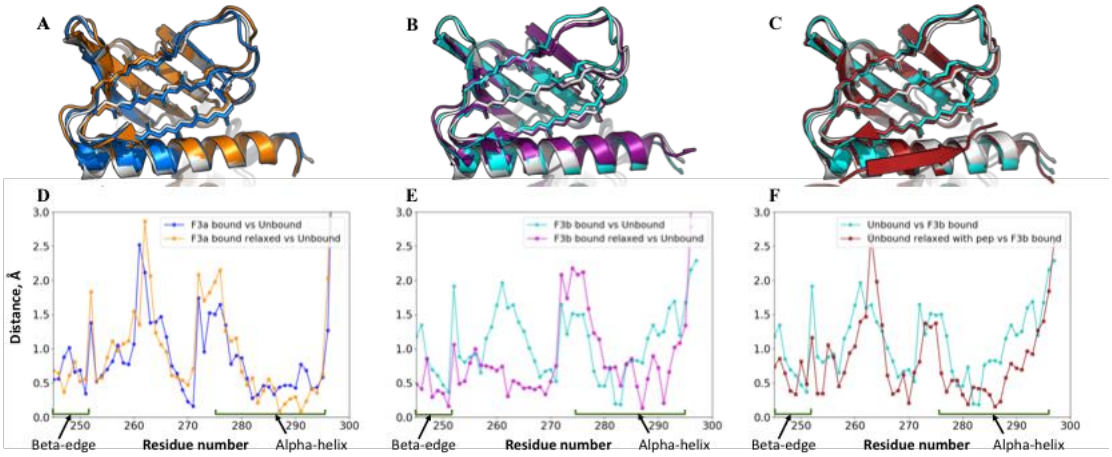


Figure 8. Conformational change at the F3b binding site of Moesin depending on the peptide bound. (A, B) Relaxing bound structures with the peptide removed brings the structures back to the unbound state (crystal structure 6TXQ, grey): Shown are **(A)** the relaxed structure model (orange), starting from a structure bound at F3a site (crystal structure 1EF1, blue), and (**B**) the relaxed structure model (purple), starting from a structure bound at F3b site (crystal structure 6TXS, cyan). **(C)**. Relaxing the unbound structure with superimposed peptide at the F3b site mimics the F3b bound conformation (6TXS, cyan): Shown are the unbound state (6TXQ, grey) and the relaxed model (red). **D-F:** Corresponding plots of the distances of different residues in the receptor between the bound and unbound states, as well as between the relaxed model and the unbound **(D,E)**, and bound **(F)** states, respectively. Note the effect of relaxing the structures on residues involved in binding the peptide (highlighted by arrows).

FERM F3 subdomain dynamics affects the affinity of the ligands

Our analysis of available FERM domain structures thus highlighted a potential allosteric communication between the F3a and F3b sites such that binding of ligands to the F3a site blocks binding to the F3b site (**Fig. 8)**, in line with the experimental results (**Table 2**). To further test this theory we checked how the use of the structures with different ligands occupying the binding distinct sites affected our simulations.

We started with the ZNF622 peptide, predicted to bind at F3a (**Fig. 6A**), and checked whether it would still reach the binding site, when using a F3b bound receptor (6TXS). We found that although some false positive conformations appeared due to the open binding site created by the co-crystallized peptide, the expected conformation at the F3a was still found

among the top-five models with peptide backbone RMSD to the conformations found in bound simulations of 3.3 Å (see **Fig. 9A, B**). For the MISP peptide that binds at F3b (**Fig. 7A**) we checked whether the peptide will reach its binding site, when an unbound structure is used for the simulation (PDB id 6TXQ). This simulation was not able to localize MISP to the F3b site (best peptide backbone RMSD among the top-five structures is 9.7 Å away from the predicted binding conformation) (**Fig. 9D**), despite successful identification of the F3b site when a bound receptor conformation was used (**Fig. 9C**). Nevertheless, using the structure generated by relaxing a structure with a superimposed peptide (**Fig. 8C**) a bound conformation was observed as the top-scoring structure at the F3b site (backbone RMSD of 2.6 Å from the top-scoring result of the bound simulation) (**Fig. 9E**). This emphasizes the importance of opening the F3b binding site prior to docking, as mentioned previously. (Note that starting from an F3a bound conformation was not possible, as that structure contains also an F3b bound crystal contact tail, which may be misleading, biasing the simulation towards the (bound) F3b site).

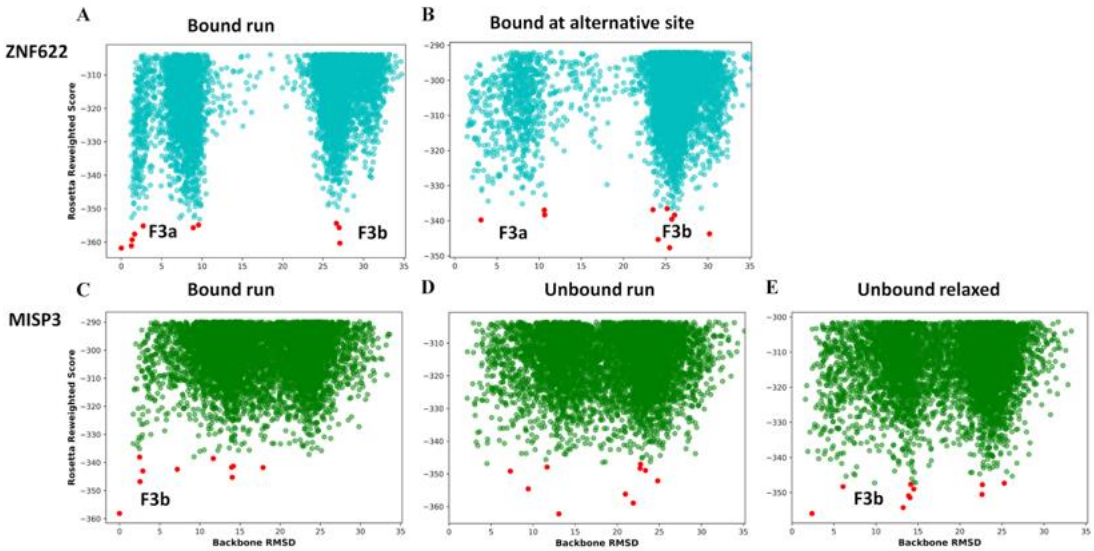


Figure 9. Energy landscapes of simulations using different receptor structures. (**A,B).** Energy landscape sampled by the ZNF622 peptide: The F3a site is accessible both in F3a bound **(A)** and F3b bound **(B)** structures. **(C-E)**: Energy landscapes demonstrate that the F3b site is inaccessible in the unbound FERM structure: Docking of the MISP peptide onto the bound **(C)**, unbound **(D)**, and unbound-relaxed structure (**E**, with superimposed peptide at the binding site) highlights the importance of conformational change for the identification of the F3b binding site. All plots are relative to the predicted F3a (A, B) and F3b (C, D, E) binding sites, with additional binding sites highlighted. Only the F3 subdomain was used in the simulations.

CONCLUDING REMARKS

ProP-PD has previously successfully been used to determine binding specificities of protein domains with single binding pockets [20, 42-44], as well as two pockets [45]. Similarly the computational peptide docking was previously shown to successfully predict the structure of protein-peptide complexes even in cases where the binding site is unknown [21]. In this study we bring together the experimental and computational approaches to unravel a complex system

of protein-peptide interaction that includes a family of modular domains with multiple binding sites, with pockets interacting with distinct binding motifs and exhibiting allosteric communication between binding sites. Such cases are challenging for both experimental and computational structural biology when used separately: to understand the system using phage display and competition assays, at least some initial information about binding sites is needed as a starting point, while for computational blind docking approaches validation steps are essential. Notably, the affinity range of the merlin FERM domain construct used here for its ligands is lower than that of moesin, possibly due to a suboptimal construct, making merlin a challenging target for affinity measurements and for ProP-PD. Using the computational docking simulations we also found it very challenging to generate a clear picture of merlin ligand preferences, suggesting that both experimental and computational approaches perform better for higher affinity interactions.

Through the ProP-PD analysis, we were able to expand the interactomes of the ERMs and merlin, and provide information of binding sites with binding site resolution. By systematic analysis, we found that the ligand sets contained representatives of at least four classes of FERM domain ligands, FYDF, xYxV, FY(D/E)L and LQE(I/L). We note that, while having similar structures and sequences, the ERMs and merlin have both shared and distinct binding preferences. Using the information obtained from the existing structures, we built high-resolution models using Rosetta FlexPepDock-based methods that were found to be in full correspondence with the experimental data, and provided a refined model of the interactions. This approach is generally applicable to SLiM-based interactions, as long as structures of the binding domains are available.

We identified for example the moesin binding site in MISP, and clarified through a combination of competition assays and docking that its preferred FERM domain pocket is F3b. Interestingly, although having similar motifs and binding at the same binding site, the MISP, TBX4 and KIRREL3 derived peptides showed some differences in their docking results: the MISP peptide identified by the phage display included three additional N-terminal residues, that were shown of little effect to the binding affinity of TBX4 and KIRREL3 (**Fig. 7G**). However, only after addition of these residues to the TBX4 and KIRREL3 were our computational approaches able to model the complexes. This could be explained by the importance of additional downstream residues of these peptides that may contribute to the binding energy of the full 16-mer peptides used in the experiment, by not having the appropriate fragments for the docking procedure due to an irregular conformation of these peptides (which would be of partially ®-strand, partially coiled conformation), or due to the fact that the C-terminal part of these peptides would remain unstructured after binding. Adding the N-terminal

residues to these peptides allowed us to observe full ®-strand pairing of these peptides, much like in crumbs and MISP peptides.

Not all interactions could be modeled as stable structures. For the NOP53 and HIF1A peptides, we were not able to identify strong binding preferences on the energy landscape (**Fig. S2**). This may suggest that the peptides bind at various locations, possibly through fuzzy interactions [46]. Binding may thus be less defined, but still of considerable strength as the affinities of NOP53 and HIF1A for moesin are in the range of other peptides measured in this study (**Table 1 and 2**). The issues could also be due to the use of too short fragments in docking, especially since the mutational analysis of the NOP53 peptide revealed that residues downstream of the core motif are critical for binding merlin. Nonetheless, the NOP53 is efficiently outcompeted both by the F2 binding LATS1 peptide as well as by F3a binding ligands (**Table 2**), suggesting that the two ligands have negative allosteric effects on NOP53 binding.

We used a modeling approach to reveal the underlying structural basis of allosteric intradomain communication between the F3a and the F3b sites (Figs. 6-8). Using the unconstrained Rosetta Fast Relax protocol we were able to model the conformational changes inside the domain, and showed that bound structures can be relaxed back to their unbound state, and in turn, that the protocol can be used to open a binding pocket on the unbound protein structure, by superimposing a ligand from another complex and relaxing the structure (**Fig. 8C,F**). The latter is very useful in cases when the prior opening of the site is needed in order to reach the pocket at the rigid-body docking step. Similar results were also shown previously in one of the most demanding CAPRI peptide docking challenges (round 28, target 121): In that interaction, successful docking of the peptide was only possible after opening the pocket by relaxing the structure with a different peptide at the potential binding site [47]. Of course, such an approach requires a preliminary assumption about the binding site that can only be derived from existing solved complexes, but this is not limited to the solved structures of the protein of interest and can be derived from homologous proteins.

Based on the results, we propose a model that explains a number of observed interdependencies (**Fig. 10**). First, ligand binding to the F3a site has a significant negative allosteric effect on binding to the F3b site, while the inverse direction is not observed (**Fig. 4** and **Table 2**). This is explained by the sensitivity of the F3b binding pocket to the context: in the unbound FERM structure, and more so, in the structure bound at the F3a site (**Fig. 8)**, this pocket is less accessible to substrate binding (**Fig. 9D**). The closed conformation of the ERMs and merlin that involves the intramolecular binding of the C-terminal regions to the F2 and F3a sites would thus lead to conformational changes that block also binding to the F3b site and potentially other sites (e.g. NOP53 binding region). The FERM domain would then only be binding available for after activation by lipid binding and phosphorylation, in agreement with

the previous reports [14, 48, 49]. Most of the ProP-PD derived moesin interactions conform to a xYxV motif (**Fig. 10B, Table S3**), and their potential interaction with moesin in a cellular setting would thus be tightly regulated.

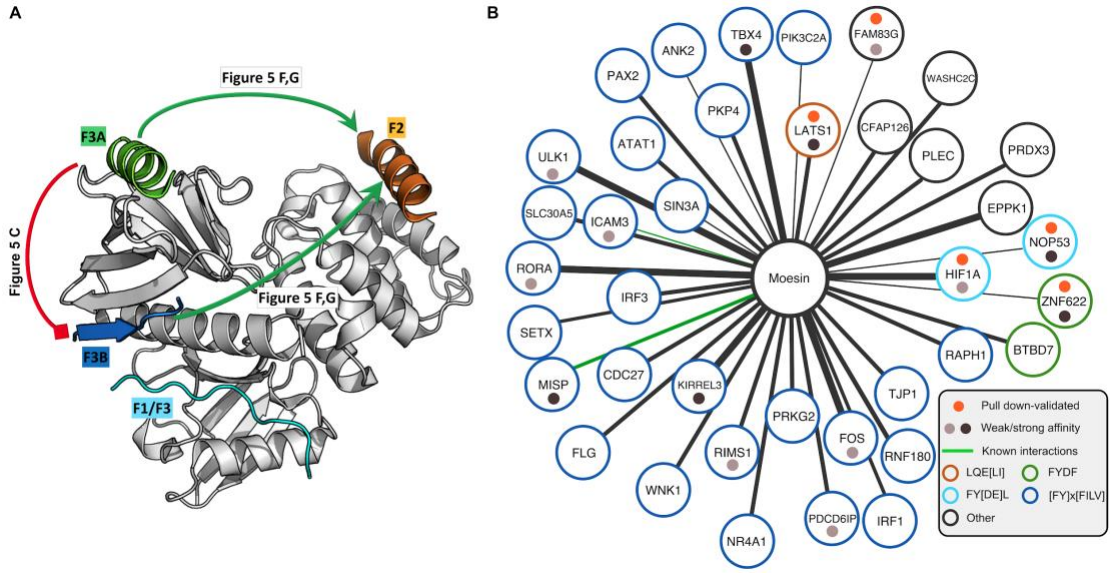


Figure 10. (A) Binding-dependent subdomain interplay. Moesin FERM domain (PDB code: 1EF1) is shown with known partners at different binding sites: the CTD at F3a and F2 sites (PDB code: 1EF1) and crumbs peptide at F3b and F1/F3 sites (PDB code: 4YL8). As demonstrated in **Figs. 4, 8** and **9**, binding of a peptide at F3a inhibits the peptide binding at the F3b site, but not the inverse: binding at the F3b site does not affect binding at F3a. In turn, binding of peptides at either F3a or F3b sites promotes binding at F2, as shown in **Fig. 5**. (**B**) Network representation of ProP-PD derived interactions of moesin FERM domain. Visualized are ligands that share GO terms with moesin that are unlikely to occur by chance, together with ligands validated experimentally. Previously reported interactions are highlighted by green edges. The colors of the nodes indicate the apparent binding motifs, as explained in the box. The dots in the nodes indicate if the interactions have been validated by means of GST-pulldown (orange dots) or by FP affinity determination (grey or black dots depending on affinity, see Tables 1 & 2 for details). The network was constructed using Cytoscape 3.7.

Moreover, the experimental results suggested that binding to either F3a or F3b may enhance LATS1 binding of F2 (**Fig. 5F, G)**. The binding of longer peptides, such as the moesin CTD, or the 40 amino acid C-terminal tail of EBP50, may thus begin with binding to F3a, promoting binding of a nearby region to F2. This interaction may be weak by itself, as suggested by the poorly resolved structure of EBP50 bound to moesin [13], where it was shown that the 14 C-terminal residues binding to F3a are the main affinity determinants. However, the experimental results further showed that binding of LATS1 to the F2 pocket had a negative, likely allosteric effect, on binding of ligands to other sites. How this is accomplished mechanistically and structurally remains to be understood, and will require a combination of modelling and kinetic binding experiment.

The present study provides an in-depth understanding of one subgroup of FERM domains. The FERM family contains more than 50 additional members, and applying the here identified principles on the whole family may allow us to learn more about conservation, and variation of the binding pocket specificities as well as the allosteric communication within the FERM domain. Since this family also binds to phosphoinositides (via the F1 domain), this platform also allows to establish a connection between protein and lipid binding, which play important regulatory roles in biological systems [48, 50].

MATERIALS AND METHODS

Plasmids, cloning and mutagenesis

Synthetic genes encoding for human merlin FERM (22-312) and radixin FERM (5-295) domains were commercially synthesized (ThermoFisher) and cloned into pETM33 using NcoI and EcoRI restriction sites. pGEX4T Moesin-FERM was a gift from Vijaya Ramesh (Addgene plasmid # 11638, [51]). Ezrin FERM domain was a kind gift from Volker Gerke [52].

Human NOP53-HA-pcDNA was a kind gift from Ronit Sarid [53]. The NOP53 gene was PCR amplified and cloned into CMV10 using HindIII and EcoRI. RRBP1-GFP was kindly provided by Alexander F Palazzo [54]. HA-HIF1-alpha-pcDNA3 was a gift from William Kaelin (Addgene plasmid # 18949, [55]). pcDNA3 LATS1 was a gift from Erich Nigg (Addgene plasmid # 41156, [56]). LATS1 was PCR amplified and cloned into CMV10 using NotI and BamHI restriction sites. GFP-FAM83G-pcDNA was a kind gift from Gopal Sapkota [57]. HA-PAK6-pcDNA was kindly provided by Michael Lu [58]. ZNF622 was a gift from Hyunjung Ha [59]. Sequences of all cloned and obtained constructs were confirmed using Sanger sequencing.

Protein expression and purification

Expression constructs encoding the FERM domains of merlin, moesin, ezrin and radixin were transformed into *E. coli* BL21(DE3). For expression, 2xYT medium was inoculated with a fresh overnight culture of the transformed cells and grown until OD_{600} reached 0.8. Protein production was induced using 0.3 mM IPTG by incubating cells overnight at 18 ˚C while shaking. Next day, cells were harvested by centrifuging at 7,000 xg for 10 minutes. For purification, the pellet was homogenized using PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) containing lysozyme, DNase I, cOmplete™ protease inhibitor cocktail, 1% Triton X-100 and 2 mM ß-mercaptoethanol and cells were lysed for 30 minutes while shaking at 4˚C. Cell debris was removed by centrifuging at 20,000 xg for 45 minutes and supernatant containing His-GST-FERM was incubated with glutathione (GSH) sepharose resin (GE) for 1 hour. Beads were collected and washed using PBS until contaminants are removed. The protein was eluted using 10 mM reduced GSH in PBS, pH 8.0

and dialyzed to PBS overnight for use in phage selections and pull downs. Protein quality was checked using SDS-electrophoresis and thermal shift assays.

For FP experiments of merlin FERM, the GST tag was removed using HRV 3C protease by incubating it with protein overnight at 4˚C in dialysis buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% Tween-20 and 3 mM DTT). Following day, GST tag and protease were removed using reverse Ni-IMAC and protein was concentrated to working concentration. For FP experiments using moesin FERM, thrombin (Sigma) was used to cleave the GST-tag from moesin FERM domain on beads and incubated overnight at 4 ˚C with under gentle rotation. Next day, the beads bound with GST were collected by gentle centrifugation while the supernatant containing thrombin and moesin FERM domain was passed though a HiTrap® benzamidine column (GE). The protein was then eluted using high salt (0.7-1 M NaCl) concentration buffer. The buffer was then exchanged using PD-10 desalting column (GE) into FP buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% Tween-20 and 3 mM DTT).

ProP-PD selections

Phage selections were performed for four days using the libraries and method described elsewhere [20, 60]. Briefly, 25 µg of target protein (GST-tagged FERM domains of merlin, moesin, ezrin and radixin) and GST-control were immobilized overnight to a 96-well maxisorp plate wells while shaking at 4˚C. Next day, the wells were blocked using 0.5% BSA in PBS for an hour. Naïve phage library containing 10^{11} CFU was precipitated using 1/5th volume of PEG/NaCl followed by centrifugation at 10,000 xg for 10 minutes and dissolved in 100 µl of PBS for each well. Control wells were washed four times using PBS containing 0.05% Tween-20 (PBST) and incubated with phages for one hour at 4˚C with shaking to remove the non-specific binders. The target wells were washed as before and phage solution was transferred to them and incubated for two hours similarly. Following that, unbound phages were removed by washing similarly. The bound phages were eluted by incubating them with log phase *E.coli* OmniMax for 30 minutes at 37 ˚C. M13KO7 helper phages (10^{11} pfu/ml) were added to each well and incubated again at 37 ˚C for 45 minutes. After the helper phage infection, the growing cultures were transferred to 10 ml of 2xYT containing kanamycin, carbenicillin and 0.3 mM IPTG and incubated overnight with shaking at 37˚C. Following day, phages were precipitated using PEG/NaCl as before.

Competitive ProP-PD selections were performed in presence of saturating concentration (20x K_d) of competing peptide. This concentration was also maintained during washing steps accordingly.

To determine the selection performance, a sandwich ELISA was performed on phage pools from each day. Briefly, 10 µg of target and control proteins were immobilized overnight on 96-well maxisorp plate as before. Wells were blocked using BSA. 100 µl of phage solution

from each pool was added to control and target wells and incubated for 1 hour at 4 ˚C with shaking. Wells were washed 5x using PBST and were incubated with anti-M13 coat HRP-conjugated antibody (1:5000) dilution for one hour. Unbound antibody was washed away as before and TMB substrate was added and allowed to develop the blue color. Reaction was stopped using 0.6 M H_2SO_4 and absorbance intensity was determined at 450 nm.
Binding-enriched phage pools were selected based on ELISA results and their peptide-coding regions were amplified and barcoded using PCR. These DNA pools were sequenced using next generation sequencing and results were analyzed using a pipeline described elsewhere [20, 60].

Fluorescence polarization assay

To obtain the saturation data, moesin FERM was titrated using FP buffer and then equal volume of 10 nM FITC-labelled peptide was added, mixed and FP signal was recorded using SpectraMax iD5 (Molecular Devices). To perform the competition experiments, a pre-complex containing moesin FERM at 2x K_D concentration and 10 nM respective FITC-labelled peptide was made and 25 µl of it was mixed with 25 µl of titrated target peptide. FP signal was determined as before. All measurements were performed in triplicates to achieve statistically significant data.

Cell culture and pull-down assays

HEK293 cells (Sigma:85120602) were cultured using DMEM (Gibco™) supplemented with 10% bovine FBS and NEAA (Gibco™) in a humid environment at 37˚C while maintaining 5% CO_2. For pull-down experiments, the cells were transiently transfected with tagged-target proteins using Fugene® HD (Promega) and following manufacturer's recommendations. Cells were allowed to grow and express the proteins for 48 hours post-transfection. Cells were washed with ice-cold washing buffer (PBS, Halt™ Protease Inhibitor Cocktail, pH 7.4) and then incubated with lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10 mM sodium pyrophosphate, 10 mM sodium orthovanadate, 10 mM sodium fluoride, 1x cOmpleate™ EDTA-free protease inhibitor tablet and 0.5% Nonidet P-40) for 30 minutes at 4 ˚C with gentle shaking. Cell debris was removed by centrifuging at 16,000 xg for 20 minutes at 4˚C and protein concentration was determined using BCA Protein Assay (Pierce™). Supernatant containing 0.5 mg of total protein was mixed with target FERM domain or GST (negative control) and one of the beads: GSH magnetic agarose beads (Pierce™), magnetic GFP-Trap® (Chromotek) and anti-FLAG® M2 magnetic beads (Sigma) acccording to manufecturer's recommendations. This mixture was incubated overnight with end-over-end rotation at 4 ˚C. Following day, the beads were collected and washed with lysis buffer 3 times. Samples were eluted using SDS-sample buffer by boiling at 95 ˚C for 5 minutes.

Eluted samples were resolved by SDS-electrophoresis and then transferred to nitrocellulose membrane. Immunoblotting was performed using anti-FLAG (sigma), anti-GFP (ab6556), anti-HA (Sigma) anti-GST (Sigma) and anti-Myc (ab9106) antibodies followed by anti-rabbit HRP-conjugated secondary antibodies (GE). Membrane was exposed to Amersham ECL™ western blotting detection reagent (GE) for a minute and signals were imaged using ChemiDoc™ Imaging system (Bio-Rad). Proteins with mutated putative binding sites were also immuo-precipitated and blotted similarly.

Global Blind Peptide Docking using PIPER-FlexPepDock

Global docking was performed using the PIPER-FlexPepDock protocol [21]. In brief, the peptide conformation is represented as an ensemble of fragments extracted from the PDB, based on sequence and (predicted) secondary structure using the Rosetta Fragment picker (with the vall 2011 fragment library) [61]. These fragments are mutated to the target peptide sequence with the Rosetta fixed backbone design protocol [62]. 50 fragments are rigid body docked onto the receptor protein using the PIPER rigid body docking program. The top 250 models for each fragment are then further refined using the Rosetta FlexPepDock protocol [36], including receptor backbone minimization, and top-scoring models are clustered. In this study all Rosetta simulations were performed using Rosetta version 2019.14. The protocol is freely available for noncommercial use as an online server: https://piperfpd.furmanlab.cs.huji.ac.il.

Peptide threading with Rosetta FlexPepBind

The FlexPepBind protocol [37] uses a template structure of a protein-peptide interaction to thread a list of peptides each onto the template, and refines each peptide using FlexPepDock. In this study structural minimization only was used to refine the complexes, including both peptide and receptor backbone minimization. In some cases, where the template peptide was longer than the threaded peptide, the peptide sequences were threaded onto possible overlapping windows in the template.

Modeling conformational changes with Rosetta FastRelax protocol

The Rosetta Relax protocol is used for full-atom refinement of protein structures. In this study the FastRelax protocol [41] was applied with default parameters (no constraints were enforced) to imitate structure relaxation after ligand release, or alternatively, to open a pocket by superimposing the ligand to its binding site on an unbound structure. In each simulation 200 decoys were generated, and the top scoring models were taken for further analysis, in which backbone Cα-Cα distances between binding regions of the original and “relaxed” structures were calculated.

Acknowledgements

This study was funded in part by grants from the Swedish research council (2016-04965 to YI), from the Carl Trygger foundation (YI, CTS14:209) and by the Israel Science Foundation, founded by the Israel Academy of Science and Humanities (grant number 717/2017 to OSF) and the US-Israel Binational Science Foundation 2015207 (to OSF). Muhammad Ali was the recipient of a PhD fellowship from the Sven and Lilly Lawski foundation. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomic Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The authors acknowledge the kind support of Dr. Norman Davey and Dr. Izabella Krystkowiak related to annotation of peptide hits and generation of consensus motifs, and Dr. Leandro Simonetti for managing the NGS data.

Author contributions

MA performed affinity determinations, molecular biology and cell-based experiments. MA and VKY performed phage experiments. AK performed the computational analysis,
All authors analyzed results and conceived experiments. MA, AK, VKY, OSF and YI wrote the manuscript.

Conflict of interest

The author declares no competing interests.

References

[1] Fehon RG, McClatchey AI, Bretscher A. Organizing the cell cortex: the role of ERM proteins. Nat Rev Mol Cell Biol. 2010;11:276-87.
[2] Neisch AL, Fehon RG. Ezrin, Radixin and Moesin: key regulators of membrane-cortex interactions and signaling. Curr Opin Cell Biol. 2011;23:377-82.
[3] Gary R, Bretscher A. Ezrin self-association involves binding of an N-terminal domain to a normally masked C-terminal domain that includes the F-actin binding site. Mol Biol Cell. 1995;6:1061-75.
[4] Ishikawa H, Tamura A, Matsui T, Sasaki H, Hakoshima T, Tsukita S, et al. Structural conversion between open and closed forms of radixin: low-angle shadowing electron microscopy. J Mol Biol. 2001;310:973-8.
[5] Ben-Aissa K, Patino-Lopez G, Belkina NV, Maniti O, Rosales T, Hao JJ, et al. Activation of moesin, a protein that links actin cytoskeleton to the plasma membrane, occurs by phosphatidylinositol 4,5-bisphosphate (PIP2) binding sequentially to two sites and releasing an autoinhibitory linker. J Biol Chem. 2012;287:16311-23.
[6] Sher I, Hanemann CO, Karplus PA, Bretscher A. The tumor suppressor merlin controls growth in its open state, and phosphorylation converts it to a less-active more-closed state. Dev Cell. 2012;22:703-5.
[7] Hennigan RF, Fletcher JS, Guard S, Ratner N. Proximity biotinylation identifies a set of conformation-specific interactions between Merlin and cell junction proteins. Sci Signal. 2019;12.
[8] McCartney BM, Fehon RG. Distinct cellular and subcellular patterns of expression imply distinct functions for the Drosophila homologues of moesin and the neurofibromatosis 2 tumor suppressor, merlin. J Cell Biol. 1996;133:843-52.
[9] Cooper J, Giancotti FG. Molecular insights into NF2/Merlin tumor suppressor function. FEBS Lett. 2014;588:2743-52.
[10] Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. Mol Biosyst. 2012;8:268-81.
[11] Pearson MA, Reczek D, Bretscher A, Karplus PA. Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. Cell. 2000;101:259-70.
[12] Li Y, Zhou H, Li F, Chan SW, Lin Z, Wei Z, et al. Angiomotin binding-induced activation of Merlin/NF2 in the Hippo pathway. Cell Res. 2015;25:801-17.
[13] Finnerty CM, Chambers D, Ingraffea J, Faber HR, Karplus PA, Bretscher A. The EBP50-moesin interaction involves a binding site regulated by direct masking on the FERM domain. J Cell Sci. 2004;117:1547-52.
[14] Terawaki S, Maesaki R, Hakoshima T. Structural basis for NHERF recognition by ERM proteins. Structure. 2006;14:777-89.
[15] Mori T, Kitano K, Terawaki S, Maesaki R, Fukami Y, Hakoshima T. Structural basis for CD44 recognition by ERM proteins. J Biol Chem. 2008;283:29602-12.
[16] Hamada K, Shimizu T, Yonemura S, Tsukita S, Tsukita S, Hakoshima T. Structural basis of adhesion-molecule recognition by ERM proteins revealed by the crystal structure of the radixin-ICAM-2 complex. EMBO J. 2003;22:502-14.
[17] Wei Z, Li Y, Ye F, Zhang M. Structural basis for the phosphorylation-regulated interaction between the cytoplasmic tail of cell polarity protein crumbs and the actin-binding protein moesin. J Biol Chem. 2015;290:11384-92.
[18] Li Y, Wei Z, Zhang J, Yang Z, Zhang M. Structural basis of the binding of Merlin FERM domain to the E3 ubiquitin ligase substrate adaptor DCAF1. J Biol Chem. 2014;289:14674-81.
[19] Teyra J, Sidhu SS, Kim PM. Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains. FEBS Lett. 2012;586:2631-7.
[20] Davey NE, Seo MH, Yadav VK, Jeon J, Nim S, Krystkowiak I, et al. Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. FEBS J. 2017;284:485-98.

[21] Alam N, Goldstein O, Xia B, Porter KA, Kozakov D, Schueler-Furman O. High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. PLoS Comput Biol. 2017;13:e1005905.
[22] Davey NE, Haslam NJ, Shields DC, Edwards RJ. SLiMFinder: a web server to find novel, significantly over-represented, short protein motifs. Nucleic Acids Res. 2010;38:W534-9.
[23] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28:235-42.
[24] Takai Y, Kitano K, Terawaki S, Maesaki R, Hakoshima T. Structural basis of PSGL-1 binding to ERM proteins. Genes Cells. 2007;12:1329-38.
[25] Liu X, Salokas K, Tamene F, Jiu Y, Weldatsadik RG, Ohman T, et al. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. Nat Commun. 2018;9:1188.
[26] Serrador JM, Vicente-Manzanares M, Calvo J, Barreiro O, Montoya MC, Schwartz-Albiez R, et al. A novel serine-rich motif in the intercellular adhesion molecule 3 is critical for its ezrin/radixin/moesin-directed subcellular targeting. J Biol Chem. 2002;277:10400-9.
[27] Heiska L, Alfthan K, Gronholm M, Vilja P, Vaheri A, Carpen O. Association of ezrin with intercellular adhesion molecule-1 and -2 (ICAM-1 and ICAM-2). Regulation by phosphatidylinositol 4, 5-bisphosphate. J Biol Chem. 1998;273:21893-900.
[28] Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. Nature. 2017;545:505-9.
[29] Zhu W, Begum G, Pointer K, Clark PA, Yang SS, Lin SH, et al. WNK1-OSR1 kinase-mediated phospho-activation of Na+-K+-2Cl- cotransporter facilitates glioma migration. Mol Cancer. 2014;13:31.
[30] Garzon-Muvdi T, Schiapparelli P, ap Rhys C, Guerrero-Cazares H, Smith C, Kim DH, et al. Regulation of brain tumor dispersal by NKCC1 through a novel role in focal adhesion regulation. PLoS Biol. 2012;10:e1001320.
[31] Yin F, Yu J, Zheng Y, Chen Q, Zhang N, Pan D. Spatial organization of Hippo signaling at the plasma membrane mediated by the tumor suppressor Merlin/NF2. Cell. 2013;154:1342-55.
[32] Chen H, Mei L, Zhou L, Zhang X, Guo C, Li J, et al. Moesin-ezrin-radixin-like protein (merlin) mediates protein interacting with the carboxyl terminus-1 (PICT-1)-induced growth inhibition of glioblastoma cells in the nucleus. Int J Biochem Cell Biol. 2011;43:545-55.
[33] Ivanov AA, Gonzalez-Pecchi V, Khuri LF, Niu Q, Wang Y, Xu Y, et al. OncoPPi-informed discovery of mitogen-activated protein kinase kinase 3 as a novel binding partner of c-Myc. Oncogene. 2017;36:5852-60.
[34] Cummins TD, Wu KZL, Bozatzi P, Dingwell KS, Macartney TJ, Wood NT, et al. PAWS1 controls cytoskeletal dynamics and cell migration through association with the SH3 adaptor CD2AP. J Cell Sci. 2018;131.
[35] Luck K, Trave G. Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions. Bioinformatics. 2011;27:899-902.
[36] Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins. 2010;78:2029-40.
[37] Alam N, Zimmerman L, Wolfson NA, Joseph CG, Fierke CA, Schueler-Furman O. Structure-Based Identification of HDAC8 Non-histone Substrates. Structure. 2016;24:458-68.
[38] DOI:10.2210/pdb6txq/pdb.
[39] DOI:10.2210/pdb6txs/pdb.
[40] Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 2019;47:W402-W7.
[41] Tyka MD, Keedy DA, Andre I, Dimaio F, Song Y, Richardson DC, et al. Alternate states of proteins revealed by detailed energy landscape mapping. J Mol Biol. 2011;405:607-18.
[42] Ueki Y, Kruse T, Weisser MB, Sundell GN, Larsen MSY, Mendez BL, et al. A Consensus Binding Motif for the PP4 Protein Phosphatase. Mol Cell. 2019;76:953-64 e6.

[43] Jespersen N, Estelle A, Waugh N, Davey NE, Blikstad C, Ammon YC, et al. Systematic identification of recognition motifs for the hub protein LC8. Life Sci Alliance. 2019;2.
[44] Sundell GN, Arnold R, Ali M, Naksukpaiboon P, Orts J, Guntert P, et al. Proteome-wide analysis of phospho-regulated PDZ domain interactions. Mol Syst Biol. 2018;14:e8129.
[45] Wigington CP, Roy J, Damle NP, Yadav VK, Blikstad C, Resch E, et al. Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling. Mol Cell. 2020;79:342-58 e12.
[46] Olsen JG, Teilum K, Kragelund BB. Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. Cell Mol Life Sci. 2017;74:3175-83.
[47] Khramushin A, Marcu O, Alam N, Shimony O, Padhorny D, Brini E, et al. Modeling beta-sheet peptide-protein interactions: Rosetta FlexPepDock in CAPRI rounds 38-45. Proteins. 2020;88:1037-49.
[48] Hamada K, Shimizu T, Matsui T, Tsukita S, Hakoshima T. Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain. EMBO J. 2000;19:4449-62.
[49] Fievet BT, Gautreau A, Roy C, Del Maestro L, Mangeat P, Louvard D, et al. Phosphoinositide binding and phosphorylation act sequentially in the activation mechanism of ezrin. J Cell Biol. 2004;164:653-9.
[50] Chinthalapudi K, Mandati V, Zheng J, Sharff AJ, Bricogne G, Griffin PR, et al. Lipid binding promotes the open conformation and tumor-suppressive activity of neurofibromin 2. Nat Commun. 2018;9:1338.
[51] Gonzalez-Agosti C, Wiederhold T, Herndon ME, Gusella J, Ramesh V. Interdomain interaction of merlin isoforms and its influence on intermolecular binding to NHE-RF. J Biol Chem. 1999;274:34438-42.
[52] Nammalwar RC, Heil A, Gerke V. Ezrin interacts with the scaffold protein IQGAP1 and affects its cortical localization. Biochim Biophys Acta. 2015;1853:2086-94.
[53] Borodianskiy-Shteinberg T, Kalt I, Kipper S, Nachum N, Katz S, Pauker MH, et al. The nucleolar PICT-1/GLTSCR2 protein forms homo-oligomers. J Mol Biol. 2014;426:2363-78.
[54] Cui XA, Zhang H, Palazzo AF. p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. PLoS Biol. 2012;10:e1001336.
[55] Kondo K, Klco J, Nakamura E, Lechpammer M, Kaelin WG, Jr. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. Cancer Cell. 2002;1:237-46.
[56] Chan EH, Nousiainen M, Chalamalasetty RB, Schafer A, Nigg EA, Sillje HH. The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. Oncogene. 2005;24:2076-86.
[57] Vogt J, Dingwell KS, Herhaus L, Gourlay R, Macartney T, Campbell D, et al. Protein associated with SMAD1 (PAWS1/FAM83G) is a substrate for type I bone morphogenetic protein receptors and modulates bone morphogenetic protein signalling. Open Biol. 2014;4:130210.
[58] Liu X, Busby J, John C, Wei J, Yuan X, Lu ML. Direct interaction between AR and PAK6 in androgen-stimulated PAK6 activation. PLoS One. 2013;8:e77367.
[59] Seong HA, Gil M, Kim KT, Kim SJ, Ha H. Phosphorylation of a novel zinc-finger-like protein, ZPR9, by murine protein serine/threonine kinase 38 (MPK38). Biochem J. 2002;361:597-604.
[60] Ali M, Simonetti L, Ivarsson Y. Screening Intrinsically Disordered Regions for Short Linear Binding Motifs. Methods Mol Biol. 2020;2141:529-52.
[61] Gront D, Kulp DW, Vernon RM, Strauss CE, Baker D. Generalized fragment picking in Rosetta: design, protocols and applications. PLoS One. 2011;6:e23294.
[62] Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A. 2000;97:10383-8.

TABLES

Table 1. Fluorescent polarization (FP) affinity measurements of direct binding affinities to moesin and merlin obtained using FITC-labeled peptides (see also **Figs 5A,B**). The known or proposed binding pockets are indicated.

Name	**Peptide**	**moesin** K_d **(µM)**	**merlin** K_d **(µM)**	**Known/ Proposed pocket**
LATS1$_{73\text{-}88}$[i]	THHKA**LQEI**RNSLLPF	1.04 ± 0.05	-	F2
EBP50$_{343\text{-}358}$[ii]	APQMDWSKKNEL**FSNL**-COO-	3.0 ± 0.3	-	F3a
ZNF622$_{73\text{-}88}$	LEFAD**FYDF**RSSYPDH	1.0 ± 0.1	4.1 ± 0.2	F3a
KIRREL3$_{637\text{-}652}$,	**YYSV**NTFKEHHSTPTI	0.080 ± 0.002	-	F3b
TBX4$_{428\text{-}443}$	**SYSV**QTMETVPYQPFP	0.75 ± 0.03	-	F3b
NOP53$_{192\text{-}207}$	**FYDL**WASDNPLDRPLV	0.71 ± 0.04	17 ± 1	F1-F3 cleft
FAM83G$_{111\text{-}126}$	AEPLPSL**EYWP**QKSDR	29 ± 3	78 ±7	-

[i,ii] known to bind to F2, and F3a sites, respectively; **-** not measured

Table 2. Affinity determinations through FP competition experiments. The FITC-labeled probe peptides are indicated with names, and the competing peptides are indicated with names and peptide sequences (see also Fig. 4).

Known/ Proposed pocket	Name	Peptide	FITC-EBP50 K_I (μM)	FITC-ZNF622 K_I (μM)	FITC-KIRREL3 K_I (μM)	FITC-TBX4 K_I (μM)	FITC-NOP53 K_I (μM)
F3a	EBP50$_{343-358}$	APQMDWSKKNEL**F**SN**L**-COO-	**2.1 ± 0.1**	3.5 ± 0.3	2.5 ± 0.1	-	3.1 ± 0.3
	FYDF						
F3a	ZNF622$_{341-356}$	LEFAD**FYDF**RSSYPDH	5.2 ± 0.1	**1.03 ± 0.04**	1.61 ± 0.06	-	4.13 ± 0.04
F3a	BTBD7$_{940-950}$	QEYPD**FYDF**SNAAARP	2.6 ± 0.1	2.53 ± 0.04	1.34 ± 0.06	-	2.7 ±0.1
	xYxV						
F3b	KIRREL3$_{637-652}$	Y**Y**S**V**NTFKEHHSTPTI	14 ±1	51.9 ± 0.7	**0.09 ± 0.01**	0.15 ± 0.09	59.3 ±0.3
F3b	TBX4$_{428-443}$	S**Y**S**V**QTMETVPYQPFP	176 ±10	148 ± 6	33 ± 1	**1.04 ± 0.05**	144 ± 5
F3b	MISP$_{595-610}$	ITGS**Y**S**V**SESPFFSPI	34.4 ±0.5		2.3 ± 0.1	1.8 ± 0.05	
F3b	PDCD6IP$_{835-868}$	S**Y**P**F**PQPPQQSYYPQQ	136 ± 3		35.2 ± 0.2		
F3b	ICAM3$_{516-531}$	SGS**Y**H**V**REESTYLPLT	990 ±30		113 ± 3		
F3b	ULK1$_{630-645}$	S**F**D**F**PKTPSSQNLLAL	N.D		112 ± 5		
F3b	LATS1$_{421-436}$	L**Y**N**I**SVPGLQTNWPQS			149 ± 5		
F3b	RORA$_{185-200}$	T**Y**N**I**SANGLTELHDDL	N.D		134 ± 3		
F3b	FOS$_{238-253}$	A**F**T**L**PLLNDPEPKPSV	N.D		328 ± 4		
F3b	RIMS1$_{1405-1420}$	M**Y**T**L**EHNDGSQSDTAV	N.D		564 ± 5		
	FY(D/E)L						
F1-F3 cleft	NOP53$_{192-207}$	**FYDL**WASDNPLDRPLV	670 ± 90	197 ± 5	117 ± 3		**0.78 ± 0.05**

F1-F3 cleft	HIF1A$_{37-52}$	**FYEL**AHQLPLPHNVSS	N.D.	248 ± 7	430 ± 20		3.0 ± 0.4

N.D: Not determinable.