Identifying 14-3-3 interactome binding sites with deep learning

Laura van Weesep[1+], Rıza Özçelik[1,2,3+], Marloes Pennings[1,4], Emanuele Criscuolo[1], Christian Ottmann[1,4], Luc Brunsveld[1,4]* and Francesca Grisoni[1,2,3]*

[1]Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven, The Netherlands.
[2] Eindhoven AI Systems Institute (EAISI), Eindhoven University of Technology, Eindhoven, The Netherlands.
[3]Centre for Living Technologies, Alliance TU/e, WUR, UU, UMC Utrecht, Utrecht, The Netherlands.
[4]Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

[+]These authors contributed equally.
*Corresponding authors: l.brunsveld@tue.nl, f.grisoni@tue.nl

Abstract

Protein-protein interactions are at the heart of biological processes. Understanding how proteins interact is key for deciphering their roles in health and disease, and for therapeutic interventions. However, identifying protein interaction sites, especially for intrinsically disordered proteins, is challenging. Here, we developed a deep learning framework to predict protein binding sites to 14-3-3 – a 'central hub' protein holding a key role in cellular signaling networks. After systematically testing multiple deep learning approaches to predict sequence binding to 14-3-3, we developed an ensemble model achieving a 75% balanced accuracy on external sequences. Our approach was applied prospectively to identify putative binding sites across medically relevant proteins (ranging from highly structured to intrinsically disordered) for a total of approximately 300 sequences. The top eight predictions were experimentally validated in the wet-lab, and binding to 14-3-3 was confirmed for five out of eight sequences (K_d ranging from 1.6 ± 0.1 μM to 70 ± 5 μM). The biological relevance of our results was further confirmed by X-ray crystallography and molecular dynamics simulations. These sequences represent potential new binding sites within the 14-3-3 interactome (*e.g.*, Tau, relating to Alzheimer's disease), and provide opportunities to investigate their functional relevance. Our results highlight the ability of deep learning to capture intricate patterns underlying protein-protein interactions, even for challenging cases like intrinsically disordered proteins. To further the understanding and targeting of 14-3-3/protein interactions, our model was provided as a freely accessible web resource at the following URL: https://14-3-3-bindsite.streamlit.app/.

Introduction

Protein-protein interactions (PPIs) are fundamental to all biological processes, from maintaining cellular homeostasis[1] to driving disease mechanisms.[2] Among the numerous protein families facilitating PPIs, the family of 14-3-3 proteins stands out due to their ubiquity and high conservation across isoforms.[3,4] These 'central hub'[5] proteins hold a key role in cellular signaling networks, as they are known to interact with over 1200 protein clients,[6,7] and are involved in pathways related to metabolism, apoptosis, cell signaling and tumor development. Protein interaction with 14-3-3 can yield a multitude of effects[8], *e.g.*, the structural stabilization of the client protein[9,10], the masking of functional sequences[11,12], or bringing two proteins together.[13,14] Owed to these reasons, elucidating the 14-3-3 interactome (protein clients and/or their binding sites) has a key relevance to gain insights into cellular regulation and mechanisms of disease, as well as to provide new avenues for therapeutic intervention.

While it is important to identify 14-3-3 binding partners and their binding sites, it is a daunting task. Proteins can interact with each other in a wide variety of ways,[15] and the exact protein interaction sites and corresponding interaction effects are often unknown.[15] Combinatorial exploration in the wet-lab is both costly and time intensive.[16–18] Owed to these reasons, deep learning[19] – a subfield of artificial intelligence based on neural networks – has gained significant traction to predict PPIs.[20–23] Deep learning, thanks to its ability to extract complex and non-linear information from large and high-dimensional data,[19] bears promise to accelerate the identification of unknown binding sites involved in PPIs. To date, however, deep learning approaches have found only limited experimental validation in exploring protein interactomes,[20,24] and only a few approaches have focused on protein interactions with 14-3-3,[25,26] or phosphorylated proteins in general. Furthermore, 14-3-3 proteins interact with multiple and diverse intrinsically disordered phosphorylated targets,[27] which challenges the usage of established deep learning approaches that rely on protein structure to perform a prediction.[23]

Stemming from these observations, this work aims to explore the 14-3-3 protein interactome, by leveraging deep learning on protein sequence data. Our approach was designed to predict putative sites of protein binding to 14-3-3. After training our model on publicly available data, and benchmarking it in comparison with existing models,[26] we validated it prospectively in the wet-lab. Via a combination of model interpretation, crystal structure determination, and molecular dynamics, we show the potential of the proposed approach to prioritize putative interaction sites of proteins with 14-3-3.

Results and Discussion

Predicting binding to 14-3-3 with deep learning

Study setup. Predicting PPI sites with machine learning is a challenging endeavor, especially when dealing with intrinsically disordered proteins, like the typical 14-3-3 binding partners.[27] In these cases, structure-based approaches inevitably fail.[28] To this end, the prediction task was cast into modeling the 14-3-3 interaction with the individual binding sites of known clients using their amino acid (AA) sequences (Fig. 1a). We used an existing dataset[26] (Table 1), and represented each binding site as a peptide sequence comprising the seven amino acids preceding and seven amino acids following the phosphorylated site (for a total of 15 AA per binding site, Fig. 1a). Moreover, an additional set of 92 phosphopeptides measured in-house for their binding to 14-3-3 was used for model validation (Table 1).

Table 1. Datasets used in this study, along with the number of AA sequences contained, and their labels (binders, non-binders). The publicly available dataset was used for model training and selection, and the in-house set for external validation.

Dataset	No.	Binders	Non-binders
Training/validation set[26]	715	360 (50%)	355* (50%)
In-house set	92	58 (63%)	34 (37%)

*93 experimentally determined (26%), and 262 (83%) likely non-binders.

Model training and benchmarking. The publicly available data was used for model training, and it was split ten times into training (67.5%), validation (22.5%), and test sets (10%). We represented the AA sequences numerically using four approaches (Fig. 1b):

- *One-hot encoding*, where each amino acid in the peptide is represented as a binary vector indicating its type. This representation captures the information in the sequence without introducing prior assumptions or additional knowledge about the amino acids[18].
- *Learnable embeddings*, where amino acids are represented using vectors learned by the model during training.[16] These embeddings are updated during training to help capturing contextual and relational information about the amino acids in the sequence.
- *BLOcks SUbstitution Matrix (BLOSUM 62)*[29], where amino acids are encoded with substitution matrix scores, which reflect evolutionary conservation and property similarities between amino acids. This representation incorporates biochemically relevant information about amino acid substitutions. Phosphorylated amino acids were indicated via a dedicated binary flag in the corresponding position (*see* Materials and Methods).
- *Physico-chemical descriptors*, where each amino acid in the peptide is represented by 18 pre-computed numerical features[30] (Supp. Table S1). For each peptide, the computed amino acid features were concatenated and used for the prediction.

Each representation was combined with the following deep learning architectures (Fig. 1c):

- *Multilayer perceptron* (MLP)[31], where complex peptide features are progressively extracted through multiple layers of fully-connected neurons[32], without considering positional information.
- *Convolutional neural network* (CNN)[33], in which windows ('kernels') slide over an input sequence, and learn to weight input elements at each window. CNNs capture local patterns in sequences, which are combined to predict the global properties of a sequence (*e.g.*, binding).
- *Recurrent neural network with gated recurring units* (GRU)[34], which iterates over the input sequence and encodes information from the N- to the C-terminus, compresses the information into a 'hidden state', which is then used to provide a prediction.

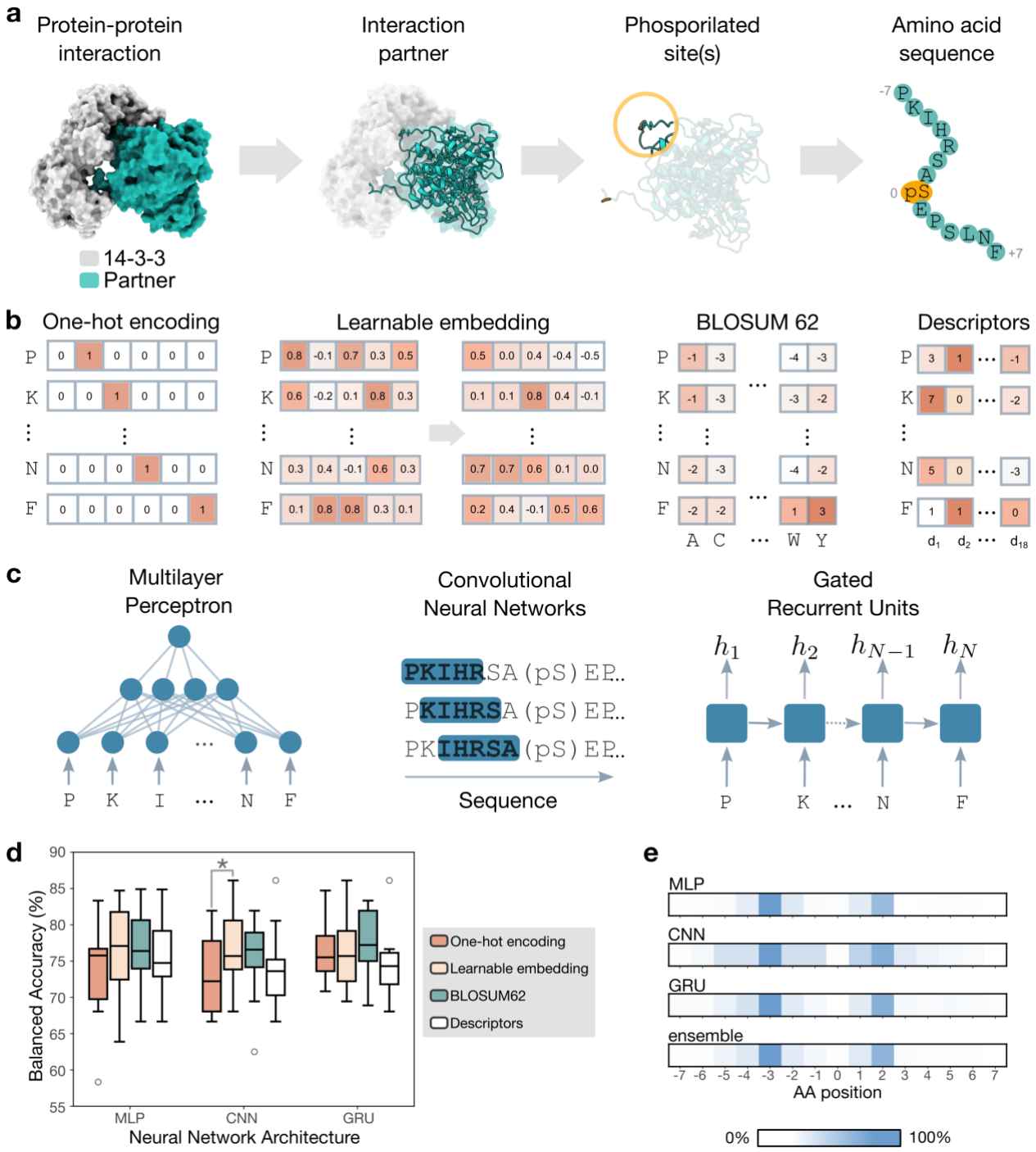


Figure 1. Predicting peptide sequence binding to 14-3-3 with deep learning. (**a**) The information on tested 14-3-3 protein interactions (Table 1) was converted into a 'machine readable' format. Seven amino acids before and after the phosphorylated site were used to form a sequence for deep learning, which is labeled according to its binding to 14-3-3 (yes/no). (**b**) Representations of amino acid (AA) sequences for model training. One-hot encoding represents each AA with a unique binary vector capturing its position in the sequence. Learnable encoding starts with a random numerical vector per amino acid and updates the vectors during training. BLOSUM62 uses substitution scores derived from evolutionary conservation. Descriptors are pre-defined features capturing the physico-chemical properties of each amino acid. (**c**) Neural network architectures. Multilayer perceptron (MLP) consists of fully connected layers. Convolutional neural networks (CNNs) slide windows over the input sequences, and Gated recurrent units (GRU) iterate over the input AAs in a stepwise manner. (**d**) Balanced accuracy per architecture-representation combination (computed on 10 test sets obtained via repeated splitting). Statistically significant differences are marked with "*" (paired Wilcoxon test, $\alpha = 0.05$). (**e**) Interpretation of the best models via input perturbation. By randomly shuffling all amino acids in any given position, we computed the relative change in the model predictions. Color indicates the relevance of the perturbation in each position, normalized by maximum achieved change, ranging from 0% (white: no impact) to 100% (blue: maximum impact).

For each representation-architecture combination, we performed hyperparameter tuning and evaluated the model on the 10 test sets (obtained via stratified splitting). The best model for each combination was evaluated on the test sets using balanced accuracy (BA), which captures the global model performance (Materials and Methods, Eq. 4). In general, no statistically significant difference between model architectures was observed (Wilcoxon signed rank test, $\alpha = 0.05$). Moreover, the chosen sequence representations were the main drivers of performance, with different trends based on the chosen architecture (Fig. 1d). For each architecture, we chose the representation leading to the highest balanced accuracy (average over 10 test-set splits), resulting in: (a) MLP with learnable embedding (BA = 77±4%); (b) CNN with BLOSUM 62 encoding (BA = 73±5%); and (c) GRU with BLOSUM 62 encoding (BA = 78±6%). Moreover, an ensemble model was obtained by averaging the prediction of each model. While this model did not improve the overall balanced accuracy (BA = 77±5%), it increased the capacity to correctly recognize binding sequences (Supp. Table S2).

The models were then retrained with all available data. Their performance was benchmarked in comparison with 14-3-3-Pred.[26] 14-3-3-Pred also combines three machine learning approaches (MLP, support vector machine [SVM], and position-specific scoring matrix [PSSM]) into an ensemble model. Both 14-3-3Pred and our models were validated on the in-house set (92 peptides, Table 1), as it comprises peptides external to all considered models and exhibiting diverse recurring AA motifs (Supp. Fig. S1) – hence allowing us to assess the potential for prospective validation. In addition to balanced accuracy, we calculated the capacity of the models to minimize false positives (precision) and to correctly recognize binding and non-binding sequences (recall and specificity, *see* Methods, Eq. 1-3). The models developed in this work systematically outperformed 14-3-3-Pred in global performance (balanced accuracy), and in most cases in terms of identification of true positives (recall, Table 2). Moreover, they consistently ranked second-best in the ability to minimize false positives (precision and specificity, Table 2). Finally, the ensemble approach balanced the strengths and weaknesses of each individual deep learning model.

Table 2. Model benchmarking on an external test set. Our model was compared with an existing one (14-3-3Pred) on a set of 92 external peptides, across various classification metrics: balanced accuracy (BA), Precision (Pr), Recall (Rc), and Specificity (Sp) (Methods, Eq. 1-4). For each classification metric, the best and second-best performance are highlighted in boldface and with underlining, respectively.

Model	Approach	BA (%)	Pr (%)	Rc (%)	Sp (%)
This work	MLP (learnable)	71	81	<u>84</u>	59
	CNN (BLOSUM 62)	71	<u>82</u>	75	<u>67</u>
	GRU (BLOSUM 62)	<u>73</u>	<u>82</u>	84	63
	Ensemble	**75**	<u>82</u>	**91**	59
14-3-3-Pred[26]	MLP	60	74	71	48
	SVM	61	**89**	29	**93**
	PSSM	60	74	71	48
	Ensemble	65	<u>82</u>	64	<u>67</u>

<u>Model interpretation.</u> To shed light onto the binding patterns learned by the models, we conducted a virtual mutation study. We randomly shuffled (15 times) amino acids occurring in each position, except for the phosphorylated amino acid, of the training peptides and used each model to predict the binding probability of the 'virtually mutated' sequences (Fig. 1e). The AAs comprised between -5 and +3 positions contributed the most to the predictions across

models, in alignment with previous findings.[26] Moreover, the amino acids in the -3 and +2 positions yielded the largest average change in predictions when perturbed. This is in line with structural biology observations, as the occurrence of arginine and proline at these positions is the most common binding motif for the interaction with 14-3-3.[35] Finally, each modeling approach has its own 'prediction hallmark', with different AA positions having different relevance (albeit relatively similar for CNN and GRU, both based on BLOSUM62 representation, Fig. 1e). This underscores that, although the individual models are trained on the same data, they capture distinct sequence-binding information, suggesting that this diversity enables the ensemble approach to balance their respective strengths and weaknesses.

Prospective model application

Experimental validation of binding sites to 14-3-3. We applied our model prospectively to identify putative, previously unidentified, binding sites with 14-3-3. As a case study, we selected seven medically relevant proteins: forkhead box O3 (FOXO3),[36] Tau,[37] Myc,[38] Bcl-2-associated agonist of cell death (BAD),[39] Notch-4,[40] Cystic fibrosis transmembrane conductance regulator (CFTR),[41] and p53.[42] These proteins contribute to a wide array of cellular processes[36,39,43] (*e.g.*, metabolism, cell survival and death) and are involved in diseases like cancer[44,45] (*e.g.*, BAD, p53 and Notch-4), Alzheimer's (Tau) and cystic fibrosis (CFTR).[46] The structures of these proteins range from ordered (p53, CFTR and Notch-4: experimental/predicted disorder ratio[47,48] between 0% to 26%) to partially and highly disordered (Myc, FOXO3, BAD, Tau; experimental/predicted disorder ratio[47,48] ranging from 58% to 95%, Supp. Table S3). Hence, they constitute an interesting and diverse test case for the 14-3-3 interactome.

For the selected proteins, their amino acid sequence was obtained from UniProt.[49] All serine and threonine residues were localized and a sequence window of 15 amino acids was obtained (-7 and +7 around such amino acids), leading to a total of 830 sequences. These sequences were further analyzed with PhosphositePlus[50] to predict whether they are phosphorylated in vivo. Only sequences labeled as phosphorylated (either according to literature[40] or to PhosphositePlus) were retained, resulting in a library of 296 peptides. These sequences were ranked by the ensemble model for their predicted binding to 14-3-3. Importantly, our model identified known binding sites for all proteins (13 in total, across Tau, BAD, FOXO3, Notch-4, CFTR, Myc and p53; Supp. Table S4), further corroborating the predictive ability and applicability of our approach.

From the model predictions, we filtered out the known binders, and selected eight top-scoring sequences, first ranked based on the majority vote of the ensemble model, and then by average prediction score across the three models (**1**-**8**, Table 3). Moreover, two bottom-scoring sequences were picked as negative controls (**9**-**10**, Table 3). These peptide sequences were obtained with a N-terminal fluorescent label to measure their binding affinity to 14-3-3γ via fluorescence anisotropy (FA) assays (Figure 2). Three out of eight 'positive' peptides (37%) showed strong, low-micromolar binding affinities (as measured via their dissociation constant [K_d], Table 3), equal to K_d = 1.6 ± 0.1 μM (**1**, FOXO3 pS413), K_d = 8.6 ± 0.8 μM (**2**, Tau pT245), and K_d = 15.9 ± 1.9 μM (**6**, BAD pS134). The remaining positive sequences showed binding, albeit weaker (K_d ranging from 70 μM to larger than 100 μM), except for the CFTR-pS422 peptide, which showed no binding in the FA assay (Figure 2). As expected, the negative controls **9** and **10** did not bind, confirming the correctness of the model-based ranking.

Table 3. Peptide selection and validation. Eight putative binding sites and two negative controls were selected for experimental validation, using the model predictions. Peptides **1**-**8** were selected by maximizing the predicted binding, while peptides **9**-**10** were selected as negative controls (predicted to be non-binding with high certainty). The protein, phosphosite, AA sequence (pS = phosphoserine, pT = phosphothreonine) and model predictions are reported, along with the experimentally determined constant of dissociation (K_D [mean ± SD, *n*=3]). Dose-response curves are reported in Supp. Fig. S2.

ID	Protein	Phosphosite	AA sequence	Model		K_D (μM)
				Outcome	Prediction	
1	FOXO 3	413	GLMQRSS(pS)FPYTTKG	Binding	0.98 ± 0.02	1.6 ± 0.1
2	Tau	245	SAKSRLQ(pT)APVPMPD	Binding	0.94 ± 0.05	8.6 ± 0.8
3	Notch 4	1847	FPRARTV(pS)VSVPPHG	Binding	0.87 ± 0.08	70 ± 1
4	Tau	198	SGDRSGY(pS)SPGSPGT	Binding	0.85 ± 0.06	71 ± 11
5	CFTR	422	NNNNRKT(pS)NGDDSLF	Binding	0.69 ± 0.14	–
6	BAD	134	KGLPRPK(pS)AGTATQM	Binding	0.68 ± 0.14	15.9 ± 1.9
7	BAD	118	GRELRRM(pS)DEFVDSF	Binding	0.65 ± 0.11	> 100
8	Myc	294	APGKRSE(pS)GSPSAGG	Binding	0.61 ± 0.04	> 100
9	Tau	111	EEAGIGD(pT)PSLEDEA	No binding	0.000 ± 0.0005	–
10	Myc	262	LHEETPP(pT)TSSDSEE	No binding	0.001 ± 0.0009	–

Interestingly, peptide **2** (Tau-pT245) showed higher binding affinities than the known 14-3-3 interaction sites:[51,52] Tau-pS214 (K_d = 16.4 ± 0.9 μM) and Tau-pS324 (K_d > 100 μM; Supp. Figure S3). This is especially interesting considering that Tau-pT245 is reported to be only phosphorylated in a normal, non-Alzheimer brain.[53]

Moreover, when comparing our predictions with those of 14-3-3Pred[26] and with 14-3-3 Site Finder, we observe moderate to no correlations between the predictions on the selected sequences (ranging from 0.20 to 0.58). Finally, the ranking obtained by our model correlates well with the observed K_d predictions (0.73, Supp. Table S5). These results corroborate the predictivity of our approach and its relevance to rationalize sequence binding to 14-3-3.

X-ray crystallography. The binding of the selected sequences was further confirmed and molecularly probed by X-ray crystallography through co-crystallization of 14-3-3 with peptides **1**-**8** (Fig. 2). Crystal structures were obtained for nearly all peptides that demonstrated binding in the FA assay, except for BAD pS134. These experiments validated the interaction of the newly discovered phosphorylated peptides to 14-3-3, as evident from the electron density maps which reveal the conformation of the peptides within the 14-3-3 binding pocket (Fig. 3a-f). Structural overlays with previously characterized 14-3-3/peptide complexes show that the binding modes of these predicted peptides are comparable to known interactions, indicating that these sequences are likely physiologically relevant rather than artificial (Fig. 2g-l).

The FOXO3 pS413 peptide exhibits an ‘open’ binding mode, bending outward from the 14-3-3 pocket due to a proline residue at the +2 position (Fig. 2a). A similar binding conformation was observed for the GAB2 pT391 peptide,[54] which aligns perfectly at its +2 proline with FOXO3, and for p53 pT387,[55] and bends out of the pocket due to glycine and proline residues

at the +2 and +3 positions, respectively (Fig. 2g). The high affinity of FOXO3 pS413 can be attributed to key interactions at the protein-peptide interface, including hydrogen bonds between FOXO3 residues S411 and S412 and 14-3-3 residues D225, N226 and W230 (Fig. S5a). Additionally, FOXO3 F414 interacts with the hydrophobic roof of the 14-3-3 pocket composed of L218, I219, and L222. A network of water-mediated hydrogen bonds is formed between the FOXO3 backbone and K49, K122 and N175 of 14-3-3. The phosphorylated residue of FOXO3 (pS413) is also involved in this hydrogen bond network, thereby stabilizing the bent conformation of the peptide (Supp. Fig. S4a). The high-affinity binding of FOXO3 pS413 was further corroborated by molecular dynamics simulations on the peptide (and the sequence extended by 40 AAs within the full-FOXO3 protein, see Materials and Methods), showing consistently low root mean squared fluctuation (RMSF) values (Supp. Fig. S5).

Despite also containing a +2 proline, the Tau pT245 peptide adopts a distinct binding mode, extending further into the 14-3-3 pocket (Fig. 2b). The 'extended' binding mode is similar to peptides such as CRAF pS259[56] and TFEB pS211,[57] all of which fold back into the pocket after a minor turn induced by the +2 proline (Fig. 2h). Conformational variations at the N-terminal side of the phospho-residue were observed, though the electron density in this region was not well-defined. Notably, all newly identified peptide sequences contained a positively charged arginine at the -3 or -4 position, consistent with many known 14-3-3 client peptides. The binding mode of Tau pT245 was also confirmed by molecular dynamics simulations on the tested peptide sequence and its extended version (with 40 additional amino acids, see Materials and Methods and Supp. Fig. S5b-c). In this context, Tau pT245 exhibited limited fluctuations in its interactions with 14-3-3 over time, as assessed by RMSF analysis (Supp. Fig. S5a).

For NOTCH4 pS1847, electron density was only observed up to the +2 serine, suggesting that the remaining residues are disordered (Fig. 2c). Similar C-terminal disorder has been reported in crystal structures of the 14-3-3 clients CIC pS173[58] and Nedd4-2 pS342[59] (Fig. 2i). In addition, only the +1 and +2 residues were resolved in the Tau pS198 crystal structure (Fig. 2d). The -1 tyrosine residue of Tau pS198 was observed in previously reported structures of USP8 pS718[60] and CRAF pS233[61], where it fits into a pocket at the top of the 14-3-3 binding groove (Fig. 2j).

Although BAD pS118 and Myc pS294 exhibited the weakest binding affinities among the tested peptides, their crystal structures displayed more ordered C-terminal regions compared to Tau pS198 and NOTCH4 pS1847 (Fig. 2e-f). The +1 aspartate residue of BAD pS118 is interacting with K122 of 14-3-3, followed by a +3 phenylalanine that shields the negatively charged aspartate (Supp. Fig. S5e) – an arrangement that appears unique among known 14-3-3 binding proteins, as far as we know. Therefore, the structural overlay for the BAD pS118 crystal structure shows more variation in the C-terminal side of the peptide (Fig. 2k). Nevertheless, some similarities were revealed in the overlay with BRAF pS726[62] and LKB1 pT336,[63] where BRAF's +1 glutamate aligns with BAD's +2 glutamate, and LBK1's +3 proline and +5 leucine occupy the same pocket as BAD's +3 phenylalanine. Moreover, molecular dynamics simulations on the extended version of BAD pS118 (by 20 residues on the N- and -C terminus of the original BAD sequence) show improved stabilization, compared to the shorter peptide, of the interactions (~2 Å reduction in RMSF) across all residues, and especially visible from the -4 leucine to the +1 aspartate residues. These analyses further support pS118 as a putative binding site of BAD to 14-3-3.

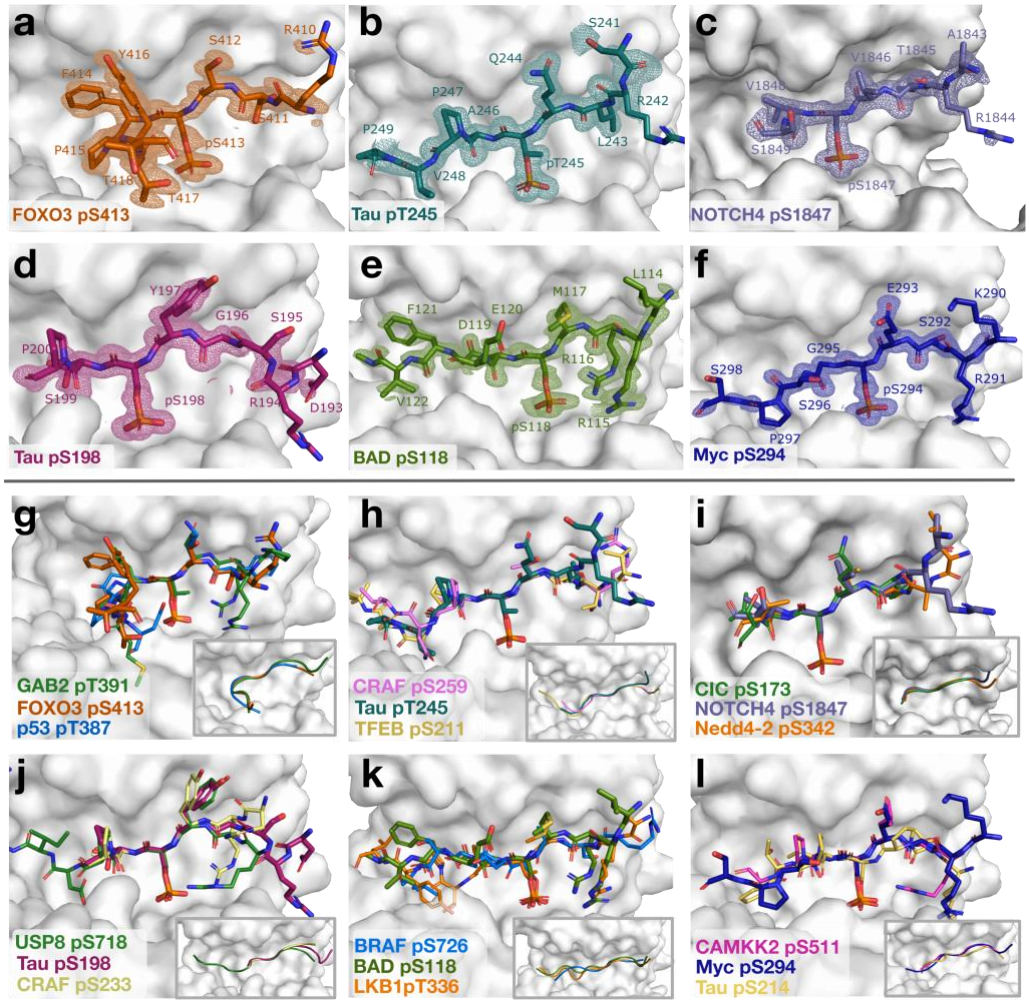


Figure 3. X-ray crystallography on selected peptide binders in comparison with known binders. (**a**-**f**) Crystal structures of the predicted peptide sequences (colored sticks) in complex with 14-3-3sigma (white surface). Final $2F_o$-F_c electron density contoured at 1.0σ. (**a**) FOXO3 pS413 (orange), (**b**) Tau pT245 (cyan), (**c**) NOTCH pS1847 (purple), (**d**) Tau pS198 (pink), (**e**) BAD pS118 (green), (**f**) Myc pS294 (blue). (**g**-**l**) Crystallographic overlay of predicted peptide sequences (**g**) FOXO3 pS413 (orange), (**h**) Tau pT245 (cyan), (**i**) NOTCH pS1847 (purple), (**j**) Tau pS198 (pink), (**k**) BAD pS118 (green), (**l**) Myc pS294 (blue) with two known 14-3-3 binding peptides (colored sticks) in the 14-3-3 pocket (white surface). Each figure includes a representation of the peptide backbones.

Finally, the Myc pS294 peptide forms a slight turn within the 14-3-3 pocket due to its +1 glycine, similar to CAMKK2 pS511[64] and Tau pS214[65], where this turn is induced by a +2 proline (Fig. 2f, 4f). This leads to a highly comparable binding mode among the 14-3-3 client peptides. In conclusion, the predicted binding sites of clinically relevant 14-3-3 client proteins demonstrated direct interactions with 14-3-3, exhibiting binding modes consistent with previously characterized 14-3-3/peptide complexes. This highlights the potential of our approach for identifying physiologically relevant phosphorylated binding sites within 14-3-3 client proteins.

Conclusions and Outlook

In this work, we developed and validated a deep learning approach for predicting protein-protein interactions between 14-3-3 proteins and phosphorylated client protein binding sites. By leveraging different amino-acid sequence representations and neural network architectures, we demonstrated that our models outperform existing tools in terms of balanced accuracy, recall, and precision. When combined within an ensemble model, our approach provided a robust predictive framework, enhancing the identification of novel

binding sites for prospective applications by minimizing false positives compared to the state-of-the art.

Our model was applied to identify novel putative binding sites on biologically relevant 14-3-3 client proteins (FOXO3, Myc, BAD, Notch-4, CFTR and p53). The model was used to screen 300 potential binding sites and to select eight peptide sequences for follow-up assays. Experimental validation confirmed the predictive power of our model, with three out of eight newly predicted phosphopeptides exhibiting low-micromolar binding affinities to 14-3-3, two moderate binders and two weak binders. Structural characterization via X-ray crystallography further substantiated our findings, revealing binding modes consistent with known 14-3-3-client interactions. This includes an 'open' binding mode, where peptides bend out of the 14-3-3 pocket, an 'extended' binding mode, in which peptides occupy the entire 14-3-3 pocket, and peptides featuring a disordered C-terminus. The identification of such structurally representative 14-3-3 binding motifs, without having provided such structural information to our models, testifies to the strength of our deep learning approach. These findings were further corroborated by molecular dynamics simulations on longer peptide versions of the putative binding sites. Our study not only advances computational predictions for 14-3-3 interactions but also underscores the importance of integrating deep learning with experimental validation. The results demonstrate that deep learning models can reliably predict biologically relevant binding sites, paving the way for more efficient exploration of the 14-3-3 interactome.

Several challenges and opportunities for future research remain. First, expanding the training dataset with additional experimentally validated binding and non-binding sequences will likely improve model generalizability. Incorporating sequence context beyond the immediate phosphosite region may further enhance predictive accuracy. While our model effectively predicts linear phosphopeptide binding motifs, potentially ideal for disordered binding partners undergoing protein-protein interactions, future work could integrate structural data more comprehensively, potentially by incorporating protein tertiary and especially quaternary structure information. Combined, this might strongly aid addressing the challenge of refining interaction predictions for disordered regions and transient interactions.

Applying this predictive framework to other phospho-dependent interactions beyond 14-3-3 proteins could broaden its impact, aiding in the discovery of new regulatory mechanisms and therapeutic targets. Additionally, prospective validation of predicted binding sites in cellular models and in vivo systems will be necessary to fully establish the physiological relevance of our findings. Ultimately, our approach contributes to a deeper understanding of cellular signaling and it facilitates the rational design of modulators targeting 14-3-3 interactions. Furthermore, by making our model freely available on an online platform (https://14-3-3-bindsite.streamlit.app/), without requiring expert deep learning knowledge, we provide an accessible tool for researchers to explore 14-3-3 interactions in their own studies, fostering further discoveries in the field.

Materials and Methods

Data collection and curation

Publicly available data. We used a previously curated 14-3-3 binding site dataset,[26] comprising 338 experimentally determined binding phosphosites,[66] 93 experimentally determined non-binding phosphosites[67] and 22 known binding sequences from the literature.[26] Moreover, 230 likely non-binding phosphosites obtained randomly were added from proteins of which already two 14-3-3-binding sites were defined. In total the data contained 360 sequences labelled as binding and 355 labelled as non-

binding. Sequences were centered around the phosphorylated amino acid and truncated or padded to 15 amino acids, if necessary.

In-house test set. An in-house dataset of 92 phosphopeptides tested for binding to 14-3-3 was used for model evaluation. 58 of those phosphopeptides are annotated as binders (K_d < 200 µM) and 34 were annotated as non-binders (K_d > 200 µM). In cases with multiple affinity scores for different 14-3-3 isoforms, the strongest binding affinity was picked. Last, we centered the sequences around the phosphorylated residues to comply with the training set format, considered 15 amino acids, and applied padding when necessary.

Model training

Training and hyperparameter tuning. The dataset was split using 10-fold stratified cross-validation splitting. 10% was used as the test set and the remaining data was split into training and validation (67.5% and 22.5% of the total dataset, respectively). Test peptides with an edit distance on the AA sequence equal to or lower than four were removed to avoid data leakage or overestimation of model performance. We used a two-staged approach for hyperparameter tuning. First, a 'broad' hyperparameter space was tested (as recently suggested[68]), and, later, the top hyperparameter configurations (216 for GRU, 324 for MLP, and 1500 for CNN) were further fine-tuned (Supporting Table S6). Early stopping on F1 score was used starting from the fifth epoch, with a patience of five epochs. The model with the highest F1 score (Eq. 5) in 10-fold validation was selected. The final hyperparameters for each model are reported in Supp. Table S7.

Evaluation metrics. The capacity of the model to correctly classify positive (binding) and negative (non-binding) peptides was quantified via Recall (Rc), Precision (Pr) and Specificity (Sp), computed as follows:[69]

$$Rc = \frac{TP}{TP + FN} \quad (1)$$

$$Sp = \frac{TN}{TN + FP} \quad (2)$$

$$Pr = \frac{TP}{TP + FP} \quad (3)$$

where true negatives (TN) and true positives (TP) represent the number of correctly identified non-binders and binders, respectively. Conversely, false negatives (FN) and false positives (FP) refer to the number of binders and non-binders that are misclassified. Recall (Eq. 1) indicates the proportion of actual binders that the model successfully identifies, specificity (Eq. 2) assesses the reliability of non-binding predictions, and precision (Eq. 3) measures the capability to minimize false positives.

Moreover, the models were assessed for their global prediction ability, via balanced accuracy (BA) and F1-score:

$$BA = \frac{Rc + Sp}{2} \quad (4)$$

$$F1 = 2 \cdot \frac{Pr + Rc}{Pr + Rc} \quad (5)$$

Balanced accuracy captures the model performance (correct predictions) normalized by the class imbalance, and F1 scores provide an overall evaluation of the model's performance in terms of minimizing false positives and negatives.

Peptide representation. The following settings were used for each peptide representation.

- One-hot-encoding. Each amino acid is assigned a unique vector with a single 1 corresponding to the respective index of that amino acid in the amino acid alphabet, with values of 0 in the remaining

elements. Phosphorylated amino acids were encoded as a distinct token (and the corresponding sparse vector).

- Learnable sequence embeddings. Amino acids were label encoded with one number per amino acid. Phosphorylated amino acids got their own unique label.
- BLOSUM 62 representation was tested in two formats: (a) by treating phosphorylated amino acids as their non-modified versions, and (b) by appending an additional numerical flag ([1,0] for phosphoserine and [0,1] for phosphothreonine). Preliminary results showed that the flagged version performed better in terms of F1 and balanced accuracy and hence it was used for this study.
- Peptide descriptors. For each amino acid, 18 descriptors were computed using the peptidy[30] software. Descriptors were concatenated, obtaining a matrix of 15 x 18 descriptors per peptide. (see Supp. Table S1).

Prospective screening on selected proteins

Library preparation. The amino acid sequence for the selected proteins was obtained from UniProt[49] (UniProt IDs: Tau = P10636-8; Myc = P01106; FOXO3 = O43524; Notch4 = Q99466; BAD = Q92934; CFTR = P13569; p53 =Q761V2). All serine and threonine residues were located and a window of 15 amino acids was obtained (-7 and +7 around such amino acids), leading to a total of 830. These sequences were further analyzed with PhosphoSitePlus,[50,70] to predict whether they are phosphorylated in vivo. Only sequences phosphorylated were according to literature or PhosphoSitePlus were retained, resulting in a library of 296 peptides. All sequences were predicted with the ensemble model and ranked by scores (average predictions across the models). The top scoring predictions were manually inspected, and known binding sites identified according to existing literature were excluded from the wet-lab validation (Supp. Table S4).

Experimental validation

Peptide materials. Selected peptides were ordered from GenScript[71] with a minimal purity of 85% with a N-terminal 6-Aminohexanoic Acid (Ahx) linker followed by the fluorescent dye 5-FAM. A C-terminal amidation served to mimic the lack of a free C-terminus in the amino-acid sequence when it is part of a larger protein. One of the top-scoring sequences was not tested due to failed synthesis by the commercial provider, and the next top-ranking sequence was picked instead. Peptide sequences are reported in Supp. Table S8.

Fluorescence Anisotropy assay. To study the binding of the fluorescently labelled peptides to 14-3-3, Fluorescence Anisotropy (FA) assays were carried out[23]. In the case of binding, tumbling of the peptide with the attached fluorophore will slow down and the emitted light will be polarized. This will lead to a higher anisotropy.[72] For all experiments, 14-3-3*γ* was used as it was shown in multiple experiments to be the strongest binding variant suitable for experimental screening.[73] The FAM-labeled peptides and the 14-3-3γ FL protein were diluted in buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.1% Tween20, 1 mg/mL BSA).

Dilution series of 14-3-3γ proteins (starting at 500 µM) were made to 10 nM of the FAM-labeled peptides in black, round-bottom 384-microwell (Corning) in a final sample volume of 10 µL. Fluorescence anisotropy values were measured using a Tecan Spark Control at room temperature (filter set lex: 485 ± 20 nm, lem: 535± 25 nm, mirror: Dichroic 510, flashes: 30, integration time: 40 *µ*s, settle time: 1 *µ*s; gain: optimized per peptide, and Z-position: calculated from well. Wells containing only FAM-peptide were used to set as G-factor. The K_D values were obtained from fitting the data using Origin 2020 with a Sigmoid Hill1 function. Data shown is the average and standard deviation of triplicates.

Co-Crystallization

The 14-3-3σΔC protein and the acetylated client peptides were dissolved in complexation buffer (25 mM HEPES pH=7.5, 2 mM MgCl2 and 100 µM TCEP) and mixed in a 1:2 or 1:4 molecular stoichiometry

(protein: peptide) with a final protein concentration of 12 mg/mL. The complex was set-up for sitting-drop crystallization at 4 °C, in a custom crystallization liquor (0.05 M HEPES (pH 7.1, 7.3, 7.5, 7.7), 0.19 M CaCl2, 24-29% PEG400, and 5% (v/v) glycerol). Crystals grew within 10-14 days at 4 °C. Crystals were fished and flash-cooled in liquid nitrogen. X-ray diffraction (XRD) data were collected at the European Synchrotron Radiation Facility (ESRF Grenoble, France, beamline ID23-2). Data was processed using CCP4i2 suite (version 8.0.019). After indexing and integrating the data, scaling was done using AIMLESS. The data was phased with MolRep, using PDB 3IQU as template. Model rebuilding and refinement was performed using REFMAC5. The PDB REDO server (pdb-redo.edu) was used to complete the model building and refinement. The images were created using the PyMOL Molecular Graphics System (Schrödinger LLC, version 4.6.0). See Supporting Table S9 for data collection and refinement statistics.

Molecular Dynamics Simulations

To investigate how our experimental results could be extended beyond the experimentally determined peptides, we selected three sequences: FOXO3 pS413 (**1**, K_d = 1.6 ± 0.1 μM), TAU pT245 (**2**, K_d = 8.5 ± 0.2 μM), and BAD pS118 (**7**, K_d>100 μM). For each of these sequences, we obtained an 'extended' sequence from the corresponding full-protein sequence from UniProt,[49] by elongating the tested sequences with 20 amino acids in both N- and C-terminal directions (55 amino acids in total). Our goal was to assess whether the additional flanking residues could alter the binding properties of the peptide within the 14-3-3 binding pocket and hence infer the plausibility of the predicted binding sites. To this end, we performed molecular dynamics (MD) to compare the stability of both sequence versions, ultimately to assess how additional flanking residues influence stability and to gain insights into these PPIs. For each peptide, molecular dynamics (MD) simulations were performed using GROMACS 2023[74] with three independent replicates. The simulations were divided into three stages: energy minimization, equilibration, and production. Energy minimization was performed using the steepest descent algorithm until a convergence criterion of 1000 kJ/mol/nm was reached. The equilibration phase was conducted under position-restrained dynamics in the NVT and NPT ensembles, using the V-rescale thermostat to maintain a temperature of 303.15 K and the Parrinello-Rahman barostat to regulate pressure at 1 atm. The production phase involved MD simulations for 300 ns with a 2-fs integration time step. The initial peptide structures were extended to a sequence of 55 amino acids using PyMOL.[75] MD simulations were performed following a five-step protocol to ensure proper system relaxation and equilibration. The first step involved an energy minimization using steepest descent, applying positional restraints on the backbone (force constant = 400 kJ/mol/nm²) and side chains (force constant = 40 kJ/mol/nm²). The peptide was frozen along all spatial dimensions during this phase. In the second step, a 5 ns MD simulation was carried out under NVT conditions, with positional restraints on the backbone and side chains. A time step of 1 fs was used, and the system was maintained at 303.15 K using the V-rescale thermostat. The peptide remained frozen along all spatial dimensions. Following this, a second round of energy minimization was performed using the same parameters as in the first phase to allow for further relaxation of the solvent environment around the peptide. The fourth phase involved a 5 ns MD simulation under NPT conditions to equilibrate the system. Positional restraints were again applied to the peptide backbone and side chains. Pressure was controlled isotropically at 1 bar using the Parrinello-Rahman barostat, and temperature was held at 303.15 K using the V-rescale thermostat. Finally, in the fifth phase, a 300 ns production MD simulation was carried out with a 2-fs time step, during which positional restraints were removed, allowing the peptide to move freely. Temperature (303.15 K) and pressure (1 bar) were controlled using the V-rescale thermostat and Parrinello-Rahman barostat,[76] respectively. For the analysis of the root mean square fluctuation (RMSF) of the peptides, the first 15 ns of the production phase were excluded from the calculation to allow for system equilibration. The RMSF values were then computed over the remaining trajectory, considering the fluctuations across all three replicates.

Software and code

The Python code and the data to replicate and extend our study are available on GitHub at the following URL: https://github.com/molML/14-3-3-bindsite. To further apply our approach prospectively, the software can be used via a freely accessible webpage at the following URL: https://14-3-3-bindsite.streamlit.app/.

Acknowledgements

This research was co-funded by the European Union (ERC, ReMINDER, 101077879 to FG and PPI-Glue, 101098234 to LB), the Netherlands Organization for Scientific Research (ENW M-grant OCENW.M20.200) and the Centre for Living Technologies. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors are grateful to the Chemical Biology and the Molecular Machine Learning groups at the TU/e for their contribution to this work via scientific discussions.

Author contributions

Conceptualization: all authors; Methodology (computational): RÖ, LvW, EC, FG; Methodology (experimental): LvW, MP, CO, LB; Software: LvW, RÖ; Validation (computational): LvW, RÖ, EC; Validation (experimental): LvW, MP; Formal analysis: all authors; Visualization: LvW, FG, RÖ, and EC; Resources and funding acquisition: FG and LB; Supervision: FG, LB, CO; Writing - Original Draft: LvW, FG, RÖ, with contributions from MP and EC; Writing - Review & Editing: all authors.

Conflict of Interest

CO and LB are co-founders of Ambagon Therapeutics. The other authors declare no conflict of interest.

References

(1) Westermarck, J.; Ivaska, J.; Corthals, G. L. Identification of Protein Interactions Involved in Cellular Signaling. *Mol. Cell. Proteomics* **2013**, *12* (7), 1752–1763. https://doi.org/10.1074/mcp.R113.027771.

(2) Ryan, D. P.; Matthews, J. M. Protein–Protein Interactions in Human Disease. *Curr. Opin. Struct. Biol.* **2005**, *15* (4), 441–446. https://doi.org/10.1016/j.sbi.2005.06.001.

(3) Gardino, A. K.; Smerdon, S. J.; Yaffe, M. B. Structural Determinants of 14-3-3 Binding Specificities and Regulation of Subcellular Localization of 14-3-3-Ligand Complexes: A Comparison of the X-Ray Crystal Structures of All Human 14-3-3 Isoforms. *14-3-3 Proteins Cancer* **2006**, *16* (3), 173–182. https://doi.org/10.1016/j.semcancer.2006.03.007.

(4) Fu, H.; Subramanian, R. R.; Masters, S. C. 14-3-3 Proteins: Structure, Function, and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40* (Volume 40, 2000), 617–647. https://doi.org/10.1146/annurev.pharmtox.40.1.617.

(5) Segal, D.; Maier, S.; Mastromarco, G. J.; Qian, W. W.; Nabeel-Shah, S.; Lee, H.; Moore, G.; Lacoste, J.; Larsen, B.; Lin, Z.-Y.; Selvabaskaran, A.; Liu, K.; Smibert, C.; Zhang, Z.; Greenblatt, J.; Peng, J.; Lee, H. O.; Gingras, A.-C.; Taipale, M. A Central Chaperone-like Role for 14-3-3 Proteins in Human Cells. *Mol. Cell* **2023**, *83* (6), 974-993.e15. https://doi.org/10.1016/j.molcel.2023.02.018.

(6) Uhart, M.; Bustos, D. M. Human 14-3-3 Paralogs Differences Uncovered by Cross-Talk of Phosphorylation and Lysine Acetylation. *PLOS ONE* **2013**, *8* (2), 1–16. https://doi.org/10.1371/journal.pone.0055703.

(7) Thurairajah, B.; Hudson, A. J.; Doveston, R. G. Contemporary Biophysical Approaches for Studying 14-3-3 Protein-Protein Interactions. *Front. Mol. Biosci.* **2022**, *9*. https://doi.org/10.3389/fmolb.2022.1043673.

(8) Somsen, B. A.; Cossar, P. J.; Arkin, M. R.; Brunsveld, L.; Ottmann, C. 14-3-3 Protein-Protein Interactions: From Mechanistic Understanding to Their Small-Molecule Stabilization. *ChemBioChem* **2024**, *25* (14), e202400214. https://doi.org/10.1002/cbic.202400214.

(9) Dhillon, A. S.; Yip, Y. Y.; Grindlay, G. J.; Pakay, J. L.; Dangers, M.; Hillmann, M.; Clark, W.; Pitt, A.; Mischak, H.; Kolch, W. The C-Terminus of Raf-1 Acts as a 14-3-3-Dependent Activation Switch. *Cell. Signal.* **2009**, *21* (11), 1645–1651. https://doi.org/10.1016/j.cellsig.2009.07.001.

(10) Aitken, A.; Howell, S.; Jones, D.; Madrazo, J.; Patel, Y. 14-3-3 α and δ Are the Phosphorylated Forms of Raf-Activating 14-3-3 β and ζ: IN VIVO STOICHIOMETRIC PHOSPHORYLATION IN BRAIN AT A Ser-Pro-Glu-Lys MOTIF (∗). *J. Biol. Chem.* **1995**, *270* (11), 5706–5709. https://doi.org/10.1074/jbc.270.11.5706.

(11) Muslin, A. J.; Xing, H. 14-3-3 Proteins: Regulation of Subcellular Localization by Molecular Interference. *Cell. Signal.* **2000**, *12* (11), 703–709. https://doi.org/10.1016/S0898-6568(00)00131-5.

(12) Brunet, A.; Kanai, F.; Stehn, J.; Xu, J.; Sarbassova, D.; Frangioni, J. V.; Dalal, S. N.; DeCaprio, J. A.; Greenberg, M. E.; Yaffe, M. B. 14-3-3 Transits to the Nucleus and Participates in Dynamic Nucleocytoplasmic Transport. *J. Cell Biol.* **2002**, *156* (5), 817–828. https://doi.org/10.1083/jcb.200112059.

(13) Bridges, D.; Moorhead, G. B. G. 14-3-3 Proteins: A Number of Functions for a Numbered Protein. *Sci. STKE* **2005**, *2005* (296), re10–re10. https://doi.org/10.1126/stke.2962005re10.

(14) Ottmann, C.; Marco, S.; Jaspert, N.; Marcon, C.; Schauer, N.; Weyand, M.; Vandermeeren, C.; Duby, G.; Boutry, M.; Wittinghofer, A.; Rigaud, J. L.; Oecking, C. Structure of a 14-3-3 Coordinated Hexamer of the Plant Plasma Membrane H+ -ATPase by Combining X-Ray Crystallography and Electron Cryomicroscopy. *Mol. Cell* **2007**, *25 3*, 427–440.

(15) 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**, *108* (4), 1225–1244. https://doi.org/10.1021/cr040409x.

(16) Chen, X.; Li, C.; Bernards, M. T.; Shi, Y.; Shao, Q.; He, Y. Sequence-Based Peptide Identification, Generation, and Property Prediction with Deep Learning: A Review. *Mol. Syst. Des. Eng.* **2021**, *6* (6), 406–428. https://doi.org/10.1039/d0me00161a.

(17) Plisson, F.; Ramírez-Sánchez, O.; Martínez-Hernández, C. Machine Learning-Guided Discovery and Design of Non-Hemolytic Peptides. *Sci. Rep.* **2020**, *10* (1), 16581. https://doi.org/10.1038/s41598-020-73644-6.

(18) Wen, B.; Zeng, W.-F.; Liao, Y.; Shi, Z.; Savage, S. R.; Jiang, W.; Zhang, B. Deep Learning in Proteomics. *Proteomics* **2020**, *20* (21–22), e1900335–e1900335. https://doi.org/10.1002/pmic.201900335.

(19) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.

(20) Soleymani, F.; Paquet, E.; Viktor, H.; Michalowski, W.; Spinello, D. Protein–Protein Interaction Prediction with Deep Learning: A Comprehensive Review. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 5316–5341. https://doi.org/10.1016/j.csbj.2022.08.070.

(21) Hashemifar, S.; Neyshabur, B.; Khan, A. A.; Xu, J. Predicting Protein–Protein Interactions through Sequence-Based Deep Learning. *Bioinformatics* **2018**, *34* (17), i802–i810. https://doi.org/10.1093/bioinformatics/bty573.

(22) Gainza, P.; Sverrisson, F.; Monti, F.; Rodolà, E.; Boscaini, D.; Bronstein, M. M.; Correia, B. E. Deciphering Interaction Fingerprints from Protein Molecular Surfaces Using Geometric Deep Learning. *Nat. Methods* **2020**, *17* (2), 184–192. https://doi.org/10.1038/s41592-019-0666-6.

(23) Zhang, J.; Durham, J.; Qian Cong. Revolutionizing Protein–Protein Interaction Prediction with Deep Learning. *Curr. Opin. Struct. Biol.* **2024**, *85*, 102775. https://doi.org/10.1016/j.sbi.2024.102775.

(24) Marchand, A.; Van Hall-Beauvais, A. K.; Correia, B. E. Computational Design of Novel Protein–Protein Interactions – An Overview on Methodological Approaches and Applications. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102370. https://doi.org/10.1016/j.sbi.2022.102370.

(25) Fan, Y.; Wang, X.; Wang, C. Building Random Forest QSAR Models for Affinity Identification of 14-3-3 ζ with Optimized Parameters. In *Proceedings of the 2020 9th International Conference on Bioinformatics and Biomedical Science*; ICBBS '20; Association for Computing Machinery: New York, NY, USA, 2021; pp 42–48. https://doi.org/10.1145/3431943.3431951.

(26) Madeira, F.; Tinti, M.; Murugesan, G.; Berrett, E.; Stafford, M.; Toth, R.; Cole, C.; MacKintosh, C.; Barton, G. J. 14-3-3-Pred: Improved Methods to Predict 14-3-3-Binding Phosphopeptides. *Bioinformatics* **2015**, *31* (14), 2276–2283. https://doi.org/10.1093/bioinformatics/btv133.

(27) Bustos, D. M.; Iglesias, A. A. Intrinsic Disorder Is a Key Characteristic in Partners That Bind 14-3-3 Proteins. *Proteins Struct. Funct. Bioinforma.* **2006**, *63* (1), 35–42. https://doi.org/10.1002/prot.20888.
(28) Perovic, V.; Sumonja, N.; Marsh, L. A.; Radovanovic, S.; Vukicevic, M.; Roberts, S. G. E.; Veljkovic, N. IDPpi: Protein-Protein Interaction Analyses of Human Intrinsically Disordered Proteins. *Sci. Rep.* **2018**, *8* (1), 10563. https://doi.org/10.1038/s41598-018-28815-x.
(29) Henikoff, S.; Henikoff, J. G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (22), 10915–10919. https://doi.org/10.1073/pnas.89.22.10915.
(30) Özçelik, R.; van Weesep, L.; de Ruiter, S.; Grisoni, F. Peptidy: A Light-Weight Python Library for Peptide Representation in Machine Learning. *ChemRxiv* **2024**. https://doi.org/10.26434/chemrxiv-2024-bm3lv.
(31) Chan, K. Y.; Abu-Salih, B.; Qaddoura, R.; Al-Zoubi, A. M.; Palade, V.; Pham, D.-S.; Ser, J. D.; Muhammad, K. Deep Neural Networks in the Cloud: Review, Applications, Challenges and Research Directions. *Neurocomputing* **2023**, *545*, 126327. https://doi.org/10.1016/j.neucom.2023.126327.
(32) Abdel-Nasser Sharkawy. Principle of Neural Network and Its Main Types: Review. *J. Adv. Appl. Comput. Math.* **2020**, *7*, 8–19. https://doi.org/10.15377/2409-5761.2020.07.2.
(33) LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1* (4), 541–551.
(34) Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *ArXiv Prepr. ArXiv14091259* **2014**.
(35) Yaffe, M. B.; Rittinger, K.; Volinia, S.; Caron, P. R.; Aitken, A.; Leffers, H.; Gamblin, S. J.; Smerdon, S. J.; Cantley, L. C. The Structural Basis for 14-3-3: Phosphopeptide Binding Specificity. *Cell* **1997**, *91* (7), 961–971.
(36) Stefanetti, R. J.; Voisin, S.; Russell, A.; Lamon, S. Recent Advances in Understanding the Role of FOXO3. *F1000Research* **2018**, *7*. https://doi.org/10.12688/f1000research.15258.1.
(37) Goedert, M.; Spillantini, M.; Cairns, N.; Crowther, R. Tau Proteins of Alzheimer Paired Helical Filaments: Abnormal Phosphorylation of All Six Brain Isoforms. *Neuron* **1992**, *8* (1), 159–168.
(38) Huang, Y.; Yang, M.; Huang, W. 14-3-3 σ: A Potential Biomolecule for Cancer Therapy. *Clin. Chim. Acta* **2020**, *511*, 50–58. https://doi.org/10.1016/j.cca.2020.09.009.
(39) Yang, E.; Zha, J.; Jockel, J.; Boise, L. H.; Thompson, C. B.; Korsmeyer, S. J. Bad, a Heterodimeric Partner for Bcl-xL and Bcl-2, Displaces Bax and Promotes Cell Death. *Cell* **1995**, *80* (2), 285–291. https://doi.org/10.1016/0092-8674(95)90411-5.
(40) Ramakrishnan, G.; Davaakhuu, G.; Chung, W. C.; Zhu, H.; Rana, A.; Filipovic, A.; Green, A. R.; Atfi, A.; Pannuti, A.; Miele, L.; Tzivion, G. AKT and 14-3-3 Regulate Notch4 Nuclear Localization. *Sci. Rep.* **2015**, *5* (1), 8782. https://doi.org/10.1038/srep08782.
(41) Boucher, R. New Concepts of the Pathogenesis of Cystic Fibrosis Lung Disease. *Eur. Respir. J.* **2004**, *23* (1), 146–158.
(42) Timofeev, O. Editorial: Mutant P53 in Cancer Progression and Personalized Therapeutic Treatments. *Front. Oncol.* **2021**, *11*. https://doi.org/10.3389/fonc.2021.740578.
(43) García-Gutiérrez, L.; Delgado, M. D.; León, J. MYC Oncogene Contributions to Release of Cell Cycle Brakes. *Genes* **2019**, *10* (3), 244.
(44) Harrison, H.; Farnie, G.; Howell, S. J.; Rock, R. E.; Stylianou, S.; Brennan, K. R.; Bundred, N. J.; Clarke, R. B. Regulation of Breast Cancer Stem Cell Activity by Signaling through the Notch4 Receptor. *Cancer Res.* **2010**, *70* (2), 709–718.
(45) Stickles, X. B.; Marchion, D. C.; Bicaku, E.; Al Sawah, E.; Abbasi, F.; Xiong, Y.; Bou Zgheib, N.; Boac, B. M.; Orr, B. C.; Judson, P. L.; others. BAD-Mediated Apoptotic Pathway Is Associated with Human Cancer Development. *Int. J. Mol. Med.* **2015**, *35* (4), 1081–1087.
(46) Gibson, R. L.; Burns, J. L.; Ramsey, B. W. Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis. *Am. J. Respir. Crit. Care Med.* **2003**, *168* (8), 918–951.
(47) Aspromonte, M. C.; Nugnes, M. V.; Quaglia, F.; Bouharoua, A.; DisProt Consortium; Tosatto, S. C. E.; Piovesan, D. DisProt in 2024: Improving Function Annotation of Intrinsically Disordered Proteins. *Nucleic Acids Res.* **2024**, *52* (D1), D434–D441. https://doi.org/10.1093/nar/gkad928.
(48) Piovesan, D.; Del Conte, A.; Mehdiabadi, M.; Aspromonte, M. C.; Blum, M.; Tesei, G.; von Bülow, S.; Lindorff-Larsen, K.; Tosatto, S. C. E. MOBIDB in 2025: Integrating Ensemble Properties and Function Annotations for Intrinsically Disordered Proteins. *Nucleic Acids Res.* **2025**, *53* (D1), D495–D503. https://doi.org/10.1093/nar/gkae969.
(49) The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.
(50) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res* **2015**, *43*, D512–D520.
(51) Hochmair, J.; Oetelaar, M. C. M. van den; Diez, L.; Lemmens, L. J. M.; Ponce, R.; Ravatt, L.; Franck, M. W.; Semenova, E.; Mohapatra, S.; Ottmann, C.; Brunsveld, L.; Wegmann, S. 14-3-3 Binding Regulates Tau Assembly and Microtubule Association. bioRxiv March 15, 2024, p 2024.03.15.585148. https://doi.org/10.1101/2024.03.15.585148.
(52) Neves, J. F.; Petrvalská, O.; Bosica, F.; Cantrelle, F.-X.; Merzougui, H.; O'Mahony, G.; Hanoulle, X.; Obšil, T.; Landrieu, I. Phosphorylated Full-Length Tau Interacts with 14-3-3 Proteins via Two Short Phosphorylated Sequences, Each Occupying a Binding Groove of 14-3-3 Dimer. *FEBS J.* **2021**, *288* (6), 1918–1934. https://doi.org/10.1111/febs.15574.
(53) Šimić, G.; Babić Leko, M.; Wray, S.; Harrington, C.; Delalle, I.; Jovanov-Milošević, N.; Bažadona, D.; Buée, L.; De Silva, R.; Di Giovanni, G.; Wischik, C.; Hof, P. R. Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies. *Biomolecules* **2016**, *6* (1). https://doi.org/10.3390/biom6010006.
(54) Bier, D.; Bartel, M.; Sies, K.; Halbach, S.; Higuchi, Y.; Haranosono, Y.; Brummer, T.; Kato, N.; Ottmann, C. Small-Molecule Stabilization of the 14-3-3/Gab2 Protein-Protein Interaction (PPI) Interface. *ChemMedChem* **2016**, *11* (8), 911–918. https://doi.org/10.1002/cmdc.201500484.

(55) Schumacher, B.; Mondry, J.; Thiel, P.; Weyand, M.; Ottmann, C. Structure of the P53 C-Terminus Bound to 14-3-3: Implications for Stabilization of the P53 Tetramer. *FEBS Lett.* **2010**, *584* (8), 1443–1448. https://doi.org/10.1016/j.febslet.2010.02.065.
(56) Molzan, M.; Schumacher, B.; Ottmann, C.; Baljuls, A.; Polzien, L.; Weyand, M.; Thiel, P.; Rose, R.; Rose, M.; Kuhenne, P.; Kaiser, M.; Rapp, U. R.; Kuhlmann, J.; Ottmann, C. Impaired Binding of 14-3-3 to C-RAF in Noonan Syndrome Suggests New Approaches in Diseases with Increased Ras Signaling. *Mol. Cell. Biol.* **2010**, *30* (19), 4698–4711. https://doi.org/10.1128/MCB.01636-09.
(57) Xu, Y.; Ren, J.; He, X.; Chen, H.; Wei, T.; Feng, W. YWHA/14-3-3 Proteins Recognize Phosphorylated TFEB by a Noncanonical Mode for Controlling TFEB Cytoplasmic Localization. *Autophagy* **2019**, *15* (6), 1017–1030. https://doi.org/10.1080/15548627.2019.1569928.
(58) Ren, Y.; Ouyang, Z.; Hou, Z.; Yan, Y.; Zhi, Z.; Shi, M.; Du, M.; Liu, H.; Wen, Y.; Shao, Y. CIC Is a Mediator of the ERK1/2-DUSP6 Negative Feedback Loop. *iScience* **2020**, *23* (11), 101635. https://doi.org/10.1016/j.isci.2020.101635.
(59) Pohl, P.; Joshi, R.; Petrvalska, O.; Obsil, T.; Obsilova, V. 14-3-3-Protein Regulates Nedd4-2 by Modulating Interactions between HECT and WW Domains. *Commun. Biol.* **2021**, *4* (1), 899. https://doi.org/10.1038/s42003-021-02419-0.
(60) Centorrino, F.; Ballone, A.; Wolter, M.; Ottmann, C. Biophysical and Structural Insight into the USP8/14-3-3 Interaction. *FEBS Lett.* **2018**, *592* (7), 1211–1220. https://doi.org/10.1002/1873-3468.13017.
(61) Molzan, M.; Ottmann, C. Synergistic Binding of the Phosphorylated S233- and S259-Binding Sites of C-RAF to One 14-3-3ζ Dimer. *J. Mol. Biol.* **2012**, *423* (4), 486–495. https://doi.org/10.1016/j.jmb.2012.08.009.
(62) Martinez Fiesco, J. A.; Durrant, D. E.; Morrison, D. K.; Zhang, P. Structural Insights into the BRAF Monomer-to-Dimer Transition Mediated by RAS Binding. *Nat. Commun.* **2022**, *13* (1), 486. https://doi.org/10.1038/s41467-022-28084-3.
(63) Lu, Y.; Ding, S.; Zhou, R.; Wu, J. Structure of the Complex of Phosphorylated Liver Kinase B1 and 14-3-3ζ. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2017**, *73* (Pt 4), 196–201. https://doi.org/10.1107/S2053230X17003521.
(64) Ballone, A.; Lau, R. A.; Zweipfenning, F. P. A.; Ottmann, C. A New Soaking Procedure for X-Ray Crystallographic Structural Determination of Protein-Peptide Complexes. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2020**, *76* (Pt 10), 501–507. https://doi.org/10.1107/S2053230X2001122X.
(65) Joo, Y.; Schumacher, B.; Landrieu, I.; Bartel, M.; Smet-Nocca, C.; Jang, A.; Choi, H. S.; Jeon, N. L.; Chang, K.-A.; Kim, H.-S.; Ottmann, C.; Suh, Y.-H. Involvement of 14-3-3 in Tubulin Instability and Impaired Axon Development Is Mediated by Tau. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2015**, *29* (10), 4133–4144. https://doi.org/10.1096/fj.14-265009.
(66) Tinti, M.; Madeira, F.; Murugesan, G.; Hoxhaj, G.; Toth, R.; MacKintosh, C. ANIA: ANnotation and Integrated Analysis of the 14-3-3 Interactome. *Database* **2014**, *2014*, 1–15. https://doi.org/10.1093/database/bat085.
(67) Johnson, C.; Crowther, S.; Stafford, M. J.; Campbell, D. G.; Toth, R.; MacKintosh, C. Bioinformatic and Experimental Survey of 14-3-3-Binding Sites. *Biochem. J.* **2010**, *427* (1), 69–78.
(68) Özçelik, R.; Grisoni, F. A Hitchhiker's Guide to Deep Chemical Language Processing for Bioactivity Prediction. arXiv July 16, 2024. https://doi.org/10.48550/arXiv.2407.12152.
(69) Ballabio, D.; Grisoni, F.; Todeschini, R. Multivariate Comparison of Classification Performance Measures. *Chemom. Intell. Lab. Syst.* **2018**, *174*, 33–44. https://doi.org/10.1016/j.chemolab.2017.12.004.
(70) *PhosphoSitePlus - Site Search*. https://www.phosphosite.org/siteSearchAction (accessed 2025-01-15).
(71) *GenScript*. https://www.genscript.com/ (accessed 2025-01-15).
(72) Lea, W. A.; Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin Drug Discov* **2011**, *6* (1), 17–32. https://doi.org/10.1517/17460441.2011.537322.
(73) Gogl, G.; Tugaeva, K. V.; Eberling, P.; Kostmann, C.; Trave, G.; Sluchanko, N. N. Hierarchized Phosphotarget Binding by the Seven Human 14-3-3 Isoforms. *Nat. Commun.* **2021**, *12* (1), 1677. https://doi.org/10.1038/s41467-021-21908-8.
(74) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.
(75) DeLano, W. L.; others. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl Protein Crystallogr* **2002**, *40* (1), 82–92.
(76) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45* (14), 1196–1199. https://doi.org/10.1103/PhysRevLett.45.1196.
(77) Chen, Y.; Chen, X.; Yao, Z.; Shi, Y.; Xiong, J.; Zhou, J.; Su, Z.; Huang, Y. 14-3-3/Tau Interaction and Tau Amyloidogenesis. *J. Mol. Neurosci.* **2019**, *68* (4), 620–630. https://doi.org/10.1007/s12031-019-01325-9.
(78) Sluchanko, N. N.; Tugaeva, K. V.; Gushchin, I.; Remeeva, A.; Kovalev, K.; Cooley, R. B. Crystal Structure of Human 14-3-3ζ Complexed with the Noncanonical Phosphopeptide from Proapoptotic BAD. *Biochem. Biophys. Res. Commun.* **2021**, *583*, 100–105. https://doi.org/10.1016/j.bbrc.2021.10.053.
(79) Mortenson, J. B.; Heppler, L. N.; Banks, C. J.; Weerasekara, V. K.; Whited, M. D.; Piccolo, S. R.; Johnson, W. E.; Thompson, J. W.; Andersen, J. L. Histone Deacetylase 6 (HDAC6) Promotes the Pro-Survival Activity of 14-3-3ζ via Deacetylation of Lysines within the 14-3-3ζ Binding Pocket*. *J. Biol. Chem.* **2015**, *290* (20), 12487–12496. https://doi.org/10.1074/jbc.M114.607580.
(80) Datta, S. R.; Katsov, A.; Hu, L.; Petros, A.; Fesik, S. W.; Yaffe, M. B.; Greenberg, M. E. 14-3-3 Proteins and Survival Kinases Cooperate to Inactivate BAD by BH3 Domain Phosphorylation. *Mol. Cell* **2000**, *6* (1), 41–51. https://doi.org/10.1016/S1097-2765(05)00012-2.
(81) Mathivanan, S.; Lakshman, P. K. C.; Singh, M.; Giridharan, S.; Sathish, K.; Hurakadli, M. A.; Bharatham, K.; Kamariah, N. Structure of a 14-3-3ε:FOXO3apS253 Phosphopeptide Complex Reveals 14-3-3 Isoform-Specific Binding of Forkhead Box Class O Transcription Factor (FOXO) Phosphoproteins. *ACS Omega* **2022**, *7* (28), 24344–24352. https://doi.org/10.1021/acsomega.2c01700.
(82) Stevers, L.; Lam, C. V.; Leysen, S.; Meijer, F.; Scheppingen, D.; de Vries, R.; Carlile, G.; Milroy, L.-G.; Thomas, D.; Brunsveld, L.; Ottmann, C. Characterization and Small-Molecule Stabilization of the Multisite Tandem Binding between 14-3-3 and the R Domain of CFTR. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*. https://doi.org/10.1073/pnas.1516631113.
(83) Egbert, C. M.; Warr, L. R.; Pennington, K. L.; Thornton, M. M.; Vaughan, A. J.; Ashworth, S. W.; Heaton, M. J.; English, N.; Torres, M. P.; Andersen, J. L. The Integration of Proteome-Wide PTM Data with Protein Structural and Sequence

Features Identifies Phosphorylations That Mediate 14-3-3 Interactions. *J. Mol. Biol.* **2023**, *435* (2), 167890. https://doi.org/10.1016/j.jmb.2022.167890.

Supporting Information

Identifying 14-3-3 interactome binding sites with deep learning

Laura van Weesep, Rıza Özçelik, Marloes Pennings, Emanuele Criscuolo, Christian Ottmann, Luc Brunsveld and Francesca Grisoni

Supplementary Tables

Supplementary Table S1. List of psycho-chemical properties computed per amino-acid, using peptidy.[30]

1. Aliphatic Index	7. No. Carbon Atoms	13. Molecular Weight
2. Aromaticity	8. No. Hydrogen Atoms	14. No. Hydrogen Bond Donors
3. Charge	9. No. Nitrogen Atoms	15. No. Hydrogen Bond Acceptors
4. Charge Density	10. No. Oxygen Atoms	16. Topological Polar Surface Area
5. Hydrophobic AA Ratio	11. No. Sulfur Atoms	17. Energy Based on LogP
6. Isoelectric Point	12. No. Phosphor Atoms	18. Average No. Rotatable Bonds

Supplementary Table S2. Machine learning performance evaluation on the test set. Classification metrics are reported as the mean and standard deviation across 10 training and test splits after hyperparameter tuning.

Architecture	Representation	BA (%)	Pr (%)	Recall (%)	Specificity (%)
MLP	Learnable	77±4	**76±6**	80±10	74±9
CNN	BLOSUM	73±5	71±5	80±7	66±8
GRU	BLOSUM	**78±6**	77±6	79±10	**77±8**
Ensemble	**//**	77±5	75±6	**82±9**	72±8

Supplementary Table S3. Intrinsic disorder ratio, as prediced by AlphaFold (and retrieved on MobiDB; https://mobidb.org, accessed February 2025) or experimentally determined (retrieved via DistProt; https://disprot.org/, accessed February 2025). UniProtID was used for retrieval

Protein	UniProtID	MobiDB (AlphaFold)	DistProt (experimental)
p53	Q761V2	0%	n.a.
CFTR	P13569	23.20%	12.50%
Notch4	Q99466	25.50%	n.a.
Myc	P01106	69.60%	57.86%
FOXO3	O43524	85.30%	n.a.
BAD	Q92934	85.70%	n.a.
Tau	P10636-8	95.10%	n.a.

Supplementary Table S4. Known binding sites that were correctly predicted by our model.

Protein	Phospohosite	Model Prediction	Reference
Tau	214	0.98±0.03	[42]
	324	0.66±0.14	[42]
	356	0.70±0.21	[77]
BAD	74	0.77±0.18	[78]
	75	0.96±0.05	[79] (mice)
	99	1.0±0.0	[80]
FOXO3	32	0.88±0.14	[80]
	253	0.99±0.01	[81]
Notch-4	1865	0.89±0.10	[40]
CFTR	737	0.70±0.19	[82]
	768	0.88±0.09	[82]
Myc	358	0.74±0.15	(truncated)[42]
p53	366	0.85±0.14	[42]

Supplementary Table S5 Ranking of the ordered peptides according to our model and the model available on the 14-3-3 Site Finder. For the other models the output from 14-3-3 site finder was used. Correlation (*r*) is computed between our ranking and those of the other methods. *Rank out the 296 phosphorylated peptides based on the 7 proteins tested. *Out of 393 top 50% entries. For all models, the Uniprot ID input used to retrieve the full protein sequences is: P10636-8,P01106,O43524,Q99466,Q92934,P13569,Q761V2. *** Correlation with the ranking based on Kd values.

Protein	**Phospho-site**	**This work*** [rank, ID]	**PTM and disorderedness****[83]	**14-3-3 Pred****[26]	**Adapted 14-3-3 Pred score****[83]	**K_d (μM)**
FOXO 3	413	3 (**1**)	12	14	2	1.6 ± 0.1
Tau	245	6 (**2**)	17	54	10	8.6 ± 0.8
Notch 4	1847	10 (**3**)	37	21	23	70 ± 1
Tau	198	11 (**4**)	40	174	98	71 ± 11
CFTR	422	20 (**5**)	68	18	39	–
BAD	134	21 (**6**)	13	73	14	15.9 ± 1.9
BAD	118	23 (**7**)	32	78	19	> 100
Myc	294	24 (**8**)	89	88	24	> 100
Corr.		1.00	0.58	0.20	0.05	0.73***

Supplementary Table S6. Hyperparameter space analyzed for the second round of hyper-parameter tuning each model (n.a. = not applicable). The search strategy (random or exhaustive) along with the number of tested hyperparameter combinations was reported.

Hyperparameter	MLP	CNN	GRU
Activation	ReLu	ReLu	ReLu
No. (dense) layers	1, 2, 3	1, 2, 3, 4	1, 2
No. of neurons per (dense) layer	8, 20, 64, 128	16, 64, 256	32, 64, 128, 256
Learning rate	1E-2, 1E-3, 5E-3	1E-2, 1E-3, 5E-3	1E-2, 1E-3, 5E-3
Batch size	32, 64, 128	32, 64, 128	32, 64, 256
Dropout	0.0, 0.1, 0.25	0.0, 0.1, 0.25	0.0, 0.1, 0.25
Epochs	200	200	200
Loss	BCE	BCE	BCE
Embedding size	32, 64	32, 64	32, 64
No. of 1D layers	n.a.	Other: 2, 3, 4	n.a.
Kernel size of the 1D layers	n.a.	5, 7, 9	n.a.
Number of filters	n.a.	3, 5, 7	n.a.
Search strategy	Exhaustive (except embedding)	Random	Exhaustive (except embedding)
No. HP combinations	324	1500	216

Supplementary Table S7. Selected model hyperparameters (n.a. = not applicable).

Hyperparameter	MLP	CNN	GRU
No. dense layers	2	4	1
No. neurons per dense layer	20	256	1
Learning rate	0.001	0.01	0.001
Batch size	128	64	128
Dropout	0.0	0.25	0.25
Epochs	18	37	25
Loss	BCE	BCE	BCE
Batch normalization	False	True	False
Embedding size	32	*n.a.*	n.a.
No. 1D conv. layers	n.a.	3	n.a.
Kernel size of the 1D layers	n.a.	5	n.a.
Number of filters	n.a.	5	n.a.

Supplementary Table S8. Ordered peptide sequences.

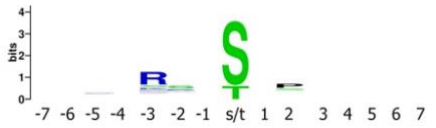
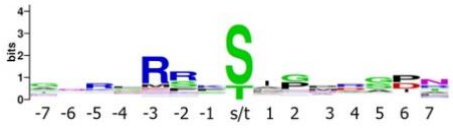
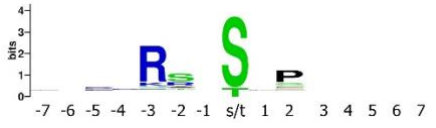
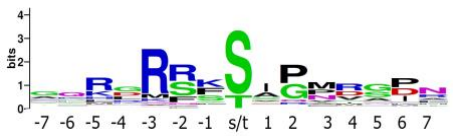
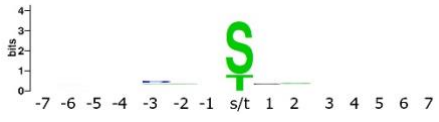
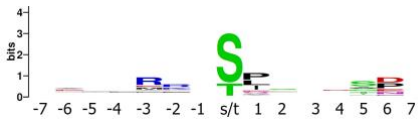
FOXO pS413	5-FAM-Ahx-GLMQRSS**pS**FPYTTKG-CONH2
Tau pT245	5-FAM-Ahx-SAKSRLQ**pT**APVPMPD-CONH2
NOTCH4 pS1847	5-FAM-Ahx-FPRARTV**pS**VSVPPHG-CONH2
Tau pS198	5-FAM-Ahx-SGDRSGY**pS**SPGSPGT-CONH2
Myc pS294	5-FAM-Ahx-APGKRSE**pS**GSPSAGG-CONH2
BAD pS134	5-FAM-Ahx-KGLPRPK**pS**AGTATQM-CONH2
BAD pS118	5-FAM-Ahx-GRELRRM**pS**DEFVDSF-CONH2
CFTR pS422	5-FAM-Ahx-NNNNRKT**pS**NGDDSLF-CONH2
Tau pT111	5-FAM-Ahx-EEAGIGD**pT**PSLEDEA-CONH2
Myc pT262	5-FAM-Ahx-LHEETPP**pT**TSSDSEE-CONH2
CFTR T1019	5-FAM-Ahx-QPYIFVA**pT**VPVIVAF-CONH2 (*Synthesis failed*)

Supplementary Table S9. XRD data collection and refinement statistics for 14-33σ/peptide structures.

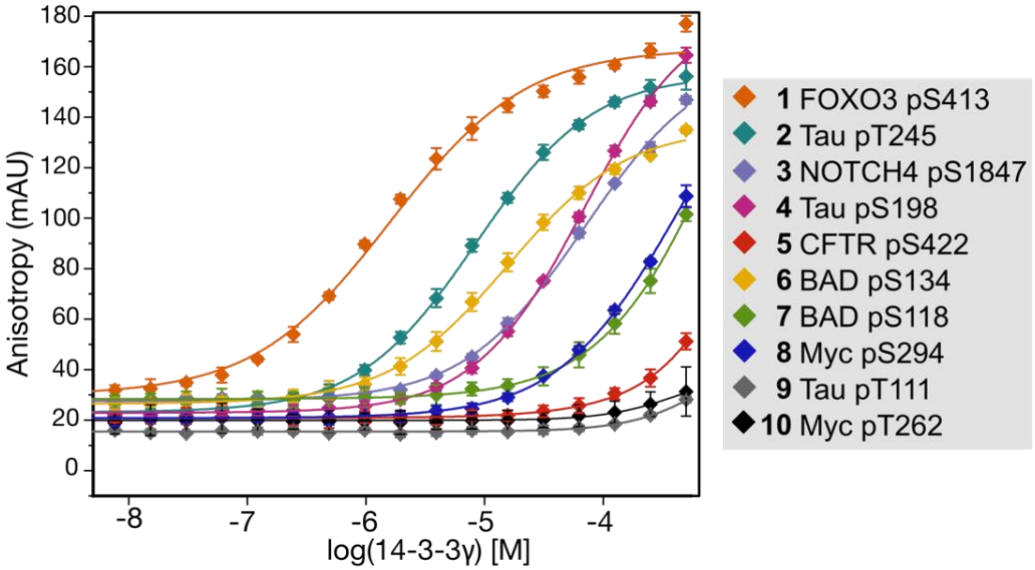
PDB	**9QNG**	**9QNK**	**9QNI**
Protein	14-3-3σΔC	14-3-3σΔC	14-3-3σΔC
Peptide	FOXO3 pS413	Tau pT245	NOTCH4 pS1847
Beam	ESRF ID23-2	ESRF ID23-2	ESRF ID23-2
DOI	10.2210/pdb9qng/pdb	10.2210/pdb9qnk/pdb	10.2210/pdb9qni/pdb
Data collection			
Wavelength (Å)	0.873128	0.873128	0.873128
Space group	C 2 2 21	C 2 2 21	C 2 2 21
Cell dimensions a, b, c (Å) α, β, γ (°)	 82.1, 112.0, 63.1 90, 90, 90	 82.9, 113.0, 63.2 90, 90, 90	 83.1, 113.0, 63.3 90, 90, 90
Resolution (Å)	63.07 – 1.35 (1.37 – 1.35)	45.92 – 1.6 (1.63 – 1.6)	66.93 – 1.8 (1.84 - 1.8)
I / σ(I)	14.3 (1.0)	14.6 (2.4)	16.5 (2.8)
Completeness (%)	100.0 (100.0)	86.5 (100.0)	82.3 (100)
Redundancy	12.9 (12.7)	12.3 (12.0)	9.0 (9.6)
$CC_{1/2}$	0.999 (0.394)	0.999 (0.777)	0.999 (0.805)
Refinement			
No. reflections	63042	34167	23160
R_{work}/R_{free}	0.161/0.189	0.201/0.224	0.185/0.220
No. atoms Protein Ligand/ion Water	 1992 6 252	 1959 4 184	 1913 5 122
B-factors Protein Ligand/ion Water	 23.36 34.04 35.47	 22.54 29.35 34.68	 34.16 44.49 35.83
R.m.s. deviations Bond lengths (Å) Bond angles (°)	 0.012 1.07	 0.016 1.37	 0.011 1.19
Ramachandran favored (%) outliers (%)	 98.74 0.00	 98.30 0.00	 97.85 0.43

PDB	**9QNJ**	**9QNL**	**9QNH**
Protein	14-3-3σΔC	14-3-3σΔC	14-3-3σΔC
Peptide	Tau pS198	BAD pS118	Myc pS294
Beam	ESRF ID23-2	ESRF ID23-2	ESRF ID23-2
DOI	10.2210/pdb9qnj/pdb	10.2210/pdb9qnl/pdb	10.2210/pdb9qnh/pdb
Data collection			
Wavelength (Å)	0.873128	0.873128	0.873128
Space group	C 2 2 21	C 2 2 21	C 2 2 21
Cell dimensions a, b, c (Å) α, β, γ (°)	 82.6, 112.4, 63.0 90, 90, 90	 82.6, 112.1, 63.0 90, 90, 90	 82.9, 112.7, 63.0 90, 90, 90
Resolution (Å)	66.57 – 1.3 (1.32 – 1.3)	45.72 – 1.3 (1.33 – 1.3)	66.76 – 1.3 (1.33 – 1.3)
I / σ(I)	24.7 (2.7)	19.4 (1.5)	31.2 (5.4)
Completeness (%)	100 (100)	89.1 (97.9)	88.3 (98.1)
Redundancy	13.0 (12.8)	13.0 (12.5)	13.5 (13.1)
$CC_{1/2}$	1.000 (0.848)	1.000 (0.589)	1.000 (0.948)
Refinement			
No. reflections	72103	62166	63470
R_{work}/R_{free}	0.163/0.183	0.170/0.202	0.189/0.209
No. atoms Protein Ligand/ion Water	 1996 8 262	 2007 6 238	 1991 8 280
B-factors Protein Ligand/ion Water	 20.48 48.06 34.19	 25.25 38.83 35.68	 19.23 31.73 32.02
R.m.s. deviations Bond lengths (Å) Bond angles (°)	 0.018 1.43	 0.009 1.03	 0.009 0.99
Ramachandran favored (%) outliers (%)	 98.29 0.00	 98.32 0.00	 98.72 0.00

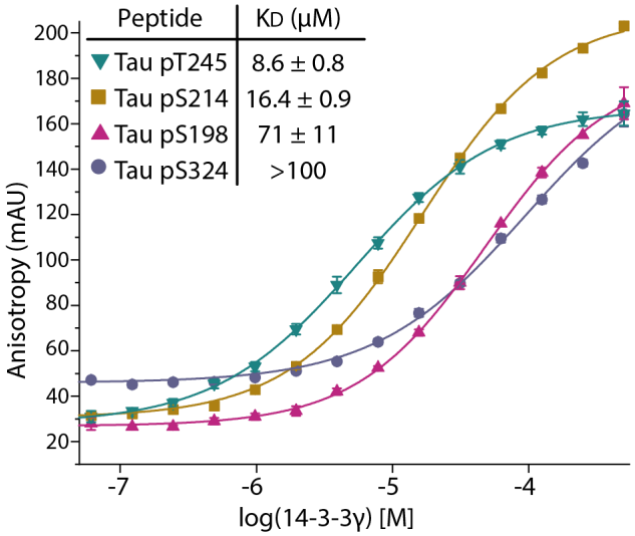
Supplementary Figures

Type	Training set	In-house
All		
Binders		
Non-binders		

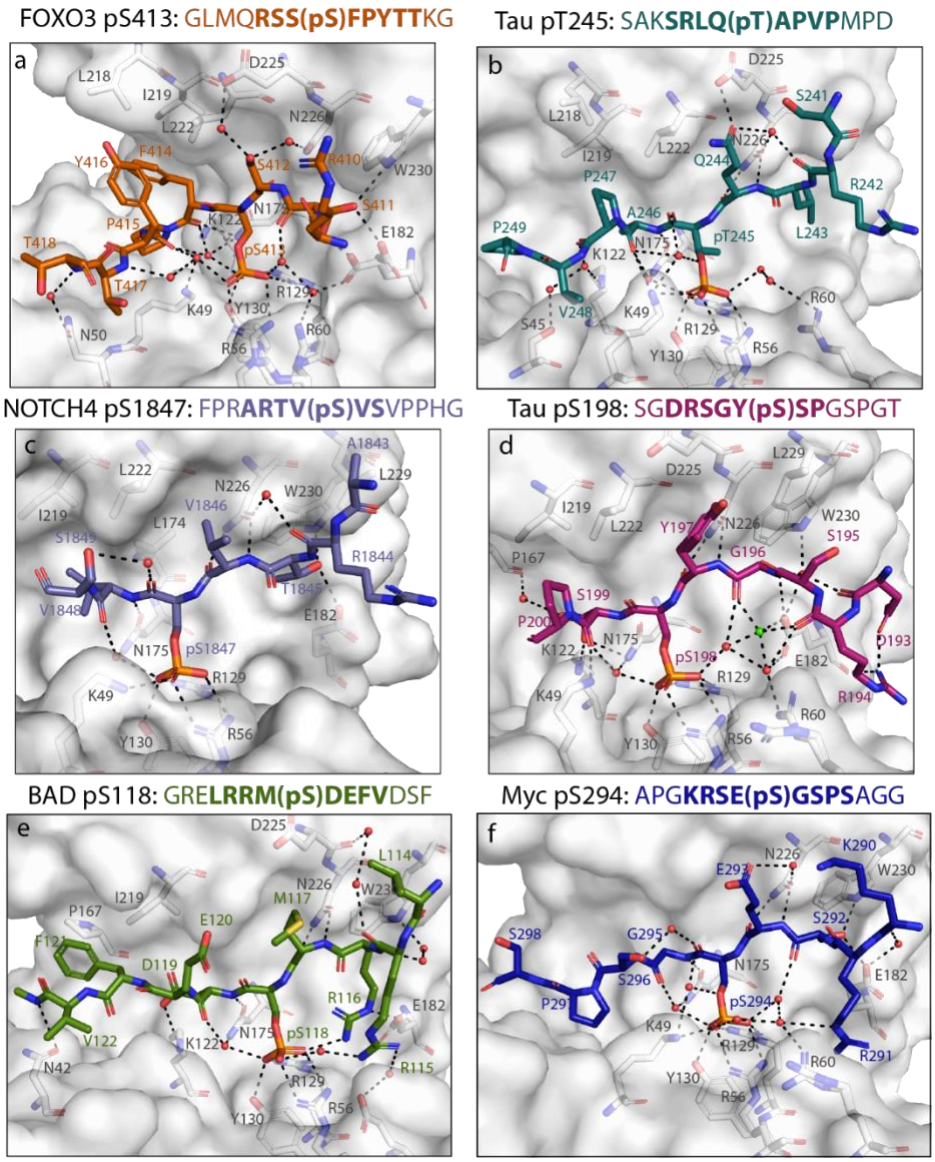
Supplementary Figure S1. LogoPlots of the sequences in literature dataset used to train and validate the models, divided by their binding label.



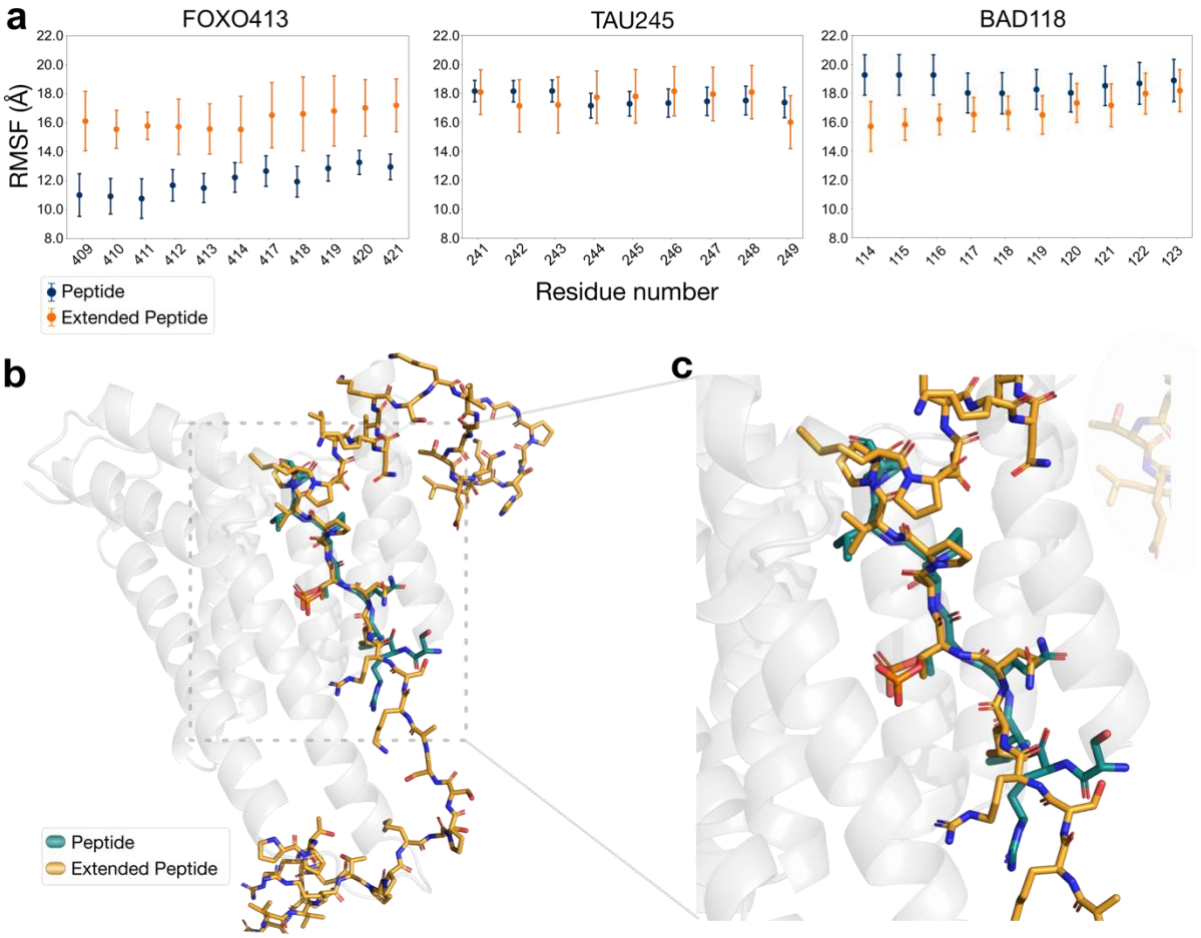
Supplementary Figure 2. Experimental validation. Eight putative binding sites (**1**-**8**) and two negative controls (**9**-**10**) were selected for experimental validation via fluorescence anisotropy (FA) assays. Dose-response curves are reported for each sequence, labeled as protein and phosphosite (pS = phosphoserine, pT = phosphothreonine), across three independent repeats.



Supplementary Figure S3. Titration of 14-3-3y to fluorescently labeled peptides Tau pS198 and pT245, and two known Tau binding sites,Tau pS214 and pS324 (10 nM). Data and K_D values are shown as mean ± SD (n=3).



Supplementary Figure S5. Interactions of predicted peptide sequences (**a**) FOXO3 pS413 (orange), (**b**) Tau pT245 (cyan), (**c**) NOTCH pS1847 (purple), (**d**) Tau pS198 (pink), (**e**) BAD pS118 (green), (**f**) Myc pS294 (blue)) with 14-3-3sigma (white surface) (relevant side chains and waters are displayed as stick and red dots, respectively, polar contacts are shown as black dashed lines).



Supplementary Figure S6. Structural insights into binding to 14-3-3. Molecular dynamics analysis to determine the stability of putative binding sites, by comparing three tested peptides (FOXO3-413, TAU-245 and BAD-118) with an extended version along the respective protein sequence (55 AAs). (**a**) Root mean squared fluctuation (RMSF [Å] – the lower, the more stable) values obtained for the selected peptides and the extended AA sequence. RMSF is reported per residue. (**b**) Structural overlay of the TAU245 peptide and its extended form into the 14-3-3 binding site after molecular dynamics simulation. (**c**) The binding poses of TAU245 at the interface with 14-3-3 are conserved among the 15 AAs and the 55 AAs versions throughout the MD simulations.