## PAPER

Check for updates

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

Laura van Weesep, [ID] †[a] Rıza Özçelik, [ID] †[abc] Marloes Pennings, [ID] [ad] Emanuele Criscuolo, [ID] [a] Christian Ottmann, [ID] [ad] Luc Brunsveld [ID] *[ad] and Francesca Grisoni [ID] *[abc]

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 potential 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 that achieved 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 predicted peptide sequences 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 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.*, relating to Alzheimer's disease as the binding to tau is not the new part), 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

*[a]Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven, The Netherlands. E-mail: l.brunsveld@tue.nl; f.grisoni@tue.nl*

*[b]Eindhoven AI Systems Institute (EAISI), Eindhoven University of Technology, Eindhoven, The Netherlands*

*[c]Centre for Living Technologies, Alliance TU/e, WUR, UU, UMC Utrecht, Utrecht, The Netherlands*

*[d]Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

† These authors contributed equally.

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, 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 aid in binding site identification to explore the 14-3-3 protein interactome, by leveraging deep learning on peptide sequences. Our approach was designed to predict putative sites of proteins 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 AAs preceding and seven AAs 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).

**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 AA 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 AAs.[29]

• Learnable embeddings, where AAs are assigned different, randomly initialized vectors.[30] These embeddings are updated during training to help capture contextual and relational information about the AAs in the sequence.

• BLOcks SUbstitution Matrix (BLOSUM 62),[31] where AAs are encoded with substitution matrix scores, which reflect evolutionary conservation and property similarities between AAs. This representation incorporates biochemically relevant information about AA substitutions. Phosphorylated AAs were indicated *via* a dedicated binary flag in the corresponding position (see Materials and methods).

• Physico-chemical descriptors, where each AA in the peptide is represented by 18 pre-computed numerical features[32] (SI Table S1). For each peptide, the computed AA features were concatenated and used for the prediction.

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

• Multilayer perceptron (MLP),[33] where complex peptide features are progressively extracted through multiple layers of fully-connected neurons,[34] without explicitly considering positional information.

• Convolutional neural network (CNN),[35] 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),[36] 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.

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, eqn (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, as shown by an increased recall (SI 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 14-3-3-Pred 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 (SI 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, eqn (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.
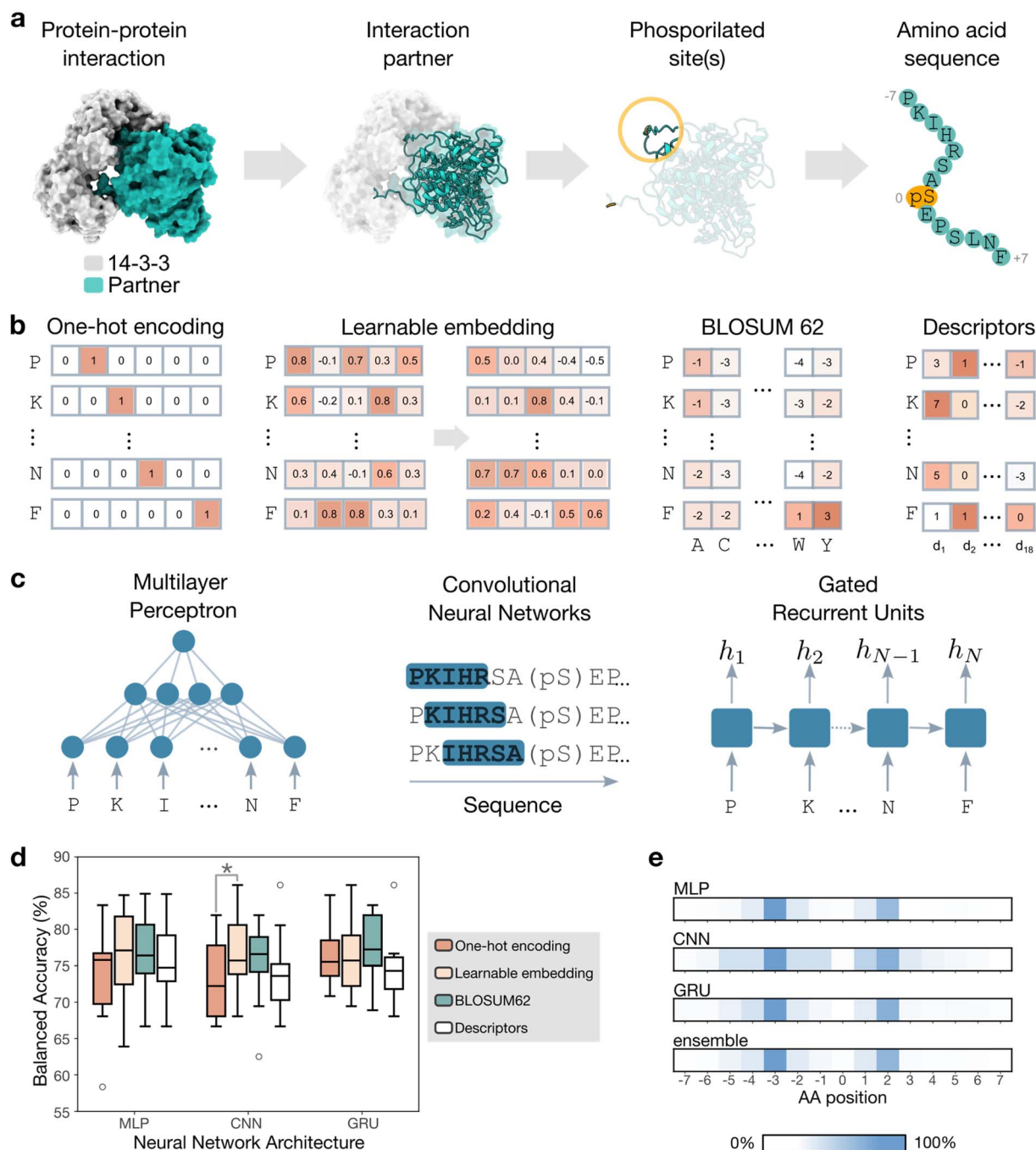
Fig. 1 Predicting peptide sequence binding to 14-3-3 with deep learning. (a) The information on tested 14-3-3 protein interactions (Table 1) was converted into a 'machine readable' format. Seven amino acids (AAs) before and after the phosphorylated site were used to form a sequence for deep learning, which is labeled according to its binding to 14-3-3 (yes/no). (b) Representations of AA sequences for model training. One-hot encoding represents each AA with a unique binary vector. Learnable embedding starts with a random numerical vector per AA 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 AA. (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 AAs 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).

**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[a] (50%) |
| In-house set | 92 | 58 (63%) | 34 (37%) |

[a] 93 experimentally determined (26%), and 262 (74%) likely non-binders.

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

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

**Model interpretation.** To shed light onto the binding patterns learned by the models, we conducted a virtual mutation study. We randomly shuffled (15 times) AAs occurring in each position, except for the phosphorylated AA, 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 AAs 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.[16] Finally, each modeling approach has a 'prediction hallmark', with a main focus on the AAs in the −5 and +3 position. Additional differences exist, albeit they are not particularly marked for CNN and GRU (both based on BLOSUM62 representation, Fig. 1e). This suggests that, although the individual models are trained on the same data, they might capture slightly different sequence-binding relationships. This might contribute to the increased performance of the ensemble model for most metrics (Table 2).

### 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),[37] Tau,[38] Myc,[39] Bcl-2-associated agonist of cell death

(BAD),[40] Notch-4,[41] Cystic fibrosis transmembrane conductance regulator (CFTR),[42] and p53.[43] These proteins contribute to a wide array of cellular processes[37,40,44] (*e.g.*, metabolism, cell survival and death) and are involved in diseases like cancer[45,46] (*e.g.*, BAD, p53 and Notch-4), Alzheimer's (Tau) and cystic fibrosis (CFTR).[47] The structures of these proteins range from ordered (CFTR and Notch-4: experimental/predicted disorder ratio[48,49] between 0% to 26%) to partially and highly disordered (p53, Myc, FOXO3, BAD, Tau; experimental/predicted disorder ratio[48,49] ranging from 38% to 95%, SI Table S3). Hence, they constitute an interesting and diverse test case for the 14-3-3 interactome.

For the selected proteins, their AA sequence was obtained from UniProt.[50] All serine and threonine residues were localized and a sequence window of 15 AAs was obtained (−7 and +7 around such AAs), leading to a total of 830 sequences. These sequences were further analyzed with PhosphositePlus[51] to verify whether they are phosphorylated *in vivo*. Only sequences labeled as phosphorylated (either according to literature[41] 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; SI 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 (Fig. 2a). 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 $K_d = 1.6 \pm 0.1$ μM (**1**, FOXO3 pS413), $K_d = 8.6 \pm 0.8$ μM (**2**, Tau pT245), and $K_d = 15.9 \pm 1.9$ 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 (Fig. 2a). As expected, the negative controls **9** and **10** did not bind, confirming the correctness of the model-based ranking.

Interestingly, peptide **2** (Tau-pT245) showed higher binding affinities than the known 14-3-3 interaction sites:[52,53] Tau-pS214 ($K_d = 16.4 \pm 0.9$ μM) and Tau-pS324 ($K_d > 100$ μM; SI Fig. S3). This is especially interesting considering that Tau-pT245 is reported to be only phosphorylated in a normal, non-Alzheimer brain.[54] The FOXO pS413 peptide showed similar affinity to the known FOXO pS253 site.[55] The p253 site is located near the nuclear locator site, whilst pS413 is located close to the nuclear exclusion site,[56] suggesting a potential dual role by 14-3-3.

The experimental $K_d$ values correlate well with the predicted binding scores ($r = 0.74$). Discrepancies exist for peptides with moderate predictions (scores ranging from 0.69 to 0.61), such as **5** (CFTR pS422) and **6** (BAD pS134). In this case, no evident

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 score, while peptides 9–10 were selected as negative controls (predicted to be non-binding with high confidence). Predicted outcome (binding and non binding, using a threshold above 0.5 in the predicted binding score) is also reported. 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 $\pm$ SD, $n = 3$]). Binding curves are reported in SI Fig. S2

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

relationship between AA similarity to training sequences and predicted binding scores exists. While this success ratio is in line with literature on machine learning for peptide discovery,[57–59] it underscores intrinsic model limitations, e.g., due to coverage of sequence-binding relationships, experimental error, and the presence of likely non-binders in the training set (Table 1). Moreover, binding data used for training come from diverse experimental assays, potentially contributing to differences with the reported FA measurements.

Moreover, when comparing our predictions with those of 14-3-3-Pred[26] and with 14-3-3 Site Finder, we observe moderate to no correlations between the predictions on the selected sequences (ranging from 0.17 to 0.61). Finally, the ranking obtained by our model correlates well with the observed $K_d$ predictions (0.74, SI Table S5). Additionally, when comparing the peptides to the most common 14-3-3 binding motifs (RSXpSXP,[16] RXY/FXpSXP,[16] pS/pT (X1–2)-COOH,[60] with X being any AA), seven out of eight sequences (except for sequence 1) would not have been found. These results corroborate the predictivity of our approach and its relevance to rationalize sequence binding to 14-3-3 beyond known common motifs.

**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. 2b–g). 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. 2h–m).

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. 2b). A similar binding conformation was observed for the GAB2 pT391 peptide,[61] which aligns perfectly at its +2 proline with FOXO3, and for p53 pT387,[62] which bends out of the pocket due to glycine and proline residues at the +2 and +3 positions, respectively (Fig. 2h). 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 (SI 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 (SI 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. 2c). The 'extended' binding mode is similar to peptides such as CRAF pS259[63] and TFEB pS211,[64] all of which fold back into the pocket after a minor turn induced by the +2 proline (Fig. 2i). 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
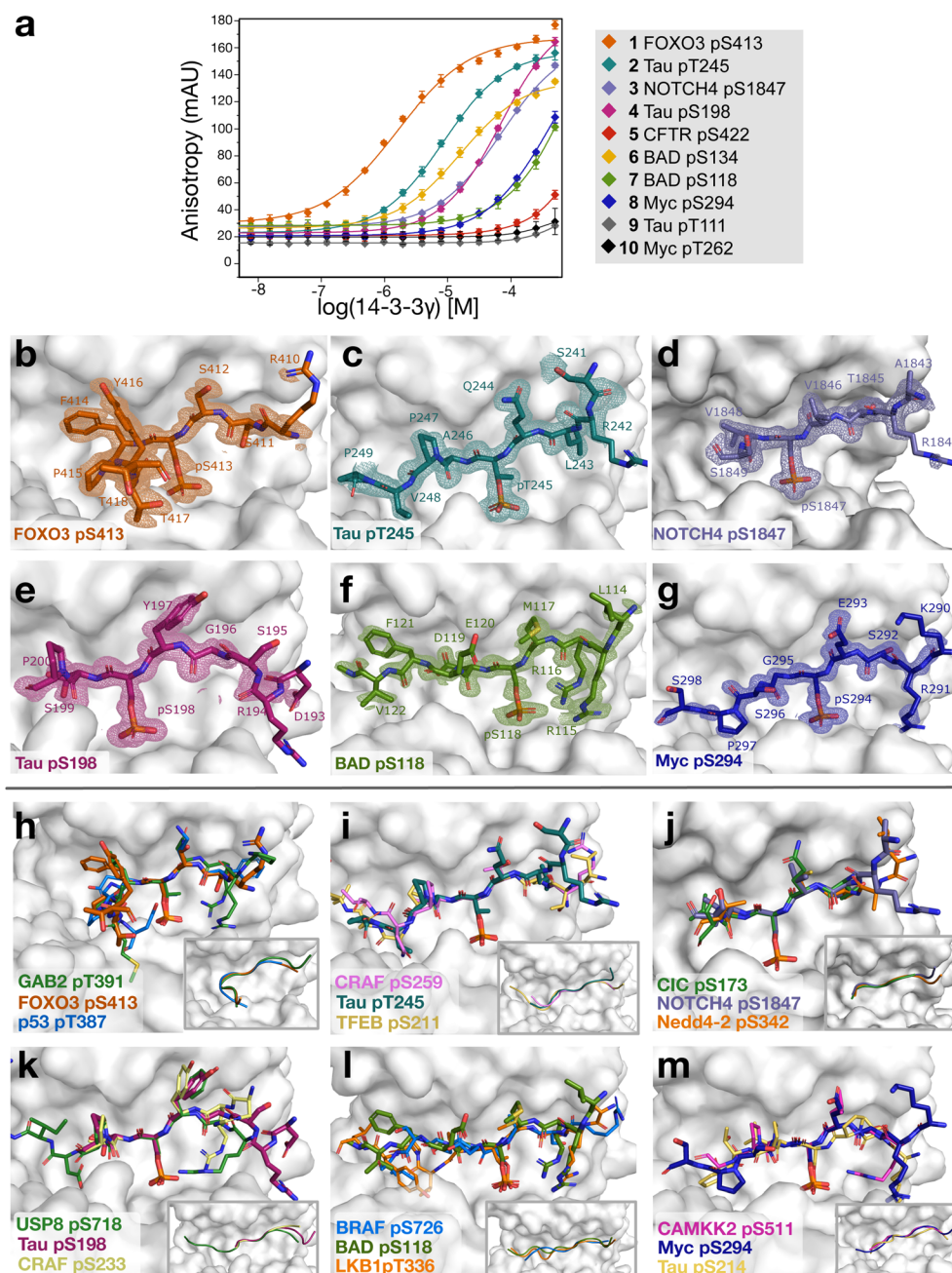
Fig. 2 Experimental characterization of the selected peptides. (a) Characterization *via* a fluorescence anisotropy assay. Binding curves are reported on a logarithmic scale for each sequence, labeled as protein and phosphosite (pS = phosphoserine, pT = phosphothreonine), across three independent repeats. (b–g) X-ray crystallography on selected peptide binders in comparison with known binders. 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 $1.0\sigma$. (b) FOXO3 pS413 (orange), (c) Tau pT245 (cyan), (d) NOTCH pS1847 (purple), (e) Tau pS198 (pink), (f) BAD pS118 (green), (g) Myc pS294 (blue). (h–m) Crystallographic overlay of predicted peptide sequences (h) FOXO3 pS413 (orange), (i) Tau pT245 (cyan), (j) NOTCH pS1847 (purple), (k) Tau pS198 (pink), (l) BAD pS118 (green), (m) 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.

additional AAs, see Materials and methods and SI Fig. S5b and c). In this context, Tau pT245 exhibited limited fluctuations in its interactions with 14-3-3 over time, as assessed by RMSF analysis (SI Fig. S5a).

For NOTCH4 pS1847, electron density was only observed up to the +2 serine, suggesting that the remaining residues are

disordered (Fig. 2d). Similar C-terminal disorder has been reported in crystal structures of the 14-3-3 clients CIC pS173[65] and Nedd4-2 pS342[66] (Fig. 2j). In addition, only the +1 and +2 residues were resolved in the Tau pS198 crystal structure (Fig. 2e). The −1 tyrosine residue of Tau pS198 was observed in previously reported structures of USP8 pS718[67] and CRAF pS233,[68]

where it fits into a pocket at the top of the 14-3-3 binding groove (Fig. 2k).

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. 2f and g). 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 (SI 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. 2i). Nevertheless, some similarities were revealed in the overlay with BRAF pS726 [69] and LKB1 pT336,[70] 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, especially visible from the −4 leucine to the −2 arginine residues. These analyses further support pS118 as a putative binding site of BAD to 14-3-3.

Finally, the Myc pS294 peptide forms a slight turn within the 14-3-3 pocket due to its +1 glycine, similar to CAMKK2 pS511 [71] and Tau pS214,[72] where this turn is induced by a +2 proline (Fig. 2g and m). 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 putative protein–protein interaction sites between 14-3-3 proteins and phosphorylated client proteins. By leveraging different amino-acid sequence representations and neural network architectures, we demonstrated that our models outperform existing tools in terms of global performance, as captured by balanced accuracy. 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 296 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 weak binders and two binders with marginal affinity. 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 potentially relevant binding sites for follow-up biological characterization, 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 and facilitate the translation into biologically relevant insights. 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. Our approach contributes to a deeper understanding of peptide–14-3-3 interactions – supporting the rational design of modulators, and expanding the available hypotheses on 14-3-3 related cellular signaling. 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,[73] 93 experimentally determined non-binding phosphosites[74] and 22 known binding sequences from the literature.[26] Moreover, 262 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 AA and truncated or padded to 15 AAs, 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 AAs, 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[75]), and, later, the top hyperparameter configurations (216 for GRU, 324 for MLP, and 1500 for CNN) were further fine-tuned (SI Table S6). Early stopping on $F$1 score was used starting from the fifth epoch, with a patience of five epochs. The model with the highest $F$1 score (eqn (5)) in 10-fold validation was selected. The final hyperparameters for each model are reported in SI 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:[76]

$$Rc = \frac{TP}{TP + FN} \tag{1}$$

$$Sp = \frac{TN}{TN + FP} \tag{2}$$

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

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 (eqn (1)) indicates the proportion of actual binders that the model successfully identifies, specificity (eqn (2)) assesses the reliability of non-binding predictions, and precision (eqn (3)) measures the capability to minimize false positives.

Moreover, the models were assessed for their global prediction ability, *via* balanced accuracy (BA) and $F$1-score:

$$BA = \frac{Rc + Sp}{2} \tag{4}$$

$$F1 = 2\frac{Pr + Rc}{Pr + Rc} \tag{5}$$

Balanced accuracy captures the model performance (correct predictions) normalized by the class imbalance, and $F$1 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 AA 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 embeddings. AAs are assigned a unique and randomly initialized vector. Phosphorylated AAs are assigned their own random vector. Vectors are then updated during model training.

● BLOSUM 62 representation was tested in two formats: (a) by treating phosphorylated AAs 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 $F$1 and balanced accuracy and hence it was used for this study.

● Peptide descriptors. For each AA, 18 descriptors were computed using the peptidy[32] software. Descriptors were concatenated, obtaining a matrix of 15 × 18 descriptors per peptide (see SI Table S1).

### Prospective screening on selected proteins

**Library preparation.** The AA sequence for the selected proteins was obtained from UniProt[50] (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 AAs was obtained (−7 and +7 around such AAs), leading to a total of 830 sequences. These sequences were further analyzed with PhosphoSitePlus,[51,77] 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 and validated according to existing literature were excluded from the wet-lab validation (SI Table S4).

### Experimental validation

**Peptide materials.** Selected peptides were ordered from GenScript[78] 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. Acetylated peptides were ordered for crystallography with a minimal purity of 95%. Peptide sequences are reported in SI Table S8.

**Protein expression and purification.** The full-length (FL) human 14-3-3γ protein was expressed from a pPROEX plasmid after transformation to BL21(DE3) competent *E. coli* (Novagen). Cultures were incubated at 37 °C, 140 rpm until $OD_{600} \sim 0.8$ was reached. Protein expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.4 mM) and cells were harvested by centrifugation (10 min, 4 °C, $16\,000 \times g$) after overnight expression (18 °C, 140 rpm). Pellets were resuspended in wash buffer (50 mM Tris pH 8.0, 300 mM NaCl, 12.5 mM imidazole and 2 mM β-mercaptoethanol (βME)). After homogenizing the cells (40 bar, Emulsiflex-C3 homogenizer), the soluble fraction was collected by centrifugation (30 min, 4 °C, $40\,000 \times g$) and loaded onto a $Ni^{2+}$-affinity column pre-equilibrated with wash buffer. After a washing step (wash buffer + 20 mM imidazole), the bound protein was eluted with 200 mM imidazole followed by dialysis overnight at 4 °C (25 mM HEPES pH 8.0, 200 mM NaCl, 10 mM MgCl$_2$, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)). The His$_6$-tag of the ΔC variant (14-3-3σ truncated after S232) for crystallography was cleaved with TEV-protease during dialysis and subjected to an additional purification by size exclusion chromatography (SEC; Superdex 75; buffer 20 mM HEPES pH 7.5, 100 mM NaCl, 10 mM MgCl$_2$, 0.5 mM TCEP). The pure protein was concentrated, aliquoted, flash-frozen in liquid N$_2$, and stored at −80 °C. Purity and exact mass were determined (SI Fig. S6) using a high-resolution liquid chromatography coupled with mass spectrometry (LC/MS) system comprised of an I-Class Acquity UPLC (Waters) with a Polaris C18A reverse-phase column $2.0 \times 100$ mm (Agilent), coupled to a Xevo G2 Quadrupole Time of Flight mass spectrometer (Waters). A flow rate of 0.3 mL min$^{-1}$ was used with a gradient of acetonitrile + 0.1% formic acid (FA) in water + 0.1% FA (acetonitrile 15–75%). Deconvolution of the *m/z* spectra was done using the MaxENT$_I$ algorithm in the Masslynx v4.1 (SCN862) software.

**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.[79] For all experiments, 14-3-3γ was used as it was shown in multiple experiments to be the strongest binding variant suitable for experimental screening.[80] 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$^{-1}$ 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 μs, settle time: 1 μs; 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 (using the Hill equation). Data shown

is the average and standard deviation of triplicates. For dose–response curves on a linear scale seep SI Fig. S2.

## 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 MgCl$_2$ 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$^{-1}$. 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 CaCl$_2$, 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 (https://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 SI Table S9 for data collection and refinement statistics. The structures were deposited in the protein data bank (PDB) with IDs: **9QNG** (FOXO pS413), **9QNH** (Myc pS294), **9QNI** (NOTCH4 pS1847), **9QNJ** (Tau pS198), **9QNK** (Tau pT245), **9QNL** (BAD pS118).

## Molecular dynamics simulations

To investigate how our experimental results could be extended beyond the experimentally determined peptides, we selected three sequences: FOXO3 pS413 (**1**, $K_d = 1.6 \pm 0.1$ μM), TAU pT245 (**2**, $K_d = 8.5 \pm 0.2$ 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,[50] by elongating the tested sequences with 20 AAs in both N- and C-terminal directions (55 AAs 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 (initial sequence), molecular dynamics (MD) simulations were performed using GROMACS 2023[81] 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$^{-1}$ nm$^{-1}$ 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.

Following, the initial peptide structures were extended to a sequence of 55 AAs using PyMOL.[82] 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 = $400 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) and side chains (force constant = $40 \text{ kJ mol}^{-1} \text{ nm}^{-2}$). 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,[83] 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.

## 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: FG, LvW, MP, 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.

## Conflicts of interest

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

## Data availability

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/**. The code and data associated with this work have been deposited in Zenodo at the time of publication; they can be accessed *via* the following DOI: **https://doi.org/10.5281/zenodo.16420029**.

Supplementary tables (Tables S1–S9) and figures (Fig. S1–S6) are available. See DOI: **https://doi.org/10.1039/d5dd00132c**.

## Acknowledgements

## References

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

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

3 A. K. Gardino, S. J. Smerdon and M. B. Yaffe, 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, DOI: **10.1016/j.semcancer.2006.03.007**.

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

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

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

7 B. Thurairajah, A. J. Hudson and R. G. Doveston, Contemporary Biophysical Approaches for Studying 14-3-3 Protein-Protein Interactions, *Front. Mol. Biosci.*, 2022, **9**, 1043673, DOI: 10.3389/fmolb.2022.1043673.

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

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

10 A. Aitken, S. Howell, D. Jones, J. Madrazo and Y. Patel, 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, DOI: 10.1074/jbc.270.11.5706.

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

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

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

14 C. Ottmann, S. Marco, N. Jaspert, C. Marcon, N. Schauer, M. Weyand, C. Vandermeeren, G. Duby, M. Boutry, A. Wittinghofer, J. L. Rigaud and C. Oecking, 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 O. Keskin, A. Gursoy, B. Ma and R. Nussinov, Principles of Protein–Protein Interactions: What Are the Preferred Ways For Proteins To Interact?, *Chem. Rev.*, 2008, **108**(4), 1225–1244, DOI: 10.1021/cr040409x.

16 M. B. Yaffe, K. Rittinger, S. Volinia, P. R. Caron, A. Aitken, H. Leffers, S. J. Gamblin, S. J. Smerdon and L. C. Cantley, The Structural Basis for 14-3-3: Phosphopeptide Binding Specificity, *Cell*, 1997, **91**(7), 961–971.

17 M. Brandt, J. C. Madsen, J. Bunkenborg, O. N. Jensen, S. Gammeltoft and K. J. Jensen, On-Bead Chemical Synthesis and Display of Phosphopeptides for Affinity Pull-Down Proteomics, *ChemBioChem*, 2006, **7**(4), 623–630, DOI: 10.1002/cbic.200500378.

18 A. Aitken, 14-3-3 Proteins: A Historic Overview, *Semin. Cancer Biol.*, 2006, **16**(3), 162–172, DOI: 10.1016/j.semcancer.2006.03.005.

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

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

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

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

23 J. Zhang, J. Durham and C. Qian, Revolutionizing Protein–Protein Interaction Prediction with Deep Learning, *Curr. Opin. Struct. Biol.*, 2024, **85**, 102775, DOI: 10.1016/j.sbi.2024.102775.

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

25 Y. Fan, X. Wang and C. Wang, 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, doi: DOI: 10.1145/3431943.3431951.

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

27 D. M. Bustos and A. A. Iglesias, Intrinsic Disorder Is a Key Characteristic in Partners That Bind 14-3-3 Proteins, *Proteins:Struct., Funct., Bioinf.*, 2006, **63**(1), 35–42, DOI: 10.1002/prot.20888.

28 V. Perovic, N. Sumonja, L. A. Marsh, S. Radovanovic, M. Vukicevic, S. G. E. Roberts and N. Veljkovic, IDPpi: Protein-Protein Interaction Analyses of Human Intrinsically Disordered Proteins, *Sci. Rep.*, 2018, **8**(1), 10563, DOI: 10.1038/s41598-018-28815-x.

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

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

31 S. Henikoff and J. G. Henikoff, Amino Acid Substitution Matrices from Protein Blocks, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**(22), 10915–10919, DOI: 10.1073/pnas.89.22.10915.

32 R. Özçelik, L. van Weesep, S. de Ruiter and F. Grisoni, Peptidy: a light-weight Python library for peptide

representation in machine learning, *Bioinform. adv.*, 2025, 5(1), vbaf058, DOI: **10.1093/bioadv/vbaf058**.

33 K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. D. Ser and K. Muhammad, Deep Neural Networks in the Cloud: Review, Applications, Challenges and Research Directions, *Neurocomputing*, 2023, **545**, 126327, DOI: **10.1016/j.neucom.2023.126327**.

34 S. Abdel-Nasser, Principle of Neural Network and Its Main Types: Review, *J. Adv. Appl. Comput. Math.*, 2020, **7**, 8–19, DOI: **10.15377/2409-5761.2020.07.2**.

35 Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Comput.*, 1989, **1**(4), 541–551.

36 K. Cho, B. Van Merriënboer, D. Bahdanau and Y. Bengio, On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, *arXiv*, 2014, preprint, arXiv:14091259, doi: 10.48550/arXiv.14091259.

37 R. J. Stefanetti, S. Voisin, A. Russell and S. Lamon, Recent Advances in Understanding the Role of FOXO3, *F1000Research*, 2018, **7**, 1372, DOI: **10.12688/f1000research.15258.1**.

38 M. Goedert, M. Spillantini, N. Cairns and R. Crowther, Tau Proteins of Alzheimer Paired Helical Filaments: Abnormal Phosphorylation of All Six Brain Isoforms, *Neuron*, 1992, **8**(1), 159–168.

39 Y. Huang, M. Yang and W. Huang, 14-3-3 σ: A Potential Biomolecule for Cancer Therapy, *Clin. Chim. Acta*, 2020, **511**, 50–58, DOI: **10.1016/j.cca.2020.09.009**.

40 E. Yang, J. Zha, J. Jockel, L. H. Boise, C. B. Thompson and S. J. Korsmeyer, Bad, a Heterodimeric Partner for Bcl-xL and Bcl-2, Displaces Bax and Promotes Cell Death, *Cell*, 1995, **80**(2), 285–291, DOI: **10.1016/0092-8674(95)90411-5**.

41 G. Ramakrishnan, G. Davaakhuu, W. C. Chung, H. Zhu, A. Rana, A. Filipovic, A. R. Green, A. Atfi, A. Pannuti, L. Miele and G. Tzivion, AKT and 14-3-3 Regulate Notch4 Nuclear Localization, *Sci. Rep.*, 2015, **5**(1), 8782, DOI: **10.1038/srep08782**.

42 R. Boucher, New Concepts of the Pathogenesis of Cystic Fibrosis Lung Disease, *Eur. Respir. J.*, 2004, **23**(1), 146–158.

43 O. Timofeev, Editorial: Mutant P53 in Cancer Progression and Personalized Therapeutic Treatments, *Front. Oncol.*, 2021, **11**, 740578, DOI: **10.3389/fonc.2021.740578**.

44 L. García-Gutiérrez, M. D. Delgado and J. León, MYC Oncogene Contributions to Release of Cell Cycle Brakes, *Genes*, 2019, **10**(3), 244.

45 H. Harrison, G. Farnie, S. J. Howell, R. E. Rock, S. Stylianou, K. R. Brennan, N. J. Bundred and R. B. Clarke, Regulation of Breast Cancer Stem Cell Activity by Signaling through the Notch4 Receptor, *Cancer Res.*, 2010, **70**(2), 709–718.

46 X. B. Stickles, D. C. Marchion, E. Bicaku, E. Al Sawah, F. Abbasi, Y. Xiong, N. Bou Zgheib, B. M. Boac, B. C. Orr and P. L. Judson, others. BAD-Mediated Apoptotic Pathway Is Associated with Human Cancer Development, *Int. J. Mol. Med.*, 2015, **35**(4), 1081–1087.

47 R. L. Gibson, J. L. Burns and B. W. Ramsey, Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis, *Am. J. Respir. Crit. Care Med.*, 2003, **168**(8), 918–951.

48 M. C. Aspromonte, M. V. Nugnes, F. Quaglia, A. Bouharoua, C. DisProt, S. C. E. Tosatto and D. Piovesan, DisProt in 2024: Improving Function Annotation of Intrinsically Disordered Proteins, *Nucleic Acids Res.*, 2024, **52**(D1), D434–D441, DOI: **10.1093/nar/gkad928**.

49 D. Piovesan, A. Del Conte, M. Mehdiabadi, M. C. Aspromonte, M. Blum, G. Tesei, S. von Bülow, K. Lindorff-Larsen and S. C. E. Tosatto, MOBIDB in 2025: Integrating Ensemble Properties and Function Annotations for Intrinsically Disordered Proteins, *Nucleic Acids Res.*, 2025, **53**(D1), D495–D503, DOI: **10.1093/nar/gkae969**.

50 The UniProt Consortium, UniProt: A Worldwide Hub of Protein Knowledge, *Nucleic Acids Res.*, 2019, **47**(D1), D506–D515, DOI: **10.1093/nar/gky1049**.

51 P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham and E. Skrzypek, PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations, *Nucleic Acids Res.*, 2015, **43**, D512–D520.

52 J. Hochmair, M. C. M. van den Oetelaar, L. Ravatt, L. Diez, L. J. M. Lemmens, R. Ponce-Lina, R. Sankar, M. Franck, G. Nolte, E. Semenova, S. Mohapatra, C. Ottmann, L. Brunsveld and S. Wegmann, Stoichiometric 14-3-3ζ Binding Promotes Phospho-Tau Microtubule Dissociation and Reduces Aggregation and Condensation, *Commun. Biol.*, 2025, **8**, 1139, DOI: **10.1038/s42003-025-08548-0**.

53 J. F. Neves, O. Petrvalská, F. Bosica, F.-X. Cantrelle, H. Merzougui, G. O'Mahony, X. Hanoulle, T. Obšil and I. Landrieu, 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, DOI: **10.1111/febs.15574**.

54 G. Šimić, M. Babić Leko, S. Wray, C. Harrington, I. Delalle, N. Jovanov-Milošević, D. Bažadona, L. Buée, R. De Silva, G. Di Giovanni, C. Wischik and P. R. Hof, Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies, *Biomolecules*, 2016, **6**(1), 6, DOI: **10.3390/biom6010006**.

55 S. Mathivanan, P. K. C. Lakshman, M. Singh, S. Giridharan, K. Sathish, M. A. Hurakadli, K. Bharatham and N. Kamariah, 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, DOI: **10.1021/acsomega.2c01700**.

56 K. Psenakova, K. Kohoutova, V. Obsilova, M. J. Ausserlechner, V. Veverka and T. Obsil, Forkhead Domains of FOXO Transcription Factors Differ in Both Overall Conformation and Dynamics, *Cells*, 2019, **8**(9), 966, DOI: **10.3390/cells8090966**.

57 F. Grisoni, C. S. Neuhaus, M. Hishinuma, G. Gabernet, J. A. Hiss, M. Kotera and G. Schneider, De Novo Design of Anticancer Peptides by Ensemble Artificial Neural

Networks, *J. Mol. Model.*, 2019, **25**(5), 112, DOI: **10.1007/s00894-019-4007-6**.

58 P. Szymczak, M. Możejko, T. Grzegorzek, R. Jurczak, M. Bauer, D. Neubauer, K. Sikora, M. Michalski, J. Sroka, P. Setny, W. Kamysz and E. Szczurek, Discovering Highly Potent Antimicrobial Peptides with Deep Generative Model HydrAMP, *Nat. Commun.*, 2023, **14**(1), 1453, DOI: **10.1038/s41467-023-36994-z**.

59 M. D. T. Torres, M. C. R. Melo, L. Flowers, O. Crescenzi, E. Notomista and C. de la Fuente-Nunez, Mining for Encrypted Peptide Antibiotics in the Human Proteome, *Nat. Biomed. Eng.*, 2022, **6**(1), 67–75, DOI: **10.1038/s41551-021-00801-1**.

60 S. Ganguly, J. L. Weller, A. Ho, P. Chemineau, B. Malpaux and D. C. Klein, Melatonin Synthesis: 14-3-3-Dependent Activation and Inhibition of Arylalkylamine N-Acetyltransferase Mediated by Phosphoserine-205, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(4), 1222–1227.

61 D. Bier, M. Bartel, K. Sies, S. Halbach, Y. Higuchi, Y. Haranosono, T. Brummer, N. Kato and C. Ottmann, Small-Molecule Stabilization of the 14-3-3/Gab2 Protein-Protein Interaction (PPI) Interface, *ChemMedChem*, 2016, **11**(8), 911–918, DOI: **10.1002/cmdc.201500484**.

62 B. Schumacher, J. Mondry, P. Thiel, M. Weyand and C. Ottmann, 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, DOI: **10.1016/j.febslet.2010.02.065**.

63 M. Molzan, B. Schumacher, C. Ottmann, A. Baljuls, L. Polzien, M. Weyand, P. Thiel, R. Rose, M. Rose, P. Kuhenne, M. Kaiser, U. R. Rapp, J. Kuhlmann and C. Ottmann, 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, DOI: **10.1128/MCB.01636-09**.

64 Y. Xu, J. Ren, X. He, H. Chen, T. Wei and W. Feng, YWHA/14-3-3 Proteins Recognize Phosphorylated TFEB by a Noncanonical Mode for Controlling TFEB Cytoplasmic Localization, *Autophagy*, 2019, **15**(6), 1017–1030, DOI: **10.1080/15548627.2019.1569928**.

65 Y. Ren, Z. Ouyang, Z. Hou, Y. Yan, Z. Zhi, M. Shi, M. Du, H. Liu, Y. Wen and Y. Shao, CIC Is a Mediator of the ERK1/2-DUSP6 Negative Feedback Loop, *iScience*, 2020, **23**(11), 101635, DOI: **10.1016/j.isci.2020.101635**.

66 P. Pohl, R. Joshi, O. Petrvalska, T. Obsil and V. Obsilova, 14-3-3-Protein Regulates Nedd4-2 by Modulating Interactions between HECT and WW Domains, *Commun. Biol.*, 2021, **4**(1), 899, DOI: **10.1038/s42003-021-02419-0**.

67 F. Centorrino, A. Ballone, M. Wolter and C. Ottmann, Biophysical and Structural Insight into the USP8/14-3-3 Interaction, *FEBS Lett.*, 2018, **592**(7), 1211–1220, DOI: **10.1002/1873-3468.13017**.

68 M. Molzan and C. Ottmann, 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, DOI: **10.1016/j.jmb.2012.08.009**.

69 J. A. Martinez Fiesco, D. E. Durrant, D. K. Morrison and P. Zhang, Structural Insights into the BRAF Monomer-to-Dimer Transition Mediated by RAS Binding, *Nat. Commun.*, 2022, **13**(1), 486, DOI: **10.1038/s41467-022-28084-3**.

70 Y. Lu, S. Ding, R. Zhou and J. Wu, Structure of the Complex of Phosphorylated Liver Kinase B1 and 14-3-3ζ, *Acta Crystallogr., Sect. F:Struct. Biol. Commun.*, 2017, **73**(4), 196–201, DOI: **10.1107/S2053230X17003521**.

71 A. Ballone, R. A. Lau, F. P. A. Zweipfenning and C. Ottmann, A New Soaking Procedure for X-Ray Crystallographic Structural Determination of Protein-Peptide Complexes, *Acta Crystallogr., Sect. F:Struct. Biol. Commun.*, 2020, **76**(10), 501–507, DOI: **10.1107/S2053230X2001122X**.

72 Y. Joo, B. Schumacher, I. Landrieu, M. Bartel, C. Smet-Nocca, A. Jang, H. S. Choi, N. L. Jeon, K.-A. Chang, H.-S. Kim, C. Ottmann and Y.-H. Suh, Involvement of 14-3-3 in Tubulin Instability and Impaired Axon Development Is Mediated by Tau, *FASEB J.*, 2015, **29**(10), 4133–4144, DOI: **10.1096/fj.14-265009**.

73 M. Tinti, F. Madeira, G. Murugesan, G. Hoxhaj, R. Toth and C. MacKintosh, ANIA: ANnotation and Integrated Analysis of the 14-3-3 Interactome, *Database*, 2014, **2014**, 1–15, DOI: **10.1093/database/bat085**.

74 C. Johnson, S. Crowther, M. J. Stafford, D. G. Campbell, R. Toth and C. MacKintosh, Bioinformatic and Experimental Survey of 14-3-3-Binding Sites, *Biochem. J.*, 2010, **427**(1), 69–78.

75 R. Özçelik and F. A. Grisoni, Hitchhiker's Guide to Deep Chemical Language Processing for Bioactivity Prediction, *arXiv*, 2024, preprint, arXiv:2407.12152, doi: DOI: **10.48550/arXiv.2407.12152**.

76 D. Ballabio, F. Grisoni and R. Todeschini, Multivariate Comparison of Classification Performance Measures, *Chemom. Intell. Lab. Syst.*, 2018, **174**, 33–44, DOI: **10.1016/j.chemolab.2017.12.004**.

77 *PhosphoSitePlus – Site Search*, **https://www.phosphosite.org/siteSearchAction**, accessed 2025-01-15.

78 *GenScript*, **https://www.genscript.com/**, accessed 2025-01-15.

79 W. A. Lea and A. Simeonov, Fluorescence Polarization Assays in Small Molecule Screening, *Expert Opin. Drug Discovery*, 2011, **6**(1), 17–32, DOI: **10.1517/17460441.2011.537322**.

80 G. Gogl, K. V. Tugaeva, P. Eberling, C. Kostmann, G. Trave and N. N. Sluchanko, Hierarchized Phosphotarget Binding by the Seven Human 14-3-3 Isoforms, *Nat. Commun.*, 2021, **12**(1), 1677, DOI: **10.1038/s41467-021-21908-8**.

81 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, GROMACS: Fast, Flexible, and Free, *J. Comput. Chem.*, 2005, **26**(16), 1701–1718, DOI: **10.1002/jcc.20291**.

82 W. L. DeLano, others. Pymol: An Open-Source Molecular Graphics Tool, *CCP4 Newsl. Protein Crystallogr.*, 2002, **40**(1), 82–92.

83 M. Parrinello and A. Rahman, Crystal Structure and Pair Potentials: A Molecular-Dynamics Study, *Phys. Rev. Lett.*, 1980, **45**(14), 1196–1199, DOI: **10.1103/PhysRevLett.45.1196**.