DEVELOPMENT OF PREDICTIVE QSAR MODELS ON THE PHOSHOPEPTIDE BINDING AFFINITY AGAINST 14-3-3 ISOFORMS

A PREPRINT

Ying Fan

Shenzhen Institute of
Information Technology
Shenzhen, 518172
China

Xiaojun Wang

Shenzhen Institute of
Information Technology
Shenzhen, 518172
China

Chao Wang *

HKBU Institute for Research
and Continuing Education
Shenzhen 518057
China

July 24, 2020

ABSTRACT

14-3-3s present in multiple isoforms in human cells and mediate signal transduction by binding to phosphoserine-containing proteins. Recent findings have demonstrated that 14-3-3s act as a key factor in promoting chemoresistance of cancer. Here, we aimed to develop the predictive models that can determine the binding affinity of phosphopeptide fragments against 14-3-3s. It is found that the hydrophobic property of residues in phosphopeptides has a significant contribution to the binding affinity for most of 14-3-3 isoforms. The conserved patterns of 14-3-3 binding motifs were verified by our predictions. A group of peptide sequences were predicted with high binding affinity and high sequence conservation, which had agreement with 14-3-3s ligands. Overall, our results demonstrate that how the residues are likely to function in 14-3-3s interaction and the computational methods we introduced may contribute to the further research.

Keywords 14-3-3s · bioinformatics · QSAR · peptide microarray

1 Introduction

14-3-3 proteins constitute a family of highly conserved acidic proteins with the size of around 30 kDa [2, 53, 1, 4]. These proteins express in high abundance in a wide range of eukaryotic cells and exist in many tissues [16]. It is found that 14-3-3s present in multiple isoforms in human cells and function as either homo- or hetero-dimer formation [16, 38]. The dimer presents a conserved groove that served as the phosphopeptide binding pocket and positively charged residues (e.g. K49, R56, and R127) in the pocket interact with the phosphorylated binding partners/ligands. 14-3-3s binding to specific phosphoserine-containing motifs could result in the assembly of important signaling complexes [39]. The interactions of 14-3-3s with signaling proteins are critical for the activation of signal transduction processes [11]. All of these processes involve multiple isoforms of 14-3-3. Through interaction with their binding partners, 14-3-3s involve in the control of diverse cellular processes, including metabolism, transcription, cell cycle, differentiation, migration and apoptosis [50, 25]. Recent findings have demonstrated that 14-3-3s as a key factor in promoting chemoresistance of cancer cell [40, 46, 51, 19]. 14-3-3s mediated tumorigenesis and chemoresistance by gene overexpression has been observed by the animal studies [42]. Also, high expression level of 14-3-3s was found to be associated with disease recurrence and poor survival in the cancer patients received chemotherapy [32, 59, 22]. These evidences suggest that 14-3-3s involve in chemoresistance in multiple human cancers as key regulators. This significant association of 14-3-3s overexpression with poor patient outcomes makes this protein an attractive candidate targets for effective cancer management. Hence, 14-3-3s can be a novel biomarker for the diagnosis and a potential therapeutic target in cancer therapy.

*Corresponding author: wangchao@hkbu.edu.hk

The human 14-3-3 protein family consists of seven distinct but related isoforms, which are named β , γ , ϵ , η , σ , τ , ζ . The seven isoforms are highly conserved in sequence and structure, each expressed by a particular gene [53]. Hence, there is functional overlap between 14-3-3s, and the ability of one isoform to compensate for the loss of another has been demonstrated in some cases [55]. A comparison of crystal structures from different human 14-3-3s shows striking structural similarities [18]. However, the distinct differences existed for each 14-3-3 isoform existed in the preference of the dimerization and binding specificity [18]. Isoform γ have both two kind of dimeric formations, but mainly with ϵ for heterodimers [55]. Only heterodimer but not homodimer formation is found in the isoform ϵ [52, 55]. Both homo- and hetero-dimeric formations have been found in 14-3-3s. One exception is that isoforms σ prefer to form homodimers [55, 3, 18]. These preferences derive from the slightly structural and sequence variances. Also, the binding partners of 14-3-3s show a distinct preference for a particular isoform. The studies to determine 14-3-3s binding sites reveal that 14-3-3s bind to phosphoserine/ threonine (*pS/T*) in a sequence-specific manner [60, 36]. There are three modes of motifs found in 14-3-3s binding sites: mode-1 (RSX*p*SXP), mode-2 (PXF/YX*p*SXP) and mode-3 (SW*p*TX) at C-terminal [28]. While 14-3-3s are known as the family of phosphor-binding proteins, several exceptions related to out of motif or even without phosphor-specific manner were also identified [28]. The motif of phosphoserine-dependent interaction is important for high affinity 14-3-3s binding. This indicated that both phosphoserine and its flanking residues take part into the binding of 14-3-3s with its target ligand.

The binding affinity of 14-3-3s with its ligand can be obtained by high throughput combinatorial peptide microarray method [31]. Upon this, the 14-3-3s binding specificity to a wide variety of phosphopeptides can be determined in an experiment. But there is a lot of limitations exists still. For example, for a tripeptide, the number of possible sequence arrangements are 20^3 . Owing to the extremely large amount of possible arrangements, predicting the binding affinity of 14-3-3s with its ligand may be a challenging research topic in life science. Therefore, it is vitally important to develop a computational method for efficiently predicting peptides. At sequence level, the quantitative structure-activity relationship (QSAR) have been broadly used to infer the biological function and to predict high bioactivity peptide [10]. One of QSAR based approach is amino acid-based peptide prediction, in which the QSAR is used for the protein and peptide activity prediction [10]. For a given 14-3-3 protein, the property of binding affinity can be seen as a kind of bioactivity of the peptide sequence. In QSAR studies, linear regression has been commonly used to link the biological activities as a response variable to the molecular descriptors as predictor variables for data analysis [65]. Variable selection using penalized methods plays a vital role in statistical modeling with high dimensional QSAR data [49]. It aims to select only a subset of important descriptors from a large number of molecular descriptors, and thereby to improve the performance of QSAR models in terms of obtaining higher prediction accuracy of the model and easy interpretation. Penalized regression methods provide an estimate QSAR model that has lower prediction errors than multiple linear regression [44].

The details of the interaction between 14-3-3s and its ligands are not fully understood still, including the isoforms specificity, sequence patterns of the ligands, and the structural features of the interaction. Considering there are a great number of proteins that found to bind 14-3-3s *in vitro* and *in vivo*, it is meaningful to determine 14-3-3s binding interactions at large scale. Here, the computational predictive models to identify the binding affinity of the phosphopeptide sequences in seven 14-3-3 isoforms were reported. The raw affinity data was retrieved for literatures and processed before analysis. The divide physicochemical property scores (DPPS) were applied to generate variables matrix. The QSAR models were developed by linear regression method for phosphopeptide sequence prediction. The prediction of 14-3-3s binding affinity based on the data integrated with QSAR modelling. Based on this, target-specificity of seven mammalian 14-3-3 proteins were determined and the putative 14-3-3 motifs were identified.

2 Results

The purpose of this study is to develop the predictive QSAR models that can determine the binding affinity of phosphopeptide fragments against seven mammalian 14-3-3 isoforms. It helps us to decode the ligand interaction at residue level. The dataset contained the raw affinity values for 14-3-3s with phosphopeptide fragments was transformed and processed. The variables for describing the binding affinity of phosphopeptide fragments were extracted by DPPS. Penalized regression method was utilized to perform QSAR study and the models for seven isoforms were created by elastic net. Based on the established QSAR models, the contributions of various residues were analyzed. Finally, we evaluated the relative binding affinity of all possible sequence arrangements from the N- and C-terminal.

2.1 Dataset

Although the overall binding profiles of seven 14-3-3 isoforms with the phosphopeptide fragments library were quite resemble, there were distinctive features that differed from one isoform to another found, which shown a non-normal distribution and resulting in ineffective discrimination of different peptide sequences. Before modeling, all the data are

pre-processed, and quality checked. First of all, the relative fluorescent values were log transformed and the peptide sequences with relative low values were removed from the library. Even the raw data is imbalance distributed, the transformed affinity data are normally distributed and but not consistent for each sub-library (Fig.1 A). Among of that, the affinity of isoform ζ is higher than the rest of six isoforms. After pre-processed, the mean values of affinity range from 5 to 6 for the seven isoforms. The pre-processed data are used for the rest of the analysis. From the analysis of high binding affinity fragments in this library, it could be found that the frequency of occurrence of a specific amino acid was different at each position. Similar finding can also be found in other reports [60].

Figure 1: (A) The distribution of the binding affinity data after processing. The mean (μ) and standard deviation (σ) values for the seven isoforms are listed in the legend of the plot. (B) and (C) The conserved peptide sequences in N- and C-sublibrary with low binding affinity against 14-3-3s.

Totally 482 of log-transformed data that shown low binding affinity (less than zero) are discarded (210 in N-terminal sublibrary and 272 in C-terminal sublibrary). The low binding affinity value indicates the corresponding peptide sequence has no contribution in protein interaction, which could cause negative effects on model development. A small portion of sequences with low affinity was found (less than 7%) for each isoform except the isoforms γ (15.3%) and σ (15.0%). Most of the peptide sequences with low binding affinity exist only in one or two isoform libraries. Only three peptide sequences commonly exist in six isoform libraries, which is "EEPpSXXX" in N-terminal library, "XXXpSPEE" and "XXXpSPLE" in C-terminal library. And the multiple alignments shown that the peptide sequences in low affinity data are conserved. The 384 out of 482 peptide sequence have at least one Glu residue at any positions (Fig.1 B and C).

2.2 QSAR modeling on the relative affinity data

All of the 30 variables derived from DPPS descriptors were used to characterize each phosphopeptide fragments. For each sublibrary, 80% fragments with the relative binding affinity data from the experiments are treated as training data. The details on the identification of the relative binding affinity for all isoform sublibraries are shown (Fig.2). It clearly shows the robustness of the QSAR model developed in this study. For the 14-3-3 isoform ζ , the modelling results has overall R^2 and $RMSE$ values of 0.7897926 and 0.619047 in the N-terminal sublibrary, 0.7444364 and 0.7037071 in the C-terminal sublibrary. The R^2 and $RMSE$ values vary between seven isoforms. This might due to the fact that the relative binding affinity values distribute differently in the peptide library.

The coefficients of the QSAR models were used to explore the contributions of amino acid residues. The overall contribution of every physicochemical parameter at each position based on the normalized coefficients of the QSAR model was calculated and plotted to explore the preferred residue for 14-3-3 interaction (Fig.3). Relative to phosphoserine, the contribution of position -2 and -1 is associated with electronic property of residues, in addition to hydrophobicity. At position +1, the contribution also come from the partial of electronic property and hydrophobicity. Meanwhile, it can be

Figure 2: Scatter plot of the relative binding affinity values (real) versus prediction by the QSAR model (prediction) of 14-3-3 seven isoforms with the phosphopeptide fragments library. The title of the plot represents the isoform type and the sublibrary type of the data used in the analysis. The overall R^2 and $RMSE$ values are shown in the inset of the plots.

found that the contribution is not common between seven isoforms. For example, electronic property at position -2 was only advantageous for the isoform σ , τ , and ζ to form high binding affinity. Arg and Lys were preferred at the position -1 relative to pS . And nonpolar residues are preferred at the position -1 and +1.

Figure 3: Heatmap of regression coefficients with N- and C- terminal sublibrary. For each amino acid residue position P , there are 10 coefficients corresponding to four types of physiochemical properties, which are electronic property (1-4), steric property (5-6), hydrophobicity (7-8) and hydrogen bond (9-10). The color scale of the heatmap is illustrated at right side.

The performance of the QSAR models for each isoform are evaluated by two ways. First, the consensus sequences of amino acids from the high binding affinity phosphopeptide fragments are identified. There is significant preference for Arg in position -3 relative to pS . The same selectivity of amino acid can be found in position -1 in addition to Lys. Phe and Ala is evidently selected in the position -2. For the C-terminal side, the slightly preference of amino acids are also determined. The selection for Ala in the position +1 and Phe in the position +2 is evident. These preferences of amino acids are highly consistent with the experiment data here we used and previous published work [60, 31]. Second, the distance and similarity of affinity among the seven isoforms are evaluated and compared with the experiment data. The distance of predicted binding affinity between the isoforms are changed slightly, which indicated that the accurate predictions are obtained by the models we developed.

2.3 Prediction of the relative binding affinity of peptides with 14-3-3s

The QSAR models developed from N- and C-terminal sublibrary against all seven 14-3-3 isoforms were used to predict the binding affinity of all possible tripeptide sequences ($20^3 = 8000$) to 14-3-3s. The generated data was standardized by isoforms so that the mean was zero and the variance was one (Supplementary data). The standardized data was treated as the predicted results for the following analysis. The ranges of standardized data are -4.27 to 3.09 in N-terminal sublibrary and -4.46 to 3.67 in C-terminal sublibrary. Moreover, the mammalian proteins that have direct interactions with 14-3-3s from previous reports were collected. The relevant interaction sites with phosphopeptide were extracted and compared with the corresponding prediction results in this study (Tab.1). Disintegrin and metalloproteinase domain-containing protein 22, also known as ADAM22, is a member of membrane protein in human [20]. The interaction of ADAM22 with 14-3-3s isoforms have confirmed by yeast two-hybrid screen [66]. There are two sites found to be conserved with consensus motif mode 1 in the peptide sequence. But both of these two sites are lack of Pro at position +2 relative to phosphoserine residue. This is consistent with the reports that an alanine substitution at +2 in synthetic

peptide binding studies shows only a moderate effect on 14-3-3s binding [62]. On the basis of prediction in our results, it can be found that the imbalance binding affinity exist between N- and C-terminal.

Table 1: The predicted binding affinity of peptide sequences in the proteins that have direct interactions against 14-3-3s from literatures.

Sequences		β	ϵ	η	γ	σ	τ	ζ	Protein Name ¹
RRH _p TLPA	N	0.82	1.13	0.68	1.03	1.60	1.41	1.02	Serotonin N-acetyltransferase (Q16613)
	C	0.55	1.34	0.44	1.37	1.27	1.21	1.06	[37, 41, 17]
RSV _p TLPR	N	0.82	1.13	0.68	1.03	1.60	1.41	1.02	Tyrosine-protein kinase ABL1 (P00519) [33, 9]
	C	1.20	2.20	1.23	1.66	1.71	1.66	1.30	
RSN _p SWQG	N	2.41	1.54	2.78	1.50	2.39	2.07	1.86	Disintegrin and metalloproteinase
	C	-0.90	-0.62	-1.01	0.26	-0.05	0.24	0.43	domain-containing protein 22 (Q9P0K1) [20]
RSN _p STET	N	2.41	1.54	2.78	1.50	2.39	2.07	1.86	
	C	-1.11	-1.21	-1.09	-0.82	-1.30	-1.17	-0.86	
RRH _p SWGp	N	0.82	1.13	0.68	1.03	1.60	1.41	1.02	A-kinase anchor protein 13 (Q12802) [27, 8]
	C	0.51	0.52	-0.02	0.73	0.73	0.53	1.16	
YAK _p SLPV	N	2.04	1.22	1.75	1.66	1.51	1.68	1.40	Proline-rich AKT1 substrate 1 (Q96B36)
	C	1.03	1.87	0.71	1.70	1.70	1.53	1.39	[29, 61, 13]
RLN _p TSDf	N	1.92	1.66	1.82	1.28	1.41	1.88	1.77	
	C	1.08	-0.05	0.03	-0.50	-0.71	0.71	-0.45	
RSP _p SMPE	N	0.11	-0.07	0.37	-0.35	0.68	0.08	-0.38	M-phase inducer phosphatase 3(P30307)
	C	0.26	0.26	0.43	0.63	0.27	0.20	0.33	[47, 6, 21]
RSAP _p SEPS	N	1.64	1.35	1.71	1.32	2.04	1.83	1.45	RAF proto-oncogene serine/ threonine-protein
	C	-0.14	-0.13	-0.04	0.30	0.00	-0.31	0.39	kinase (P04049)[14, 26, 12, 35]

¹The literatures that confirmed the 14-3-3s binding of the proteins are listed here.

The QSAR models developed from N- and C-terminal sublibrary against all seven 14-3-3 isoforms were used to predict the binding affinity of all possible tripeptide sequences ($20^3 = 8000$) to 14-3-3s. The generated data was standardized by isoforms so that the mean was zero and the variance was one (Supplementary data). The standardized data was treated as the predicted results for the following analysis. The ranges of standardized data are -4.27 to 3.09 in N-terminal sublibrary and -4.46 to 3.67 in C-terminal sublibrary. Moreover, the mammalian proteins that have direct interactions with 14-3-3s from previous reports were collected. The relevant interaction sites with phosphopeptide were extracted and compared with the corresponding prediction results in this study (Tab.1). Disintegrin and metalloproteinase domain-containing protein 22, also know ADAM22, is a member of membrane protein in human [20]. The interaction of ADAM22 with 14-3-3s isoforms have confirmed by yeast two-hybrid screen [66]. There are two sites found to be conserved with consensus motif mode 1 in the peptide sequence. But both of these two sites are lack of Pro at position +2 relative to phosphoserine residue. This is consistent with the reports that an alanine substitution at +2 in synthetic peptide binding studies shows only a moderate effect on 14-3-3s binding [62]. On the basis of prediction in our results, it can be found that the imbalance binding affinity exist between N- and C-terminal.

The predicted results confirm that high conserved binding affinity among seven 14-3-3 isoforms. The sequences of tripeptide RST (c-Raf-1, A-Raf), RDS (Cdc25a), RPS (Cdc25b), RAA (PKC- ϵ), RAK (PCTAIRE-2), RSH (mT), RHA (TH), RHS (TPH) and RSK (A20) at N-terminal sublibrary matched motif mode 1 all are shown high binding affinity (standardized affinity > 1.5), except RSP (Cdc25c). It was reported that 14-3-3 participate in the cell cycle regulation by interacting with Cdc25c [21]. The phosphopeptide sequence in the flanking regions of Cds25c residue 216 is RSP_pSMPE, which partially match motif mode 1. Although the predicted binding affinity of RSP was low. It can be found that MPE at C-terminal present high binding affinity in our result, which suggested that the contribution of interaction of Cdc25c against 14-3-3s might be different with the other ligands. RIH (Cdc25a), RFQ (Cdc25b), CVR (PKC γ), PTR (IRS-1), SYT (CK-8), LYR (PICALM) at N-terminal sublibrary matched motif mode 2 are also shown high binding affinity in predicted results.

2.4 Cluster analysis of the relative binding affinity

The two-way hierarchical cluster analysis was utilized based on the normalized binding affinity data (Fig.4). The distinct clustering of the isoform-specific group was not found in the clustering result. However, the clustering identified a group of peptide sequences with high binding affinity in the N-terminal sublibrary. The number of tripeptides in the group are 253. The values of binding affinity in this group are significantly higher than the rest of data from all seven isoforms ($p < 0.0005$).

Figure 4: Hierarchical cluster analysis of the standardized binding affinity data predicted by the QSAR models. The color scale is illustrated at the bottom. The group of sequences showed significant binding affinity is highlighted with black contours.

The pattern of sequences in this group are enriched in the binding motif mode 1. The tripeptide sequences that is commonly found from previous reports are clustered in this group, e.g. RDS, RST, RFQ, RAA, RAK, RSK, RPS and so on (Supplementary data). The position specific scoring matrix produced by the peptide sequences indicated that strong selection for peptides containing Arg or Lys in the position -3. This finding was consistent with the patterns found in the experiment dataset[31]. And non-polar amino acid at the position -2, e.g. Ala, Pro, Cys, Trp, Tyr, Phe, Ser. Ser, Phe or Tyr in the *pS* -2 position have been reported in the degenerate peptide library [60]. And in the position -1, Arg, Lys, Asn are the most common residues, which have been illustrated before[28].

3 Discussion

For the development of QSAR models, it is important to screen the amino acid descriptors. To date, there are diverse types of amino acid descriptors are available from published literatures or database, e.g. and so on. For improving the performance of modeling, several types of descriptors commonly used in the QSAR modeling were collected and used in this study to obtain a suitable descriptor. These descriptors were z-scales [24], VHSE[34], HESH [45], G8 [54] and DPPS[56]. It is commonly known that every type of descriptor has their own advantages and limitations. The z-scale is widely applied in and proven to be useful for the prediction of peptide activity [23]. But the limitations in the modeling of tripeptide sequences were also reported [57]. Although VHSE and z-scale were constructed from different kinds of physicochemical parameters, both of them have three types of variables (hydrophobicity, steric property and electronic property). Previous study found the different results generated using VHSE and z-scale [30]. And in the activity prediction of angiotensin-converting enzyme inhibitors, bradykinin-potentiating pentapeptides, bitter tasting dipeptides, the application of VHSE descriptors could acquire satisfactory results [54, 34]. This suggested the advantageous use of VHSE for describing the structural variability in the protein-protein interactions. More physicochemical parameters were included both by DPPS ($n = 171$) and HESH ($n = 119$) than VHSE ($n = 50$). The property of hydrogen bond was added into DPPS and HESH to describe the peptides. These two types of amino acid descriptors have been applied to build some good models with different regression methods [48, 45, 64]. Even DPPS and HESH were newly published descriptors, the quality of the models (R^2 and $RMSE$) developed by HSEH and DPPS were not improved in our preliminary study. However, the performance of prediction with DPPS were better than the rest of descriptors in the preliminary results of this study. Thus, DPPS was selected as the amino acid descriptor here. At the same time, various regression methods (including support vector regression, random forest, partial least squares, multiple linear regression, lasso and ridge) were evaluated by the quality of the models in the preliminary studies. Even the R^2 and $RMSE$ of the

models by these methods were totally different, it can be found that the models by elastic net method had the most powerful fitting capacity. Hence, elastic net was used here for developing the QSAR models.

The ten types of amino acids used in the peptide library only account for 50% of total, especially only five types at $P_{\pm 3}$. In this case, we could not get the full views on the 14-3-3s binding preference due to the limitations existed in the experiments. The QSAR modeling method provide an alternative way to extend the knowledge on 14-3-3s interactions. According to the coefficients from QSAR modeling, the important properties of residues were identified. The hydrophobic property of residues at N-terminal has a significant contribution to the binding affinity of phosphopeptides. The conserved patterns of phosphopeptide fragments in the ligands were verified in our predicted results. In addition, most of the proteins that reported to bind to 14-3-3s through experiment were also detected in our study. The binding affinity of the phosphopeptide sequences in these well know proteins were not determined yet by the combinatorial peptide library [31]. Our predictions provide the evidences that the QSAR modeling method can be applied to such type of research and insightful information could be obtained. Meanwhile, new peptide sequences with high affinity value were also determined, which have great potentials to be found at the next step of research. After being screened and validated by in vitro and in vivo studies, these peptides that was not found before are plausible to be developed as the inhibitors or agents that interfere with 14-3-3 related pathways and as novel therapeutic molecules for targeted cancer therapy with purpose of clinical research.

4 Materials and Methods

4.1 Dataset

The dataset was extracted from a fingerprint peptide library coupled with seven mammalian 14-3-3 isoforms [31]. The different source of isoforms was used in this peptide library. The 14-3-3s binding affinity data in the library was generated by the fragment-based combinatorial peptide microarray and consisted of 1000-member phosphopeptide fragments with seven 14-3-3 isoforms, respectively. The data was split into two sublibraries. There were 500 members of N-terminal $P_{-3}P_{-2}P_{-1}pS-X_{+1}X_{+2}X_{+3}$ and 500 members of C-terminal $X_{-3}X_{-2}X_{-1}pS-P_{+1}P_{+2}P_{+3}$ sublibraries each. Peptide residues in the phosphopeptide fragments were designated as P and X, then numbered at subscript according to their proximity to the phosphoserine and negative if they are located N-terminal of it. Ten representative amino acids (Ala, Glu, Phe, Gly, Lys, Leu, Pro, Gln, Arg, Val) were used at the position of $P_{\pm 1/2}$, and among them five amino acids (Glu, Phe, Leu, Pro, Arg) at position of $P_{\pm 3}$. X represents an isokinetic mixture of rest 14 amino acids. The relative fluorescence values were generated by fluorescence signal of Cy3-labeled 14-3-3s, which represented the affinity between phosphopeptide and 14-3-3s. To improve the data discrimination, raw relative fluorescence values were transformed by taking the logarithm to the base of two before analysis [58]. The processed data represented the relative binding affinity of 14-3-3 isoforms with each phosphopeptide fragment.

4.2 Amino acids descriptors

The tripeptide sequences in the phosphopeptide fragments were extracted and used for the rest of the analysis. The relative binding affinity of each tripeptide can be characterized by describing the residues with descriptors. Nonbonding effects, like electrostatic, van der Waals, hydrophobic interactions and hydrogen bond play central roles in phosphopeptide interaction with all type of 14-3-3 isoforms [43]. DPPS descriptor is a kind of suitable way to be used as the descriptor here [48]. The DPPS descriptors were made up of ten variables for describing the electronic property, steric property, hydrophobicity and hydrogen bond of all amino acids (Tab.2). For tripeptide sequences in both N- and C- terminal sublibraries, physicochemical parameters of amino acid residue at each position were characterized by 10 DPPS descriptors. By assigning DPPS to tripeptide, totally of 30 variables were generated. Due to the different contributions of residue position and its physicochemical parameters to binding affinity, variable selection is required prior to QSAR modeling, expecting to improve modeling qualities and to decrease complexity. The key variables for the binding affinity were screened by stepwise regression. Then based on the screened variable, we can build a QSAR model for each sub-library of seven isoforms.

4.3 QSAR modelling

The initial dataset is divided into two parts, one for training and the other for testing. For example, there are 500 relative affinity data in one of isoform sublibraries. 400 samples of the relative affinity data are randomly selected as training dataset and the rest of 100 samples as testing dataset. The training datasets are used for model construction and testing datasets are used for performance test. After construction of modeling finished, the test datasets are used for the confirmations of the performance. Linear regression is the commonly used modeling methods for QSAR studies. It has advantages in its simplicity, accuracy and ease of interpretation. Hence linear regression is chosen for modeling here.

Table 2: Amino acids descriptors

AA	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Ala	-1.02	-2.88	-0.56	0.36	-6.15	-1.68	0.04	-2.51	-1.94	-0.01
Arg	1.99	4.13	-4.41	-1.02	4.78	3.04	-9.06	6.71	4.41	0.07
Asn	-2.19	1.86	0.38	-0.13	-2.30	1.41	-5.71	-1.11	1.73	-0.19
Asp	-6.60	3.32	1.61	0.36	-3.25	1.95	-7.36	0.14	1.24	-0.15
Cys	0.21	1.12	3.42	-0.68	-2.27	-1.22	3.11	-2.98	-1.70	1.57
Gln	-0.47	1.16	-0.57	0.69	0.39	1.93	-5.46	-0.84	1.93	0.85
Glu	-5.39	0.65	-0.98	1.39	-0.23	2.51	-6.84	-0.68	1.41	1.28
Gly	-2.86	-5.00	-2.97	0.53	-11.45	1.89	-2.11	-3.99	-2.16	-0.76
His	0.73	2.68	-0.66	-1.89	1.60	1.13	-1.94	-0.11	0.44	0.15
Ile	1.91	-3.13	0.01	1.14	2.70	-4.55	8.93	0.18	-1.10	-0.76
Leu	1.64	-2.57	0.00	1.35	2.62	-2.65	7.72	0.05	-1.03	-1.81
Lys	2.47	1.54	-4.28	-0.86	2.77	2.06	-6.18	2.05	2.19	-1.65
Met	1.93	-0.01	1.21	0.99	2.79	-0.56	5.33	-0.87	-0.99	-1.09
Phe	2.68	0.84	2.22	0.71	5.02	-0.30	8.60	1.13	-1.40	-0.28
Pro	0.45	-2.89	1.77	-5.81	-3.79	-0.61	0.70	1.21	-1.67	1.79
Ser	-1.76	-0.19	1.06	-0.69	-5.72	0.14	-4.14	-2.42	-0.13	0.69
Thr	-0.55	-0.66	0.13	-0.31	-2.76	-1.56	-2.46	-2.12	0.17	0.08
Trp	3.88	1.78	1.68	2.00	9.31	0.89	7.53	4.27	-0.23	-1.42
Tyr	2.10	1.26	1.15	0.91	5.90	0.74	3.71	3.32	0.25	1.33
Val	0.83	-3.02	-0.22	0.97	0.05	-4.55	5.61	-1.41	-1.44	0.30

Electronic property: columns V1-V4, steric property: columns V5-V6, hydrophobicity: columns V7-V8, hydrogen bond: columns V9-V10.

The model assumes that there is a linear relationship between the peptide activity and the variables we extracted by the amino acid descriptor. For each peptide sequence, the relative binding affinity value $f(X)$ is represented by a vector of p variables:

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots \beta_p X_p \quad (1)$$

where β_0 is the model constant, $X_1, X_2, X_3 \dots X_p$ are a series of amino acid descriptors with their corresponding coefficients $\beta_1, \beta_2, \beta_3 \dots \beta_p$ based on the regression setting. Once the tripeptide sequence was translated into a two-way data matrix by amino acid descriptors, regression methods such as ordinary least squares can be used to develop the QSAR models relating sequence to bioactivity.

The panelized regression methods, including lasso, ridge and elastic net regression, can create the linear models that is panelized, that is, a constraint λ is added in the equation to minimize the sum of squared residuals[5]. The consequence of importing this penalty, is to reduce the coefficient values towards zero. This allows variables with less contributions to have a coefficient close to or equal zero. The main difference between lasso and ridge regression is the penalty term they use. Ridge uses L_2 -norm which is the sum of squared coefficients, $\lambda \sum_{t=1}^p \beta_t^2$. Lasso uses L_1 -norm which is the sum of the absolute coefficients, $\lambda \sum_{t=1}^p |\beta_t|$. L_1 -norm imposes sparsity among the coefficients, which makes the fitted model more interpretable [15]. Then elastic net was introduced and has a penalty which is a mix of L_1 - and L_2 -norms, $\lambda((1 - \alpha) \sum_{t=1}^p \beta_t^2 / 2 + \alpha \sum_{t=1}^p |\beta_t|)$. Since α is the mixing parameter between ridge($\alpha = 0$) and lasso($\alpha = 1$), elastic net is regarded as a compromise between lasso and ridge. Here, elastic net regression was selected to establish the QSAR models since it solves the limitations of both lasso and ridge. The 10-fold cross-validation methods were used to choose the tuning parameter λ and α , which is able to minimize the cross validation error. The training dataset is divided randomly into ten parts and then each of parts are used as testing for the model on the other parts. Then the best value pair for λ and α was selected for the resulting models.

To get the measures of goodness-of-fit for QSAR models, the Pearson's product-moment correlation coefficient squared (R^2) and the root mean squared error ($RMSE$) are calculated between prediction and real dataset. $RMSE$ can be calculated by

$$RMSE = \sqrt{\sum_{i=1}^n (e_i - p_i)^2} \quad (2)$$

where e_i is i -th binding affinity value from the experiment dataset with n samples, p_i is i -th value from the predicted results by the models. And R^2 can be calculated by

$$R^2 = \frac{\sum_{i=1}^n (e_i - p_i)^2}{\sum_{i=1}^n (e_i - \bar{e})^2} \quad (3)$$

where $\bar{e}$ is the average binding affinity from the experimental dataset. Ideally, lower $RMSE$ and higher R^2 values are indicative of a good model. The most ideal result would be an $RMSE$ value of zero and R^2 value of 1.

4.4 Data analysis

Cluster analysis of the relative binding affinity was performed using euclidean measure to obtain distance matrix and complete agglomeration method for clustering (*heatmap.2* in R package *gplots*). The relative binding affinity values were scaled before the analysis. The sequence conservation analysis was performed by uploading the peptide sequences to weblogo web server [7]. The default parameters were used in this analysis.

5 Conclusions

In this study, the development of the predictive QSAR models and prediction of the binding affinity of phosphopeptide fragments against 14-3-3s were described. According to the models, the important properties of residues in the phosphopeptide were identified. The conserved patterns of peptide sequences were verified in our predicted results. The well-known ligands against 14-3-3s were also found in our study. The predictions provide the evidences that the conserved peptides existed in the 14-3-3s interaction and could be applied in the further studies.

14-3-3s are regarded as the novel therapeutic targets for cancer therapy. A small portion of positive charged residues located in the characterized phosphopeptide binding domain in 14-3-3s. This domain is required for binding and makes 14-3-3s an ideal molecular target. Currently, there are not a small molecule inhibitor of 14-3-3s developed in the clinic. It will be important to elucidate the underlying mechanism of 14-3-3s interaction that could be mined in cancer study by the efforts from computational and biological works.

References

- [1] Alastair Aitken. 14-3-3 proteins: A historic overview. *Seminars in Cancer Biology*, 16(3):162–172, jun 2006.
- [2] Alastair Aitken, D. B. Collinge, B. P.H. van Heusden, T. Isobe, P. H. Roseboom, G. Rosenfeld, and J. Soll. 14-3-3 proteins: a highly conserved, widespread family of eukaryotic proteins, 1992.
- [3] Anne Benzinger, Grzegorz M Popowicz, Joma K Joy, Sudipta Majumdar, Tad A Holak, and Heiko Hermeking. The crystal structure of the non-liganded 14-3-3sigma protein: insights into determinants of isoform specific ligand binding and dimerization. *Cell research*, 15(4):219–227, 2005.
- [4] Dave Bridges and Greg B G Moorhead. 14-3-3 Proteins: a Number of Functions for a Numbered Protein. *Science's STKE : signal transduction knowledge environment*, 2005(296):re10, 2005.
- [5] Peter Bruce and Andrew Bruce. *Practical Statistics for Data Scientists: 50 Essential Concepts*. " O'Reilly Media, Inc.", 2017.
- [6] Dmitry V Bulavin, Yuichiro Higashimoto, Ian J Popoff, William A Gaarde, Venkatesha Basrur, Olga Potapova, Ettore Appella, and Albert J Fornace. Initiation of a G2/M checkpoint after ultraviolet radiation requires p38 kinase. *Nature*, 411(6833):102–107, 2001.
- [7] Gavin E. Crooks, Gary Hon, John Marc Chandonia, and Steven E. Brenner. WebLogo: A sequence logo generator. *Genome Research*, 14(6):1188–1190, 2004.
- [8] Dario Diviani, Liliane Abuin, Susanna Cotecchia, and Laetitia Pansier. Anchoring of both PKA and 14-3-3 inhibits the Rho-GEF activity of the AKAP-Lbc signaling complex. *The EMBO journal*, 23(14):2811–2820, 2004.
- [9] S Dong, S Kang, S Lonial, H J Khoury, J Viallet, and J Chen. Targeting 14-3-3 sensitizes native and mutant BCR-ABL to inhibition with U0126, rapamycin and Bcl-2 inhibitor GX15-070. *Leukemia*, 22(3):572–577, 2008.
- [10] Qi-Shi Du, Ri-Bo Huang, and Kuo-Chen Chou. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Current protein & peptide science*, 9(3):248–60, 2008.

- [11] Robert J. Ferl, Michael S. Manak, and Matthew F. Reyes. The 14-3-3s, jun 2002.
- [12] Andreas Fischer, Angela Baljuls, Joerg Reinders, Elena Nekhoroshkova, Claudia Sibilski, Renate Metz, Stefan Albert, Krishnaraj Rajalingam, Mirko Hekman, and Ulf R Rapp. Regulation of Raf activity by 14-3-3 proteins Raf kinases associate functionally with both homo-and heterodimeric forms of 14-3-3 proteins. *Journal of Biological Chemistry*, 284(5):3183–3194, 2009.
- [13] Bruno D Fonseca, Vivian H-Y Lee, and Christopher G Proud. The binding of PRAS40 to 14-3-3 proteins is not required for activation of mTORC1 signalling by phorbol esters/ERK. *Biochemical Journal*, 411(1):141–149, 2008.
- [14] Ellen Freed, Marc Symons, Susan G Macdonald, Frank McCormick, and Rosamaria Ruggieri. Binding of 14-3-3 proteins to the protein kinase Raf and effects on its activation. *Science*, 265(5179):1713–1716, 1994.
- [15] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Springer series in statistics New York, 2001.
- [16] Haian Fu, Romesh R Subramanian, and Shane C Masters. 14-3-3 P PROTEINS : Structure , Function , and Regulation. *Annual Review of Pharmacology and Toxicology*, 40:617–47, 2000.
- [17] Surajit Ganguly, Joan L Weller, Anthony Ho, Philippe Chemineau, Benoit Malpoux, and David C Klein. Melatonin synthesis: 14-3-3-dependent activation and inhibition of arylalkylamine N-acetyltransferase mediated by phosphoserine-205. *Proceedings of the National Academy of Sciences*, 102(4):1222–1227, 2005.
- [18] Alexandra K. Gardino, Stephen J. Smerdon, and Michael 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, 2006.
- [19] Cathie Garnis, Bradley P Coe, Adrian Ishkanian, Lewei Zhang, Miriam P Rosin, and Wan L Lam. Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. *Genes, Chromosomes and Cancer*, 39(1):93–98, 2004.
- [20] Nathan J Gödde, M D Giovanna, Lucy Paradiso, and Ulrike Novak. Efficient ADAM22 surface expression is mediated by phosphorylation-dependent interaction with 14-3-3 protein family members. *Journal of cell science*, 119(16):3296–3305, 2006.
- [21] Paul R Graves, Christine M Lovly, Geoffrey L Uy, and Helen Piwnica-Worms. Localization of human Cdc25C is regulated both by nuclear export and 14-3-3 protein binding. *Oncogene*, 20(15):1839–1851, 2001.
- [22] Y Gu, K Xu, C Torre, M Samur, B G Barwick, M Rupji, J Arora, P Neri, J Kaufman, A Nooka, and Others. 14-3-3 ζ binds the proteasome, limits proteolytic function and enhances sensitivity to proteasome inhibitors. *Leukemia*, 32(3):744–751, 2018.
- [23] SVEN HELLBERG, LENNART ERIKSSON, JÖRGEN JONSSON, FREDRIK LINDGREN, MICHAEL SJÖSTRÖM, BERT SKAGERBERG, SVANTE WOLD, and PETER ANDREWS. Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships. *International journal of peptide and protein research*, 37(5):414–424, 1991.
- [24] Sven Hellberg, Michael Sjoestroem, Bert Skagerberg, and Svante Wold. Peptide quantitative structure-activity relationships, a multivariate approach. *Journal of medicinal chemistry*, 30(7):1126–1135, 1987.
- [25] Heiko Hermeking. The 14-3-3 cancer connection. *Nature reviews. Cancer*, 3(12):931–943, 2003.
- [26] Kenji Irie, Yukiko Gotoh, Beverly M Yashar, Beverly Errede, Eisuke Nishida, and Kunihiro Matsumoto. Stimulatory effects of yeast and mammalian 14-3-3 proteins on the Raf protein kinase. *Science*, 265(5179):1716–1719, 1994.
- [27] Jing Jin, F Donelson Smith, Chris Stark, Clark D Wells, James P Fawcett, Sarang Kulkarni, Pavel Metalnikov, Paul O'Donnell, Paul Taylor, Lorne Taylor, and Others. Proteomic, functional, and domain-based analysis of in vivo 14-3-3 binding proteins involved in cytoskeletal regulation and cellular organization. *Current biology*, 14(16):1436–1450, 2004.
- [28] Catherine Johnson, Sandra Crowther, Margaret J Stafford, David G Campbell, Rachel Toth, and Carol MacKintosh. Bioinformatic and experimental survey of 14-3-3-binding sites. *The Biochemical journal*, 427(1):69–78, mar 2010.
- [29] Kristina S Kovacina, Grace Y Park, Sun Sik Bae, Andrew W Guzzetta, Erik Schaefer, Morris J Birnbaum, and Richard A Roth. Identification of a proline-rich Akt substrate as a 14-3-3 binding partner. *Journal of Biological Chemistry*, 278(12):10189–10194, 2003.
- [30] Yao-Wang Li, Bo Li, Jiguo He, and Ping Qian. Quantitative structure–activity relationship study of antioxidative peptide by using different sets of amino acids descriptors. *Journal of Molecular Structure*, 998(1-3):53–61, 2011.

- [31] Candy H S Lu, Hongyan Sun, Farhana B. Abu Bakar, Mahesh Uttamchandani, Wei Zhou, Yih Cherng Liou, and Shao Q. Yao. Rapid affinity-based fingerprinting of 14-3-3 isoforms using a combinatorial peptide microarray. *Angewandte Chemie - International Edition*, 47:7438–7441, 2008.
- [32] Jing Lu, Hua Guo, Warapen Treekitkarnmongkol, Ping Li, Jian Zhang, Bin Shi, Chen Ling, Xiaoyan Zhou, Tongzhen Chen, Paul J Chiao, and Others. 14-3-3 ζ cooperates with ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition. *Cancer cell*, 16(3):195–207, 2009.
- [33] Manuela Mancini, Valentina Corradi, Sara Petta, Giovanni Martinelli, Enza Barbieri, and Maria Alessandra Santucci. mTOR inhibitor RAD001 (Everolimus) enhances the effects of imatinib in chronic myeloid leukemia by raising the nuclear expression of c-ABL protein. *Leukemia research*, 34(5):641–648, 2010.
- [34] H U Mei, Zhi H Liao, Yuan Zhou, and Shengshi Z Li. A new set of amino acid descriptors and its application in peptide QSARs. *Peptide Science: Original Research on Biomolecules*, 80(6):775–786, 2005.
- [35] Neil R Michaud, John R Fabian, Karen D Mathes, and Deborah K Morrison. 14-3-3 is not essential for Raf-1 function: identification of Raf-1 proteins that are biologically activated in a 14-3-3-and Ras-independent manner. *Molecular and Cellular Biology*, 15(6):3390–3397, 1995.
- [36] Tomas Obsil, Rodolfo Ghirlando, D. Eric Anderson, Alison Burgess Hickman, and Fred Dyda. Two 14-3-3 Binding Motifs Are Required for Stable Association of Forkhead Transcription Factor FOXO4 with 14-3-3 Proteins and Inhibition of DNA Binding. *Biochemistry*, 42(51):15264–15272, 2003.
- [37] Tomas Obsil, Rodolfo Ghirlando, David C. Klein, Surajit Ganguly, and Fred Dyda. Crystal structure of the 14-3-3 ζ : serotonin N-acetyltransferase complex: A role for scaffolding in enzyme regulation. *Cell*, 105(2):257–267, 2001.
- [38] Tomas Obsil and Veronika Obsilova. Structural basis of 14-3-3 protein functions. *Seminars in Cell & Developmental Biology*, 22(7):663–672, 2011.
- [39] KI Pennington, Ty Chan, Mp Torres, and JI Andersen. The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein–protein interactions. *Oncogene*, 37(42):5587–5604, 2018.
- [40] Jonathan R Pollack, Therese Sørli, Charles M Perou, Christian A Rees, Stefanie S Jeffrey, Per E Lonning, Robert Tibshirani, David Botstein, Anne-Lise Børresen-Dale, and Patrick O Brown. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proceedings of the National Academy of Sciences*, 99(20):12963–12968, 2002.
- [41] Nikita Pozdeyev, Carla Taylor, Rashidul Haque, Shyam S Chaurasia, Amy Visser, Aamera Thazyeen, Yuhong Du, Haian Fu, Joan Weller, David C Klein, and Others. Photic regulation of arylalkylamine N-acetyltransferase binding to 14-3-3 proteins in retinal photoreceptor cells. *Journal of Neuroscience*, 26(36):9153–9161, 2006.
- [42] Sumaiyah K Rehman, Shau-Hsuan Li, Shannon L Wyzomierski, Qingfei Wang, Ping Li, Ozgur Sahin, Yi Xiao, Siyuan Zhang, Yan Xiong, Jun Yang, and Others. 14-3-3 ζ Orchestrates Mammary Tumor Onset and Progression via miR-221-Mediated Cell Proliferation. *Cancer research*, 74(1):363–373, 2014.
- [43] Katrin Rittinger, Joe Budman, Jian Xu, Stefano Volinia, Lewis C Cantley, Stephen J Smerdon, Steven J Gamblin, and Michael B Yaffe. Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Molecular cell*, 4(2):153–166, 1999.
- [44] Gisbert Schneider. Modeling structure-activity relationships. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience, 2013.
- [45] Mao Shu, Hu Mei, Shanbin Yang, Limin Liao, and Zhiliang Li. Structural parameter characterization and bioactivity simulation based on peptide sequence. *QSAR & Combinatorial Science*, 28(1):27–35, 2009.
- [46] Pranav Sinha, Sandra Kohl, Jochen Fischer, Gero Hütter, Monika Kern, Eckard Köttgen, Manfred Dietel, Hermann Lage, Martina Schnölzer, and Dirk Schadendorf. Identification of novel proteins associated with the development of chemoresistance in malignant melanoma using two-dimensional electrophoresis. *ELECTROPHORESIS: An International Journal*, 21(14):3048–3057, 2000.
- [47] Elphine Telles, Amol S Hosing, Samrat T Kundu, Prasanna Venkatraman, and Sorab N Dalal. A novel pocket in 14-3-3 ϵ is required to mediate specific complex formation with cdc25C and to inhibit cell cycle progression upon activation of checkpoint pathways. *Experimental cell research*, 315(8):1448–1457, 2009.
- [48] F Tian, L Yang, F Lv, Q Yang, and P Zhou. In silico quantitative prediction of peptides binding affinity to human MHC molecule: an intuitive quantitative structure–activity relationship approach. *Amino Acids*, 36(3):535, 2009.

- [49] Robert Tibshirani. Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(3):273–282, 2011.
- [50] Guri Tzivion and Joseph Avruch. 14-3-3 Proteins: Active cofactors in cellular regulation by serine/threonine phosphorylation. *Journal of Biological Chemistry*, 277(5):3061–3064, 2002.
- [51] Alexei Vazquez, Lukasz F Grochola, Elisabeth E Bond, Arnold J Levine, Helge Taubert, Thomas H Müller, Peter Würfl, and Gareth L Bond. Chemosensitivity profiles identify polymorphisms in the p53 network genes 14-3-3 τ and CD44 that affect sarcoma incidence and survival. *Cancer research*, 70(1):172–180, 2010.
- [52] Berlinda Verdoodt, Anne Benzinger, Grzegorz M. Popowicz, Tad A. Holak, and Heiko Hermeking. Characterization of 14-3-3sigma dimerization determinants: Requirement of homodimerization for inhibition of cell proliferation. *Cell Cycle*, 5(24):2920–2926, 2006.
- [53] W Wang and D C Shakes. Molecular evolution of the 14-3-3 protein family. *Journal of molecular evolution*, 43(4):384–398, 1996.
- [54] Xiao-Yu Wang, Juan Wang, Yong Hu, Yong Lin, Mao Shu, Li Wang, Xiao-Ming Cheng, and Zhi-Hua Lin. Predicting the activity of peptides based on amino acid information. *Journal of the Chinese Chemical Society*, 58(7):877–883, 2011.
- [55] Erik W. Wilkert, Robert a. Grant, Stephen C. Artim, Michael B. Yaffe, Erik W Wilker, Robert a. Grant, Stephen C. Artim, and Michael B. Yaffe. A structural basis for 14-3-3 σ functional specificity. *Journal of Biological Chemistry*, 280(19):18891–18898, 2005.
- [56] S. Wold, J. Jonsson, M. Sjöström, M. Sandberg, and S. Rännar. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. *Analytica Chimica Acta*, 277(2):239–253, 1993.
- [57] Jianping Wu, Rotimi E Aluko, and Shuryo Nakai. Structural requirements of angiotensin I-converting enzyme inhibitory peptides: Quantitative structure- activity relationship study of di-and tripeptides. *Journal of agricultural and food chemistry*, 54(3):732–738, 2006.
- [58] Jin Xiong. *Essential bioinformatics*. Cambridge University Press, 2006.
- [59] Jia Xu, Sunil Acharya, Ozgur Sahin, Qingling Zhang, Yohei Saito, Jun Yao, Hai Wang, Ping Li, Lin Zhang, Frank J Lowery, and Others. 14-3-3 ζ turns TGF- β 's function from tumor suppressor to metastasis promoter in breast cancer by contextual changes of Smad partners from p53 to Gli2. *Cancer cell*, 27(2):177–192, 2015.
- [60] Michael B. Yaffe, Katrin Rittinger, Stefano Volinia, Paul R. Caron, Alastair Aitken, Henrik Leffers, Steven J. Gamblin, Stephen J. Smerdon, and Lewis C. Cantley. The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell*, 91(7):961–971, dec 1997.
- [61] Fengxue Zhang, Zanna M Beharry, Thurl E Harris, Michael B Lilly, Charles D Smith, Sandeep Mahajan, and Andrew S Kraft. PIM1 protein kinase regulates PRAS40 phosphorylation and mTOR activity in FDCP1 cells. *Cancer biology & therapy*, 8(9):846–853, 2009.
- [62] Weiping Zheng, Dirk Schwarzer, Aaron LeBeau, Joan L Weller, David C Klein, and Philip A Cole. Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *Journal of Biological Chemistry*, 280(11):10462–10467, 2005.
- [63] P. Zhou, C. Wang, Y. Ren, C. Yang, and F. Tian. Computational Peptidology: A New and Promising Approach to Therapeutic Peptide Design. *Current Medicinal Chemistry*, 20(15):1985–1996, 2013.
- [64] Peng Zhou, Xiang Chen, Yuqian Wu, and Zhicai Shang. Gaussian process: an alternative approach for QSAM modeling of peptides. *Amino acids*, 38(1):199–212, 2010.
- [65] Peng Zhou and Jian Huang. *Computational peptidology*. Springer, 2015.
- [66] Peng Cheng Zhu, Yubo Sun, Renner Xu, Yingying Sang, Jing Zhao, Gang Liu, Liang Cai, Changben Li, and Shouyuan Zhao. The interaction between ADAM 22 and 14-3-3 ζ : regulation of cell adhesion and spreading. *Biochemical and biophysical research communications*, 301(4):991–999, 2003.