International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:8, No:6, 2014

Systematic Identification and Quantification of Substrate Specificity Determinants in Human Protein Kinases

Manuel A. Alonso-Tarajano, Roberto Mosca, Patrick Aloy

Abstract—Protein kinases participate in a myriad of cellular processes of major biomedical interest. The *in vivo* substrate specificity of these enzymes is a process determined by several factors, and despite several years of research on the topic, is still far from being totally understood. In the present work, we have quantified the contributions to the kinase substrate specificity of i) the phosphorylation sites and their surrounding residues in the sequence and of ii) the association of kinases to adaptor or scaffold proteins. We have used position-specific scoring matrices (PSSMs), to represent the stretches of sequences phosphorylated by 93 families of kinases. We have found negative correlations between the number of sequences from which a PSSM is generated and the statistical significance and the performance of that PSSM. Using a subset of 22 statistically significant PSSMs, we have identified specificity determinant residues (SDRs) for 86% of the corresponding kinase families. Our results suggest that different SDRs can function as positive or negative elements of substrate recognition by the different families of kinases. Additionally, we have found that human proteins with known function as adaptors or scaffolds (kAS) tend to interact with a significantly large fraction of the substrates of the kinases to which they associate. Based on this characteristic we have identified a set of 279 potential adaptors/scaffolds (pAS) for human kinases, which is enriched in Pfam domains and functional terms tightly related to the proposed function. Moreover, our results show that for 74.6% of the kinase–pAS association found, the pAS colocalize with the substrates of the kinases they are associated to. Finally, we have found evidence suggesting that the association of kinases to adaptors and scaffolds, may contribute significantly to diminish the *in vivo* substrate crossed-specificity of protein kinases. In general, our results indicate the relevance of several SDRs for both the positive and negative selection of phosphorylation sites by kinase families and also suggest that the association of kinases to pAS proteins may be an important factor for the localization of the enzymes with their set of substrates.

Keywords—Kinase, phosphorylation, substrate specificity, adaptors, scaffolds, cellular colocalization.

I. Introduction

PHOSPHORYLATION is the most common post-translational modification of proteins, and is also an important mechanism for the regulation of protein function.[1] Protein phosphorylation is a reversible and fast reaction that have been conserved in evolution as a mechanism for regulating proteins in a non transcription-dependent manner.[2] The addition (or removal) of a phosphate group, can regulate different characteristics and properties of the affected protein such as its conformation, its activation state, its interactions with other proteins or its cellular localization.[3]

Protein kinases are the enzymes that catalyze the phosphorylation reaction. In human there have been described 518 protein kinases, which constitutes one of the largest families of proteins and accounts for nearly 2% of our genes.[4] Kinases are key players in several cellular processes and their deregulation have been tightly related to pathologies such as cancer [2], [5] and diabetes [6], [7]. Most protein kinases share a common fold of the catalytic domain, but despite their similarities at the catalytic region, kinases have achieved a remarkable sequence diversity by combining different classes of protein domains.[4], [8] Indeed, this diversity plays a major role in the substrate specificity and functional aspects observed *in vivo* for these enzymes.[9], [10], [11] In general, the *in vivo* substrate specificity observed in kinases, is known to be determined by several contextual factors such as the sequence vicinity of the phosphorylation site, cellular localization, cell-type specific coexpression and interactions of kinases and their substrates with adaptor or scaffold proteins. [12], [13]

Advances in high-throughput phosphoproteomic methodologies, have provided valuable data of experimentally determined phosphorylation sites for hundreds of kinases from yeast, human and other organisms.[14], [11], [15], [16], [17] Based on the aforementioned data, several authors have studied the kinase specificity by analyzing different sequence motifs that are targeted by the kinases in their substrates.[18], [13], [9], [19] These motifs — often termed phosphorylation motifs — have been generally represented in position-specific scoring matrices (PSSMs), which allow the probabilistic modeling of signals in sequence alignments.[20] PSSMs have been previously used for the prediction of novel phosphorylation sites and for the assignment of experimentally determined phosphorylation sites to kinases.[21], [22] Other more sophisticated methods for the prediction of phosphorylation sites implement complex algorithms such as hidden Markov models, artificial neural networks or expert systems to integrate several sources of

Manuel A. Alonso-Tarajano former Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, Barcelona 08028, Spain. (e-mail: alonso.tarajano@gmail.com).

Roberto Mosca is with the Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, Barcelona 08028, Spain. (e-mail: roberto.mosca@irbbarcelona.org).

Patrick Aloy is with the Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, Barcelona 08028, Spain. (Tel: +34 93 40 39690; e-mail: patrick.aloy@irbbarcelona.org).

information (*e.g.*, structural disorder, sequence conservation, positional correlations of residues).[23] However, in those cases it is more difficult to infer the decisions that support the predictions, as opposed to the cases of PSSMs, where it is much easier to pinpoint the determinant residues of a functional phosphorylation site.[23]

Adaptors and scaffolds are multidomain proteins involved in the dynamic spatio-temporal organization of large signaling complexes and cellular structures. [24] Due to their roles in cellular signaling, some of these proteins have be implicated in cancer and tumorigenesis. [25], [26] The specificity of many signal transduction events is modulated by adaptors and/or scaffolds, which can recruit signaling enzymes to proper cellular locations [27], [28]. Indeed, the associations of kinases with adaptors and scaffolds can enhance efficient catalytic activation and accurate substrate selection. This is the case of the PKA kinase, which is targeted to discrete cellular environments by the A-kinase anchoring protein (AKAP) [29]. Other two examples are the kinase suppressor of Ras (KSR) and IQGAP, which function as platforms and regulators of the mitogen-activated protein kinase (MAPK) pathway. [30], [25] Adaptors and scaffolds are extremely diverse proteins which lack common sequence signature motifs. Therefore, their identification based only on sequence is currently not possible. Nevertheless, these proteins often contain protein-protein interaction domains (*e.g.*, SH2, SH3 and PD) and it has been suggested that some scaffolds interact with at least two signaling proteins [24]. Based on these characteristics, Ramirez and Albrecht devised a computational method from which they identified 250 potential human signaling scaffolds [31]. However, in their analysis the authors excluded proteins with intrinsic catalytic activity as potential scaffolds, a criteria that may constitute a limitation of their method [32], [33].

In this work we have studied these two elements of the substrate specificity of human protein kinases. First, we focused on the identification of SDRs in the sequences phosphorylated by several kinase families families, and we quantified their contribution to the specificity of those families. Second, we studied how the association of kinases to adaptor and scaffold proteins may influence the cellular colocalization of kinases and their substrates, and also how these associations may diminish the substrate crossed-specificity of kinases.

II. Materials and Methods

A. Integration of human phosphorylation data

We compiled a local database of experimentally determined phosphorylation sites by integrating data from the public resources HPRD [34], PhosphoSitePlus [35] and Phospho-ELM [36]. We kept only those phosphorylation sites for which the responsible kinase was known and we filtered out those without a supporting publication. Our integrated set increases by 18%, 58% and 59% the numbers of kinases, substrates and phosphorylation sites (respectively), with respect to the average contained in the three source databases (see Appendix A).

B. Construction of the position-specific scoring matrices

For generating the PSSM for each kinase family, we used sequence alignments of peptides with a length of nine residues, which contain the phsophorylation site in the central position. For computing the score of each residue we used (1), which is based in the log-odds of the residues at each position of the alignment and considers the frequencies of the residues in the human proteome. [37] The cut-off *p*-value for matches to the PSSMs was set to $1e^{-04}$.

$$S_{rp} = \log(\frac{q_{rp}}{f_r}),\ p = 1\ to\ w \tag{1}$$

S: score of residue r at position p; q: frequency of residue r at position p; f: frequency of residue r in the reference proteome and w: length of the sequence alignment.

C. Evaluation of the position-specific scoring matrices

We have used the information content (IC) the percent of recall (recall) and the area under the receiver operating characteristic curve (AUC-ROC) to evaluate the statistical significance and the performance of the PSSMs. By percent recall we mean the fraction of seed phosphorylation sequences that match the cognate PSSM with a statistically significant score. The statistical significance tests were based on empirical *p*-values for both the IC and the recall. For this, we used sets of 100'000 PSSMs that have been generated from random sequences, and that also respect the cardinality of seed sequences of the PSSM being assessed. For computing the IC we have used the KullbackLeibler distance [38], where the IC is the sum of the expected self-information of each element, see (2).

$$IC = -\sum_{r,p} q_{rp} \times \log(\frac{q_{rp}}{f_r}) \tag{2}$$

IC: information content; q: frequency of residue r in position p of the sequence alignment; f: frequency of the residue r in the reference proteome.

D. Identification of specificity-determinant residues

The number of phosphorylation events available for each kinase family in our set is not uniform. In order to count with enough data to conduct the identification of SDRs, we have selected a subset 22 kinase families with at least 100 phosphorylation events. For each family, we have attempted the identification of residues that could contribute significantly to the specificity of the corresponding kinases (*i.e.*, the SDRs). Based on the corresponding PSSMs, we have classified as SDRs those residues with a score equal or higher than half the score of the phospho-acceptor residue. Finally, we computed the frequency of each SDR across the phosphorylation events of each family in the experiment.

E. Identification and analysis of known adaptors and scaffolds

We collected from UniProtKB/Swiss-Prot all human proteins annotated with either adaptor or scaffold terms in their Function field. By using the high confidence human interactome from Interactome3D [39], we filtered out the

adaptors and scaffolds without evidence of binary interaction with at least one human protein kinase. Finally, we obtained a set of 191 known adaptor or scaffold (kAS) proteins, which associate to 287 human protein kinases.

Some adaptors and scaffolds are known to associate to both the kinases and their substrates. We have tested whether the kAS proteins interact with a statistically significant number of the substrates of the kinases to which they are associated. We have used as the test statistic the number of interactions of proteins in a subnetwork of the interactome. For constructing the backgrounds for the statistical test, we first selected the kinases with at least five substrates (156 kinases in total) and using those substrates as seeds we generated a first level subnetwork of the human interactome. We generated different backgrounds depending on the cardinality of substrates (*S*) of each kinase. For generating the backgrounds we started by randomly selecting a node (*K*) having at least *S* partners. Later, for a number *S* of randomly selected *K*'s partners, we identified the first neighbors (*P*). Finally, we counted the number of interactions between each *P* and all *K*'s partners. While randomly rewiring the subnetwork, we repeated the process 10'000 times for each background set. For testing the initial hypothesis, we conducted a right tale Fisher's exact test.

F. Identification of kinases sharing a significant number of substrates

We have investigated if there exist a relationship between the association to common adaptors or scaffolds and the substrate cross-specificity of kinases. We have approached the identification of kinases sharing at least one kAS protein and also sharing a significant number of *in vivo* substrates. The size of the overlap between the sets of substrates was used as a test statistic. For estimating the statistical significance of the overlaps, we computed empirical *p*-values for sets of kinases with cardinality two and three. We performed the analysis only for the 111 kinases for which we known at least five *in vivo* substrates.

III. Results and Discussion

A. Position-specific scoring matrices

We have analyzed the performance and statistical significance of the PSSMs corresponding to the 93 kinase families for which we count with at least one phosphorylation site. Regarding their performance, we have found negative correlations between the number of seed phosphorylation sites and i) the recall (R = –0.48, *p*-value = $1.2e^{-06}$), ii) the IC (R = –0.33, *p*-value = 0.0013) and iii) the AUC-ROC (R = –0.47, *p*-value = $1.6e^{-06}$). These results suggest that in our data, the increase of the sequence diversity generated by the increase of the number of seed phosphorylation sites, can exert a negative effect in both the performance and the level of self-information of a PSSM (see Fig. 1). We suggest that the substrate specificity of some kinases and kinase families might be represented best by multiple PSSMs, a concept that have been previously applied in the analysis of DNA recognition by transcription factors [40]. Although not covered in the work here presented, we consider that in such cases, multiple PSSMs could be useful for modeling fairly different phosphorylation motifs that are targeted by the same kinase or kinase family.

TABLE I
Comparison of significant and not significant sets of PSSMs

	Total PSSMs	IC	% recall	AUC-ROC	Psites
Significant	69	7.13	46.60	0.77	52.00
Not significant	24	9.49	100.00	1.00	4.50
p-value		0.177	$1.89e^{-04}$	$1.36e^{-04}$	$8.57e^{-09}$

The table shows the median values of the parameters used for comparing the two sets of PSSMs. The last row shows the results of the Mann-Whitney *U* test, which is based on the differences of the medians (significance level $\alpha < 0.05$). Psites stands for seed phosphorylation sites.

The IC can be used as a statistic to estimate how different is that PSSM from a uniform distribution. From our analysis, 69/93 (74.2%) of the PSSMs were found to be statistically significant; and the two sets of PSSMs — significant and not significant — differ in their median values of the percent recall, the AUC-ROC and the number of seed phosphorylation sites. Our results show that PSSMs with a statistically significant IC were generated from sets of seed phosphorylation sites larger than the ones from not statistically significant PSSMs. In agreement to what was previously mentioned, significant PSSMs show significantly lower values of recall and AUC-ROC (see Table I and Fig. 2). Surprisingly, we have not found significant differences between the two sets of PSSMs based on their median IC values. However, in an equivalent comparison using PSSMs from independent kinases, we have found significant differences if the median IC values between sets of significant and non significant PSSMs (Mann-Whitney *U* test *p*-value = $6.2e^{-04}$).

Based on the results of the current analysis, we selected a subset of significant PSSMs to conduct the identification of specificity-determinant residues (SDRs) for the corresponding families of kinases.

B. Specificity-determinant residues

From the previously identified group of significant PSSMs, we selected 22 for which we count with at least 100 phosphorylation events. For 19/22 (86.4%) of the families analyzed we identified at least one SDR. For these 19 families we have successfully classified as SDRs residues that have been reported to play important roles in the specificity of the corresponding kinases (*e.g.*, $\text{MAPK}_{\text{P+1}}$, $\text{PIKK}_{\text{Q+1}}$, $\text{AKT}_{\text{R-3}}$ and $\text{CK2}_{\text{E+3}}$) ⋆. The quantification of the relevance of the SDRs — based on their frequency among the phosphorylation events of each family — shows a wide variation across the different families. For example, the four SDRs previously mentioned have relatively high frequencies that range between 88.86% and 45.83%; however, other SDRs show much lower frequencies (*e.g.*, $\text{PKC}_{\text{K+2}}$ = 19.79%, $\text{CAMKL}_{\text{N+3}}$ = 18.03% and $\text{CK2}_{\text{D+2}}$ = 15.54%, see Table II).

Based on our data, we hypothesize that the combination of multiple SDRs of low frequencies contribute in an additive way to the recognition of the phosphorylation sites by the

⋆SDRs are shown as the acronym of the kinase family, followed by the residue (one letter code) and its position relative to the phosphorylation site.

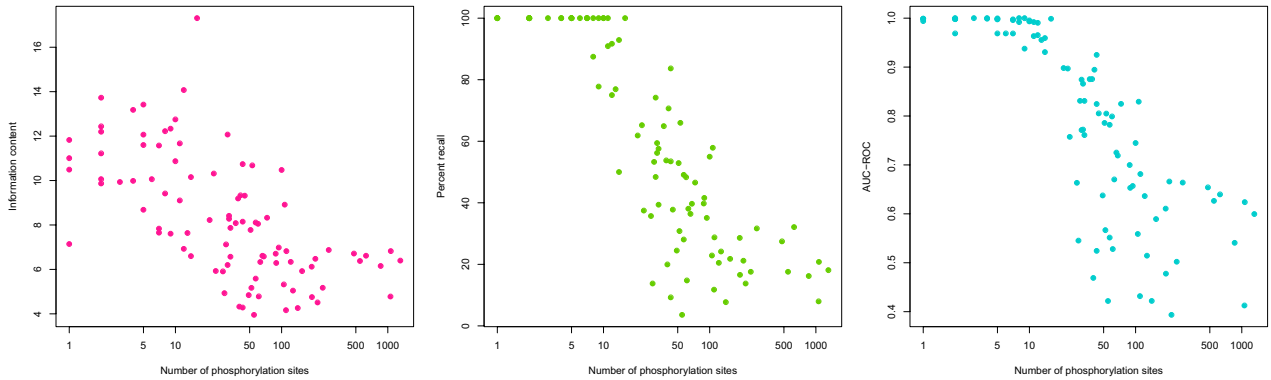

Fig. 1. Correlations of measurements with the number of seed phosphorylation sites. The IC, percent recall and AUC-ROC display negative correlations with the number of seed phosphorylation sites. X-axes are displayed in logarithmic scale.

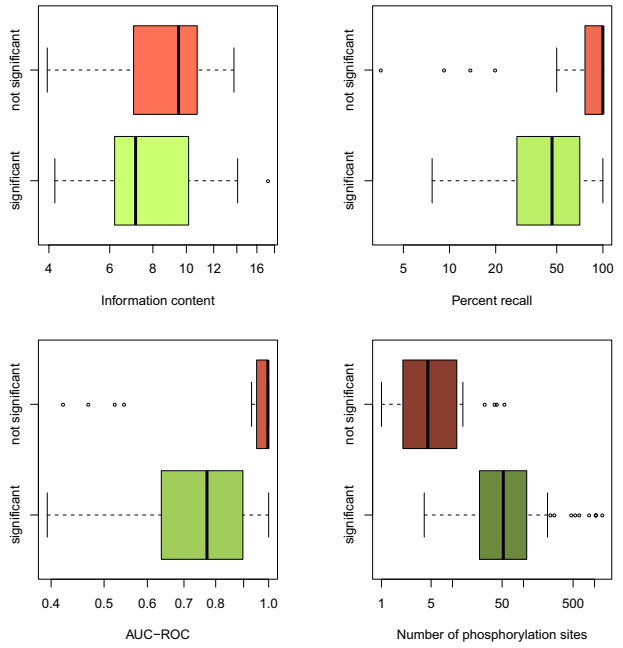

Fig. 2. The PSSMs were classified based on the significance of their IC. The two groups of PSSMs were later compared based on their IC, percent recall, AUC-ROC and number of seed phosphorylation sites. The thick lines in the boxes represent the medians.

TABLE II
SPECIFICITY-DETERMINANT RESIDUES OF KINASE FAMILIES

Kinase family	SDR	% freq.	% cross-freq.
CDK	P+1	81.72	5.95
GSK	S-4	38.49	14.01
GSK	P+1	53.96	5.95
GSK	S+4	48.56	11.51
MAPK	P-2	31.37	5.83
MAPK	P+1	88.86	5.95
PIKK	Q+1	80.83	3.98
AKT	R-3	84.13	5.14
AKT	W+1	3.85	0.66
CAMKL	R-3	31.15	5.14
CAMKL	K-3	21.31	5.65
CAMKL	N+3	18.03	3.83
PKC	R-3	23.19	5.14
PKC	R-2	24.32	5.09
PKC	R+2	27.71	4.57
PKC	K+2	19.79	3.88
CK1	S-3	28.5	8.08
CK1	S+3	31.09	9.85
CK2	D+2	15.54	6.17
CK2	E+3	45.83	6.15

In the table, **% freq.**: frequency of the SDR among the phosphorylation events of current kinase family. **% cross-freq.**: frequency of the SDR among the complementary phosphorylation events, that is, the ones from kinase families without the current SDR.

kinases. In contrast, we consider that SDRs of high frequencies have a larger contribution to the kinase specificity. Moreover, we have noted that the frequency of any given SDR is low — 6.0% on average — among the phosphorylation events of the kinase families that do not count with that SDR. To our opinion, this suggests that SDRs may also function as elements of negative selection to avoid the phosphorylation of non-cognate sequences.

We have identified SDRs that, to the best of our knowledge, have not been previously reported as determinants of the specificity for the corresponding kinase families. These are the cases of $CAMKL_{N+3}$ and AKT_{W+1}, with frequencies of 18.03% and 3.85% respectively. The SDR N+3, is present in the sequences targeted by the microtubule affinity-regulating kinases (MARK) — CAMKL family members — within the repeat regions of the human TAU protein, which is implicated in Alzheimer's disease [41]. Besides, N+3 have a low frequency (3.83%) among the phosphorylation events of the other 21 kinase families in the analysis. Given that the repeat regions of TAU are responsible for the binding to the microtubules [42]; we consider that the presence of N+3 in these regions is an important element for the recognition by MARK kinases, and therefore for the regulation of the association of TAU to the microtubules. The identification of W+1 as an SDR for the AKT family is an interesting result, given that tryptophan is rarely found in the close sequence vicinity of phosphorylation sites — 0.66% among the phosphorylation events of non AKT kinase families —. W+1 was identified as an SDR even when occurring at low frequency (3.85%) among the phosphorylation events of AKT

kinases, which prompted us to research further about the biological relevance of the finding. Interestingly we found reports in the literature showing that, by phosphorylating sequences containing a conserved W+1, some AKT kinases are implicated in the regulation of transcription factors of the FOXO family [43]. To our opinion, this result supports the utility of our approach for the identification of SDRs, even for residues that occur at low frequency among the phosphorylation sites of the kinase of interest.

C. Association of kinases to known adaptors and scaffolds

As previously described, we have compiled a set (kAS) of 191 human proteins that are known to function as adaptors or scaffolds (available on request). These 191 kAS proteins associate to 287 (55.4%) — via 1281 binary PPIs —- protein kinases, which represent a total of 94 (72.3%) kinase families and also comprise the nine major groups in which human kinases are classified. To our opinion, these findings suggest that the association to adaptors or scaffolds is a widespread mechanism among human protein kinases. The results from the analysis of enrichments in Pfam domain families [44] and molecular function terms (MF) of the Gene Ontology [45] show that 14/23 (60.8%) of the enriched Pfam domains are known to be directly involved in promoting PPIs (*e.g.*, PDZ, SH2 and SH3); and that 100% of the enriched MF terms are related to protein binding, adaptor or scaffolding functions (see Fig. 5 and Fig. 6 in Appendix B). Together, these results support the biological role as adaptors or scaffolds of the proteins in the kAS set.

Adaptors and scaffolds can function as linking elements between the kinases and their substrates by recruit the enzymes to cellular compartments where they gain spatial proximity to its relevant set of substrates. We have searched for evidence supporting that kAS proteins could interact with a large number of the substrates of the kinases to which they associate. The result of our analysis suggests that, compared to any random kinase partner, kAS proteins are five times more likely to interact with a significantly large number of the substrates of their corresponding kinases (p-value = $1.08e^{-15}$). This result supports our initial assumption and therefore we decided to use this property of kAS proteins to identify potential adaptors and scaffolds of proteins kinases in the human interactome.

D. Potential adaptors and scaffolds of protein kinases

We have identified a total of 706 associations kinase–potential adaptor/scaffold (K–pAS, available on request). These include 279 pAS proteins — 25.4% of them is present in the kAS set — that are known to interact with 78 (50%) of the 156 kinases initially considered for the experiment. The 78 kinases cover 44 (33.8%) of all human kinase families. Analysis of Pfam domains composition show enrichment 10 Pfam families, all of them known to mediate PPIs or to be present in proteins involved in cellular signaling (see Fig. 7 in Appendix B). Half of these ten Pfam families enriched in the set of pAS proteins were also enriched in the kAS set, a finding that supports the hypothesis of common biological functions. Additionally, we have found

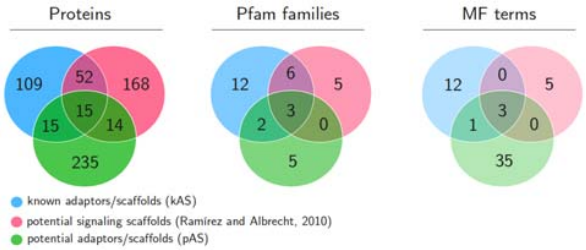

Fig. 3. Comparison of the sets of adaptors and scaffolds.

39 MF terms to be overrepresented in the pAS set (see Fig. 8 in Appendix B). A considerable fraction of these terms (24/39, 61.5%) refer to 'protein binding' functions of signaling-related molecules such as receptors, kinases, phosphatases and transcription factors. This suggests pAS proteins could be able to mediate PPIs for different classes of signaling-related proteins. In contrast with the kAS set, for the pAS proteins we do not find enrichments in MF terms directly related to adaptor nor scaffolding functions; a result that supports the pAS proteins as a novel set of potential adaptors and scaffolds.

We have compared the three sets of adaptors and scaffolds (*i.e.*, kAS, pAS and the set identified by Ramírez and Albrecht) in terms of their protein composition and enriched MF terms and Pfam domains (see Fig. 3). We have found a relatively low average overlap of proteins between the three sets (18.4%), which highlights the lack of a consensus criteria for the computational identification of adaptors and scaffolds. In contrast to our methods, Ramírez and Albrecht considered that scaffolds lack intrinsic enzymatic activity [46]. We consider this criteria to be inaccurate, given the cases of the focal adhesion kinase (FAK) [33] and the kinase suppressor of Ras (KSR) [32], which are both scaffolds with reported catalytic activity. Differences in the Pfam families and the MF terms enriched can be partially attributed to differences in sets of proteins defined as the backgrounds. Nevertheless, for all the three sets the Pfam domains and MF terms enriched support the hypothesis of adaptor or scaffolding roles. Finally, differences in the definition of human interactome can also influence the results of the identification strategies.

Taken together, we consider that our strategy have been able to suggest a set of potential adaptor and scaffold of human human protein kinases, whose functional annotations are in agreement with the proposed biological roles.

E. Cellular colocalization of kinases, adaptors, scaffolds and substrates

Adaptors and scaffolds can play a fundamental role in the *in vivo* specificity of protein kinases by promoting the cellular colocalization of these enzymes with their cognate substrates. Here we have searched for evidence of colocalization of the pAS proteins with the substrates of the associated kinases. For this, we have used the 706 K–pAS relations previously identified, and we have evaluated whether a given pAS is annotated to a cellular component term (CC) — from the Gene

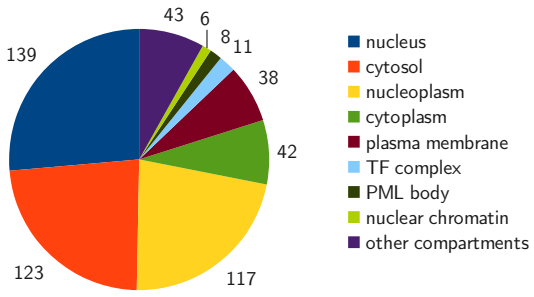

Fig. 4. Cellular component terms shared by substrates and pAS proteins. The slices represent the number of K–pAS pairs where the pAS protein is annotated to the the given CC term. TF and PML stand for transcription factor and nuclear bodies respectively.

Ontology database — that have been previously found to be enriched in the set of substrates of its associated kinase.

For 527/706 (74.6%) of the K–pAS pairs, we found evidence of colocalization between the pAS and the substrates (available on request). This set of 527 K–pAS pairs accounts for 41 kinases, 156 pAS proteins — corresponding to 52.6% and 55.9% (respectively) of the ones in the initial 706 kinase-pAS pairs— and 35 unique CC terms. In Fig. 4 we show a pie chart representation of the CC terms shared by the pAS proteins and the sets of substrates; while in the Table III we show cases of pAS proteins that are found to colocalize with substrates of their corresponding kinases. For example, the pair formed by the β-adrenergic receptor kinase 1 (ARBK1) and the Na(+)/H(+) exchange regulatory cofactor NHE-RF (NHRF1), where the later it has been reported to be involved in the scaffolding of β-adrenergic receptors — substrates of ARBK1 — at the plasma membrane [47]. Another example is the case of the checkpoint kinase-1 (CHK1) and the 14-3-3 protein zeta (1433Z), where the later it has been reported to be required for the nuclear retention of CHK1 [48]. A third case is casein kinase α-1 (KC1A), for which we identified the catenin β-1 (CTNB1) as a pAS. KC1A phosphorylates CTNB1 at serine 45, both proteins are components of the canonical Wnt signaling pathway and they are also part of the large APC–Axin-1–β-catenin complex [49]. Interestingly, CTNB1 contains 12 repeats of the Armadillo (ARM) domain, which is implicated in mediating PPIs. It has been recently suggested that proteins containing ARM repeats, constitute an attractive modular system as scaffolds for peptide-mediated PPIs [50]. Therefore, we consider that CTNB1 may constitute a plausible scaffold that may promote spatial proximity between KC1A and its substrates.

To our opinion, these results suggest that the association to pAS proteins might play an important role in the colocalization of the analyzed kinases with their cognate sets of substrates. Nevertheless, we are aware that in many cases, the CC shared by the substrates and the pAS proteins are too broad (*e.g.*, cytosol, nucleoplasm, cytoplasm) and can not fully justify, based on spatial constrains, the substrate specificity of the kinases.

F. Association to adaptors and scaffolds diminish substrate cross-specificity of kinases

We have analyzed the role that potential adaptors and scaffolds may play in kinase specificity by promoting spatial proximity between the enzymes and their substrates. However, different kinases may associate to the same adaptors and scaffolds and this could lead to substrate cross-specificity. Here we have tested whether the association to common kAS proteins would promote significant substrate cross-specificity between kinases. For this, we have used the subset of K–kAS associations where the kinases have at least five substrates and for which the kAS in the analysis are known to interact with at least two kinases. In total we analyzed 23 cases of two or more kinases that associate to a common adaptor or scaffold, and for non of the cases the kinases shared a number of *in vivo* substrates larger than what would be expected due to chance (see Table IV). Nevertheless, we found the case of the kinases MK01 and MK03 — ERK2 and ERK1 MAP kinases, respectively — which share 73 *in vivo* substrates. Even when it was not statistically significant, the number substrates in common was very large when compared to other sets of kinases in out analysis, and therefore we decided to explore this particular case in more detail. In fact, ERK1 and ERK2 are very closely related kinases, with 82% and 89% of identity in their full and catalytic domain sequences. ERK1 and ERK2 share many if not all functions [51] and despite numerous efforts to establish differences, the detection of such distinctive functions it has been difficult to pinpoint [52]. Therefore, we consider that their large sequence identity, together with their almost identical functions can explain the large substrate overlap reflected in our data. To our opinion, these results support the hypothesis that adaptors and scaffolds are able to diminish *in vivo* substrate cross-specificity by recruiting the kinases to specific macromolecular complexes or cellular locations.

IV. Conclusion

Protein kinases constitute one of the largest and more diverse superfamilies of proteins in human and they are implicated in several cellular processes and pathologies [6], [1]. Despite most kinases share a highly conserved catalytic domain, the observed *in vivo* substrate specificity of these enzymes show little correlation with their primary sequences. In this sense, is known that the *in vivo* specificity of protein kinases is regulated by several factors. Here we have approached the identification and the quantification of the contribution of different elements to the substrate specificity of human protein kinases. For this we have analyzed the residues in the close neighborhood of the the phosphorylation site, the association of kinases to adaptors and scaffolds and the cellular colocalization of kinases and their substrates.

We have generated PSSMs from the sequences targeted by 93 families of kinases and we have analyzed their statistical significance and performance. We have found negative correlations between the number of seed phosphorylation sites

TABLE III
CELLULAR COMPONENT TERMS SHARED BY POTENTIAL ADAPTOR/SCAFFOLD PROTEINS AND SUBSTRATES

Kinase	CC description	Enrichment ratio	Adj. *p*-value	pAS co-annot.
ARBK1	apical plasma membrane	15.7	$4.08e^{-02}$	NHRF1
CDK4	chromatin	16.12	$3.76e^{-02}$	EP300
CDK4	transcription factor complex	20.92	$1.67e^{-02}$	E2F4,EP300
CDK9	PML body	61.38	$4.19e^{-03}$	PIAS4
CHK1	nucleoplasm	10.16	$1.30e^{-04}$	1433Z,CHK2,EP300,MDM2,UBC
CHK2	PML body	29.86	$3.25e^{-03}$	RB,SIRT1,SUMO1
CSK	membrane raft	33.87	$2.74e^{-03}$	ERBB2
EGFR	endosome	8.47	$6.48e^{-04}$	FYN,GRB2,NTRK1
FYN	cell junction	3.98	$3.15e^{-02}$	PTN12
INSR	cytosol	12.11	$5.41e^{-04}$	ABI1,GRB2,IRS1,P85A,SRC,UBC
KC1A	lateral plasma membrane	38.75	$3.46e^{-02}$	CTNB1
KC1A	APC-Axin-1-beta-catenin complex	275.94	$4.18e^{-02}$	CTNB1
KCC2G	vesicle membrane	19.29	$7.14e^{-04}$	GRB2,NCK1
PDPK1	mitochondrion	6.81	$2.48e^{-02}$	1433Z,CASP3,MAD1,PDK1
PLK1	nucleus	4.33	$7.49e^{-04}$	ABL1,ANDR,GRB2,P53,VHL

CC description, description of the CC term enriched in the set of substrates of the kinase; **Enrichment ratio**, ratio of enrichment of the CC term; **Adjusted *p*-value**, multiple test correction by Bonferroni's method; **pAS co-annot.**, pAS proteins associated to the current kinase, that are annotated to the corresponding CC term. Kinases and pAS proteins are represented by their UniProt IDs. See full table of results in Supplementary Materials.

TABLE IV
STATISTICAL SIGNIFICANCE OF THE NUMBER OF SHARED SUBSTRATES FOR KINASE-KNOWN ADAPTOR/SCAFFOLD PAIRS

Adap./Scaff.	Assoc. kinases	Shared subst.	*p*-value
APBB1	EGFR, ERBB2	2	1.00
BIRC5	AURKA, AURKB	4	1.00
CD2AP	ABL1, FYN	1	1.00
DAG1	FYN, SRC	13	0.52
DOK4	EGFR, ERBB2	2	1.00
DOK6	EGFR, ERBB2	2	1.00
ELP1	GSK3B, MK08	7	0.78
FRS3	MK01, FGFR1	1	1.00
FYB	ABL1, FYN	1	1.00
IMA2	SGK1, CHK2	1	1.00
JIP2	EGFR, ERBB2	2	1.00
KHDR1	LCK, SRC	14	0.43
NCK1	ABL1, EGFR	3	1.00
PAR6A	KPCI, KPCZ	3	1.00
PAR6B	KPCI, KPCZ	3	1.00
PKHO1	AKT1, CSK21	5	1.00
SCRIB	MK01, MK03	73	0.10
SH2B1	EGFR, INSR	4	1.00
SHC1	EGFR, INSR	4	1.00
SHC2	EGFR, ERBB2	2	1.00
SHC3	EGFR, ERBB2	2	1.00
SQSTM	KPCI, KPCZ	3	1.00
TGFI1	FAK1, FAK2	1	1.00

Proteins are represented by their UniProt Ids. **Shared substrates**, number of *in vivo* substrates shared by the kinases; ***p*-value**, statistical significance of the number of substrates shared by the kinases.

and *a*) the percent recall, *b*) the information content and *c*) the AUC-ROC of the PSSMs. Based on the IC we have estimated the statistical significance of the PSSMs. We have observed that statistical and non-statistically significant PSSMs show significant differences in the number of seed phosphorylation sites and on their performance parameters (*i.e.*, the percent recall and the AUC-ROC). Our results show the negative effect that the sequence degeneracy caused by the increase of the seed phosphorylation sites can impose on the performance and on the level of self-information of the PSSMs.

Starting from 22 statistically significant PSSMs, we have identified several SDRs that function as positive (or negative) elements for the substrate recognition by different kinases families. The SDRs identified among the different kinase families show high diversity in terms of the type of residue, the position relative to the phosphorylation site and the frequency among phosphorylated sequences available. Some kinase families are very specific towards particular SDRs, which occur in more than 80% of the sequences they target (*e.g.*, $AKT_{R\text{-}3}$, CDK_{P+1}, $MAPK_{P+1}$ and $PIKK_{Q+1}$). We have observed that multiple SDRs are generally identified in families for which the frequencies of the SDRs range approximately between 15% and 55% of the target sequences (*e.g.*, $CK1_{S\text{-}3}$, $CK2_{D\text{-}1}$, $GSK_{S\text{-}4}$, GSK_{P+1} and $PLK_{E\text{-}2}$). Our opinion is that in such cases cases, multiple SDRs may contribute cooperatively to the recognition of the phosphorylation site. We have also noted that the SDRs occur at low frequency (6.01% on average) among the complementary target sequences (*i.e.*, the phosphorylation sites corresponding to those kinase families that do not count with the given SDR). To our opinion, this suggests that an SDR contribute as a negative selection factors for non-cognate phosphorylation sites.

We have compiled a set of 191 proteins with known roles as adaptors or scaffolds and that associate to 55% of the human kinases, which account for 72.3% of all human kinase families. When compared to random proteins in the human interactome, this set of proteins was five times more likely to interact with a large fraction of the substrates of the human kinases to which they associate. To our opinion, these results suggest that the association to adaptors or scaffolds is a common mechanism among human kinases and also supports the concept of adaptors and scaffolds as mediators in the encounter of kinases with their cognate substrates.

We have devised a strategy for the identification of potential adaptors and scaffolds of human protein kinases. For 50% of the initial kinases in the analysis we identified a total of 279 potential adaptors/scaffolds. This set of proteins is enriched in functional terms and in domain families that suggest a

tight link to protein-protein binding functions involved in cellular signalling events. We have also found that for 74.6% of the kinase–potential adaptor/scaffold associations identified, the adaptor/scaffold is annotated under cellular compartment terms found to be enriched among the set of substrates of the associated kinase. We consider that these results put forward a role for the potential adaptors/scaffolds in promoting the colocalization of the kinases and their sets of substrates.

Finally, we analyzed whether the association of different kinases to common adaptors/scaffolds, may relate with the *in vivo* substrate cross-specificity of that kinases. We have not found any case of two or more kinases that, having an adaptor or scaffold in common, also share a number of *in vivo* substrates larger than what would be expected by chance. To our opinion, these results suggest that the association of kinases to adaptors and/or scaffolds may play important roles in the localization of the enzymes with their set of cognate substrates and also in diminishing substrate cross-specificity *in vivo*.

Appendix A

TABLE V
PHOSPHORYLATION DATA OF HUMAN PROTEIN KINASES

Database	Kinases	Substrates	P.Sites	P.Events
HPRD	291	938	3382	5896
Phospho-ELM	218	924	3125	2378
PhosphoSitePlus	318	1664	4711	4711
SBNB_PhosphoDB	325	1856	5946	8880

The first three rows contain the data from the source databases, while the last row correspond to our integrated data. Kinases: number of kinases, Substrates: number of substrates, P.Sites: total (non-redundant) number of distinct residues phosphorylated in distinct substrates. P.Events: total number of phosphorylation events.

Appendix B

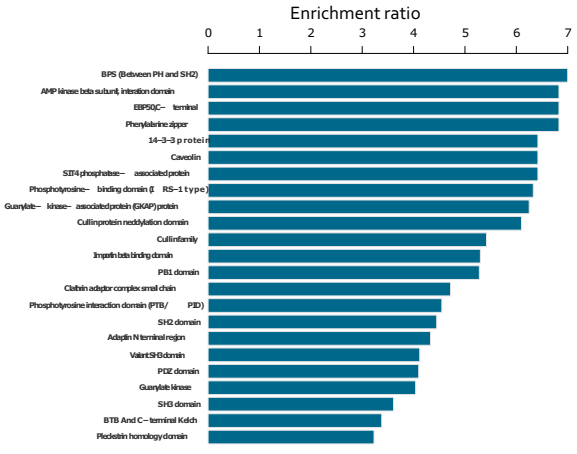

Fig. 5. Pfam domains enriched among the known adaptors and scaffolds.

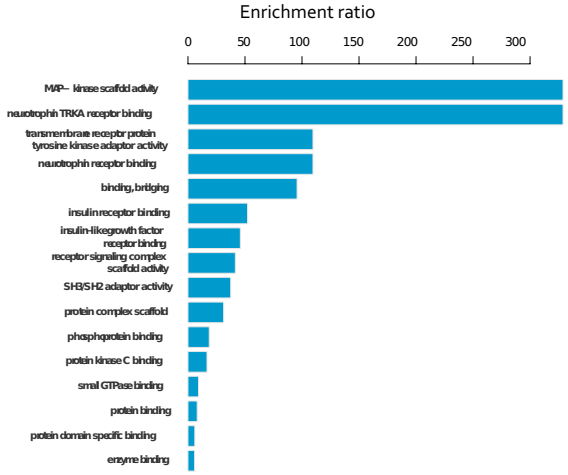

Fig. 6. Molecular function terms of the Gene Ontology enriched among the known adaptors and scaffolds.

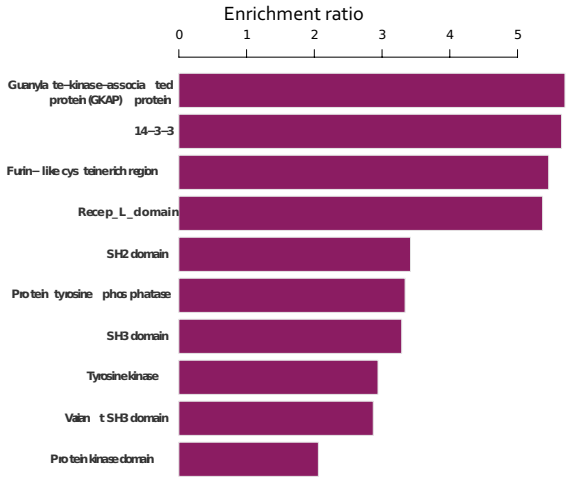

Fig. 7. Pfam domains enriched among the potential adaptors and scaffolds.

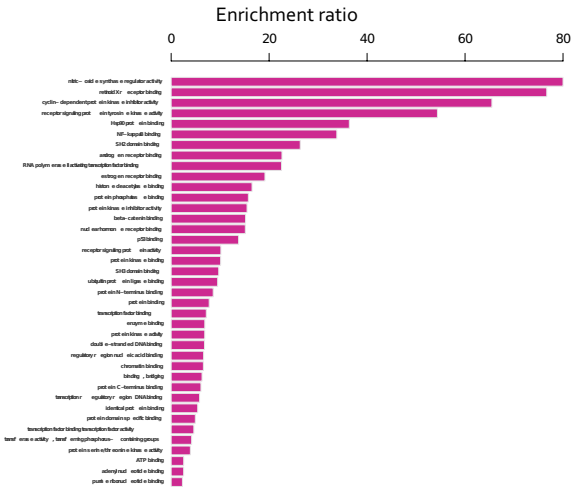

Fig. 8. Molecular function terms of the Gene Ontology enriched among the potential adaptors and scaffolds.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish Ministry of Economy and Competitivity (BIO 2010 22073). MAAT is supported by a 'la Caixa'/IRB Barcelona International Ph.D. Programme Fellowship (01/09/FLC). MAAT would like to thank M. Duran-Frigola (IRB Barcelona), A. Stein and R.A. Pache (UCSF, USA) and A. Zanzoni (TAGC-U1090 Inserm, France) for helpful discussions.

REFERENCES

[1] P. Cohen, "The origins of protein phosphorylation." *Nature cell biology*, vol. 4, no. 5, pp. E127–30, May 2002.

[2] S. Arena, S. Benvenuti, and a. Bardelli, "Genetic analysis of the kinome and phosphatome in cancer." *Cellular and molecular life sciences : CMLS*, vol. 62, no. 18, pp. 2092–9, Sep. 2005.

[3] A. Forrest, T. Ravasi, D. Taylor, T. Huber, D. Hume, and S. Grimmond, "Phosphoregulators: protein kinases and protein phosphatases of mouse," *Genome research*, vol. 13, no. 6b, p. 1443, 2003.

[4] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, "The protein kinase complement of the human genome." *Science (New York, N.Y.)*, vol. 298, no. 5600, pp. 1912–34, Dec. 2002.

[5] A. Torkamani, G. Verkhivker, and N. J. Schork, "Cancer driver mutations in protein kinase genes." *Cancer letters*, vol. 281, no. 2, pp. 117–27, Aug. 2009.

[6] P. Cohen, "The role of protein phosphorylation in human health and disease," *Eur J Biochem.*, vol. 268, no. 19, pp. 5001–5010, 2001.

[7] L. R. Pearce, D. Komander, and D. R. Alessi, "The nuts and bolts of AGC protein kinases." *Nature reviews. Molecular cell biology*, vol. 11, no. 1, pp. 9–22, Jan. 2010.

[8] K. Deshmukh, K. Anamika, and N. Srinivasan, "Evolution of domain combinations in protein kinases and its implications for functional diversity." *Progress in biophysics and molecular biology*, vol. 102, no. 1, pp. 1–15, Jan. 2010.

[9] J. Alexander, D. Lim, B. a. Joughin, B. Hegemann, J. R. a. Hutchins, T. Ehrenberger, F. Ivins, F. Sessa, O. Hudecz, E. a. Nigg, A. M. Fry, A. Musacchio, P. T. Stukenberg, K. Mechtler, J.-M. Peters, S. J. Smerdon, and M. B. Yaffe, "Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling." *Science signaling*, vol. 4, no. 179, p. ra42, Jan. 2011.

[10] A. N. Kettenbach, D. K. Schweppe, B. K. Faherty, D. Pechenick, A. a. Pletnev, and S. a. Gerber, "Quantitative phosphoproteomics identifies substrates and functional modules of aurora and polo-like kinase activities in mitotic cells." *Science signaling*, vol. 4, no. 179, p. rs5, Jan. 2011.

[11] J. Ptacek, G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, R. R. McCartney, M. C. Schmidt, N. Rachidi, S.-J. Lee, A. S. Mah, L. Meng, M. J. R. Stark, D. F. Stern, C. De Virgilio, M. Tyers, B. Andrews, M. Gerstein, B. Schweitzer, P. F. Predki, and M. Snyder, "Global analysis of protein phosphorylation in yeast." *Nature*, vol. 438, no. 7068, pp. 679–84, Dec. 2005.

[12] J. A. Ubersax and J. E. Ferrell, "Mechanisms of specificity in protein phosphorylation." *Nature reviews. Molecular cell biology*, vol. 8, no. 7, pp. 530–41, Jul. 2007.

[13] B. Kobe, T. Kampmann, and J. K. Forwood, "Substrate specificity of protein kinases and computational prediction of substrates," *Biochimica et biophysica acta*, vol. 1754, pp. 200 – 209, 2005.

[14] H. Zhu, J. F. Klemic, S. Chang, P. Bertone, a. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. a. Reed, and M. Snyder, "Analysis of yeast protein kinases using protein chips." *Nature genetics*, vol. 26, no. 3, pp. 283–9, Nov. 2000.

[15] J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann, "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks." *Cell*, vol. 127, no. 3, pp. 635–48, Nov. 2006.

[16] J. Mok, P. M. Kim, H. Y. K. Lam, S. Piccirillo, X. Zhou, G. R. Jeschke, D. L. Sheridan, S. a. Parker, V. Desai, M. Jwa, E. Cameroni, H. Niu, M. Good, A. Remenyi, J.-L. N. Ma, Y.-J. Sheu, H. E. Sassi, R. Sopko, C. S. M. Chan, C. De Virgilio, N. M. Hollingsworth, W. a. Lim, D. F. Stern, B. Stillman, B. J. Andrews, M. B. Gerstein, M. Snyder, and B. E. Turk, "Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs." *Science signaling*, vol. 3, no. 109, p. ra12, Jan. 2010.

[17] B. Hegemann, J. R. a. Hutchins, O. Hudecz, M. Novatchkova, J. Rameseder, M. M. Sykora, S. Liu, M. Mazanek, P. Lenart, J.-K. Heriche, I. Poser, N. Kraut, a. a. Hyman, M. B. Yaffe, K. Mechtler, and J.-M. Peters, "Systematic Phosphorylation Analysis of Human Mitotic Protein Complexes," *Science Signaling*, vol. 4, no. 198, pp. rs12–rs12, Nov. 2011.

[18] A. Kreegipuu, N. Blom, S. Brunak, and J. Ja, "Statistical analysis of protein kinase specificity determinants," *FEBS letters*, vol. 430, pp. 45–50, 1998.

[19] R. H. Newman, J. Hu, H.-S. Rho, Z. Xie, C. Woodard, J. Neiswinger, C. Cooper, M. Shirley, H. M. Clark, S. Hu, W. Hwang, J. Seop Jeong, G. Wu, J. Lin, X. Gao, Q. Ni, R. Goel, S. Xia, H. Ji, K. N. Dalby, M. J. Birnbaum, P. a. Cole, S. Knapp, A. G. Ryazanov, D. J. Zack, S. Blackshaw, T. Pawson, A.-C. Gingras, S. Desiderio, A. Pandey, B. E. Turk, J. Zhang, H. Zhu, and J. Qian, "Construction of human activity-based phosphorylation networks," *Molecular Systems Biology*, vol. 9, no. 655, pp. 1–12, Apr. 2013.

[20] G. Z. Hertz and G. D. Stormo, "Identifying DNA and protein patterns with statistically significant alignments of multiple sequences." *Bioinformatics (Oxford, England)*, vol. 15, no. 7-8, pp. 563–77, 1999.

[21] J. C. Obenauer, "Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3635–3641, Jul. 2003.

[22] N. F. W. Saunders, R. I. Brinkworth, T. Huber, B. E. Kemp, and B. Kobe, "Predikin and PredikinDB : a computational framework for the prediction of protein kinase peptide specificity and an associated database of phosphorylation sites," *BMC Bioinformatics*, vol. 11, pp. 1–11, 2008.

[23] N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, and S. r. Brunak, "Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence." *Proteomics*, vol. 4, no. 6, pp. 1633–49, Jun. 2004.

[24] M. C. Good, J. G. Zalatan, and W. a. Lim, "Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information," *Science*, vol. 332, no. 6030, pp. 680–686, May 2011.

[25] C. D. White, M. D. Brown, and D. B. Sacks, "IQGAPs in cancer: a family of scaffold proteins underlying tumorigenesis." *FEBS letters*, vol. 583, no. 12, pp. 1817–24, Jun. 2009.

[26] H. Zhang, A. Photiou, G. Arnhild, J. Stebbing, and G. Giamas, "The role of pseudokinases in cancer." *Cellular signalling*, vol. 24, no. 6, pp. 1173–1184, Feb. 2012.

[27] A. S. Shaw and E. L. Filbert, "Scaffold proteins and immune-cell signalling." *Nature reviews. Immunology*, vol. 9, no. 1, pp. 47–56, Jan. 2009.

[28] A. Alexa, J. Varga, and A. Reményi, "Scaffolds are 'active' regulators of signaling modules." *The FEBS journal*, vol. 277, no. 21, pp. 4376–82, Nov. 2010.

[29] M. Colledge and J. D. Scott, "AKAPs: from structure to function." *Trends in cell biology*, vol. 9, no. 6, pp. 216–21, Jun. 1999.

[30] M. M. McKay, D. a. Ritt, and D. K. Morrison, "Signaling dynamics of the KSR1 scaffold complex." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 27, pp. 11 022–7, Jul. 2009.

[31] F. Ramírez and M. Albrecht, "Finding scaffold proteins in interactomes." *Trends in cell biology*, vol. 20, no. 1, pp. 2–4, Jan. 2010.

[32] D. F. Brennan, A. C. Dar, N. T. Hertz, W. C. H. Chao, A. L. Burlingame, K. M. Shokat, and D. Barford, "A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK." *Nature*, vol. 472, no. 7343, pp. 366–9, Apr. 2011.

[33] W. G. Cance, E. Kurenova, T. Marlowe, and V. Golubovskaya, "Disrupting the Scaffold to Improve Focal Adhesion Kinase-Targeted Cancer Therapeutics," *Science Signaling*, vol. 6, no. 268, pp. pe10–pe10, Mar. 2013.

[34] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, "Human Protein Reference Database–2009 update." *Nucleic acids research*, vol. 37, no. Database issue, pp. D767–72, Jan. 2009.

[35] P. V. Hornbeck, I. Chabra, J. M. Kornhauser, E. Skrzypek, and B. Zhang, "PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation." *Proteomics*, vol. 4, no. 6, pp. 1551–61, Jun. 2004.

[36] H. Dinkel, C. Chica, A. Via, C. M. Gould, L. J. Jensen, T. J. Gibson, and F. Diella, "Phospho.ELM: a database of phosphorylation sites–update 2011." *Nucleic acids research*, vol. 39, no. November 2010, pp. 261–267, Nov. 2010.

[37] J. M. Claverie and S. Audic, "The statistical significance of nucleotide position-weight matrix matches." *Computer applications in the biosciences : CABIOS*, vol. 12, no. 5, pp. 431–9, Oct 1996.

[38] G. D. Stormo, "DNA binding sites: representation and discovery." *Bioinformatics (Oxford, England)*, vol. 16, no. 1, pp. 16–23, Jan. 2000.

[39] R. Mosca, A. Céol, and P. Aloy, "Interactome3D: adding structural details to protein networks." *Nature methods*, vol. 10, no. 1, pp. 47–53, Dec. 2013.

[40] G. Badis, M. F. Berger, A. a. Philippakis, S. Talukder, A. R. Gehrke, S. a. Jaeger, E. T. Chan, G. Metzler, A. Vedenko, X. Chen, H. Kuznetsov, C.-F. Wang, D. Coburn, D. E. Newburger, Q. Morris, T. R. Hughes, and M. L. Bulyk, "Diversity and complexity in DNA recognition by transcription factors." *Science (New York, N.Y.)*, vol. 324, no. 5935, pp. 1720–3, Jun. 2009.

[41] D. Matenia and E.-M. Mandelkow, "The tau of MARK: a polarized view of the cytoskeleton." *Trends in biochemical sciences*, vol. 34, no. 7, pp. 332–42, Jul. 2009.

[42] A. Alonso, T. Zaidi, M. Novak, I. Grundke-Iqbal, and K. Iqbal, "Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 12, pp. 6923–8, Jun. 2001.

[43] H. Matsuzaki, A. Ichino, T. Hayashi, T. Yamamoto, and U. Kikkawa, "Regulation of intracellular localization and transcriptional activity of FOXO4 by protein kinase B through phosphorylation at the motif sites conserved among the FOXO family." *Journal of biochemistry*, vol. 138, no. 4, pp. 485–91, Oct. 2005.

[44] M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn, "The Pfam protein families database." *Nucleic acids research*, vol. 40, no. Database issue, pp. D290–301, Jan. 2012.

[45] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nature genetics*, vol. 25, no. 1, pp. 25–9, May 2000.

[46] A. Zeke, M. Lukács, W. a. Lim, and A. Reményi, "Scaffolds: interaction platforms for cellular signalling circuits." *Trends in cell biology*, vol. 19, no. 8, pp. 364–74, Aug. 2009.

[47] S. Karthikeyan, T. Leung, and J. A. A. Ladias, "Structural determinants of the Na+/H+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors." *The Journal of biological chemistry*, vol. 277, no. 21, pp. 18 973–8, May 2002.

[48] K. Jiang, E. Pereira, M. Maxfield, B. Russell, D. M. Goudelock, and Y. Sanchez, "Regulation of Chk1 includes chromatin association and 14-3-3 binding following phosphorylation on Ser-345." *The Journal of biological chemistry*, vol. 278, no. 27, pp. 25 207–17, Jul. 2003.

[49] G. A. Penman, L. Leung, and I. S. Näthke, "The adenomatous polyposis coli protein (APC) exists in two distinct soluble complexes with different functions." *Journal of cell science*, vol. 118, no. Pt 20, pp. 4741–50, Oct. 2005.

[50] C. Reichen, S. Hansen, and A. Plckthun, "Modular peptide binding: From a comparison of natural binders to designed armadillo repeat proteins," *Journal of Structural Biology*, vol. In press, no. 0, p. http://dx.doi.org/10.1016/j.jsb.2013.07.012, 2013.

[51] R. Roskoski, "ERK1/2 MAP kinases: structure, function, and regulation." *Pharmacological research : the official journal of the Italian Pharmacological Society*, vol. 66, no. 2, pp. 105–43, Aug. 2012.

[52] A. C. Lloyd, "Distinct functions for ERKs?" *Journal of Biology*, vol. 5, no. 5, p. 13, Jan. 2006.