BioMol Concepts, Vol. 1 (2010), pp. 185–195 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BMC.2010.013

Variable contribution of protein kinases to the generation of the human phosphoproteome: a global weblogo analysis

Mauro Salvi[1,a], Luca Cesaro[1,a] and Lorenzo A. Pinna[1,2,*]

[1] Department of Biological Chemistry, University of Padova, V.le G. Colombo 3, 35131 Padova, Italy
[2] Venetian Institute of Molecular Medicine (VIMM), Padova, Italy

* Corresponding author
e-mail: lorenzo.pinna@unipd.it

Abstract

In an attempt to evaluate the contribution of individual protein kinases to the generation of the human phosphoproteome, we performed a global weblogo analysis exploiting a database of 45641 phosphosites (80% pSer, 11% pTyr, 9% pThr). The outcome of this analysis was then interpreted by comparison with similar logos constructed from bona fide phospoacceptor sites of individual pleiotropic kinases. The main conclusions that were drawn are as follows: (i) the hallmarks surrounding phosphorylated Ser/Thr residues are more pronounced than and sharply different from those found around phosphorylated Tyr, which is consistent with the view that local consensus sequences are particularly important for substrate recognition by Ser/Thr protein kinases. (ii) Only six residues are positively selected around phosphorylated Ser/Thr residues, notably Pro (particularly at n+1), Glu, and to a lesser extent Asp, at various positions with special reference to n+3, Arg (and to a much lesser extent Lys), particularly at n-3 and n-5, and Ser, at various positions, particularly n+4 and n-4. (iii) This composite signature reflects the contribution of kinases whose bona fide substrates exhibit logos partially overlapping that of the whole phosphoproteome. These are Pro-directed kinases belonging to the CMGC group, some basophilic kinases belonging to the ACG and CAMK groups, phosphate-directed kinases such as GSK3 and members of the CK1 group and the individual highly acidophilic CK2. Collectively taken our data support the concept that a relatively small number of highly pleiotropic kinases contribute to the generation of the great majority of the human Ser/Thr phosphoproteome.

Keywords: iceLogo; kinase motif; phosphoproteome; protein kinase; two-sample logo.

Introduction

Reversible protein phosphorylation affecting Ser, Thr and Tyr residues is the most frequent and general mechanism by which nearly all biological functions are regulated. It is generally held that not less than 30% of the cellular proteome undergoes this type of post-translational modification, which is catalyzed by the members of an individual large family of enzymes, protein kinases, numbering over 500 in human and often referred to as the human 'kinome' (1). Moreover, considering that many proteins are phosphorylated at multiple residues (2, 3) it is expectable that in a single eukaryotic cell many thousands of phosphosites are persistently generated over time, an inference which is fully consistent with the almost 60000 non-redundant phosphopeptides derived from approximately 10000 non-redundant proteins presently retrieved in PhosphoSitePlus® (4), the most comprehensive and updated phosphosite database (www.phosphosite.org). All these are generated by the concerted action of approximately 500 protein kinases, which in turn are under the control of an exceedingly complicated signaling network. Alterations in the phosphoproteome will therefore ultimately reflect physiopathological events taking place in the cell.

It is well known for a long time that the pleiotropicity of protein kinases is extremely variable, some of them being dedicated enzymes, affecting only one or few substrates (e.g., phosphorylase kinase and wee1), whereas others are very pleiotropic, with hundreds of protein targets subjected to their control. It is expectable that the latter will be the main contributors of the phosphoproteome as a whole. A formidable challenge in the postgenomic era would be to track out the functional links between every protein kinases and its targets, which in turn would provide a global enlightening over the signaling network responsible for cell regulation under physiological conditions or its dysregulation whenever a pathogenic event occurs.

In the past, when a limited number of protein kinases were known and proteomics was still in its infancy, the link between kinases and their targets was generally inspected by a 'downward' analytical strategy, focusing on the kinase of interest and searching for its 'natural' substrates. These studies took enormous advantage of so-called 'consensus sequences' defined with the aid of artificial peptide substrates, whose latest repertoires date back to the end of the past century [e.g., Refs. (5, 6), see also Ref. (7)]. More recently however, the identification of the whole kinome, on one side, and the spectacular development of mass spectrometry technologies on the other, providing endless repertoires of *in vivo* generated 'orphan' phosphosites, have made possible, and in a way mandatory, a global 'upward' approach,

[a] These authors contributed equally to this work.

exploiting bioinformatics to try to ascribe each phosphosite to its putative kinase, or at least to evaluate the contribution of individual kinases, or groups of them, to the generation of a given phosphoproteome.

Several computational approaches have been developed to attain this goal, as reviewed in Ref. (8). Among these the ones based on 'weblogo' (9), a graphical representation of kinase site specificity, has the limit of ignoring specificity determinants which are not residing inside the phosphoacceptor site [e.g., remote docking sites, compartmentation, etc. (7, 10, 11)] having, however, the big advantage of being simple and rather reliable as a predictive parameter whenever a sufficiently large number of bona fide phosphosites generated by the kinase of interest are already known. In other words, a number of phosphosites present in the database can escape clear-cut assignment by weblogo analysis, whereas those displaying the signatures of the weblogo of a given kinase can be ascribed to this with good confidence, as previously shown in the case of phosphosites generated by CK2 (12). More detailed information can be drawn from improved versions of the weblogo, where the data are analyzed in a comparative, instead of absolute, manner, thus highlighting differences between two sets of sequence alignments, e.g., sequences surrounding a phosphorylated amino acid compared with those surrounding the same amino acid regardless of its phosphorylation state. This is possible by applying the 'two-sample' logo (13) and/or a more recent version of it termed 'iceLogo' (14).

In this paper, we have exploited these bioinformatic tools to visualize conserved patterns in protein sequences surrounding Ser, Thr and Tyr residues in the ~45000 non-redundant human phosphopeptides available in the PhosphoSitePlus® database, and to make inferences about the contribution of individual subfamilies of pleiotropic kinases, or groups of kinases, to their generation.

Methods

Weblogo analysis

Weblogo analysis was performed using WebLogo 2.8.2 (http://weblogo.berkeley.edu/logo.cgi) (9). IceLogo analysis was performed using IceLogo 1.0 (14). In IceLogo images the amino acids are pink-colored if this specific amino acid does not occur in the positive or reference set. Human phosphopeptides (+7, -7) were downloaded from PhosphoSitePlus® (December 2009 database). Human random peptides (+7, -7) have been extracted from the Swiss Prot database using a homemade bioperl script and unix text processing commands. Non-redundant sequences have been randomized using unix command shuff.

Results and discussion

Recurrent features of sequences surrounding phosphorylated Ser/Thr vs. Tyr residues

A weblogo analysis of all sequences surrounding the potentially phosphoacceptor residues serine and threonine in the whole human proteome (Figure 1A), where serine is slightly more abundant than threonine, reveals no significant selection at any position: frequency of individual residues (quantified in the inset by the size of each letter) generally reflects the random frequency of the same residues in the proteome as a whole, highlighting the 'rarity' of certain residues, such as W, M, C, Y, H, F as opposed to the relative abundance of others, particularly L, S, A, P, E, G.

If however the same analysis is applied to phosphorylated human Ser/Thr residues currently available in the PhosphoSitePlus® database (amounting to 35679) this scenario is significantly altered (Figure 1B) with a marked selection of some residues at a given position, notably n+1 (P and to a lesser extent S, D and E), n+3 (particularly E), n-3 (R). As better outlined by representations where all stacks are normalized to the same height (compare insets of Figure 1B and 1A) phosphoserine neatly predominates over phosphothreonine (>5-fold) and there is a general tendency of a few residues to increase their frequency at given positions, with special reference to Pro at position n+1, Arg at positions n-3 and n-5, Glu at positions n+3 and to a lesser extent at n+4, n+5, n+6 and Asp at position n-1. By passing from the random proteome to the phosphoproteome the frequency of all these residues increases from 0.7- to 4.5-fold. In contrast, other residues (e.g., Ser) display similar frequency in the two weblogos, whereas others (notably Leu and Ala) are less frequent at most positions in the phosphoproteome weblogo than they are in the random one, suggesting that they tend to be 'discarded' around phosphorylated Ser/Thr residues.

If the same type of analysis is performed with sequences surrounding tyrosine and phosphotyrosine the weblogo displayed in Figure 2 is obtained: the former one (Figure 2A) is similar to that of sequences surrounding non-phosphorylated Ser/Thr (Figure 1A) with frequency of individual residues generally reflecting the frequency they randomly have in the whole proteome (thus L, S, G, A, E are overrepresented, whereas W, M, H, Y, C are rather rare). Passing to the phospho-Tyr weblogo (Figure 2B) significant alterations in relative abundance of individual residues at given positions can be observed, but these are less pronounced and involve different residues than in the case of the pS/T vs. S/T weblogo. Apart from a significant increase in the frequency of upstream acidic residues, which now share with Ser the top position in the equally high stacks representation (Figure 2B, inset) and the partial displacement of Leu from top level previously occupied at all positions, no other deep alterations can be observed passing from the Tyr to the phospho-Tyr weblogo. Of special note is the lack of selection of Pro (particularly at position n+1), of Glu (most downstream positions) and of Arg (at nearly all upstream positions), which were the hallmarks of the phospho-Ser/Thr weblogo (compare Figures 2B and 1B). This global analysis not only highlights the fact that consensus sequences of Ser/Thr kinases sharply differentiate from those of Tyr kinases but it also corroborates the notion that although substrate targeting by the former deeply relies on local specificity determinants, the latter play a much less prominent role in determining the substrate specificity of most tyrosine kinases (6).

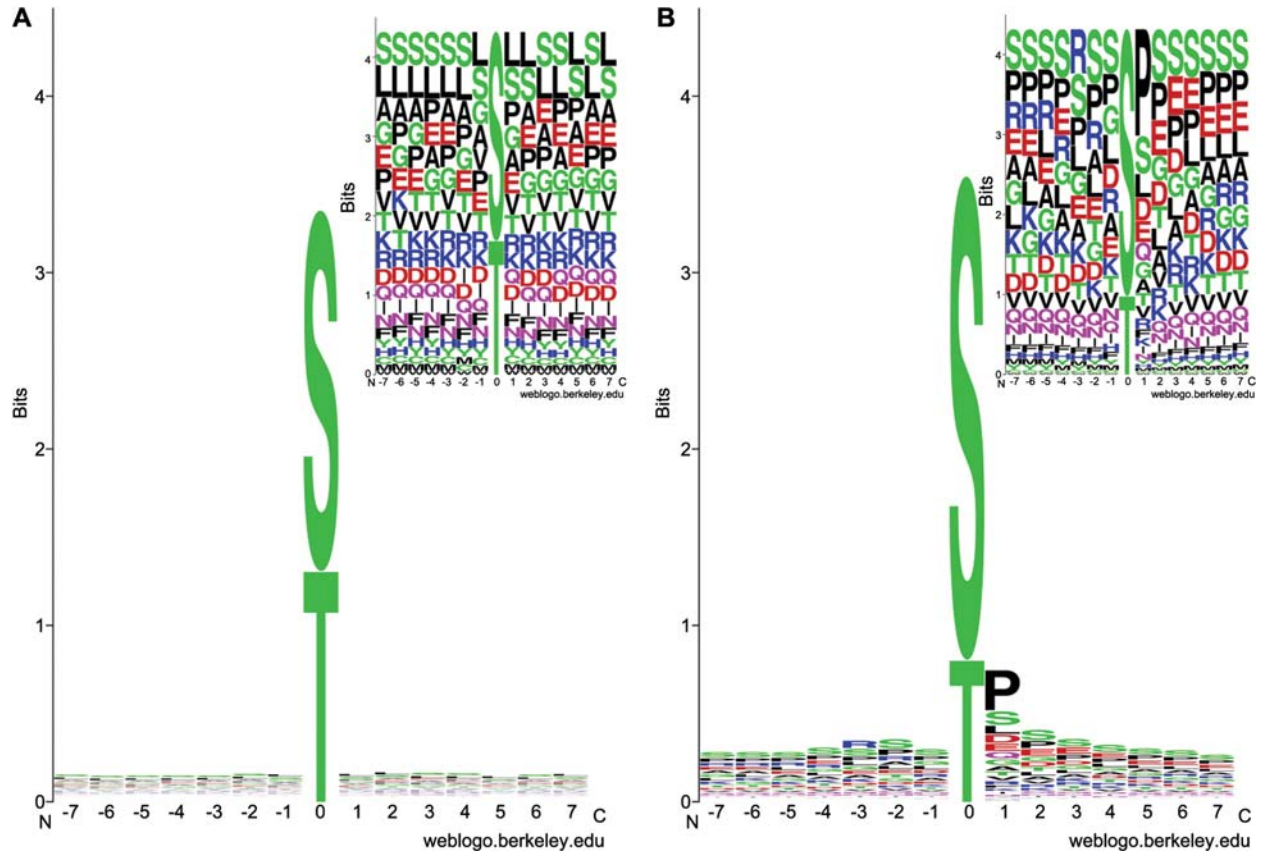

Figure 1 Weblogo analysis of Ser/Thr proteome and phosphoproteome.
(A) Weblogo representation of 35679 random peptides centered on Ser or Thr residues collected as described in the Methods section. In the inset the corresponding frequency plot is reported. (B) Weblogo representation of 35679 phospho-Ser/Thr peptides from PhosphoSitePlus®. In the inset the corresponding frequency plot is reported.

Two-sample logo ('iceLogo') analysis of the whole phosphoproteome

A more detailed evaluation of positive and negative selection of individual residues at given positions around phosphorylated serines/threonines can be provided by the two-sample logo (13), a graphical representation of the differences between two sets of sequence alignment. In our case, the two sets were sequences surrounding phosphorylated serines/threonines vs. sequences surrounding serines/threonines in general. This type of analysis gives rise to an upper section displaying a set of amino acid symbols overrepresented in the positive set (phosphoresidues) as compared to the reference one (all Ser/Thr sites no matter whether phosphorylated or not), a lower section displaying a set of amino acid symbols underrepresented in the positive set, and a middle section eventually containing amino acid symbols which are equally represented in the positive and background (negative) set.

In particular, we have adopted a recently developed visualization method of the two-sample logo, the so-called 'iceLogo' (14) to compare the human Ser/Thr phosphoproteome, based on 35679 phosphosites, to the random sequences surrounding all Ser and Thr residues in the human proteome (Figure 3). Its analysis on one side corroborates the main conclusions suggested by the weblogo (Figure 1B) and on the other it provides additional information. Particularly noteworthy are the striking overrepresentation of Pro at position n+1, where aspartic is the only other overrepresented residue, and the enrichment in basic residues (particularly Arg) at upstream positions and of acidic residues (particularly glutamic acid) at most downstream positions. It is important to note, however, also the enrichment in Pro at positions other than n+1 and of acidic residues at upstream positions, although not as overrepresented as basic residues (with the exception of position n-1). The same applies to the basic residue Arg which is also found among the positively selected residues at some positions downstream, although not as prominent as acidic residues. The iceLogo also highlights the concept that in the phosphoproteome pSer is predominant (overrepresented) over pThr (underrepresented) and that on the average residues enriched at phosphosites are rather few: serine, proline, arginine, lysine, aspartic and glutamic acids, and, only marginally and just at position n-1, glycine. The majority of residues indeed are underrepresented, suggesting a negative selection by protein kinases or at least by the most pleiotropic of these. Particularly remarkable in this respect is the negative selection of leucine and, to a lesser extent, of

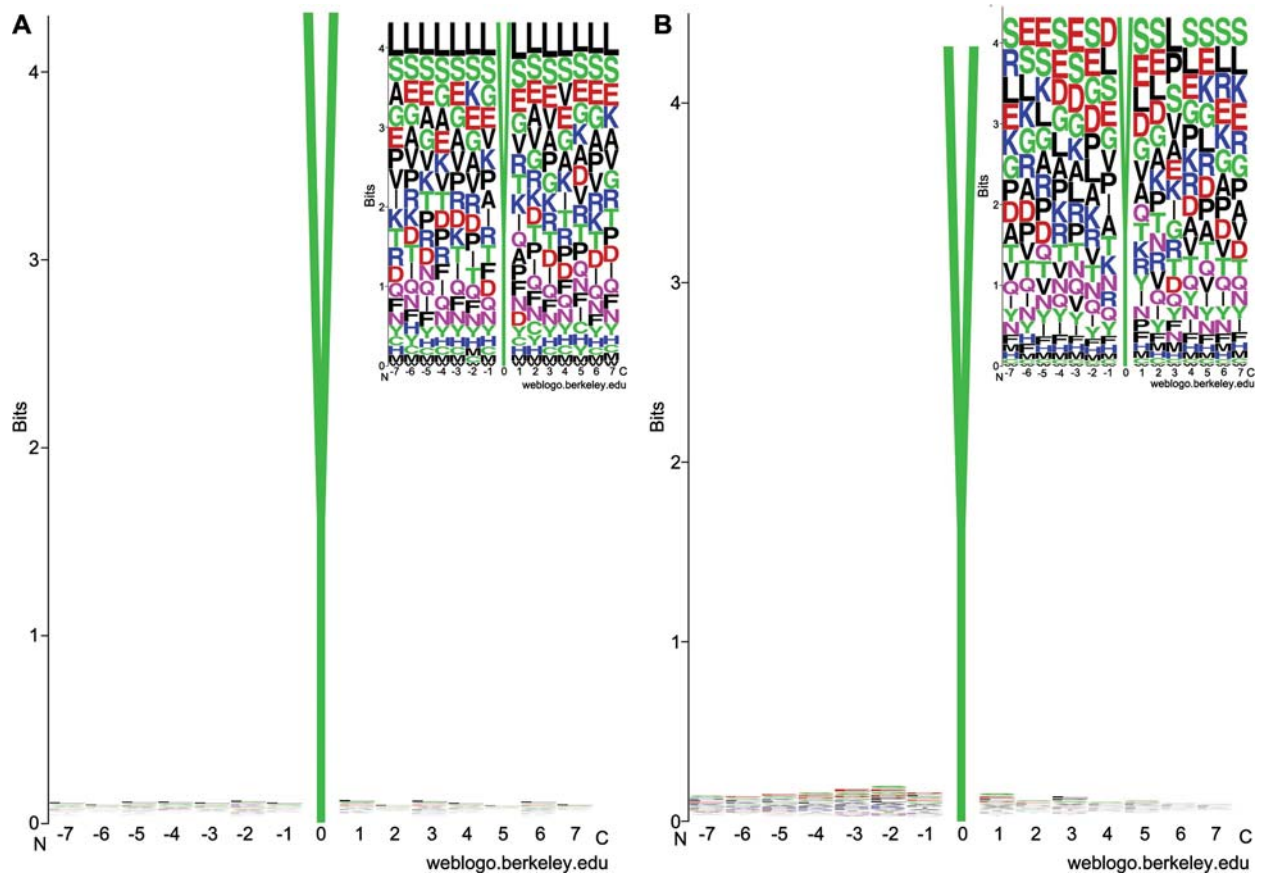

Figure 2 Weblogo analysis of Tyr proteome and phosphoproteome.
(A) Weblogo representation of 9962 random peptides centered on Y residues collected as described in the Methods section. In the inset the corresponding frequency plot is reported. (B) Weblogo representation of 9962 phospho-Y peptides from PhosphoSitePlus®. In the inset the corresponding frequency plot is reported.

isoleucine and valine at all positions surrounding the phosphoresidue, consistent with the concept that Ser/Thr embedded inside hydrophobic clusters are shielded by the 3D conformation of the protein substrate where they are not accessible to protein kinases. By contrast, serine is positively selected at nearly all positions, consistent with the observation that often phosphorylated seryl residues are clustered together (3) (see also below).

A cursory scrutiny of the iceLogo combined with current knowledge of the consensus sequence of some categories of pleiotropic protein kinases suggests that major contributions to the human phosphoproteome are provided by relatively few classes of protein kinases, notably (i) proline-directed kinases of the CMGC group, (ii) the basophilic kinases which are predominant in the AGC and CAMK groups, (iii) some phosphate-directed kinases whose targeting is primed by previously phosphorylated residues, and (iv) the very acidophilic protein kinase CK2.

IceLogo analysis of phosphosites generated by individual protein kinases

This inference has been corroborated by generating the iceLogo of a number of pleiotropic kinases whose repertoires of bona fide substrates are sufficiently numerous for rendering reliable this type of analysis, and by making an estimate of their contribution to the iceLogo of the whole phosphoproteome. The iceLogo drawn from repertoires of bona fide substrates of a number of proline-directed (cdc2, Erk1) and basophilic (PKA, PKB/Akt, PKCα, CaMKα) protein kinases, and of the acidophilic protein kinases CK2 and the Golgi Casein kinase (GCK) are displayed in Figure 4, which also includes the iceLogo of two 'phosphate-directed' kinases, CK1 and GSK3β, whose targeting is critically primed by a phosphorylated serine at definite distance from the target amino acid. It should be firstly noted that the iceLogo of individual kinases differentiate not only for the nature and position of the amino acids positively selected (generally corresponding to the 'consensus sequence') but also for the number of critical positions where positive and negative selections are particularly evident. These could be just one or two, for example, in the cases of ERK1, CDK2, PKA and PKB/Akt, whereas being more numerous with other kinases, for example, PKCα and, even more with CK2. As a general rule positions where stacks of positive selection are prominent are the same where negative selection is also most evident, highlighting the crucial relevance of some

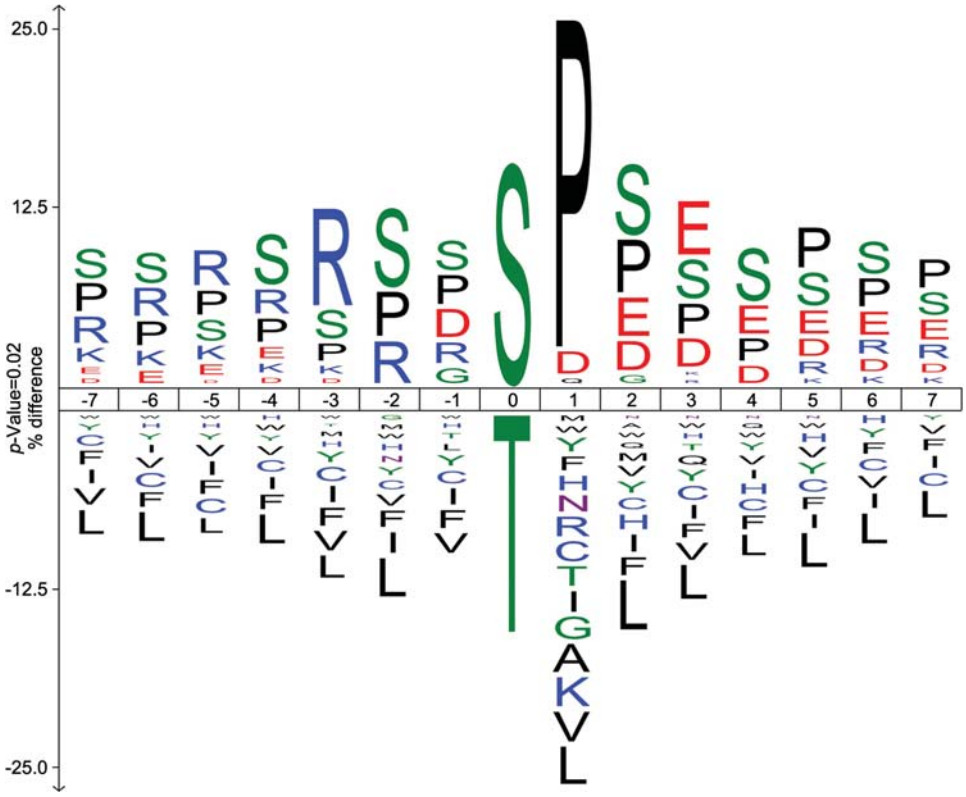

Figure 3 IceLogo analysis of S/T phosphoproteome vs. S/T proteome.
IceLogo representation highlighting the different distribution of residues surrounding pS/pT with regard to the random distribution. Random and phosphopeptides are collected and analyzed as described in Methods section.

positions where residues are neatly perceived as positive or negative determinants of recognition as compared to others where nearly every residue is equally tolerated. Thus, the very positive selection of just Pro at position n+1 and of just Arg at positions n-2/n-3 and n-3/n-5 by CDK2, PKA and PKB/Akt, respectively, mirrors high stacks of numerous negatively selected residues at the same positions. This trend is confirmed by the profile of CK2, characterized by a remarkable positive selection of acidic residues at all positions between n-1 and n+5, peaking at position n+3, reflecting in a symmetric shape of the negatively selected residues on display in the lower set. It is important to note that Leu is among the most negatively selected residues at nearly all critical positions, with the majority of protein kinases, thus accounting for the observation that this residue is the most negatively selected one at nearly all positions of the whole human Ser/Thr phosphoproteome (Figure 3). By comparing this global picture with the iceLogo of individual protein kinases it is clear that its features reflect the combination of three major contributions: those of proline directed and of basophilic protein kinases and the one of protein kinase CK2. These altogether account for five out of the six residues positively selected in the whole Ser/Thr phosphoproteome weblogo, namely P, R, K, E, D. The sixth residue found in the upper panel of the phosphoproteome iceLogo at many positions is serine: this can partially reflect the high frequency of this residue in the whole proteome by simply assuming that it is not 'discarded' (see Figures 1A and 2A); but it probably is also symptomatic of the contribution of 'primed' kinases, whose consensus is generated by hierarchical phosphorylation of a seryl residue at a given position (16). This feature is almost invariably displayed by GSK3β and to a lesser extent by CK1. The iceLogo of these two kinases is also shown in Figure 4, and it can be seen that not only are they compatible with that of the whole phosphoproteome but they can well account for the appearance in this serine among the positively selected amino acids, particularly at positions n-4 and n+4 (GSK3 contribution) and n-3/n+3 (CK1 contribution). This inference would imply of course that the serines positively selected by this type of analysis were phosphorylated at some time before the target phosphoamino acid was in its turn generated.

One conclusion which can be drawn from our iceLogo analysis is that the local specificity determinants of the most common protein kinases, established by early studies mostly performed with synthetic peptides when the kinome was still largely unknown (6), are sufficient to account for the local features surrounding phosphoresidues in the whole known phosphoproteome. Negative selection of Thr as compared to Ser at n=0 position is also consistent with the concept (confirmed by the iceLogo in Figure 3) that most protein kinases, with few exceptions, display a preference for Ser over Thr within otherwise identical sequences (6). Consequently, it is very unlikely that pleiotropic protein kinases recognizing dif-

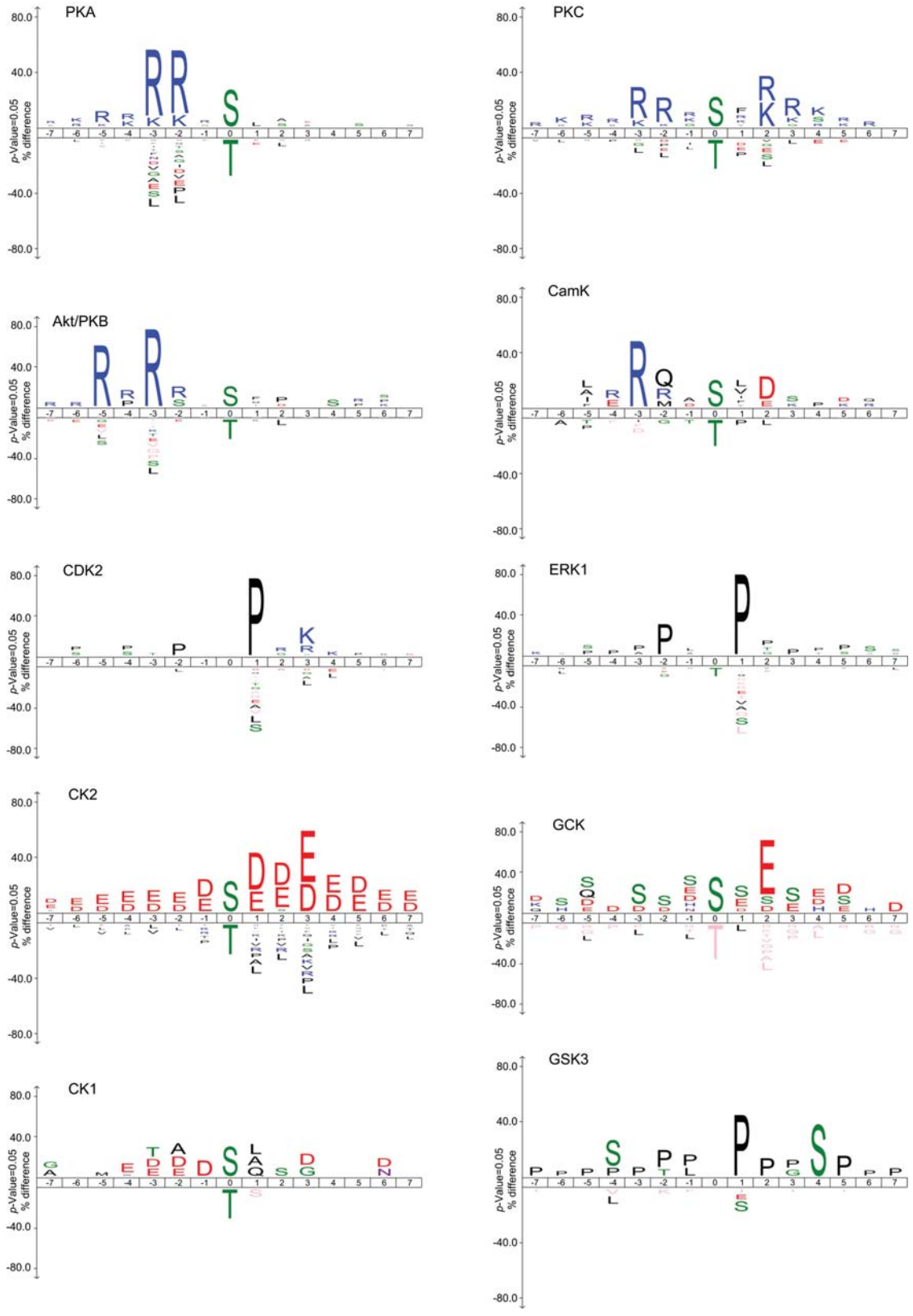

Figure 4 IceLogo analysis of phosphosites generated by individual protein kinases vs. random S/T proteome.
Bona fide substrates for each kinase considered here have been extracted from PhosphoSitePlus® and PhosphoElm (15); CK2 substrates have been collected from Ref. (12).

ferent and presently unknown local specificity determinants substantially contribute to the generation of the Ser/Thr phosphoproteome.

As mentioned above, reliance of Tyr protein kinases on local specificity determinants is, on the average, much less pronounced than that of Ser/Thr protein kinases (6, 17). This is clearly confirmed by the iceLogo generated using the ~10000 available phosphotyrosyl sites available, which were exploited to construct the weblogo shown in Figure 2. The most noticeable hallmark of the pTyr sites iceLogo is the flatness and uniformity of its profile (shown in Figure 5A) whose height never reaches 10%, as compared to those of the pSer and pThr sites (Figure 5B and C, respectively), denoting a more pronounced selection of certain residues at given positions. In particular, positive selection of Pro at position n+1, Glu at position n+3 and Arg at n-3, which are evident in both the pS and pT iceLogo, are entirely lacking in the pY one, whose only feature shared with iceLogo of pS and pT sites is negative selection of Leu (and to a lesser extent of other hydrophobic residues) at nearly all positions. It is also worthy to note that there are significant differences between the IceLogo of pS and pT phosphosites, notably a more pronounced proline selection in the latter as opposed to a more pronounced Arg (upstream) and Glu (downstream) selection of the former. These clearly reflect the unique feature of many Pro-directed protein kinases of preferring Thr over Ser, whereas the opposite applies to other protein kinases, either basophilic or acidophilic (Figure 4).

Expectedly, a large number of 'proline-directed' and 'basophilic' protein kinases cooperate with the positive selection of prolines and arginines/lysines, respectively, in the IceLogo of the whole phosphoproteome (Figure 3). A future challenge will be to dissect the contribution of individual kinases (or subsets of kinases) belonging to these two categories. We reasoned that some relevant information could be gained by recalculating the iceLogo of individual protein kinases considered in Figure 4 relative to the whole phosphoproteome rather than to the proteome as such. In this way, unique distinctive features of the individual kinase bona fide phosphosites can be better perceived and their possible contribution to the whole phosphoproteome more finely delineated. This analysis (Figure 6) allows, for example, to perceive more neatly the discriminating role of a prolylresidue at position n+1 where it is either absolutely required or rejected by Pro-directed kinases and by the other kinases, respectively. It also shows that negative selection of Thr with regard to Ser at position zero, shared by several kinases in Figure 4, is less pronounced in Figure 6, where the background set is provided by the phosphoproteome and not by the whole proteome. Likewise, this mode of representation highlights the relevance of positive and negative determinants which were disclosed in old specificity studies, but were not evident in Figure 4: this applies, for example, to the positive selection of a hydrophobic residue at n+1 in PKA sites and the negative selection of basic residues upstream from CK2 phosphoresidues, which were already inferred from studies with model peptides (6) but cannot be seen in Figure 4. The outcome of this analysis, moreover,

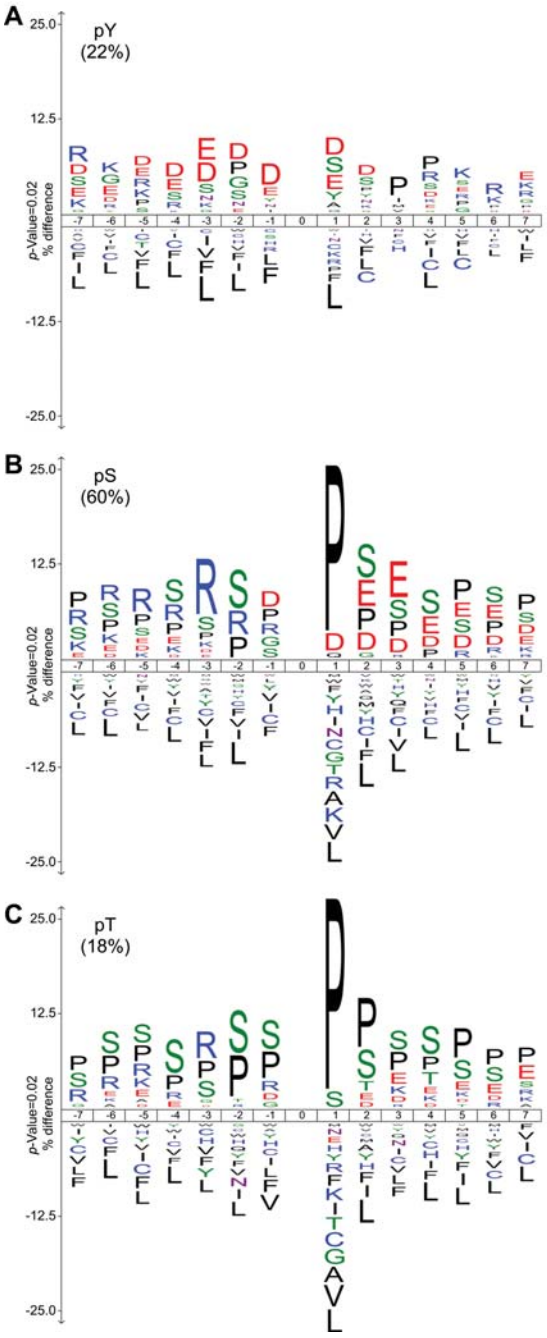

Figure 5 IceLogo analysis of pY, pS, pT sites vs. random. IceLogo representation highlighting the different residue distribution surrounding pY, (A) pS (B) and pT (C) residues regarding the random distribution. Random and phosphopeptides are collected as described in the Methods section.

highlights the contribution among basophilic kinases of PKB/Akt and of PKA, as their hallmarks (positive selection of Arg at position n-3/n-5 and n-2/n-3, respectively) are also a prominent feature of the whole phosphoproteome IceLogo (Figure 3). Apparently, however, other basophilic protein kinases are responsible for the markedly positive selection of Arg (and to a lesser extent Lys) at upstream positions other than n-2, n-3 and n-5 (see also Conclusions). It is dif-

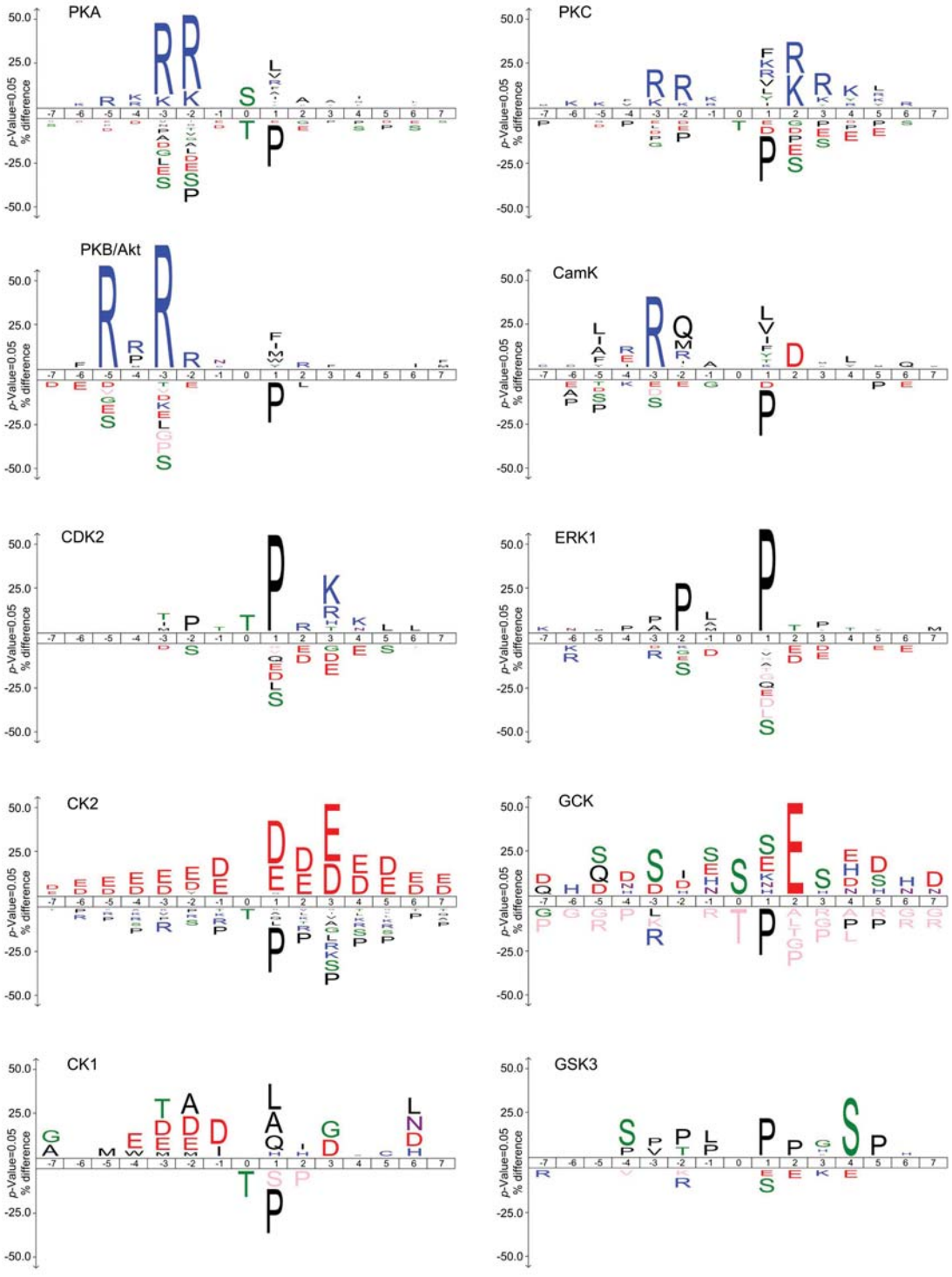

Figure 6 IceLogo analysis of phosphosites generated by individual protein kinases vs. human phospho-S/T proteome.
Bona fide substrates for each kinase considered here have been extracted from PhosphoSitePlus® and PhosphoElm (15); CK2 substrates have been collected from Ref. (12).

ficult to dissect the contributions of individual classes of Pro-directed protein kinases because the iceLogos of ERK1 and CDK2 are similar as far as positive selection of Pro at positions n+1 and n-2 are concerned. CDK2, however (and other CDKs as well), differentiates from ERKs for the positive selection of Lys at position n+2: this feature is not

	Pro-directed	PKA	AKT	CK2	GCK	GSK3	CK1	None
Pro-directed	11 008	72	149	2406	1333	263	172	
PKA	72	739	88	225	133	14	0	
AKT	149	88	793	278	137	34	0	
CK2	2406	225	278	11 361	0	232	1507	
GCK	1715	188	222	2812	4470	193	130	
GSK3	1651	83	142	1595	693	5048	715	
CK1	1224	0	0	1507	538	715	4462	
None								9033

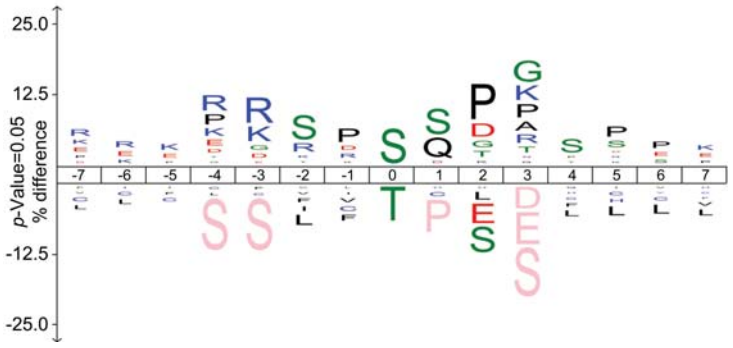

Figure 7 Grouping of human proteome Ser/Thr phosphosites according to the presence of consensus sequences specifically recognized by a number of pleiotropic protein kinases.
In total, 35679 Ser/Thr phosphosites available in the PhosphoSitePlus® database have been analyzed for the presence of motifs specifically recognized by individual (groups of) pleiotropic protein kinases as follows: pS/pT-P (Pro-directed PKs); R-R-x-pS/pT (PKA), R-x-R-x-x-pS/pT (PKB/Akt); pS/pT-x-x-E/D/S (CK2); pS-x-E/S-X′, where x′≠E/D/S (G-CK); pS/pT-x-x-x-S (GSK3); S-x-x-pS/pT (CK1). The table is constructed to highlight the entire number of phosphosites conforming to the consensus of an individual type of kinase (shaded boxes) and sharing it with the consensus of other kinases (e.g., in the case of AKT the total number of phosphosites with the consensus R-x-R-x-x-S/T is 793, of these 149 also display the motif of Pro-directed kinases, 89 of PKA, 278 of CK2, etc.). A total of 9033 phosphosites not conforming to any of the above motifs have been analyzed for their IceLogo vs. random S/T sites (lower panel).

evident in the IceLogo of the whole phosphoproteome, suggesting that contribution of CDKs is less pronounced than that of ERKs and of other Pro-directed, yet not basophilic, protein kinases. In this category also falls GSK3, although the most prominent signature of the iceLogo of these kinases is positive selection of Ser at position n+4 and to a lesser extent n-4. Certainly, this reflects the well established primed GSK3 consensus S-x-x-x-pS, and the presence of Ser as the second most selected residue at position n+4 in the phosphoproteome iceLogo (Figure 3) is most likely a signature of a conspicuous contribution of GSK3 to its generation (see also Conclusions).

Contribution of the master kinase CK2

At variance with Pro and basic residues (Arg and to a lesser extent Lys) whose abundance in the whole phosphoproteome IceLogo reflects the contribution of numerous kinases, either Pro-directed or basophilic, the overrepresentation of acidic residues at nearly all positions, with special reference to those downstream from the target amino acid, is likely to be accounted for almost exclusively by an individual, very pleiotropic kinase, CK2. This point of view is strongly supported by the shape and composition of the IceLogo of CK2, whose upper panel, composed almost exclusively of Glu and Asp residues, permeate the one of the whole phosphoproteome. It should be emphasized that CK2 is the only pleiotropic Ser/Thr kinase whose targeting is critically determined by carboxylic side chains downstream, with the one at position n+3 being nearly essential (18). Consequently, at variance with the majority of protein kinases, reliable determinations of CK2 activity can be performed in crude extracts using highly selective peptide substrates which are not appreciably affected by any other protein kinases (19). Another acidophilic kinase critically relying on the recognition of acidic residues downstream is the GCK whose consensus is S-x-E (20). GCK is still an 'orphan' enzyme whose gene(s) and structure are unknown and is not a member of the known 'kinome'. It is possible, however, that its contribution to the phosphoproteome is not negligible, although it is masked by that of CK2: it should be noted in this regard that at position n+2 in the CK2 iceLogo Asp predominates over Glu, whereas the opposite applies to the whole phosphoproteome iceLogo, suggesting the contribution of (a) kinase(s) which, unlike CK2, prefer Glu over Asp at this position. This is indeed a typical feature of GCK (6, 21), as also highlighted by the iceLogo drawn from the known phosphosites generated by this kinase (Figures 4 and 6). Indeed, as outlined below more than 4000 phosphosites (out of 35679) display the G-CK consensus (S-x-E) but not that of CK2.

Even with this limit, CK2 remains by far the main individual contributor of acidic phosphosites found in the phosphopeptides database, confirming our previous estimate that, individually considered, this kinase is the most pleiotropic

one, being possibly responsible for the generation of 20% or so of the eukaryotic phosphoproteome (12).

Conclusions

The weblogo analyses presented here, performed on a database of 45641 phosphosites now available, provide a striking *a posteriori* revaluation of some concepts that were proposed in the past when only a minor proportion of the kinome was known and the substrate specificity of individual kinases was generally inspected with the aid of artificial peptide substrates. Indeed, an even cursory scrutiny of Figures 1, 2 and 5 substantiate the 'old' concept that, at variance with Ser/Thr phosphorylation, Tyr phosphorylation only marginally relies on well defined consensus sequences recognized by individual tyrosine kinases, instead being mainly dictated by features not residing within the phosphoacceptor site. In contrast, the generation of phosphorylated serines and threonines clearly reflects the contribution of individual categories of Ser/Thr protein kinases whose consensus can be neatly appreciated within the global weblogo of the Ser/Thr phosphoproteome. This not only underscores the relevance of already known local specificity determinants for substrate recognition by Ser/Thr protein kinases but also supports the view that pleiotropic protein kinases displaying as yet not deciphered consensus sequences provide an only marginal contribution to the generation of the phosphoproteome. Indeed, only six residues are positively selected around phosphorylated Ser and Thr and all of these were known to act as key recognition elements for the four categories of Ser/Thr protein kinases grouped according to their specificity determinants, namely basophilic, Pro-directed, phosphate-directed (or 'primed') and acidophilic (6). As summarized in Figure 7A, a large proportion of the 35679 Ser/Thr phosphosites considered in our analysis display one or more of the motifs specifically recognized either by Pro-directed kinase in general (S/T-P) or by individual pleiotropic kinases of the other categories, notably, PKA (R-R-x-S/T), Akt (R-x-R-x-x-S/T), CK2 (S/T-x-x-E/D), G-CK (S-x-E), GSK3 (S-x-x-x-S/T) and CK1 (S-x-x-S/T). Only 9033 phosphosites (approximately 25%) do not conform to any of these consensus motifs and are therefore likely to be generated by different Ser/Thr kinases. An iceLogo analysis of these latter sites (Figure 7B) suggests a relevant contribution to their generation by kinases that recognize Proline at positions different from n+1 ('non-canonical' pro-directed kinases) and by a variety of basophilic protein kinases other than PKA and Akt. These basophilic kinases account for the persistent positive selection of Arg (and to a letter extent Lys) at upstream positions even after all the putative PKA and Akt phosphosites have been removed. Among these a conspicuous implication of PKC and CaMK would be consistent with the positive selection of Lys and Arg at position n+3 and of Asp at position n+2, respectively, which is typical of these two kinases (Figure 4). The only entirely new positively selected residue disclosed by this analysis is Gln at position n+1: this is a hallmark of ATM/ATR and DNAPK (22) which becomes evident only after the overwhelming positive selection of Pro at the same position has been removed.

Acknowledgments

This research was supported by the University of Padova (Progetto Ateneo 2008 to M.S.) and by the European Commission (PROKINASERESEARCH 503467), the Italian Cystic Fibrosis Research Foundation (Grant FFC#4/2009) with the contribution of Loifur s.r.l. and 'Amici per la Ricerca '09', and by the Associazione Italiana per la Ricerca sul Cancro (AIRC) to L.A.P.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science 2002; 298: 1912–34.
2. Cohen P. The regulation of protein function by multisite phosphorylation – a 25 year update. Trends Biochem Sci 2000; 25: 596–601.
3. Yachie N, Saito R, Sugahara J, Tomita M, Ishihama Y. In silico analysis of phosphoproteome data suggests a rich-get-richer process of phosphosite accumulation over evolution. Mol Cell Proteomics 2009; 8: 1061–71.
4. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B. PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics 2004; 4: 1551–61.
5. Kemp BE, Pearson RB. Protein kinase recognition sequence motifs. Trends Biochem Sci 1990; 15: 342–6.
6. Pinna LA, Ruzzene M. How do protein kinases recognize their substrates? Biochim Biophys Acta 1996; 1314: 191–225.
7. Ubersax JA, Ferrel JE Jr. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 2007; 8: 530–41.
8. Toppo S, Pinna LA, Salvi M. Matching up phosphosites to kinases: a survey of the available predictive programs. Curr Bioinform 2010; 5: 141–52.
9. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004; 14: 1188–90.
10. Biondi RM, Nebreda AR. Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. Biochem J 2003; 372: 1–13.
11. Shi Z, Resing KA, Ahn NG. Networks for the allosteric control of protein kinases. Curr Opin Struct Biol 2006; 16: 686–92.
12. Salvi M, Sarno S, Cesaro L, Nakamura H, Pinna LA. Extraordinary pleiotropy of protein kinase CK2 revealed by weblogo phosphoproteome analysis. Biochim Biophys Acta 2009; 1793: 847–59.
13. Vacic V, Iakoucheva LM, Radivojac P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 2006; 22: 1536–7.
14. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. Nat Methods 2009; 6: 786–7.
15. Diella F, Gould CM, Chica C, Via A, Gibson TJ. Phospho.ELM: a database of phosphorylation sites – update 2008. Nucleic Acids Res 2008; 36: D240–4.
16. Roach PJ. Multisite and hierarchal protein phosphorylation. J Biol Chem 1991; 266: 14139–42.
17. Donella-Deana A, Marin O, Cesaro L, Gunby RH, Ferrarese A, Coluccia AM, Tartari CJ, Mologni L, Scapozza L, Gambacorti-

Passerini C, Pinna LA. Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. Biochemistry 2005; 44: 8533–42.
18. Meggio F, Marin O, Pinna LA. Substrate specificity of protein kinase CK2. Cell Mol Biol Res 1994; 40: 401–9.
19. Ruzzene M, Pinna LA. Assay of protein kinases and phosphatases using specific peptide substrates. In: Hardie DG, editor. Protein phosphorylation: a practical approach. Oxford: Oxford University Press, 1999: 221–53.
20. Meggio F, Boulton AP, Marchiori F, Borin G, Lennon DP, Calderan A, Pinna LA. Substrate-specificity determinants for a membrane-bound casein kinase of lactating mammary gland. A study with synthetic peptides. Eur J Biochem 1988; 177: 281–4.
21. Lasa-Benito M, Marin O, Meggio F, Pinna LA. Golgi apparatus mammary gland casein kinase: monitoring by a specific peptide substrate and definition of specificity determinants. FEBS Lett 1996; 382: 149–52.
22. O'Neill T, Dwyer AJ, Ziv Y, Chan DW, Lees-Miller SP, Abraham RH, Lai JH, Hill D, Shiloh Y, Cantley LC, Rathbun GA. Utilization of oriented peptide libraries to identify substrate motifs selected by ATM. J Biol Chem 2000; 275: 22719–27.