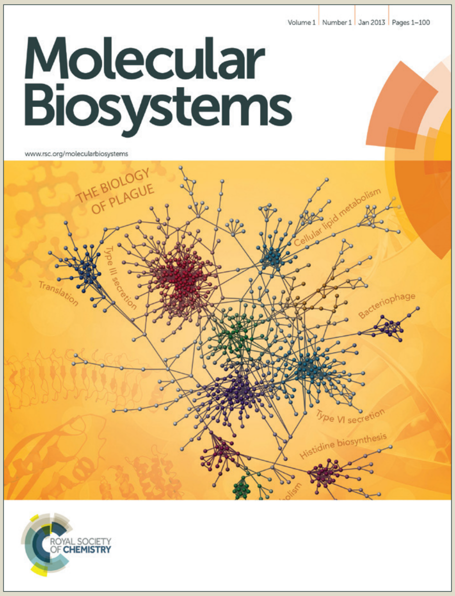


THE GENERATION OF PHOSPHOSERINE STRETCHES IN PHOSPHOPROTEINS: MECHANISM AND SIGNIFICANCE

Luca Cesaro and Lorenzo A. Pinna

Correspondence to: lorenzo.pinna@unipd.it

Department of Biomedical Sciences, University of Padova, Via Ugo Bassi 58B, 35131 Padova, Italy. Tel.: + 39 049 8276108; fax: + 39 049 8073310.

Synopsis

In the infancy of studies on protein phosphorylation the occurrence of clusters of three or more consecutive phosphoseryl residues in secreted and in cellular phosphoproteins was reported. Later however, while the reversible phosphorylation of Ser, Thr and Tyr residues was recognized to be the most frequent and general mechanism of cell regulation and signal transduction, the phenomenon of multi-phosphorylation of adjacent residues was entirely neglected. Now-a-days, in the post-genomic era, the availability of large phosphoproteomics database makes possible a comprehensive re-visitation of this intriguing aspect of protein phosphorylation, aimed at shedding light on both its mechanistic occurrence and its functional meaning. Here we describe an analysis of the human phosphoproteome disclosing the existence of more than 800 rows of 3 to >10 consecutive phosphoaminocids, composed almost exclusively of phosphoserine, while clustered phosphothreonines and phosphotyrosines are almost absent. A scrutiny of these phosphorylated rows supports the conclusion that they are generated through the major contribution of a few hierarchical protein kinases, with special reference to CK2. Also well documented is the combined intervention of CK1 and GSK3, the former acting as priming and primed, the latter as primed kinase. The by far largest proportion of proteins containing $(pS)_n$ clusters display a nuclear localization where they play a prominent role in the regulation of transcription. Consistently the molecular function of the by far largest majority of these proteins is ability to bind other macromolecules and/or nucleotides and metal ions. A “String” analysis performed under stringent conditions reveals that >80% of them are connected to each other by physical and/or functional links, and that this network of interactions mostly take place at nuclear level.

Background

According to our present knowledge almost three quarters of the human proteome is composed by phosphoproteins, a figure which is expected to further increase in the future.[1] The relative abundance of phosphorylated aminoacids is more or less the same estimated in early studies after the discovery of phosphotyrosine, with phosphoserine by far predominant (81%) over phosphothreonine (15%) and phosphotyrosine (just 4%). The total number of phosphosites presently reported in the PhosphoSitePlus database[2] is 10-fold larger than the number of phosphoproteins, reflecting the fact that singly phosphorylated proteins, once believed to be the rule, are rather the exception (just 15%, see Fig. 1) the large majority of phosphoproteins being multi-phosphorylated. Indeed a large number of phosphoproteins (26%) include more than 12 phosphoresidues in their sequence.

This raises the question as to whether the presence of multiple phosphoresidues in a protein may generate clusters of adjacent phosphorylated aminoacids. The first evidence that rows of 3 or more phosphoseryl residues are present in the secreted storage proteins casein (from milk) and phosvitin (from egg yolk) was provided in 1959 by the Nobel laureate Frederick Sanger by partial acid hydrolysis of these phosphoproteins followed by high voltage paper electrophoresis.[3] By adopting the same procedure few years later Sperti et al were able to show that similar $(pS)_n$ clusters can be isolated also from Ehrlich ascites cells incubated with ^{32}Pi.[4] No attempt however was done at that time to isolate the proteins containing $(pS)_n$ clusters and their mode of generation was not further investigated. During the seventies of the past century however the primary sequence of casein fractions was elucidated revealing that the rows of pS were composed of 3-4 consecutive residues followed by a doublet of glutamic acids: $(pS)_{3\text{-}4}$EE. This observation, in conjunction with a thorough analysis of many other phosphorylated residues in casein fractions led Marcier to postulate that the phosphates were incorporated by a “casein kinase” recognizing the consensus sequence S-x-E/pS, where glutamic acid can be replaced by a phosphoserine (but not by Asp) as specificity determinant.[5] This conclusion was confirmed by in vitro studies with casein kinase preparations from the Golgi apparatus of lactating mammary gland taking advantage of synthetic peptides.[6,7] The genuine Golgi casein kinase (often referred to as G-CK) remained an orphan enzyme until 2012, when it was shown to be identical to Fam20C, an atypical kinase belonging to the small family of Four Jointed (FJ) proteins.[8,9,10] Fam20C/G-CK is now recognized as the kinase responsible for the generation of the largest proportion of the phosphosecretome, by phosphorylating a plethora of secreted proteins at S-x-E/pS motifs.[11] Intriguingly however it doesn’t seem to be able to generate the long rows of phosphoserines present in phosvitin and firstly isolated by Sanger in 1959, since in this protein typical motifs $(pS)_n$-EE are lacking.[12]

On the other hand the rising interest for signalling via protein phosphorylation and, more recently, global phosphoproteomics analyses have disclosed the existence of rows of 3 or more adjacent phosphoresidues in many proteins: only in rare cases the mechanisms by which they are generated are known and almost invariably their biological roles remain undeciphered. While in fact a phosphate incorporated into a single residue is sufficient to alter the properties of a protein, thereby accounting for its responsiveness to a given stimulus, multiple phosphorylation sites in the same protein often make possible the cross-talk between different signalling pathways. However the rationale underlying the necessity that several phosphoresidues are clustered close to each other has not been systematically explored.

An analysis performed on the PhosphoSitePlus database reveals that there are 2893 phosphoserines clustered to form rows of 3 or more adjacent phosphoresidues, belonging to 717 proteins altogether and accounting for about 2.5% of the whole number of phosphoserines found in the human proteome (see Supplemenaty material, Table S1). As summarized in Fig. 2 there are 631 phosphoserine triplets, 199 quadruplets, 27 quintuplets and 22 rows composed of >5 consecutive phosphoserines, a few of which include more than 10 consecutive phosphoseridues. Often moreover additional phosphoserines are flanking these rows, separated by few non phosphorylated residues in between (see Table S1).

Such a situation also reflects in the weblogo of phosphoserine extracted from the PhosphoSitePlus database (Fig. 3A) disclosing the positive selection of additional phospho-serine residues (denoted by black coloured “s”) on both sides. This confirms a tendency of p-Ser residues to cluster close to each other, a feature which is not shared by either p-Thr or p-Tyr residues (see Fig.3, B and C). Intriguingly this trend is accentuated if the weblogo around triplets of phosphoserines is extracted (Fig. 3 D): in this case the selection of phosphoserines (and of serines as well, possibly representing the dephosphorylation product of phosphoserines) around the triplet becomes predominant at every position with a concomitant loss of importance of proline, basic and, to a lesser extent, acidic residues. This suggests that the main determinants underlying the formation of $(pS)_n$ clusters are different from those which specify the recognition of isolated serines by an ample spectrum of protein kinases.

Mechanistic aspects.

A basic question arising from the above observations concerns the mechanism by which such rows of adjacent phosphoserines are generated. In simpler words can we postulate the intervention of a subset of protein kinases particularly well suited to do the job, either individually or in combination, as the weblogos shown in Fig. 3 would suggest?

In principle any protein kinase can contribute to the generation of a $(pS)_n$ cluster provided that one or more of the seryl residues conform to its consensus sequence. However while many protein kinases recognize canonical consensuses that enable them to entirely phosphorylate a pS doublet, only few can generate alone triplets of phosphoserines. One of this is CK2, whose canonical consensus, S/T-x-x-E/D can specify the complete phosphorylation of seryl triplets followed by acidic triplets (S-S-S-E/D-E/D-E/D), and several examples are know where CK2 accomplishes this role (see supplementary material, Table S2, section 2). AKT/PKB, whose consensus is R-x-R-x-x-S/T, also has this potential if a triplet of serines is adjacent to the C terminal end of 5 arginines (R-R-R-R-R-S-S-S) where some arginines can be eventially replaced by lysine. This motif can justify the generation of only one or two of the $(pS)_3$ clusters listed in Table S1: a triplet found in cGMP-inhibited 3',5'-cyclic phosphodiesterase (Q14432), and one located upstream of a very long stretch of phosphoserines, found in Serine/arginine repetitive matrix protein 2 (Q9UQ35). In this latter case the phosphoserine triplet generated by the basophilic kinase can prime the phosphorylation of the whole stretch by CK1 as discussed below (see Table 2).

If now we move to $(pS)_n$ clusters longer than 3 residues, to the best of our knowledge, based on the known consensus sequences of protein kinases,[13,14] there is no kinase able to entirely do the job by itself, unless the mechanism of hierarchical (or "primed") phosphorylation is exploited. This mechanism, based on the observation that the consensus sequence of few protein kinase can be generated by a previously phosphorylated residue,[15] provides the rationale not only for the phosphorylation of long stretches of serines by an individual kinase, but also for the concerted phosphorylation of several clusterd residues by more than one kinase, as exemplified by the paradigmatic case of APC repeat 3.[16,17]

Hierarchical phosphorylation and "primed" kinases.

At least four protein kinases are known whose site recognition can be specified by phosphorylated residue(s) through a mechanism known as "hierarchical" or "primed" phosphorylation (see Table 1). They all are highly pleiotropic enzymes, responsible for the generation of significant proportions of the human phosphoproteome.

Two of them, CK2 (an acronym derived from the misnomer "casein kinase-2"[22]) and the genuine casein kinase (G-CK), recently identified with Fam20C, an atypical kinase responsible for the phosphorylation of a plethora of secreted proteins at S-x-E/pS motifs,[8,11] recognize equally well either carboxylic side chains or phosphorylated residues as specificity determinants to generate canonical non-primed and primed consensus sequences, respectively (see Table 1). In the case of CK2 it has been shown that phosphorylated and carboxylic determinants are not perfectly equivalent, especially when multiple phosphoserines are located at positions different from the very critical one (which in the case of CK2 is n+3).[19,23,24] Especially pertinent to our topic is the observation that in the absence of the crucial determinant at n+3 (either carboxylic or phosphorylated) a seryl residue adjacent to two or three phosphoserines or even embedded between phosphoserine(s) and carboxylic acids can be appreciably phosphorylated by CK2.[19]

On the other hand GSK3 and CK1 (a small branch of the kinome with 7 isoforms in human[25], whose acronym derives from the misnomer "casein kinase 1"[22]) were initially considered obligatory primed kinases, recognizing the consensuses S/T-x-x-x-pS/pT and pS/pT/pY-x-x-(x)-S/T, respectively, where the phosphorylated determinant cannot be efficiently replaced by a carboxylic one of either glutamic or aspartic acids. Later however it became clear that possibly both, and by sure CK1 can also perform non primed phosphorylations, dictated by non canonical consensus sequences whose specificity determinants in some cases remain rather conjectural[20,21,26] (see Table 1).

Sometimes these two kinases operate in a concerted manner where the priming kinase is CK1 and the primed one is GSK3, two well known examples being provided by β-catenin and the APC repeat 3. Interestingly in both cases priming phosphorylation by CK1 is critically dependent on a Leu adjacent to the N terminus of the target serine.[21]

$(pS)_n$ clusters generation: the implication of G-CK/Fam20C and CK2.

The ability of the genuine casein kinase (G-CK/Fam20C) to phosphorylate seryl rows exploiting a combination of non-primed and primed consensus sequences has been known for a long time, dating back to the elucidation of the primary structure of caseins which contain clusters of 3 and 4 consecutive phosphoserines N terminal to glutamic acid doublets.[5] In these cases the two C-terminal glutamic acids dictate the phosphorylation of the last two serines (fulfilling the non primed consensus S-x-E) which in turn primes the phosphorylation of the upstream serine(s) conforming to the hierachical consensus (S-x-pS). In principle G-CK/Fam20C has the potential to exhaustively phosphorylate rows of consecutive serines, no matter how long, provided they are adjacent to the N-terminal side of a Glutamic acid doublet (S_n-EE), as detailed in Scheme 1. To perform the same job, CK2, whose canonical non primed consensus is S/T-x-x-E/D, needs a triplet of either Glu or Asp at the N terminal end of the polyseryl stretch (S_n-E/D-E/D-E/D), as also reported in Scheme 1. If however we also take into account the ability of CK2 to phosphorylate seryl residues devoid of the canonical consensus but adjacent to pS residues, just a couple of acidic residues (either Glu or Asp) may be sufficient also for CK2 to trigger the exhaustive phosphorylation of stretches of serines located upstream.

If we now look at the real situation it appears that only seldom G-CK/Fam20C exploits its potential to phosphorylate clusters of 3 or more serines: a scrutiny of its known targets in fact reveals that most of the residues posphorylated by this kinase, with the notable exception of those present in caseins, are not clustered together. Very telling in this respect may be the example of osteopontin where 22 phosphosites conforming to the Fam20C consensus are concentrated in a relatively short sequence but they never give rise to rows of 3 or more serines. Consistently the weblogo of bona fide Fam20C phosphosites displays no evident selection for phosphoserine except at position n+2 where it shares with Glutamic acid the essential role of specificity determinant (see Fig. 4A). It can be concluded therefore that the majority of the 8 phosphoserine clusters displaying the motif $(pS)_n$-E-E-X (being X neither Glu nor Asp) shown in Table S2 (group 1) are not generated by Fam20C, but possibly by CK2, also considering that only few belong to secreted proteins. Especially telling is the case of the $(pS)_3$EEV motif representing the N terminal autophosphorylation site of the β-subunit of CK2 itself, notoriously generated by CK2 despite it is more reminiscent of a Fam20C site for its lack of a third acidic determinant.

At variance with Fam20C, phosphoserines are abundantly represented in the weblogo of bona fide CK2 phosphosites (Fig. 4B), consistent with the idea that they are often clustered together in the targets of this kinase. In agreement with this prediction 63 clusters of 3 or more adjacent phosphoserines have the signature of typical CK2 sites, specified by C terminal triplets composed either by E/D-E/D-E/D (Table S2, group 2) or E/D-E/D-pS (group 3A), or E/D-pS-E/D (group 3B), or E/D-pS-pS (group 3C). Their length ranges from 3 to 8 consecutive phosphoserines, often flanked by additional phosphoresidues nearby, also conforming to the CK2 consensus. If however non canonical priming is also considered, determining the appreciable phosphorylation of serines embedded between two or three phosphoserines (pS-S-pS, pS-S-pS-pS) though lacking the canonical determinant at position n+3),[19] 23 additional clusters of 3 to 6 phosphoserines can be entirely attributed to CK2 (Table S2 group 4).

Finally there are more than 260 additional clusters, among those not entirely attributable to CK2, that include serine(s) which individually have the signature of CK2 sites and may therefore represent partial contributions of CK2 to the generation of $(pS)_n$ stretches. These are listed in Table S3 of supplementary materials.

Therefore in more than 350 cases the $(pS)_n$ clusters listed in Table S1 can be considered as generated with the exclusive or partial contribution of CK2.

Somewhat surprisingly none of the individual phosphoserines composing the clusters (2875 altogether) conform to the ideal substrate consensus for multisite CK2 phosphorylation ([DE]-S-pS-[ADE]-pS-[DEH]-[DEpS]) as defined by St Denis et al.,[24] while six clustered phosphoserines display optimal sequence for proline directed CK2 hierarchical phosphorylation (S-X-X-[D/E]-X-pS-P).[24] Two of these belong to clusters listed in Table S2, groups 3C and 4, respectively, while the others are among those listed in Table S3.

In several instances the actual implication of CK2 in the generation of $(pS)_n$ clusters is well documented, a few notable examples being provided by clusters found in the β-subunit of CK2 itself, PHLP, PIAS 1, PML, CTDP1, DEK and XRCC1. Especially remarkable are the stretches of the β-subunit of CK2 and of CTDP1 known to be entirely phosphorylated by CK2 despite they display non canonical consensuses. Somewhat unexpectedly moreover a few clusters of 3-4 phosphoserines are reported among the bona fide CK2 substrates (e.g. RsssGSWGN in Transcription factor 4 and RssssESSHSs in Ribosomal protein S6 kinase alpha-5) although they don’t display the features of CK2 targets, and they are not included in Table S2.

Collectively taken our data indicate that CK2 is the kinase whose implication in the generation of $(pS)_n$ clusters is most frequently and convincingly documented.

$(pS)_n$ clusters generation: the implication of CK1 and GSK3.

Both CK1 and GSK3 can work in a hierarchical manner (see Table 1) and they have been often reported to operate in concert. It is conceivable therefore that they might contribute to the generation of $(pS)_n$ clusters. The weblogos of both kinases, drawn from the repertoires of their bona fide targets are shown in Fig. 4, C and D, respectively. While the weblogo of GSK3 corroborates the concept that it typically is a primed kinase whose site recognition is critically dictated by a phosphorylated residue at position n+4 (reinforced by a proline at n+1), the weblogo of CK1 suggests a mixed origin of its phosphosites, generated both through hierarchical and non-hierarchical mechanisms: the selection of phosphoserines spaced by 2 residues between each other is symptomatic of primed phosphorylation; conversely the selection of a Leu at position n+1 and acidic residues both upstream and downstream are consistent with non primed, non canonical consensuses. These are grounded either on acidic side chain(s) surrogating the phospho-residue at n-3 or on the motif SL-$X_{3\text{-}6}$-$(D/E)_n$[21] (see Table 1). Also to note is the “crowding” of adjacent phosphoserines in the CK1 weblogo, suggestive of the participation of phosphoserines generated by this kinase in $(pS)_n$ clusters.

An elegant and paradigmatic example of how a $(pS)_3$ cluster is generated by the combined intervention of CK1 and GSK3 is provided by the APC repeat 3[16,27] where a triplet of serines flanked by tree additional isolated serines is entirely phosphorylated through a “ping-pong” hierarchical mechanism whose 5 steps have been chemically dissected,[17] as illustrated in Fig. 5. As in the case of β-catenin S45, also here the priming kinase is CK1 which phosphorylates the third serine of the triplet (S1505) displaying the non canonical consensus S-L-X_5-D-E. This primes the phosphorylation by GSK3 of S1501, which is slightly outside the triplet but in turn primes the phosphorylation of the second serine of the triplet (S1504) by CK1, acting now as a primed kinase. Once phosphorylated, pS1504 promotes the phosphorylation of S1507, outside the C-terminal end of the triplet, by CK1; this phosphorylation, on one side primes the phosphorylation by CK1 of S1510 (whose phosphorhylation by CK1 is also partially accomplished in a non hierarchical manner, thanks to its SL motif) on the other primes the phosphorylation of S1503 by GSK3, thus completing the phosphorylation of the triplet.

A pattern almost identical to the one leading to the phosphorylation of APC repeat 3 seemingly accounts for the exhaustive phosphorylation of the CASC5 pS triplet (Fig. 5B), although in this case the phosphorylation mechanism has not been investigated in detail. This observation corroborates the possibility that analogous interplays between CK1 and GSK3, initiated by the priming phosphorylation of a SL motif by CK1, may be also responsible for the generation of other $(pS)_n$ stretches among those listed in Table S1. Pertinent to this may be the analysis of sequences harbouring $(pS)_n$ clusters (see Table S1): in 120 cases these sequences include also the motif pSL and their weblogo is shown in Fig. 3E. This highlights the tendency of the pSL motif to occupy positions downstream from the $(pS)_n$ stretches, where it can prime upstream phosphorylation events as observed in the cases of APC and CASC5 (see above).

It should be also noted that owing to their ability to perform hierarchical phosphorylations in opposite directions and with different spacing between phosphoresidues, CK1 and GSK3 look ideally suited to saturate with phosphate long poly-serine stretches once their intervention is primed by another kinase acting at one of the two extremities of the row. A mechanism like this could account for the formation of the three longest stretches of phosphoserines retrieved in the PhosphoSitePlus database,which are shown in Table 2.

In these cases, the triggering event could be the phosphorylation of one or more N-terminal serines by Akt, PKA and/or other basophilic kinase, whose targeting is specified by a polybasic motif located upstream. Once phosphorylated these serines will start a downstream cascade of phosphorylation events catalyzed by CK1, affecting all serines located at n+3 positions relative to those already phosphorylated until the C terminus is reached. To note that in the case of the very long stretch of protein SRRM2 CK1 alone can phosphorylate the whole cluster, as the priming phosphoserines potentially generated by the basophilic kinase(s) are three: KRKRRpSpSpS…, sufficient to trigger the hierarchical phosphorylation of all the > 30 serines down-stream. In the other two cases instead, AR6P4 and PPRC1, it is expectable that the basophilic kinase(s) will phosphorylate only 2 and one of the serines adjacent to the cluster of basic determinants, respectively (blue shadowed in Table 2). Consequently CK1 can only partially phosphorylate these two stretches of serines, proceeding downwards, skipping those not in the right n+3 position. All or part of these can be then phosphorylated by GSK3, primed by the phosphates incorporated by CK1 at n+4 positions and proceeding backward. Additional rounds of hierarchical phosphorylation events will finally saturate with phosphate the gaps still present in the PPRC1 stretch. The cooperation between CK1 and GSK3 in the generation of $(pS)_n$ clusters, as illustrated in Fig. 5 and Table 2, corroborates the concept that a cross-talk often takes place between these two kinases, a typical example being provided by the Wnt signalling where CK1primes the intervention of GSK3. A similar cross-talk between CK2 and GSK3, documented by the hierarchical phosphorylation of glycogen synthase at its sites 5 and 3 [15], does not appear to significantly contribute to the generation of $(pS)_n$ clusters.

Biological functions of $(pS)_n$ clusters.

Very little is known about the biological role of poly-phosphoserine rows. In the few cases where detailed information is available it appears that rather than regulating the biological activity of the proteins they belong to, these negatively charged stretches are critical for determining and/or tuning the interactions with other macromolecules. This is the case, e.g. of the phosphoserines clustered in the repeat 3 of APC,[27] generated by the concerted activity of CK1 an GSK3 (see above, Fig 5) and of those generated by CK2 in the transcription factor DEK[28] and in the scaffold protein XRCC1.[29] These clusters have been shown to tighten the interaction with β-catenin, to weaken binding of DNA, and to enable the assembly of DNA single strand break repair protein complexes, respectively. In the case of PML instead, the phosphorylation of its triplet of serines by CK2 is required to commit the protein to degradation.[30]

The by far great majority of the $(pS)_n$ clusters listed in Table S1 however were identified in the course of phosphoproteomics analyses, and no functional studies are available about their possible biological significance. To get a global insight into this aspect therefore we have performed comprehensive analyses of the subcellular localization, molecular functions and biological implications of those proteins where $(pS)_n$ rows are known to be present, bearing in mind that this kind of analysis does not provide the proof of concept that the observed behaviour of the proteins is necessarily conferred by the presence of the $(pS)_n$ clusters in them.

Subcellular localization

A gene-ontology analysis of the subcellular localization of the 687 proteins which include phosphoseryl clusters listed in Table S1 is summarized in Fig. 6A. It can be seen that the majority of these proteins have a nuclear localization. An over-representation of these proteins in the nuclear compartments is also supported by a BiNGO GO-slim[31] analysis (Fig. 6B orange colour) thus confirming an observation published more than 50 years ago by Sperti et al.[4] reporting that the ^{32}P-radiolabeled phosphoserine rows isolated from Erhlich ascites cells incubated with ^{32}P-ATP were predominantly present in the nuclear fraction. The GO-slim analysis also allows to conclude that proteins including in their sequences phosphoserine clusters are under-represented in membranes and extracellular regions (Fig. 6C, blue colour), consistent with the observation discussed above (also see Table S2 Group 1) that Fam20C, the genuine casein kinase committed to the phosphorylation of secreted proteins, plays an only marginal role in the generation of $(pS)_n$ clusters.

Molecular Functions and Biological Implications.

As shown in Fig. 7A ability to interact with other macromolecules and/or with nucleotides and metal ions appears to be a common denominator of nearly all the proteins harbouring phosphoserine clusters in their sequence. Almost half of them interact with other proteins but a large number are also endowed with the ability to bind nucleic acids, nucleotides and metal ions. In accordance with these molecular functions it appears that the biological processes where proteins that contain phosphoserine clusters are mainly implicated fall in the category of regulation of gene expression with special reference to transcription (Fig. 7B).

Interestingly a String[32] analysis performed with high stringency parameters reveals that the great majority of proteins with phosphoserine stretches (567 out of 687) are interconnected by physical and/or functional links (Fig. 8), suggesting that such clusters evolved to ensure their participation in a limited number of specialized roles. These are likely to deal with gene expression and other nuclear functions as judged from their biological activity (see Fig. 7B) and the nuclear localization of the majority of the interacting proteins (denoted by red colour in Fig. 8).

Given these premises it would be interesting to know how many of these proteins may be implicated in cell transformation and tumorigenesis. To this purpose advantage has been taken of a database including about 3000 oncogenes. 127 of these are also found in our repertoire of proteins containing $(pS)_n$ clusters listed in Table S1. A similar cross-check with a database of 717 oncosuppressor genes provided 49 positive matches. These data suggest that many oncogenes and oncosuppressors encode proteins equipped with $(pS)_n$ clusters which may be instrumental to their functional interactivity/plasticity.

OUTLOOK

The existence of clusters of 3 or more adjacent phosphoseryl residues was already reported in the infancy of studies on protein phosphorylation.[3,4,5] However no systematic investigation of this phenomenon, aimed at assessing its frequency, mechanistic aspects and possible functional significance, was performed to date. From time to time the growing literature on signal transduction via protein phosphorylation reported the sporadic occurrence of phosphoresidues adjacent to each other, but in general this was not considered worthy of special attention, falling in the generic huge and quickly growing repertoire of phosphosites responsive to an ample spectrum of stimuli.

Now-a-day, in the post-genomic era, thanks to the spectacular progress done in the fields of mass spectrometry and of proteomics in general, almost exhaustive repertoires of phospho-sites are available that allow to address also the issue of phosphorylated clusters in a global manner. Indeed for the time being, while the proportion of phosphorylated proteins approaches 75% of total, it is clear that only a minority of these are mono-phosphorylated, the majority including 2 to up to several dozens phosphoresidues.[1] It is generally held, and well documented in many cases, that such multi-phosphorylation represents a device by which distinct signalling pathways can cross-talk with each other and/or commit the target proteins to distinct functions/fates. But this seems not to be the case occurring in the phosphoproteins where phosphoresidues are clustered together, rather suggesting that these polyanionic structural elements confer special interacting and binding properties to the protein they belong to.

Our functional analysis corroborates this concept by showing that a common denominator of the proteins harbouring phosphoserine clusters is ability to interact with other macromolecules (either proteins or nucleic acids) and/or to bind cations and small molecules. Whether and to what extent such polyanionic stretches are instrumental to the interacting properties of the proteins they belong to, remains a matter of investigation. Pertinent to this may be the observation that phosphorylated clusters are almost exclusively composed of phosphoserines, with the exclusion of phosphothreonine and phosphotyrosine (see weblogos in Fig. 3). Interestingly moreover most of the interactions take place in the nuclear compartment, in accordance with a functional analysis revealing the massive implication of these proteins in the regulation of gene expression and transcription. Not surprisingly a substantial number of these proteins also fall in the category of oncogenes and/or tumor suppressors. Of special interest is the outcome of a string analysis, disclosing the participation of the majority of the proteins harbouring phosphoserine clusters in a tight unique network of physical and functional interactions, suggesting that they may compose an up to now uncharacterized cellular entity.

The mechanism by which phosphoserine clusters are generated has been partially unravelled by our analysis. Based on the legitimate assumption that only a handful of protein kinases recognize consensuses enabling them to phosphorylate 3 adjacent serines and that the phosphorylation of 4 or more serines in a row requires a hierarchical mechanism by one or more kinases in combination, we can conclude that 3 kinases are majorly implicated in the generation of phosphoserine clusters, with the occasional participation of few other kinases that may prime the intervention of a hierarchical one. The combined implication of CK1 and GSK3, soundly documented in a couple of cases, is likely to account for the phosphorylation of several other analogous clusters, including some of the longest phosphoserine stretches retrieved in the database, though this remains only conjectural for the time being.

The lion’s share however appears to be that of CK2, whose implication in the generation of hundreds clusters ranging between 3 to 9 phosphoserines is resting on published data and/or coincidental evidence. To perform this job CK2 often exploits its canonical non primed consensus, enabling it to generate phosphoserine triplets adjacent to the N terminal side of acidic triplets (E/D-E/D-E/D). But in the case of serine stretches longer than 3 residues and/or not entirely conforming to its non primed consensus, its targeting is often occurring hierarchically, being dictated by phosphorylated motifs conforming to its canonical or non canonical consensuses. In contrast the optimal consensus sequence for hierarchical multisite CK2 phosphorylation and for proline directed CK2 hierachical phosphorylation as recently defined by a peptide array approach,[24] appear to play an only marginal role in the generation of $(pS)_n$ clusters. The predominant nuclear location of CK2, moreover, and its know implication in the regulation of gene expression, fit very well with the subcellular distribution and functional analysis of proteins harbouring phosphoserine clusters (see above).

By sharp contrast Fam20C, i.e. the genuine casein kinase sharing with CK2 a similar consensus and responsible for the first $(pS)_{3\text{-}4}$ clusters ever described in the literature, seems to be only seldom involved in the generation of phosphoserine stretches: this is clear both looking at the paucity of clusters conforming to its consensus (Table S2, group 1) and its weblogo (Fig. 4A) where phosphoserine is almost absent except in the n+2 position where it can replace Glu as specificity determinant. Consistent with this, proteins with phosphoserine clusters are under-represented in membranes and extracellular compartments (Fig. 6C) where Fam20C, secreted as an active kinase in the Golgi apparatus, is expected to reside, and plays its role as the main generator of the phosphosecretome.[11]

In summary the implication of all the four known hierarchical protein kinases, CK1, CK2, GSK3 and Fam20C in the generation of $(pS)_n$ clusters is well documented and likely to account for a large proportion of the clusters retrieved in the PhosphoSitePlus database. The individual contribution of these four kinases however appears to be quite uneven. On one side the role played by CK2 is both massive and amply documented (see Table S2); on the other, although there are several clues and coincidental evidences supporting an important role played by CK1 and GSK3 as well, either alone or in combination, only in few instances their actual implication has been unambiguously demonstrated. It should be born in mind in this respect that the by far largest majority of $(pS)_n$ clusters retrieved in the PhosphoSitePlus database were detected in the course of large scale phosphoproteomics analyses without any parallel mechanistic or functional information about their generation. So their assignment to the intervention of individual kinases is mostly grounded on our knowledge of the consensuses of individual protein kinases, which is still largely incomplete. Pertinent to this may be the observation that only in the case of about 20% of protein kinases a sufficient number of phosphosites (20 or more) are known, to allow to extract a weblogo or at least to figure out a reliable consensus. In some cases non hierachical kinases are likely to trigger the intervention of hierarchical ones, as illustrated in Fig. 3 where phosphorylation is initiated at the N terminus by basophilic kinases, priming the subsequent intervention of the hierarchical kinases CK1 and GSK3. But it is also expectable that several $(pS)_n$ clusters are partially generated by non hierarchical kinases hitting C terminal, rather than N terminal, serine(s) when these are specified e.g. by the S-P or S-Q motifs or by basic residues downstream, recognized by proline-directed, ATM/ATR kinases and by many PKC isoforms, respectively. One or the other of these features are recurrent at 243 clusters listed in Table S1. This initial C-terminal phosphorylation could prime the intervention of hierarchical kinases able to propagate the phosphorylation backward, notably CK2, GSK3 and Fam20C.

The possibility should be also explored that long stretches of polyserine might be recognized per se by hierarchical kinases, thus undergoing a preliminary phosphorylation at one or few residues (“phospho seeding”), sufficient to trigger the fast and exhaustive phosphorylation of the remaining cluster by a hierarchical mechanism. To note that in the case of egg yolk phosvitin, i.e. the most heavily phosphorylated protein known, the typical S_nEE motifs which make casein susceptible to multiphosphorylation by Fam20C (see Scheme 1) are absent, rising the possibility that “phosvitin kinase(s)” responsible for the in vivo phosphorylation of this protein are still missing.[12] The same may apply to the generation of similar stretches of phosphoserines whose priming signature is not evident, in the human proteome.

In conclusion while the generation of a large proportion (50% or so) of the known $(pS)_n$ clusters can be accounted for, either unambiguously or with fairly good confidence, to the four hierarchical kinases, either alone or primed by other kinases, the mechanism governing the formation of the other clusters remains a matter of conjecture. A more comprehensive knowledge of the consensus sequences recognized by protein kinases and biochemical studies focusing on individual examples of $(pS)_n$ clusters, will shed light on this controversial issue.

Another intriguing question concerns the turnover of the phosphate incorporated into the $(pS)_n$ clusters. Is this reaction reversible as it is expected to be in situations where phosphorylation is a transient event occurring in response to specific stimuli and representing a regulatory device? Should this be the case, are there protein phosphatases specifically committed to the dephosphorylation of $(pS)_n$ clusters? And, are all the clustered phosphates wiped off simultaneously or gradually, thus possibly conferring to the phosphatases a tuning potential which is lacking in the case of singly phosphorylated sites? The prototype protein with $(pS)_n$ clusters, casein, provides no information in this respects, being casein a secreted, storage phosphoprotein. In a few other cases, e.g. the pS triplet of PML, phosphorylation correlates with protein degradation and they may fall in the category of phosphodegrons which sometimes are specified by sequential multiphopshorylation events.[33] But no information is available about the fate of the majority of $(pS)_n$ clusters known to date. Their variable responsiveness to stimuli and/or to changes in metabolic conditions remains a matter of conjecture. A valuable tool to investigate this issue would be provided by specific antibodies able to recognize and quantify stretches composed by 3 or more phosphoseryl residues, but such reagents are still missing.

Aknowledgements

The financial support of AIRC (Grant IG 14180) is gratefully acknowledged.

References

1 K. Sharma, R.C.J. D'Souza, S. Tyanova, C. Schaab, J.R. Wiśniewski, J. Cox and M. Mann, Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling, *Cell Rep.*, 2014, **8**, 1583–1594.

2 P.V. Hornbeck, B. Zhang, B. Murray, J.M. Kornhauser, V. Latham and E. Skrzypek, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations, *Nucleic Acids Res.*, 2015, **43**:D512-20.

3 J. Williams and F. Sanger, The grouping of serine phosphate residues in phosvitin and casein, *Biochim. Biophys. Acta*, 1959, **33**, 294-296.

4 S. Sperti, M. Lorini, L.A. Pinna and V. Moret, Phosphorylserine sequences in phosphoproteins from Ehrlich ascites cells, *Biochim. Biophys. Acta*, 1964, **82**, 476-480.

5 J.C. Mercier, Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases, *Biochimie*, 1982, **63**, 1-17.

6 F. Meggio, A.P. Boulton, F. Marchiori, G. Borin, D.P. Lennon, A. Calderan and L.A. Pinna, Substrate-specificity determinants for a membrane-bound casein kinase of lactating mammary gland. A study with synthetic peptides, *Eur. J. Biochem.*, 1988, **177**, 281-284.

7 F. Meggio, J.W. Perich, H.E. Meyer, E. Hoffmann-Posorske, D.P. Lennon, R.B Johns and L.A. Pinna, Synthetic fragments of beta-casein as model substrates for liver and mammary gland casein kinases, *Eur. J. Biochem.*, 1989, **186**, 459-464.

8 V.S. Tagliabracci, J.L. Engel, J. Wen, S.E. Wiley, C.A. Worby, L.N. Kinch, J. Xiao, N.V. Grishin and J.E Dixon, Secreted kinase phosphorylates extracellular proteins that regulate biomineralization, *Science*, 2012, **336**, 1150-1153.

9 G. Lolli, G. Cozza, M. Mazzorana, E. Tibaldi, L. Cesaro, A. Donella-Deana, F. Meggio, A. Venerando, C. Franchin, S. Sarno, R. Battistutta and L.A. Pinna, Inhibition of protein kinase CK2 by flavonoids and tyrphostins. A structural insight, *Biochemistry*, 2012, **51**, 6097-6107.

10 H.O. Ishikawa, A. Xu, E. Ogura, G. Manning and K.D. Irvine, The Raine syndrome protein FAM20C is a Golgi kinase that phosphorylates bio-mineralization proteins, *PLoS One*, 2012, **7**:e42988

11 V.S. Tagliabracci, S.E. Wiley, X. Guo, L.N. Kinch, E. Durrant, J. Wen, J. Cui, K.B. Nguyen, J.L. Engel, J.J. Coon, N. Grishin, L.A. Pinna, D.J. Pagliarini and J.E. Dixon, A single kinase generates the majority of the secreted phosphoproteome, *Cell*, 2015, **161**, 1619-1632.

12 V.S. Tagliabracci, L.A. Pinna and J.E. Dixon, Secreted protein kinases, *Trends Biochem. Sci.*, 2013, **38**, 121-130.

13 L.A. Pinna and M. Ruzzene, How do protein kinases recognize their substrates?, *Biochim. Biophys. Acta*, 1996, **1314**, 191-225.

14 J.A. Ubersax and J.E. Ferrell Jr., Mechanisms of specificity in protein phosphorylation, *Nat. Rev. Mol. Cell. Biol.*, 2007, **8**, 530-541.

15 P.J. Roach, Multisite and hierarchal protein phosphorylation, *J. Biol. Chem.*, 1991, **266**, 14139-14142

16 N.C. Ha, T. Tonozuka, J.L. Stamos, H.J. Choi, W.I. Weis, Mechanism of phosphorylation-dependent binding of APC to beta-catenin and its role in beta-catenin degradation, *Mol. Cell*, 2004, **15**, 511–521.

17 A. Ferrarese, O. Marin, V.H. Bustos, A. Venerando, M. Antonelli, J.E. Allende and L.A. Pinna, Chemical dissection of the APC Repeat 3 multistep phosphorylation by the concerted action of protein kinases CK1 and GSK3, *Biochemistry*, 2007, **46**, 11902-11910.

18 A.M. Brunati, O. Marin, A. Bisinella, A. Salviati and L.A. Pinna, Novel consensus sequence for the Golgi apparatus casein kinase, revealed using proline-rich protein-1 (PRP1)-derived peptide substrates, *Biochem. J.*, 2000, **351**, 765-768.

19 J.W. Perich, F. Meggio, E.C. Reynolds, O. Marin and L.A. Pinna, Role of phosphorylated aminoacyl residues in generating atypical consensus sequences which are recognized by casein kinase-2 but not by casein kinase-1, *Biochemistry*, 1992, **31**, 5893-5897.

20 Z. Songyang, K.P. Lu, Y.T Kwon, L.H. Tsai, O. Filhol, C. Cochet, D.A. Brickey, T.R. Soderling, C. Bartleson, D.J. Graves, A.J. DeMaggio, M.F. Hoekstra, J. Blenis, T. Hunter and L.C. Cantley, A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1, *Mol. Cell. Biol.*, 1996, **16**, 6486-6493.

21 O. Marin, V.H. Bustos, L. Cesaro, F. Meggio, M.A. Pagano, M. Antonelli, C.C. Allende, A.L. Pinna and J.E. Allende JE, A noncanonical sequence phosphorylated by casein kinase 1 in beta-catenin may play a role in casein kinase 1 targeting of important signaling proteins, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 10193-10200.

22 A. Venerando, M. Ruzzene, L.A Pinna, Casein kinase: the triple meaning of a misnomer, *Biochem. J.*, 2014, **460**, 141-156.

23 F. Meggio, O. Marin and L.A. Pinna, Substrate specificity of protein kinase CK2, *Cell. Mol. Biol. Res.*, 1994, **40**, 401-409.

24 N. St-Denis, M. Gabriel, J.P. Turowec, G.B Gloor, S.S. Li, A.C. Gingras and D.W. Litchfield, Systematic investigation of hierarchical phosphorylation by protein kinase CK2, *J. Proteomics*, 2015, **118**, 49-62.

25 U. Knippschild, M. Krüger, J. Richter, P. Xu, B. García-Reyes, C. Peifer, J. Halekotte, V. Bakulev and J. Bischof, The CK1 Family: Contribution to Cellular Stress Response and Its Role in Carcinogenesis, *Front. Oncol.*, 2014, **4**:96, doi: 10.3389/fonc.2014.00096

26 H. Okamura, C. Garcia-Rodriguez, H. Martinson, J. Qin, D.M. Virshup and A. Rao, A conserved docking motif for CK1 binding controls the nuclear localization of NFAT1, *Mol. Cell. Biol.*, 2004, **24**, 4184-4195.

27 Y. Xing, W.K. Clements, I. Le Trong, T.R. Hinds, R. Stenkamp, D. Kimelman and W. Xu, Crystal structure of a beta-catenin/APC complex reveals a critical role for APC phosphorylation in APC function, *Mol. Cell*, 2004, **15**, 523-533.

28 F. Kappes, C. Damoc, R. Knippers, M. Przybylski, L.A. Pinna and C. Gruss, Phosphorylation by protein kinase CK2 changes the DNA binding properties of the human chromatin protein DEK, *Mol. Cell. Biol.*, 2004, **24**, 6011-6020.

29 J.L. Loizou, S.F. El-Khamisy, A. Zlatanou, D.J. Moore, D.W. Chan, J. Qin, S. Sarno, F. Meggio, L.A. Pinna and K.W. Caldecott, The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks, *Cell*, 2004, 117, 17-28.

30 P.P. Scaglioni, T.M. Yung, S. Choi, C. Baldini, G. Konstantinidou and P.P. Pandolfi, CK2 mediates phosphorylation and ubiquitin-mediated degradation of the PML tumor suppressor, *Mol. Cell. Biochem.*, 2008, **316**, 149-54.

31 S. Maere, K. Heymans and M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, *Bioinformatics*, 2005, **21**, 3448-3449.

32 D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen and C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.*, 2015, **43**, D447-452

33 N.A. Lyons, B.R. Fonslow, J.K. Diedrich, J.R. Yates 3rd and D.O. Morgan, Sequential primed kinases create a damage-responsive phosphodegron on Eco1, *Nat. Struct. Mol. Biol.*, 2013, **20**, 194-201.

34 T.D. Schneider and R.M. Stephens, Sequence logos: a new way to display consensus sequences, *Nucleic Acids Res.*, 1990, **18**, 6097-6100.

35 G.E. Crooks, G. Hon, J.M. Chandonia and Brenner SE, WebLogo: a sequence logo generator, *Genome Res.*, 2004, **14**, 1188-1190.

36 P. Carmona-Saez, M. Chagoyen, F. Tirado, J.M. Carazo and A. Pascual-Montano, GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists, *Genome Biol.*, 2007, **8**, R3.

37 R. Nogales-Cadenas, P. Carmona-Saez, M. Vazquez, C. Vicente, X. Yang, F. Tirado, J.M. Carazo and A. Pascual-Montano, GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information, *Nucleic Acids Res.*, 2009, **37**, W317-322.

38 D. Tabas-Madrid, R. Nogales-Cadenas and A. Pascual-Montano, GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics, *Nucleic Acids Res.*, 2012, **40**, W478-483.

Table 1 “Primed” Protein kinases implicated in hierarchical phosphorylation.

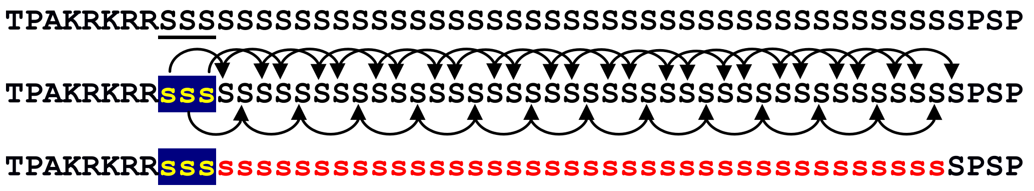
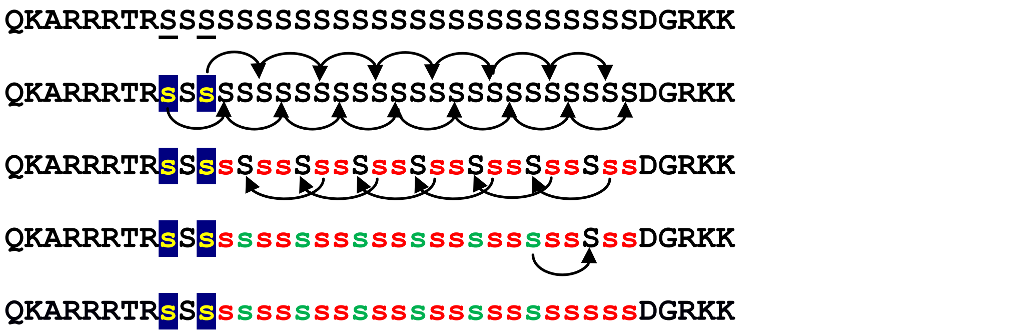
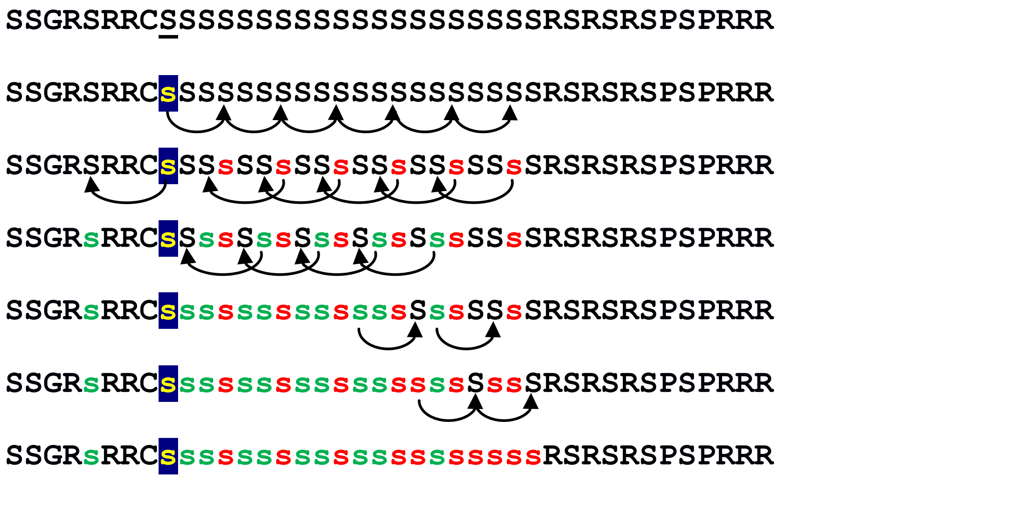
Protein Kinase	Consensus sequences[1]		Ref.
	Canonical[2]	Non canonical[3]	
G-CK/Fam20C	S-X-**E/pS**	S-Q-X_{2-4}-$(E/D)_n$	13, 18
CK2	S/T-X -X-**E/D/pS/pY**	(pS)-pS-S-pS	13, 19
		pS-pS-S-E/D	19
CK1	**pS/pT**-X-X-(X)-S/T	$(E/D)_n$-X-X-S/T	13, 20
		S-L-X_{2-4}-$(E/D)_n$	21
GSK3	S/T-X-X-X-**pS/pT**	----	13

1) Target residue(s) underlined. Phosphorylated determinants denoted by red colour. X denotes any residue.

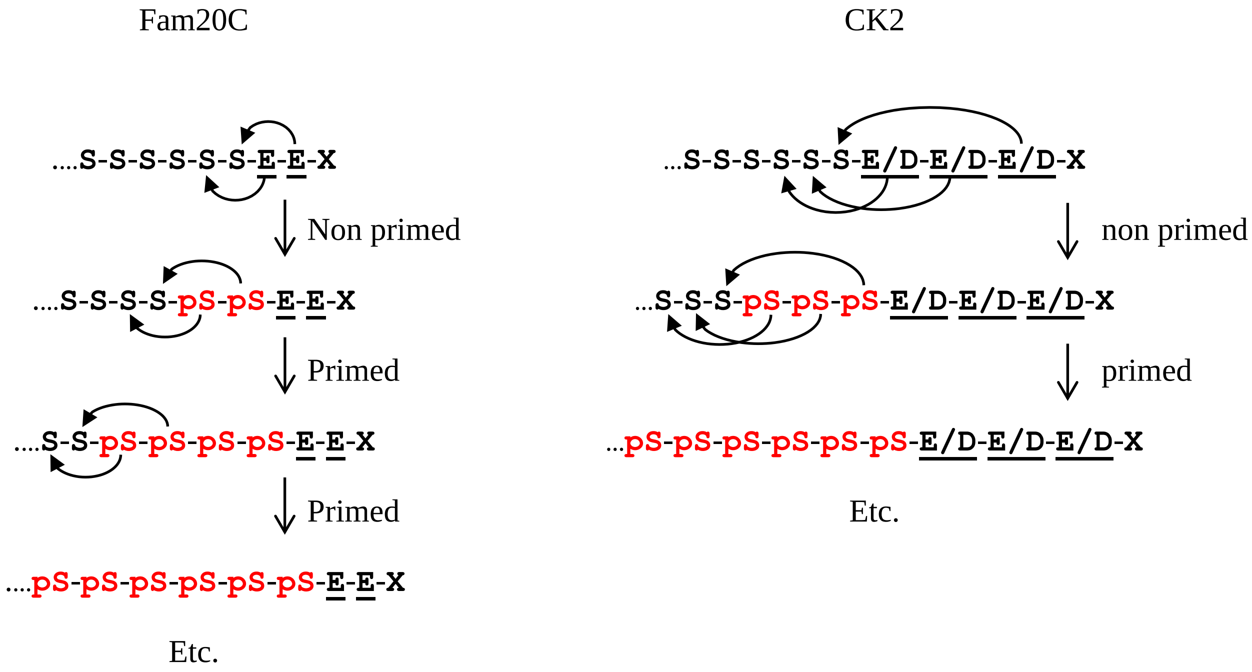
2) Crucial specificity determinants of canonical consensuses are bold typed.

3) Non canonical sequences lack the crucial specificity determinants present in the canonical consensus.

Table 2 Proposed implication of priming basophilic kinases and primed CK1 and GSK3 in the generation of the 3 largest phosphoserine stretches of the human phosphoproteome. (Phospho)serine conforming to the consensus of AKT/PKB and/or PKA are underlined/blue shadowed. Once phosphorylated they trigger hierarchical phosphorylation by CK1 of serines denoted by red colour, which in turn prime the phosphorylation by GSK3 of serines denoted by green colour. For additional details see text.

Q9UQ35 Serine/arginine repetitive matrix protein 2 GN=SRRM2	 TPAKRKRRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSPSP TPAKRKRRsssSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSPSP TPAKRKRRssssssssssssssssssssssssssssssssssssssssssSPSP
Q66PJ3 ADP-ribosylation factor-like protein 6-interacting protein 4 GN=ARL6IP4	 QKARRRTRSSSSSSSSSSSSSSSSSSSSSSSSSDGRKK QKARRRTRsSsSSSSSSSSSSSSSSSSSSSSSSDGRKK QKARRRTRsSssSssSssSssSssSssSssDGRKK QKARRRTRsSssssssssssssssssssssSssDGRKK QKARRRTRsSssssssssssssssssssssssssDGRKK
Q5VV67 Peroxisome proliferator-activated receptor gamma coactivator-related protein 1 GN=PPRC1	 SSGRSRRCSSSSSSSSSSSSSSSSSSSSRSRSRSPSPRRR SSGRSRRCsSSSSSSSSSSSSSSSSSSSRSRSRSPSPRRR SSGRSRRCsSSsSSsSSsSSsSSsSSsSRSRSRSPSPRRR SSGRsRRCsSssSssSssSssSssSSsSRSRSRSPSPRRR SSGRsRRCssssssssssssSsSSsSRSRSRSPSPRRR SSGRsRRCssssssssssssssssSssSRSRSRSPSPRRR SSGRsRRCssssssssssssssssssssRSRSRSPSPRRR

Scheme 1 Minimum requirements for the exaustive phosphorylation of Seryl stretches by Fam20C/G-CK and CK2.


Fam20C
CK2
....S-S-S-S-S-S-E-E-X
Non primed
....S-S-S-S-pS-pS-E-E-X
Primed
....S-S-pS-pS-pS-pS-E-E-X
Primed
....pS-pS-pS-pS-pS-pS-E-E-X
Etc.
...S-S-S-S-S-S-E/D-E/D-E/D-X
non primed
...S-S-S-pS-pS-pS-E/D-E/D-E/D-X
primed
...pS-pS-pS-pS-pS-pS-E/D-E/D-E/D-X
Etc.

FIGURES

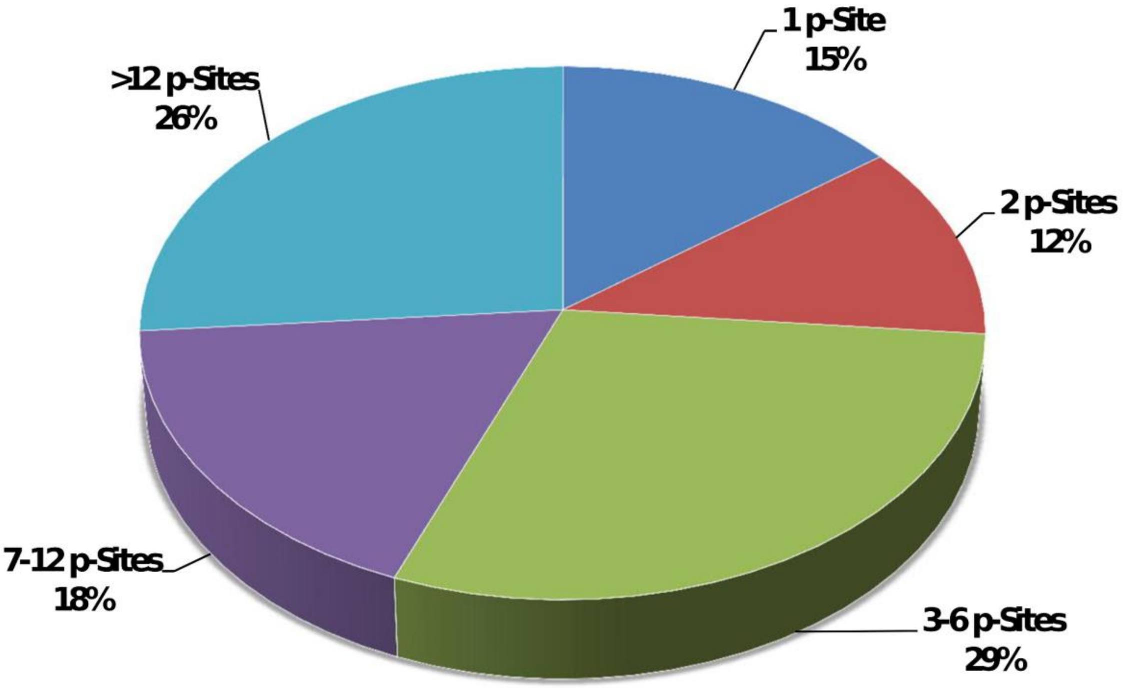


Figure 1 The majority of phosphoproteins are multiphosphorylated. Constructed with data from ref. 1.

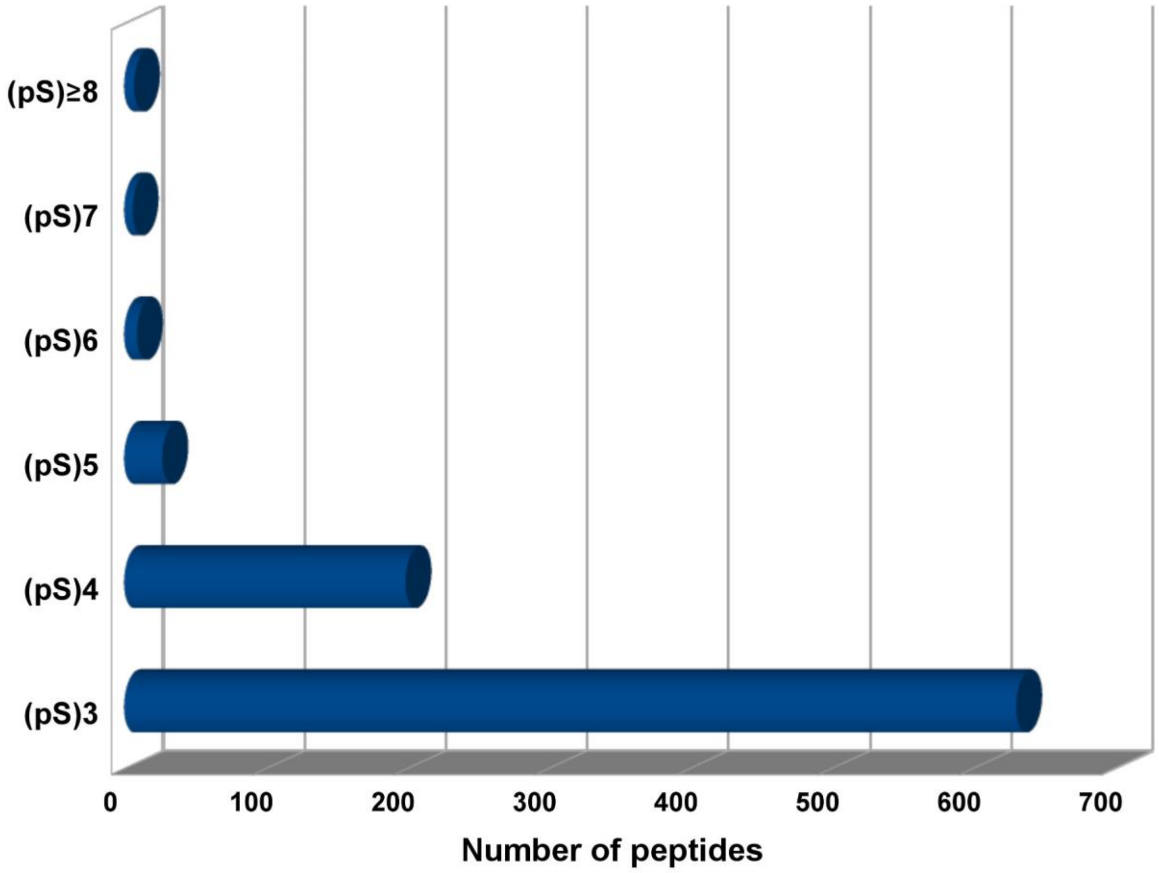


Figure 2 Phosphorylated clusters sorted by number of phosphoserine residues. Drawn from Table S1 (supplementary materials).

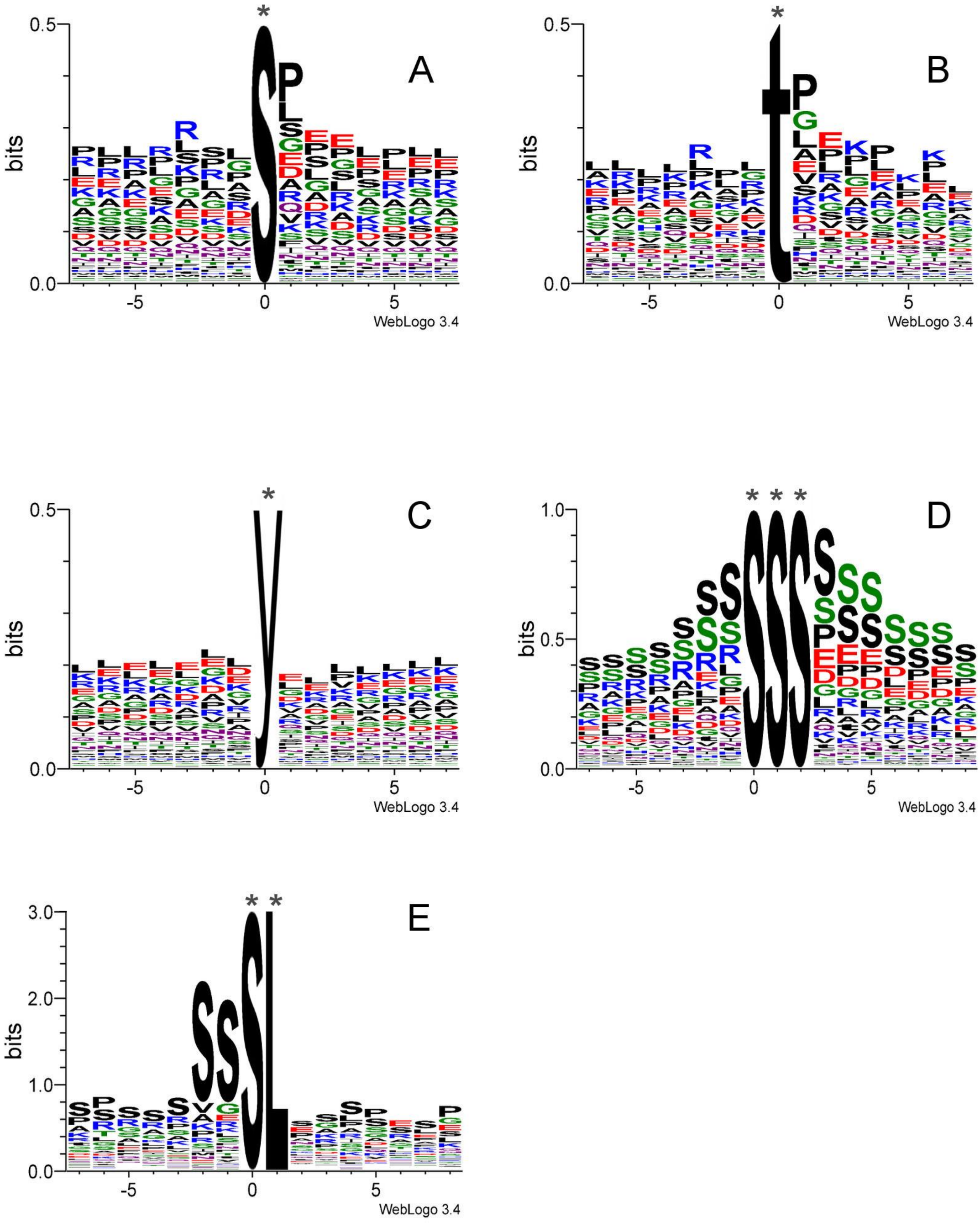


Figure 3 Weblogo of phosphoserine (A), phosphotheonine (B) and phosphotyrosine (C) extracted from PhosphoSitePlus. In D and E the weblogo of pS_3 triplets and of phosphoserines adjacent to Leu (pSL motif) in the sequences listed in Table S1 are shown, respectively. The weblogos were calculated using the command line client of Weblogo3[34,35]. The lowercase letters s, t and y denoting the phosphorylated aminoacids were added to the set of symbols to count. To avoid confusion with their non phosphorylated countparts (uppercase and green) they are black coloured. The height of the phosphoresidue at position 0 (denoted by asterisks) is arbitrarily reduced.

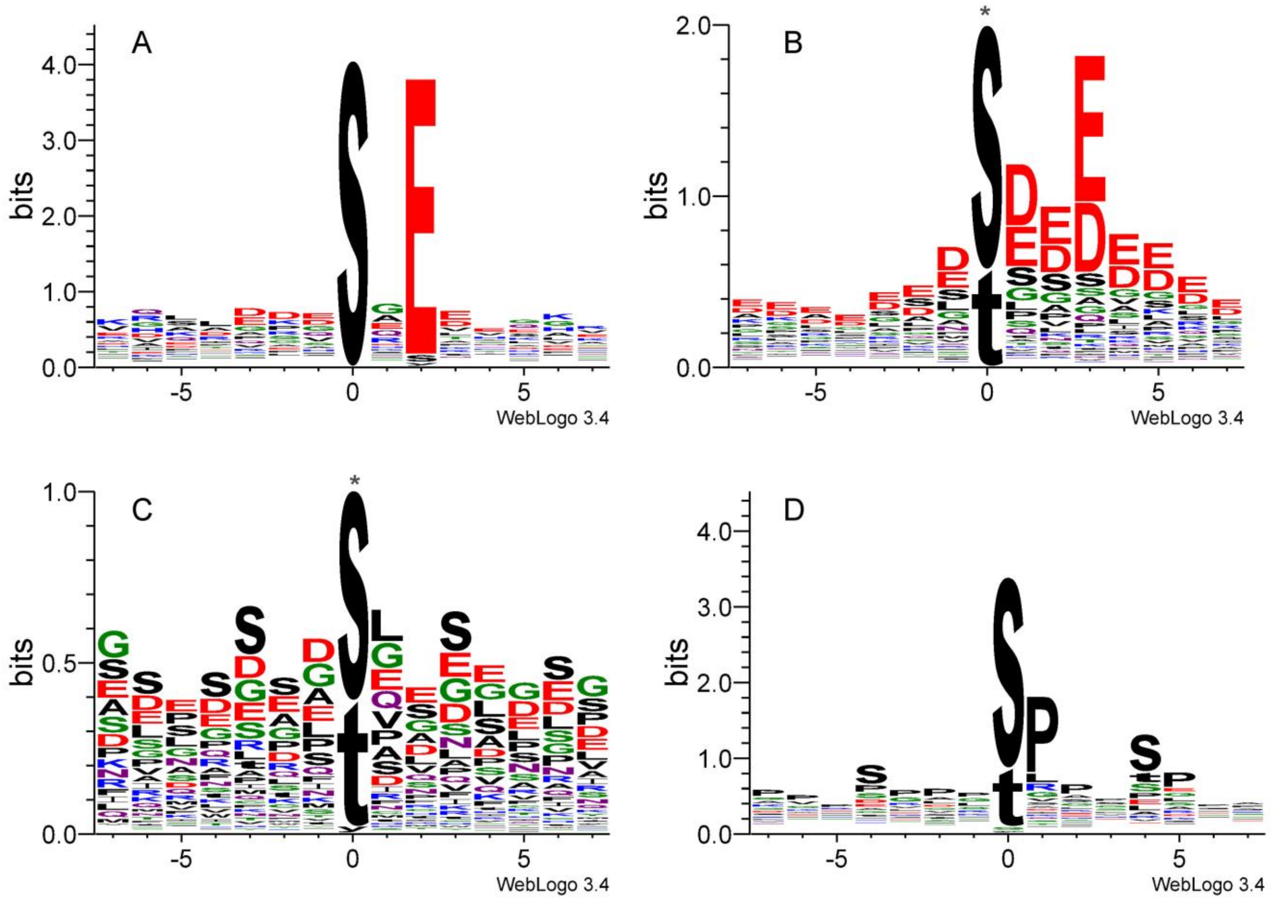


Figure 4 Weblogo of bona fide phosphosites generated by Fam20C (A), CK2 (B), CK1 (C) and GSK3 (D). Drawn from PhosphoSitePlus2 except for Fam20C.[11] Lowercase and colour conventions are as in Fig. 3. In B and C the height of the phosphoresidue at position 0 (denoted by asterisks) is arbitrarily reduced.

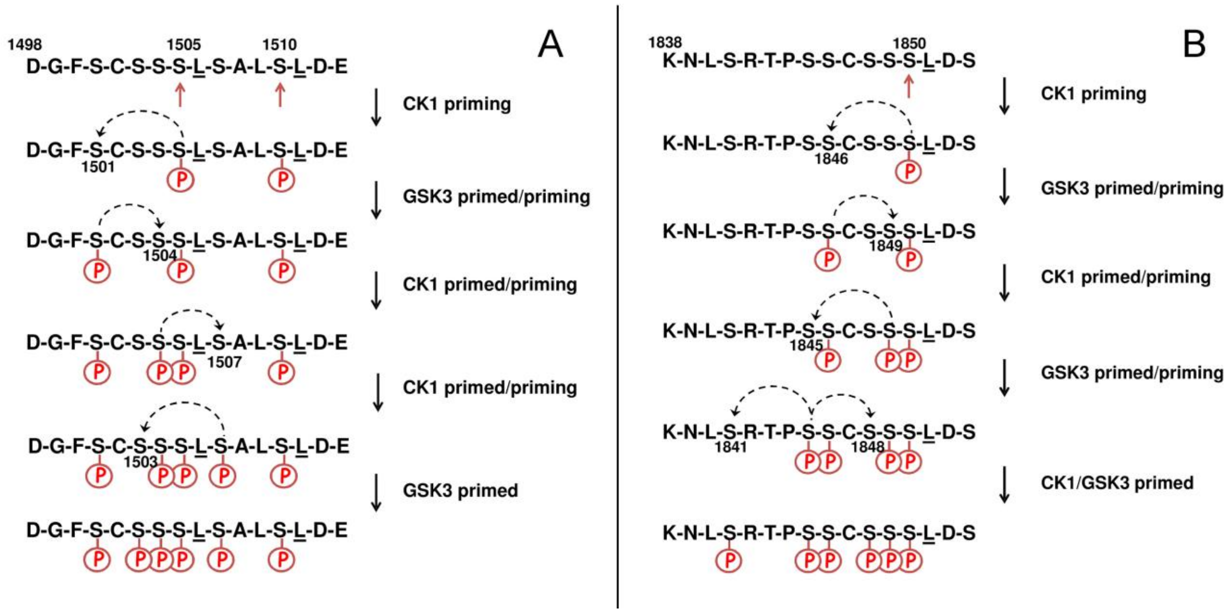


Figure 5 Hierarchical multisite phosphorylation of APC repeat 3 by the combined action of CK1 and GSK3 (A). Drawn from Ferrarese et al Biochemistry[16]. In B the analogous case of CASC5 is schematically illustrated. For details see text.

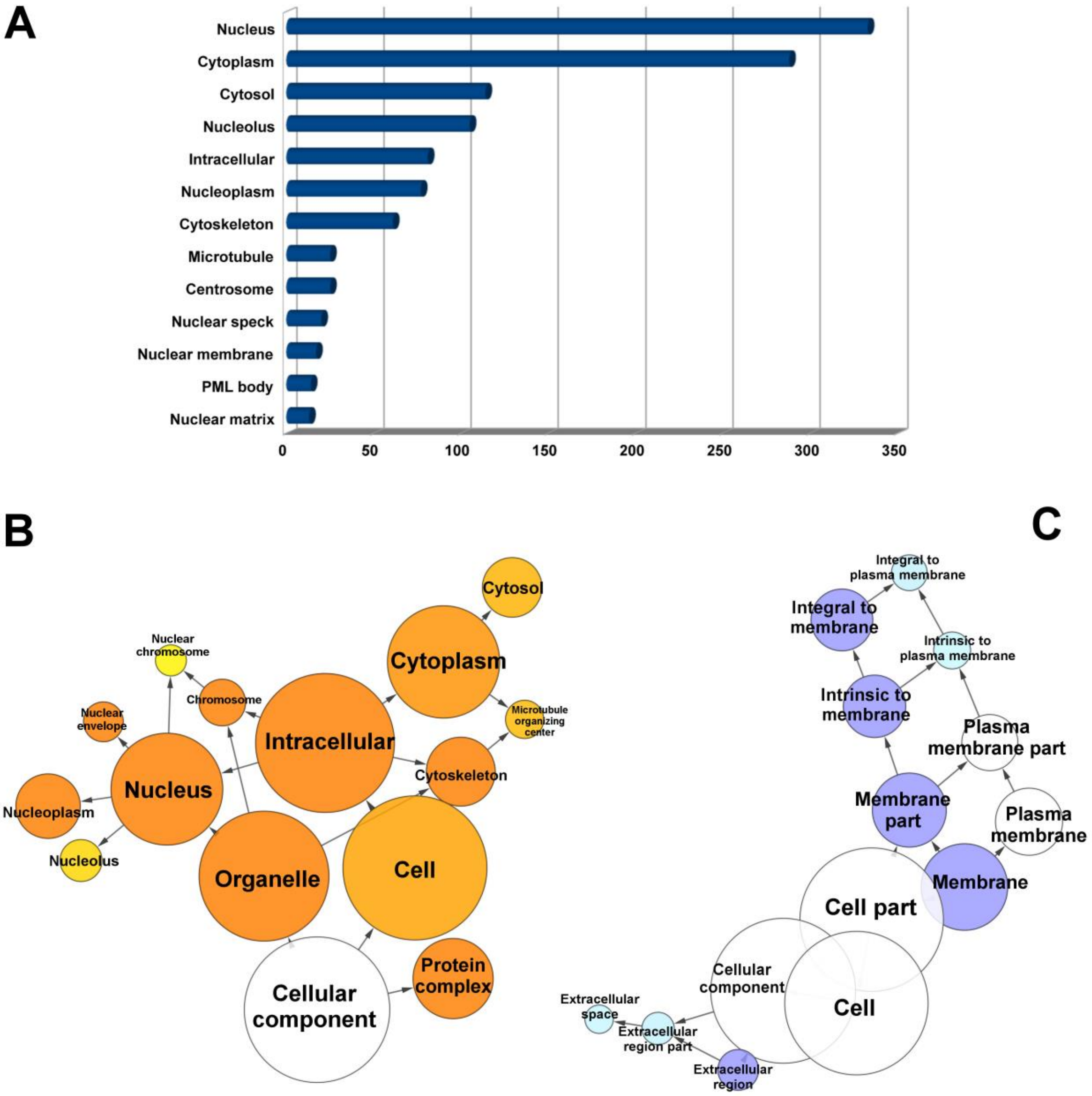


Figure 6 Localization of proteins containing $(pS)_n$ clusters (A) and their over- (B) and underrepresentation (C) in subcellular compartments. Localization of all proteins listed in Table S1 was analyzed by GeneCodis3 (http://genecodis.cnb.csic.es/analysis)[36,37,38] gene-ontology program. Overrepresentation and underrepresentation, denoted by orange and blue colors respectively were estimated using BiNGO (Biological Network Gene Ontology tool)[30]

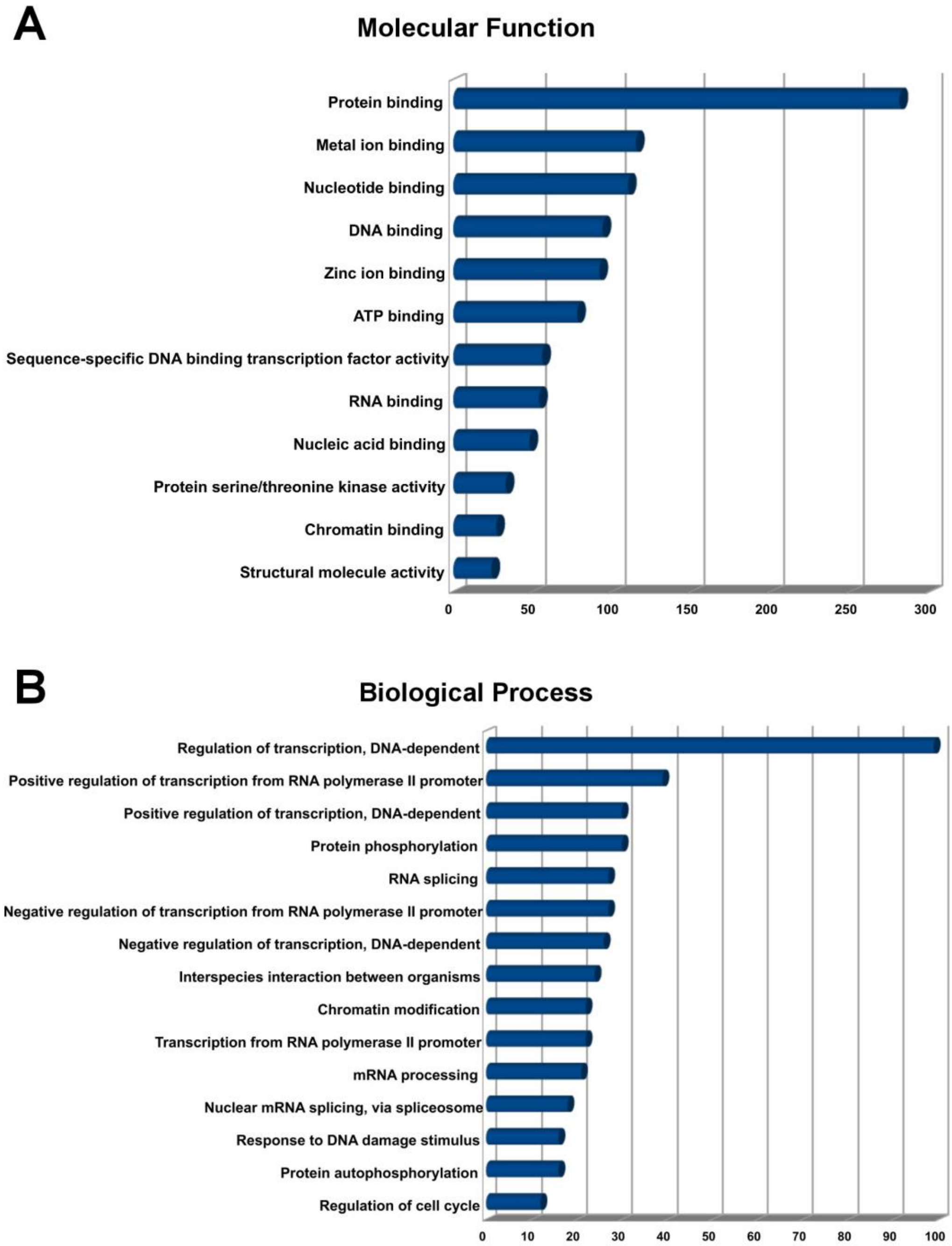


Figure 7 Proteins with $(pS)_n$ clusters: molecular functions (A) and biological implications (B). All the proteins listed in Table S1 were analyzed according to GeneCodis3 (http://genecodis.cnb.csic.es/analysis)[36,37,38] gene-ontology program.

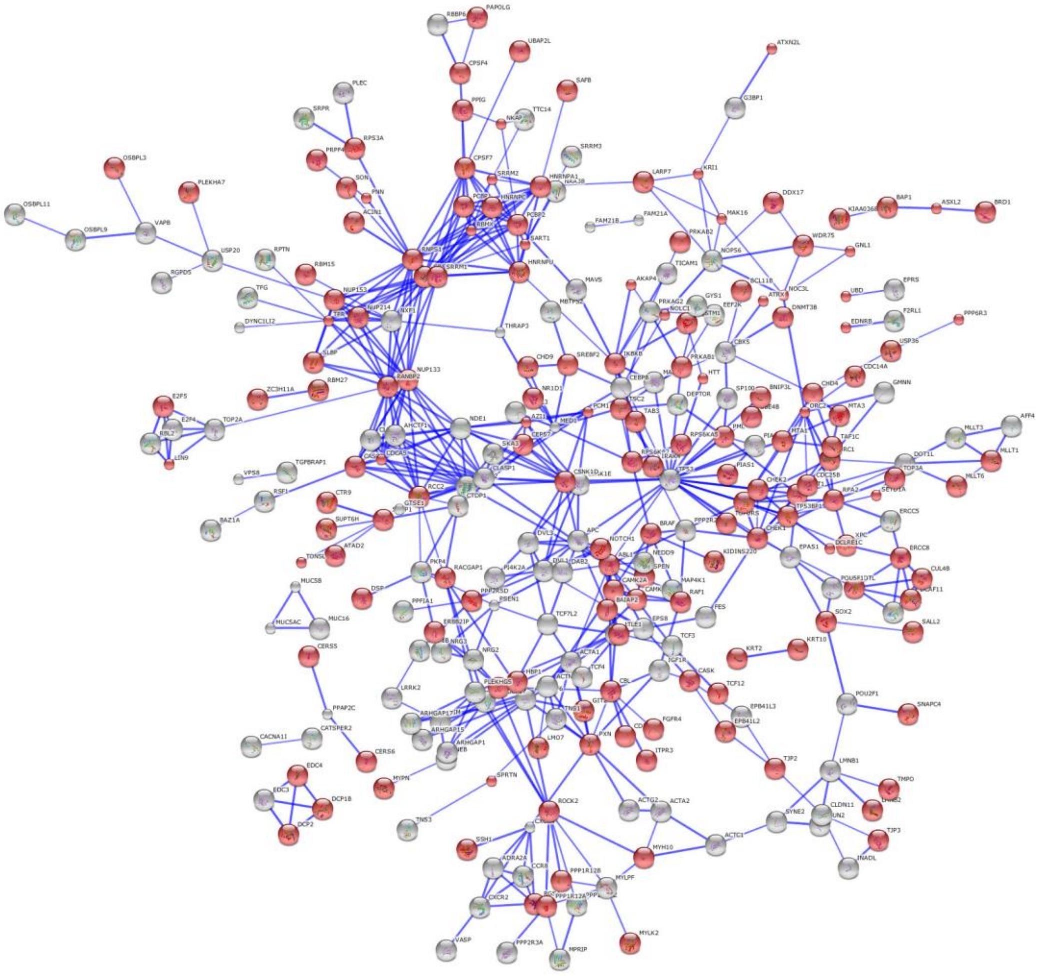


Figure 8 STRING[32] analysis of proteins containing clusters of 3 or more adjacent phosphoserine residues. Functional/physical interactions between the proteins are displayed. Proteins localized in the nucleus are denoted by red colour.