Extracellular protein phosphorylation, the neglected side of the modification

Eva Klement[1] and Katalin F. Medzihradszky[1,2]

[1]Laboratory of Proteomics Research, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary [2]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, San Francisco, USA

Corresponding author:

Katalin F. Medzihradszky

Laboratory of Proteomics Research

Biological Research Centre of the Hungarian Academy of Sciences

62. Temesvari krt.

H-6726 Szeged

Hungary

E-mail: folkl@cgl.ucsf.edu

Tel: +36-62-599-773

Keywords: extracellular phosphorylation; topology; mass spectrometry; large scale studies; Fam20C

1

Running title: Extracellular protein phosphorylation

Abbreviations

ACP	amorphous calcium phosphate
CCTOP	Constrained Consensus Topology prediction
CSF	cerebrospinal fluid
ER	endoplasmic reticulum
Fam20	Family with sequence similarity 20
Fj	Four-jointed
FGF23	Fibroblast growth factor 23
GO	gene onthology
HILIC	hydrophilic interaction liquid chromatography
HMGB1	high mobility group box 1 protein
IMAC	immobilized metal affinity chromatography
OPN	osteopontin
PTM	post-translational modification
SCX	strong cation exchange chromatography
SIBLING	small integrin-binding ligand N-linked glycoprotein
TMT	Tandem Mass Tag
VLK	vertebrate lonesome kinase

Summary

The very existence of extracellular phosphorylation has been questioned for a long time, although casein phosphorylation was discovered a century ago. In addition, several modification sites localized on secreted proteins or on extracellular or lumenal domains of transmembrane proteins have been catalogued in large scale phosphorylation analyses, though in most such studies this aspect of cellular localization was not considered. Our review presents examples when additional analyses were performed on already public datasets that revealed a wealth of information about this "neglected side" of the modification. We also sum up accumulated knowledge about extracellular phosphorylation, including the discovery of Golgi-residing kinases and the special difficulties encountered in targeted analyses. We hope future phosphorylation studies will not ignore the existence of phosphorylation outside of the cell, and further discoveries will shed more light on its biological role.

Phosphorylation is the main regulatory modification of intracellular proteins and is involved in the majority of cellular processes. The existence of this modification extracellularly has long been disputed despite the fact that caseins abundant in milk were the first phosphoproteins discovered in the late 1800s [1]. Gradually, other secreted proteins and peptide hormons were reported to contain this modification. Recently (as of November 3, 2015) the UniProt database [2] yielded 131 proteins once the entries were filtered using the following criteria: species - *homo sapiens*; subcellular localization – secreted; post-translational modification (PTM) – phospho*; and all of these were 'reviewed' (Supplemental Table 1). Approximately, ¾ of these proteins are indeed *in vivo* modified on an amino acid side-chain that normally is localized 'outside' of the cell. The lumenal side of the endoplasmic reticulum (ER) or the Golgi are also considered as such. The accumulating evidence slowly changed the old dogma that restricted phosphorylation to the intracellular milieu, yet, our understanding of the regulation and the exact role of the modification on extracellular proteins is still very limited.

Are these proteins secreted phosphorylated or modified extracellularly?

In the classical secretion pathway a short signal sequence directs the protein to the ER, then the protein travels through the Golgi network and is finally released to the extracellular space. During this transport the protein may undergo post-translational modifications. Specific kinase activity distinct from cytoplasmic kinases was detected in the Golgi apparatus isolated from lactating mammary gland, liver, spleen and other organs [3,4]. This kinase was shown to phosphorylate proteins on an S-x-E/pS motif; resist inhibitors of other protein kinases; use Mn^{2+} as activator instead of Mg^{2+}; and ATP but not GTP as the phosphate donor [4]. However, this kinase escaped identification for a long time. The first Golgi-residing kinase Four-jointed (Fj) was identified in *Drosophila* [5]. Fj was shown to phosphorylate cadherin domains of the transmembrane proteins Fat and Dachsous, yet with different site specificity. PSI-BLAST (Position-Specific Iterated Basic Local Alignment Search Tool) search performed with the human orthologue of Fj, Fjx1 identified Family with sequence similarity 20 (Fam20) and 198 (Fam198) members as other potential Golgi kinases [6]. Eventually, protein Fam20C was identified as 'the real casein kinase' featuring all the characteristics described decades earlier, as listed above. It resides in the Golgi but it can also be found in an N-terminally truncated, secreted form [6,7]. Phosphorylation of the substrate proteins occurs

5

intracellularly [7,8], yet Fam20C may exert its activity also extracellularly. Fam20C was demonstrated to phosphorylate a wide array of secreted proteins involved in biomineralization, lipid homeostasis, wound healing, cell adhesion and migration [7].

Another newly identified kinase: vertebrate lonesome kinase (VLK) - which shows preference for Tyr residues - also localizes in the secretory pathway [9,10]. High expression level of VLK was detected in platelets, and upon stimulation by the thrombin receptor activating peptide (TRAP) VLK was rapidly and quantitatively released from platelets. Besides phosphorylating substrate proteins in the secretory pathway, secreted VLK was found to modify targets in the extracellular environment. Though numerous substrates of VLK were identified, no consensus motif could be determined [10].

Other kinases have also been observed in the releasate of activated platelets that can perform phosphorylation of extracellular proteins if sufficient ATP is available. Platelet secreted casein kinase(s) were shown to phosphorylate protein S and this phosphorylation enhanced its activated protein C cofactor activity thereby affecting its anticoagulant property [11]. Upon stimulation by thrombin a cAMP-dependent protein kinase (PKA) was released from platelets which phosphorylated vitronectin [12]. Several protein kinase C isozymes were also found secreted upon thrombin stimulation from platelets [13].

Proteins lacking a signal peptide and thus not entering the classical ER/Golgi secretion pathway may be released to the extracellular space by unconventional means of protein export, either *via* direct membrane translocation or secretion in various vesicules [14]. In the secretome analysis of MCF7 human breast cancer cell line it was found that only half of the detected proteins were secreted *via* the classical ER/Golgi pathway or shed by plasma membrane, the other half of the proteins used unconventional secretion [15]. Intracellular phosphorylation of these proteins may precede their export. Actually, in certain cases it was demonstrated that phosphorylation is a prerequisite for the transport. Fibroblast growth factor 2 (FGF2) is secreted by direct translocation across plasma membranes and its phosphorylation at Tyr-82 by Tec kinase is required for the secretion [16]. Heat stress-induced translocation of annexin 2 is dependent on Tyr-23 phosphorylation [17]. The high mobility group box 1 protein (HMGB1) is a DNA chaperone that acts also extracellularly in the regulation of inflammation. Phosphorylation of HMGB1 in its nuclear localization signal (NLS) regions results in the relocation of the protein from the nucleus to the cytoplasm with

consecutive secretion [18]. However, it should be noted that detection of phosphoproteins or kinase activities outside the cell may also be an artifact from broken cells and thus, not physiologically relevant.

Large scale MS analyses significantly contribute to extracellular phosphorylation data accumulation

Most information on the phosphorylation of extracellular proteins may be retrieved from mass spectrometric data. Large scale analyses of tissues or whole cells are a major source of such data. Interestingly, the resulting huge datasets practically are never interrogated for the subcellular localization of the modification, and the extracellular phosphorylation sites detected are not reported separately. That is a pity, since such information has been available for years and has been mostly ignored. For a few datasets we performed this analysis with the following notices: i) we accepted the reported phosphorylation data at face value, ii) we somewhat trusted the UniProt cellular localization labels, iii) proteins secreted *via* alternative pathways were kept, iv) proteins that localize to the ER/Golgi were subjected to topology prediction using the Constrained Consensus Topology prediction method (CCTOP, http://cctop.enzim.ttk.mta.hu) [19] and only proteins with lumenal modification sites were kept.

Many plasma proteins are synthesized in hepatocytes, therefore, the in-depth phosphoproteome analysis of a liver homogenate performed by Bian *et al.* proved to be a rich source of extracellular phosphorylation data [20]. Their workflow of „enzyme-assisted" two dimensional reversed phase chromatography supplemented with phosphopeptide enrichment using Ti(IV)-immobilized metal affinity chromatography (IMAC) resulted in phosphosite mapping of 54 secreted and 51 ER- or Golgi-localized proteins with lumenal phosphorylation sites (Supplemental Table 2). Similarly, large amount of extracellular phosphorylation data could be retrieved from the analysis of breast cancer xenograft (human on mouse) and human ovarian tumor samples performed by Mertins *et al.* [21]. Combined, 489 secreted and 275 ER/Golgi-residing phosphoproteins were identified in that study using high pH reversed phase fractionation followed by Fe(III)-IMAC phosphopeptide enrichment and LC-MS/MS analysis (Supplemental Table 3). Strong cation exchange chromatography (SCX) with consecutive TiO_2 enrichment performed by Sharma *et al.* in the analysis of a human cancer cell line also yielded extensive phosphoproteome coverage [22]. Actually, the authors of this study paid some attention to extracellular phosphorylation. However, only Fam20C phosphorylation was considered within S-x-E/pS motifs and Pro was not permitted

in the 'middle' position. Their search yielded 101 phosphosites on 64 proteins. Unfortunately, modification within a 3-amino-acid-constrain does not guarantee the identification of the kinase responsible. The authors did not check whether the regions modified ever enter the ER/Golgi system, and thus, the list contains numerous cytoplasmic modification sites that were possibly phosphorylated by kinases other than Fam20C. For example, Golgin subfamily B member 1 (UniProt Q14789-2) was included in the list with phosphorylated Ser-128, 139, 676, 1756, 1758 and 3145. Except, according to the protein's topology all of these sites are located in the cytoplasmic domain (positions: 1-3235). Our re-evaluation which was not restricted to Fam20C phosphorylation revealed 71 secreted and an additional 90 ER/Golgi-resident phosphoproteins in their study (Supplemental Table 4).

Extracellular phosphorylation analysis from cell culture media

Though whole cell analyses may contain extensive data on the phosphorylation of secreted proteins more specific information can be extracted from targeted analyses that use the culture media of specific cell lines. Tagliabracci *et al.* compared the phosphoproteome of wild type and Fam20C knockout/knockdown HepG2 liver cells secreted into the cell culture media [7]. As Fam20C shows high expression level in lactating mammary gland and mineralized tissue, their analyses were extended to breast epithelial and osteoblast-like cell lines to map modification sites that are linked to the Fam20C kinase activity. Applying Fe(III)-IMAC phosphopeptide enrichment, over 100 secreted phosphoproteins were identified in the culture media of these cell lines and the majority of the modification sites responded to Fam20C depletion. Interestingly, about ⅓ of the phosphosites whose intensity faded upon Fam20C depletion did not match the consensus S-x-E/pS sequence. Phosphosite mapping of insulin-like growth factor-binding protein 1 *in vitro* phosphorylated by Fam20C also indicated modifications in non-S-x-E/pS motives. The authors speculated that the recognition motif of Fam20C might be significantly broader than originally suspected, however, kinase assays performed with synthetic peptides could extend Fam20C activity only to S-x-Q-x-x-D/E-D/E-D/E sequences present in proline-rich phosphoprotein 1 (PRP1) [7,23]. In addition, one has to consider the potential existence of 'phosphorylation cascades' in the Golgi where the elimination of a kinase affects not only the direct substrates but also all the 'downstream' modifications.

Cancer cells are known to display differences in intercellular communication or in their cell adhesion and migration properties that might partially be due to changes in the phosphorylation pattern of the proteins secreted into their microenvironment. In a phosphorylation analysis of the secretome of different luminal and basal type breast cancer cell lines over 1700 phosphoproteins were identified altogether combining hydrophilic interaction liquid chromatography (HILIC) with TiO_2 phosphopeptide enrichment [24]. Approximately half of these phosphoproteins were predicted to be extracellular or membrane proteins according to gene onthology (GO) annotation and bioinformatics analyses using SignalP (classical secretion), SecretomeP (unconventional secretion) and TMHMM (transmembrane domains) algorithms. The proteins that showed subtype specific phosphorylation sites were involved among others in cell signaling and interaction, antigen presentation or cellular assembly and organisation.

Analytical challenges in extracellular phosphorylation studies

While the secretome can be studied from the culture media of different cell lines, phosphorylation analysis directly from readily available body fluids has the most potential in the clinic. There is only a limited number of such reports yet and the phosphorylation data derived from these studies are far from those targeting intracellular phosphorylation. In cerebrospinal fluid (CSF) 25-45 phosphoproteins were identified from 60-200 µg starting material combining SCX fractionation with phosphoSer specific molecularly imprinted polymers or using TiO_2 enrichment [25,26]. Plasma/serum phosphoproteome analyses fared even worse in this aspect, 40-100 µl of plasma (~2-5 mg protein) yielded 30-70 phosphoproteins using TiO_2 enrichment (either alone or in combination with SCX) or Fe(III)-IMAC [27-29]. The largest phosphopeptide dataset from plasma was reported by Zawadzka *et al.* in parallel to the breast cancer cell line secretome analysis [24]. On average, 130 phosphoproteins could be identified from 1 ml plasma (~50 mg protein) of control or breast cancer patients using HILIC fractionation in combination with TiO_2 enrichment. Phosphorylation analysis of saliva also yielded moderate results, a total of 85 phosphoproteins were identified from ~50 mg protein after SCX fractionation and Fe(III)-IMAC phosphopeptide enrichment [30].

One of the main reasons for the relatively low outcome in the phosphorylation analysis of body fluids is the low level of the modification and the complexity and high dynamic range of proteins present in such

matrices. Immunodepletion was used in some cases for the removal of up to the fourteen most abundant plasma proteins to increase the depth of the analysis [24,25,28]. However, comparison of TiO_2 enrichment applied to serum with or without depletion of albumin and IgG showed considerable loss of phosphopeptides in the depleted fraction [27]. Depletion of the highly abundant proteins was also found detrimental in the phosphoproteome analysis of CSF [26]. Indeed, off-target removal was reported using albumin-specific and other immunodepletion kits [31-33]. This may be due to the role of albumin as a cargo for various compounds or a result of nonspecific binding to the affinity matrix. Additionally, ten of the fourteen proteins targeted by the MARS-14 column are actually phosphorylated [29]. Hexapeptide bead libraries (Proteominer) provide another means to narrow down the dynamic range, and that was used in the saliva phosphoproteome analysis by Stone *et al.* [30]. This approach, however, may distort the quantitative results.

A highly efficient enrichment pipeline is a prerequisite for successful phosphorylation analysis from such complex matrices like serum. Above examples show a repertoire of available tools developed for the isolation of phosphopeptides. The most preferred ones are metal affinity techniques: TiO_2 and IMAC often in combination with other orthogonal separations. Though very powerful, selectivity of these methods in CSF or serum is far below that reported for whole cell lysates. In the latter selectivity as high as 90% could be achieved [34] whereas in body fluids selectivity is rather around 30-60% [26, 27,29], a consequence of the high dynamic range and the overall low extent of the modification. Additionally, TiO_2 and recently also IMAC was shown to isolate sialylated glycopeptides under certain conditions [35,36,29]. In the HeLa phosphopeptide dataset obtained after SCX fractionation and subsequent TiO_2 enrichment, discussed earlier [22], 12.5% of the MS/MS spectra contained glycan specific oxonium ion (m/z 204.087, HexNAc) [37]. N-glycosylation is the most frequent and highly abundant modification on secreted and membrane proteins, thus, an even higher glycopeptide contribution is expected when extracellular or membrane proteins are probed. Approximately 1/3 of the MS/MS spectra contained glycan specific oxonium ions in datasets obtained after Fe(III)-IMAC phosphopeptide enrichment of serum [38,29]. Glycopeptides interfere with the phosphorylation analysis at multiple levels. First, glycopeptides affect enrichment efficiency by competing for the active sites of the affinity material. Second, co-elution of glycopeptides can lead to undersampling of the phosphopeptides in the LC-MS analysis of the enriched fraction. In selected cases

this interference of the glycopeptides was taken into account. In the CSF analysis Bahl *et al.* [26] removed N-glycans by peptide N-glycosidase F treatment prior to phosphopeptide enrichment. As deglycosylation may have an adverse effect on protein solubility, it should be performed on the digested sample. Deglycosylation, however, could be cost-prohibitive especially when large amounts of serum samples are processed. Additionally, upon removal of the N-glycans the Asn residues are converted to Asp increasing the acidic character of the peptide that can induce nonspecific binding to the affinity material. For IMAC we demonstrated that high acetonitrile containing solvents promote glycopeptide binding to the sorbent in a mixed mode of metal affinity and hydrophilic interaction [29]. Elimination of this latter component by lowering the acetonitrile content or replacing acetonitrile with methanol – a strong solvent in HILIC – abolished glycopeptide capture completely. TiO_2 enrichment is also performed in solvents with high acetonitrile ratio, therefore, a similar mixed mode of metal affinity and hydrophilic interaction is expected in the capture of the glycopeptides to this sorbent.

Bioinformatics challenges

Large scale phosphopeptide datasets are processed in an automated fashion and the database search results rarely undergo manual validation. Whether a spectrum contains sufficient information for modification site assignment is MS/MS activation and sequence specific. Even if the phosphopeptide is identified correctly, reliability of the site assignment can be questionable, especially if multiple Ser/Thr residues are in close proximity. In such cases often all sites are reported as modified that calls for some caution when handling large scale mass spectrometric data. Development of additional bioinformatic tools that calculate probability of site assignments using prior search outputs [39-42] or direct integration of a scoring algorithm into the search engine to measure the confidence of the modification site localization [43] led to an improvement in this respect. Although, presently only a single search engine, Protein Prospector indicates directly when there is no sufficient information for reliable site assignment. Additionally, there are recent initiatives to make all mass spectral data available. Despite of these initiatives, there are errors in public databases. In PhosphoSitePlus [44] over 1000 phosphoproteins are listed that according to UniProt or GO annotation may be considered extracellular or transmembrane proteins. Among the modification sites listed for these proteins the ratio of phosphorylated Tyr residues is

unusually high which may be a consequence that a large proportion of the information in this database was obtained from affinity enrichments using phosphoTyr antibodies but also reflects the uncertainties of site assignments. In addition, human or software fault may introduce errors in curated lists. For example, Tyr-225 in osteopontin and Tyr-158 in insulin-like growth factor-binding protein 1 were incorrectly listed as Fam20C phosphorylation sites in the UniProt database, although these were originally detected TMT-modified [7] (Supplemental Table 1).

Even when the sequence identification as well as the site assignments are correct we may not know where the modified protein actually came from. In this aspect we may rely on 'Subcellular localization' information from the UniProt database and/or GO annotations, but the available information varies. For certain entries the localization is straightforward, such as „Golgi apparatus membrane; Single-pass type I membrane protein". But more frequently, this information line reads like a complete list of potential cellular localizations, including everything from the nucleus to the extracellular space. There are numerous proteins where this diverse localization list reflects known translocations. For example, as mentioned above HMGB1 indeed may occur in the nucleus as well as in the cytoplasm, and also secreted. However, this uncertainty about the 'origin of the protein' certainly hinders the characterization of subcellular localization-specific modifications.

Last but not least for the characterization of extracellular phosphorylation the topology of the membrane proteins has to be considered. Luckily, such prediction-based information has been available for several proteins, and is regularly listed in UniProt. In a previous large scale glycosylation study we found this information very reliable, we encountered only 2 examples among hundreds of assignments that indicated glycosylation on a predicted cytoplasmic domain [45]. Also, there are several topology prediction servers available to assess potential localization of the modification sites, of which CCTOP was used here for topology prediction of the ER/Golgi-resident phosphoproteins.

Biological significance of the modification

Improved analytical methodologies continuously expand the class of secreted phosphoproteins leaving clarification of the biological importance of the modification far behind. The most established role of extracellular phosphorylation is related to the formation of the mineralized tissue. The structural matrix of

bones and teeth is made up of collagen. The different aspects of collagen phosphorylation on processing, fibril formation or protein interactions have recently been reviewed [46]. Other non-collagenous extracellular matrix proteins are also crucial in bone and tooth formation *via* the control of hydroxyapatite crystallization [47,48]. These secretory calcium-binding phosphoproteins (osteopontin, OPN; dentin matrix protein-1, DMP1; bone sialoprotein, BSP; matrix extracellular phosphoglycoprotein, MEPE and dentin sialophosphoprotein, DSPP) belong to the small integrin-binding ligand N-linked glycoprotein (SIBLING) family and are heavily phosphorylated [49]. Abnormal phosphorylation of these proteins is implicated in severe bone or tooth defects [6,50]. OPN is the most extensively studied member of the SIBLING family. The majority of its Ser residues (37 of 42 in human OPN) and a few Thr residues are reported to be phosphorylated, more than half of these being targets of the Fam20C kinase. The phosphorylated residues are mainly found in clusters and highly conserved in mammals [51,52]. OPN shows a variety in PTMs and processing that is linked to its diverse biological functions. Besides its essential role in inhibiting bone mineralization, it is involved in cell attachment and signaling, inflammation or tumorigenesis [52,53].

Milk caseins are also related to the secretory calcium-binding phosphoproteins [54,55]. Phosphoserine clusters within -S-S-S-E-E- sequences (called phosphate centers) present in α_{s1}-, α_{s2}- and β-caseins show high affinity for calcium resulting in sequestration of amorphous calcium phosphate (ACP) [56]. The formed nanoclusters with a core of ACP surrounded by casein molecules assemble to casein micelles. These structures ensure supersaturation of calcium and phosphate in milk without calcification [57,58]. Similar sequestration of ACP as calcium phosphate nanoclusters was demonstrated for OPN peptides suggesting that such nanoclusters of other secreted phosphoproteins might be involved in the inhibition of calcification in soft and mineralized tissues, the extracellular matrix or biofluids [59].

Phosphorylation analyses performed on plasma/serum samples revealed the modification predominantly on proteins involved in blood coagulation, lipid transport and homeostasis, complement activation, proteolysis, cell adhesion and signaling [27-29]. The details, however, how the modification affects these processes are largely unknown. Recently, phosphorylation of Fibroblast growth factor 23 (FGF23) at Ser-180 was demonstrated to inhibit glycosylation at Thr-178 in the subtilisin-like proprotein convertase cleavage region thereby affecting secretion of the full length biologically active hormone that regulates phosphate retention in the kidney [60,61]. Thus, phosphate level in the circulation is balanced by the

cross-talk of glycosylation, phosphorylation and proteolytic processing of FGF23. Previously, phosphorylation of factor XI by a platelet derived kinase was shown to increase its susceptibility to cleavage by factor XIIa or thrombin [62] suggesting a possible regulation of the coagulation cascade.

Conclusions and perspectives

Large scale phosphoproteomic analyses predominantly from whole cell lysates or conditioned media led to an expansion of extracellular phosphorylation data, although the general public is not aware of it yet. Properly cataloguing such data in public databases would facilitate our understanding what functions this PTM fulfills outside the cell. Currently, the PhosphoSitePlus database has the largest collection of phosphorylation data. This list can be interrogated on the protein level about cellular localization.

Unfortunately, results from the direct analysis of biofluids are limited due to the extreme dynamic range in these matrices that currently limits MS-based phospho-screening of biofluids. However, there are initial attempts to find biomarker candidates in plasma or to assess variation in the phosphorylation level of plasma phosphoproteins [24,63]. Continuous improvements in MS detection sensitivity will facilitate detection of lower abundance phosphoproteins. Yet, it is a question how deep we can dig in this huge dynamic range especially when unexpected post-translationally or artificially (due to sample handling) modified counterparts of the higher abundance components reach the level of low abundance proteins and further increase complexity. In addition, it has also been reported that using a single enrichment protocol reveals only a portion of the 'phosphoproteome' [64-66]. Perhaps, the combined effort of multiple groups could deliver almost comprehensive results.

Within the cell phosphorylation is a highly regulated process with a concerted action of a plethora of kinases and phosphatases affecting the vast majority of intracellular proteins. The lower abundance of extracellular phosphorylation and the limited number of kinases and phosphatases detected outside the cell doubt such a control extracellularly. However, the wealth of extracellular phosphorylation data and also recent targeted analyses anticipate discovery of further secretory pathway kinases [67]. Deciphering the actual roles of the modification on extracellular proteins is also urgently needed to clear this puzzling picture.

14

References

1. Hammarsten, O. (1883) Zur Frage ob Casein ein einheitlicher Stoff sei. *Zeitschrift für Physiologische Chemie* 7, 227

2. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43, D204-212

3. Bingham, E. W., Farrell, H. M. Jr., Basch, J. J. (1972) Phosphorylation of casein. Role of the golgi apparatus. *J Biol Chem* 247, 8193-8194

4. Lasa, M., Marin, O., Pinna, L. A. (1997) Rat liver Golgi apparatus contains a protein kinase similar to the casein kinase of lactating mammary gland. *Eur J Biochem* 243, 719-725

5. Ishikawa, H. O., Takeuchi, H., Haltiwanger, R. S., Irvine, K. D. (2008) Four-jointed is a Golgi kinase that phosphorylates a subset of cadherin domains. *Science* 321, 401-404

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

7. Tagliabracci, V. S., Wiley, S. E., Guo, X., Kinch, L. N., Durrant, E., Wen, J., Xiao, J., Cui, J., Nguyen, K. B., Engel, J. L., Coon, J. J., Grishin, N., Pinna, L. A., Pagliarini, D. J., Dixon, J. E. (2015) A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell.* 161, 1619-1632

8. Wang, S. K., Samann, A. C., Hu, J. C., Simmer, J. P. (2013) FAM20C functions intracellularly within both ameloblasts and odontoblasts in vivo. *J Bone Miner Res* 28, 2508-2511

9. Kinoshita, M., Era, T., Jakt, L. M., Nishikawa, S. (2009) The novel protein kinase Vlk is essential for stromal function of mesenchymal cells. *Development* 136, 2069-2079

10. Bordoli, M. R., Yum, J., Breitkopf, S. B., Thon, J. N., Italiano, J. E. Jr., Xiao, J., Worby, C., Wong, S. K., Lin, G., Edenius, M., Keller, T. L., Asara, J. M., Dixon, J. E., Yeo, C. Y., Whitman, M. (2014) A secreted tyrosine kinase acts in the extracellular environment. *Cell* 158, 1033-1044

11. Stavenuiter, F., Gale, A. J., Heeb, M. J. (2013) Phosphorylation of protein S by platelet kinases enhances its activated protein C cofactor activity. *FASEB J* 27, 2918-2925

12. Korc-Grodzicki, B., Tauber-Finkelstein, M., Chain, D., Shaltiel, S. (1988) Vitronectin is phosphorylated by a cAMP-dependent protein kinase released by activation of human platelets with thrombin. *Biochem Biophys Res Commun* 157, 1131-1138

13. Hillen, T. J., Aroor, A. R., Shukla, S. D. (2001) Selective secretion of protein kinase C isozymes by thrombin-stimulated human platelets. *Biochem Biophys Res Commun* 280, 259-264

14. Rabouille, C., Malhotra, V., Nickel, W. (2012) Diversity in unconventional protein secretion. *J Cell Sci* 125, 5251-5255

15. Villarreal, L., Méndez, O., Salvans, C., Gregori, J., Baselga, J., Villanueva, J. (2013) Unconventional secretion is a major contributor of cancer cell line secretomes. *Mol Cell Proteomics* 12, 1046-1060

16. Ebert, A. D., Laussmann, M., Wegehingel, S., Kaderali, L., Erfle, H., Reichert, J., Lechner, J., Beer, H. D., Pepperkok, R., Nickel, W. (2010) Tec-kinase-mediated phosphorylation of fibroblast growth factor 2 is essential for unconventional secretion. *Traffic* 11, 813-826

17. Deora, A. B., Kreitzer, G., Jacovina, A. T., Hajjar, K. A. (2004) An annexin 2 phosphorylation switch mediates p11-dependent translocation of annexin 2 to the cell surface. *J Biol Chem* 279, 43411-43418

18. Youn, J. H., Shin, J. S. (2006) Nucleocytoplasmic shuttling of HMGB1 is regulated by phosphorylation that redirects it toward secretion. *J Immunol* 177, 7889-7897

19. Dobson, L., Reményi, I., Tusnády, G. E. (2015) CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res* 43, W408-412

20. Bian, Y., Song, C., Cheng, K., Dong, M., Wang, F., Huang, J., Sun, D., Wang, L., Ye, M., Zou, H. (2014) An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteomics* 96, 253-262

21. Mertins, P., Yang, F., Liu, T., Mani, D. R., Petyuk, V. A., Gillette, M. A., Clauser, K R., Qiao, J. W., Gritsenko, M. A., Moore, R. J., Levine, D. A., Townsend, R., Erdmann-Gilmore, P., Snider, J. E., Davies, S. R., Ruggles, K. V., Fenyo, D., Kitchens, R. T., Li, S., Olvera, N., Dao, F., Rodriguez, H., Chan, D. W., Liebler, D., White, F., Rodland, K. D., Mills, G. B., Smith, R. D., Paulovich, A. G., Ellis, M., Carr, S. A. (2014) Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol Cell Proteomics* 13, 1690-1704

22. Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., Wiśniewski, J. R., Cox, J., Mann, M. (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* 8, 1583-1594

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

24. Zawadzka, A. M., Schilling, B., Cusack, M. P., Sahu, A. K., Drake, P., Fisher, S. J., Benz, C. C., Gibson, B. W. (2014) Phosphoprotein secretome of tumor cells as a source of candidates for breast cancer biomarkers in plasma. *Mol Cell Proteomics* 13, 1034-1049

25. Chen, J., Shinde, S., Koch, M. H., Eisenacher, M., Galozzi, S., Lerari, T., Barkovits, K., Subedi, P., Krüger, R., Kuhlmann, K., Sellergren, B., Helling, S., Marcus, K. (2015) Low-bias phosphopeptide enrichment from scarce samples using plastic antibodies. *Sci Rep* 5, 11438

26. Bahl, J. M., Jensen, S. S., Larsen, M. R., Heegaard, N. H. (2008) Characterization of the human cerebrospinal fluid phosphoproteome by titanium dioxide affinity chromatography and mass spectrometry. *Anal Chem* 80, 6308-6316

27 Zhou, W., Ross, M. M., Tessitore, A., Ornstein, D., Vanmeter, A., Liotta, L. A., Petricoin, E. F. 3rd. (2009) An initial characterization of the serum phosphoproteome. *J. Proteome Res* 8, 5523-5531

28. Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpí, E., Abian, J. (2010) Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines. *J. Proteome Res* 9, 876-884

29. Klement, E., Raffai, T., Medzihradszky, K. F. (2016) Immobilized metal affinity chromatography optimized for the analysis of extracellular phosphorylation. *Proteomics* 16, 1858-1862

30. Stone, M. D., Chen, X., McGowan, T., Bandhakavi, S., Cheng, B., Rhodus, N. L., Griffin, T. J. (2011) Large-scale phosphoproteomics analysis of whole saliva reveals a distinct phosphorylation pattern. *J Proteome Res* 10, 1728-1736

31. Gundry, R. L., Fu, Q., Jelinek, C. A., Van Eyk, J. E., Cotter, R. J. (2007) Investigation of an albumin-enriched fraction of human serum and its albuminome. *Proteomics Clin Appl* 1, 73-88

32. Bellei, E., Bergamini, S., Monari, E., Fantoni, L. I., Cuoghi, A., Ozben, T., Tomasi, A. (2011) High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. *Amino Acids* 40, 145-156

33. Stempfer, R., Kubicek, M., Lang, I. M., Christa, N., Gerner, C. (2008) Quantitative assessment of human serum high-abundance protein depletion. *Electrophoresis* 29, 4316-4323

34. Ruprecht, B., Koch, H., Medard, G., Mundt, M., Kuster, B., Lemeer, S. (2015) Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol Cell Proteomics* 14, 205-215

35. Larsen, M. R., Jensen, S. S., Jakobsen, L. A., Heegaard, N. H., (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* 6, 1778-1787

36. Zhu, J., Wang, F., Cheng, K., Dong, J., Sun, D., Chen, R., Wang, L., Ye, M., Zou, H. (2013) A simple integrated system for rapid analysis of sialic-acid-containing N-glycopeptides from human serum. *Proteomics* 13, 1306-1313

37. Medzihradszky, K. F., Kaasik, K., Chalkley, R. J. (2015) Characterizing sialic acid variants at the glycopeptide level. *Anal Chem* 87, 3064-3071

38. Gu, H., Ren, J. M., Jia, X., Levy, T., Rikova, K., Yang, V., Lee, K. A., Stokes, M. P., Silva, J. C. (2016) Quantitative Profiling of Post-translational Modifications by Immunoaffinity Enrichment and LC-MS/MS in Cancer Serum without Immunodepletion. *Mol Cell Proteomics* 15, 692-702

39. Savitski, M. M., Lemeer, S., Boesche, M., Lang, M., Mathieson, T., Bantscheff, M., Kuster, B. (2011) Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* 10, M110.003830

40. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24, 1285-1292

41. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* 10, 5354-5362

42. Saeed, F., Pisitkun, T., Hoffert, J. D., Rashidian, S., Wang, G., Gucek, M., Knepper, M. A. (2013) PhosSA: Fast and accurate phosphorylation site assignment algorithm for mass spectrometry data. *Proteome Sci* 11, S14

43. Baker, P. R., Trinidad, J. C., Chalkley, R. J. (2011) Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 10, M111.008078

44. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* 43, D512-520

45. Trinidad, J. C., Schoepfer, R., Burlingame, A. L., Medzihradszky, K.F. (2013) N- and O-glycosylation in the murine synaptosome. *Mol Cell Proteomics* 12, 3474-3488

46. Yalak, G., Olsen, B. R. (2015) Proteomic database mining opens up avenues utilizing extracellular protein phosphorylation for novel therapeutic applications. *J Transl Med* 13, 125

47. Addison, W. N., Masica, D. L., Gray, J. J., McKee, M. D. (2010) Phosphorylation-dependent inhibition of mineralization by osteopontin ASARM peptides is regulated by PHEX cleavage. *J Bone Miner Res* 25, 695-705

48. Deshpande, A. S., Fang, P. A., Zhang, X., Jayaraman, T., Sfeir, C., Beniash, E. (2011) Primary structure and phosphorylation of dentin matrix protein 1 (DMP1) and dentin phosphophoryn (DPP) uniquely determine their role in biomineralization. *Biomacromolecules* 12, 2933-2945

49. George, A., Veis, A. (2008) Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. *Chem Rev* 108, 4670-4693

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

51. Christensen, B., Nielsen, M. S., Haselmann, K. F., Petersen, T. E., Sørensen, E. S. (2005) Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *Biochem J* 390, 285-292

52. Sodek, J., Ganss, B., McKee, M. D. (2000) Osteopontin. *Crit Rev Oral Biol Med* 11, 279-303

53. Lund, S. A., Giachelli, C. M., Scatena, M. (2009) The role of osteopontin in inflammatory processes. *J Cell Commun Signal* 3, 311-322

54. Kawasaki, K., Weiss, K. M. (2006) Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Zool B Mol Dev Evol* 306, 295-316

55. Kawasaki, K., Lafont, A. G., Sire, J. Y. (2011) The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol* 28, 2053-2061

56. Holt, C., Wahlgren, N. M., Drakenberg, T. (1996) Ability of a beta-casein phosphopeptide to modulate the precipitation of calcium phosphate by forming amorphous dicalcium phosphate nanoclusters. *Biochem J* 314, 1035-1039

57. Holt. C. (2004) An equilibrium thermodynamic model of the sequestration of calcium phosphate by casein micelles and its application to the calculation of the partition of salts in milk. *Eur Biophys J* 33, 421-434

58. Holt, C., Carver, J. A., Ecroyd, H., Thorn, D. C. (2013) Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J Dairy Sci* 96, 6127-6146

59. Holt, C., Sørensen, E. S., Clegg, R. A. (2009) Role of calcium phosphate nanoclusters in the control of calcification. *FEBS J* 276, 2308-2323

60. Tagliabracci, V. S., Engel, J. L., Wiley, S.,E., Xiao, J., Gonzalez, D. J., Nidumanda Appaiah, H., Koller, A., Nizet, V., White, K. E., Dixon, J. E. (2014) Dynamic regulation of FGF23 by Fam20C phosphorylation, GalNAc-T3 glycosylation, and furin proteolysis. *Proc. Natl. Acad. Sci. U. S. A* 111, 5520-5525

61. Lindberg, I., Pang, H. W., Stains, J. P., Clark, D., Yang, A. J., Bonewald, L., Li, K. Z. (2015) FGF23 is endogenously phosphorylated in bone cells. *J Bone Miner Res* 30, 449-454

62. Ekdahl, K. N., Elgue, G., Nilsson, B. (1999) Phosphorylation of coagulation factor XI by a casein kinase released by activated human platelets increases its susceptibility to activation by factor XIIa and thrombin. *Thromb Haemost* 82 1283-1288

63. Zawadzka, A. M., Schilling, B., Held, J. M., Sahu, A. K., Cusack, M. P., Drake, P. M., Fisher, S. J., Gibson B. W. (2014) Variation and quantification among a target set of phosphopeptides in human plasma by multiple reaction monitoring and SWATH-MS2 data-independent acquisition. *Electrophoresis* 35, 3487-3497

64. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* 4, 231-237

65. Di Palma, S., Zoumaro-Djayoon, A., Peng, M., Post, H., Preisinger, C., Munoz, J., Heck, A. J. (2013) Finding the same needles in the haystack? A comparison of phosphotyrosine peptides enriched by immuno-affinity precipitation and metal-based affinity chromatography. *J Proteomics* 91, 331-337

66. Soderblom, E. J., Thompson, W. J., Farnsworth, C. L., Richardson, B. M., Silva, J. C., Moseley, A. (2013) Expanding the depth of coverage in quantitative phosphoproteomics through the combination of metal oxide and motif-specific antibody enrichment strategies. In the Proceedings of 61st Annual ASMS Conference, Minneapolis MN

67. Yang, X., Yan, W., Tian, Y., Ma, P., Opperman, L. A., Wang, X. (2016) Family with sequence similarity member 20C is the primary but not the only kinase for the small-integrin-binding ligand N-linked glycoproteins in bone. *FASEB J* 30, 121-128