

Crosstalk between Signaling Pathways Provided by Single and Multiple Protein Phosphorylation Sites

Hafumi Nishi¹, Emek Demir² and Anna R. Panchenko¹

1 - National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA

2 - Computational Biology Center, Memorial Sloan Kettering Research Center, New York, NY 10065, USA

Correspondence to Anna R. Panchenko: panch@ncbi.nlm.nih.gov

<http://dx.doi.org/10.1016/j.jmb.2014.11.001>

Edited by D. Tawfik

Abstract

Cellular fate depends on the spatiotemporal separation and integration of signaling processes that can be provided by phosphorylation events. In this study, we identify the crucial points in signaling crosstalk that can be triggered by discrete phosphorylation events on a single target protein. We integrated the data on individual human phosphosites with the evidence on their corresponding kinases, the functional consequences of phosphorylation on activity of the target protein and corresponding pathways. Our results show that there is a substantial fraction of phosphosites that can play critical roles in crosstalk between alternative and redundant pathways and regulatory outcome of phosphorylation can be linked to a type of phosphorylated residue. These regulatory phosphosites can serve as hubs in the signal flow and their functional roles are directly connected to their specific properties. Namely, phosphosites with similar regulatory functions are phosphorylated by the same kinases and participate in regulation of similar biochemical pathways. Such sites are more likely to cluster in sequence and space unlike sites with antagonistic outcomes of their phosphorylation on a target protein. In addition, we found that *in silico* phosphorylation of sites with similar functional consequences has comparable outcomes on a target protein stability. An important role of phosphorylation sites in biological crosstalk is evident from the analysis of their evolutionary conservation.

Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Recent phosphoproteomic analyses showed that almost half of all proteins in eukaryotic cells are phosphorylated and protein phosphorylation enables cells to dynamically regulate protein activity and subcellular localization and transmit signals downstream the reaction path [1,2]. Regulatory mechanisms of phosphorylation are quite diverse. It may be accompanied by changes in local site environment or global conformation, leading to protein activation or inactivation [3]. At the same time, it can modulate the nature and strength of protein interactions, thereby regulating protein binding and coordinating different pathways [4,5,50].

Many proteins contain multiple phosphorylation sites that can control different functions of the target protein and provide an expanded combinatorial repertoire for regulation of functional activity. For example, the

binding affinity of tumor suppressor protein p53 to CREB binding protein is modulated by multiple phosphorylation events and its triple phosphorylation results in a considerable increase in affinity compared to a single phosphorylation [6]. In other cases, phosphorylation at different sites might have an opposite effect on protein activity causing protein activation or inhibition [7,8]. Multiple sites can be dephosphorylated by single or different kinases or phosphatases that might serve as a basis of separation or integration of various signals and allow system control by different agonists [9] (Fig. 1). Moreover, the mechanism of phosphorylation might define the response kinetics and it is known that sequential phosphorylation may result in steeper response curves while random phosphorylation gives rise to more shallow responses [10,11].

Biological signaling is very complex, involving many states and oftentimes redundant or alternative

0022-2836/Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). *J. Mol. Biol.* (2014) xx, xxx-xxx

Please cite this article as: Nishi Hafumi, et al, Crosstalk between Signaling Pathways Provided by Single and Multiple Protein Phosphorylation Sites, *J Mol Biol* (2014), <http://dx.doi.org/10.1016/j.jmb.2014.11.001>

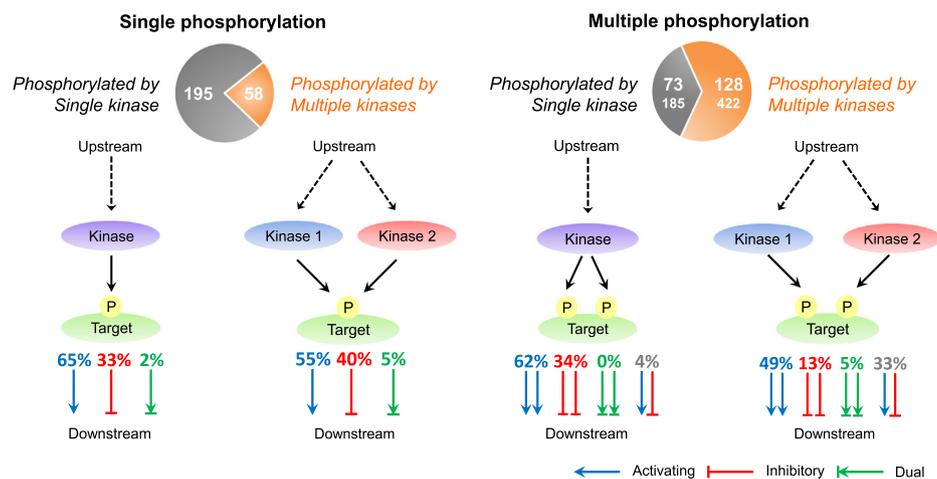


Fig. 1. Scenarios of pathway regulation by single or multiple phosphorylation. Left and right: regulation by proteins containing single or multiple phosphosites, respectively. For each panel, two scenarios are illustrated: when site/sites are phosphorylated by one kinase or multiple kinases. Functional effect on a target protein is shown by the following: activating (blue), inhibitory (red) and dual (green) arrows. In the right panel, phosphosite pairs with the same function are shown as a set of two red, blue or green arrows, respectively, while pairs with different function are shown as a combination of blue and red arrows. Homo-functional pairs may provide a pathway crosstalk by integrating activating or inhibitory signals. Hetero-functional site pairs may provide a different scenario where phosphorylation of one site leads to activation while phosphorylation of another site on the same protein leads to inhibition. Percentage of sites with given functional consequences is listed above the arrows. Pie charts show the number of proteins (sites below) phosphorylated by single (gray) and multiple (brown) kinases. The numbers were based on the intersection of Function and Kinase sets.

relationships between the systems components. The signaling complexity in turn may or may not be accompanied by modularity and hierarchical organization [12,13]. It has been argued that such a seemingly unnecessary increase in diversity of regulatory systems might compensate for the variety of inputs and disturbances to provide specific system responses [14]. Moreover, cellular fate depends on the spatiotemporal distinction between signaling processes and requires the correct integration and separation of different cellular signals that, in turn, provides signal amplification and enhances the response sensitivity. At the same time, the signal integration and separation between alternative and redundant pathways may provide better response specificity. There can be multiple points in signaling pathways that mediate such *pathway crosstalk* when the components and their functional states of one pathway may affect the function of another pathway. In some cases, pathway crosstalk may be sustained by single proteins [13] through molecular switches provided by post-translational modifications. Namely, different phosphorylation events may lead to inhibition or activation of the target protein and consequently potentially inhibit one pathway and activate another.

There have been numerous studies addressing the topic of topological properties of regulatory networks with the ultimate goal of identifying their hubs and bottlenecks [15,16]. However, a full understanding of how signal propagation is controlled requires an integration of systems and molecular levels of description. In particular, deducing the ubiquitous principles of

regulation of protein activity and signal transduction through phosphorylation of individual sites remains an unsolved problem. In this study, we try to pinpoint the crucial points in signaling and pathway crosstalk that are triggered by discrete phosphorylation events in human proteins. We integrate the data on individual phosphosites with the evidence on their corresponding kinases and functional consequences of phosphorylation on the target protein and downstream signaling. Our results show that there are certain patterns in phosphosites' locations and structural/sequence properties that point to their potential role in mediating the communication between different functional states and pathways. Namely, phosphosites having similar regulatory functions, sharing the same kinases and participating in regulation of similar pathways are more likely to cluster in sequence and space. Phosphorylation of sites with similar downstream functional consequences and phosphosites regulated by the same kinase has comparable effects on protein stability. The fundamental regulatory role of such phosphorylation sites is also evident from their evolutionary conservation patterns.

Results

Functional effect of phosphorylation can be linked to phosphorylated residue type

Molecular mechanisms of signaling through phosphorylation differ depending on the external signal

and internal properties of the regulatory system, biomolecules, their interactions and pathways. In some cases, the number, position and combinations of phosphosites determine the functional outcome, while in other cases, the identities of individual phosphorylated residues are also of extreme importance. First we asked if functional consequences of phosphorylation might be linked with the residue type for those sites that can be phosphorylated in human proteins (Ser, Thr and Tyr). Overall, there were substantially more activating than inhibitory sites for all three types of residues, which was consistent with the previous observation that activating signal flows outnumbered the inhibitory signal flows in signal transduction networks [17]. We found the strongest tendency toward activation for pTyr (the ratio of activating and inhibitory sites was 3.7 for pTyr, 2.6 for pThr and 1.7 for pSer; Fig. 2a). Phosphorylation of tyrosine was used considerably more often for activating signals compared to the two other residues taken together (Fisher exact test p -value of 4.3×10^{-6}) whereas pSer was more regularly utilized for inhibition compared to the other two residues (p -value of 1.0×10^{-6}). Although the same tyrosine site can be phosphorylated by different kinases, such cases of promiscuous phosphorylation were found to be less frequent for tyrosine compared to serine and threonine. This supports previous observations about the targeted specificity of tyrosine kinases [18]. The results were unchanged even if phosphosites were clustered based on their location in sequence (Table S4).

We analyzed local structural changes upon phosphorylation by measuring the center-of-mass displacement of the side chain of target residue before and after phosphorylation (the phosphate group was not considered in the calculation of the center of mass). We found that, for large majority of sites, phosphorylation produced only small changes in side-chain conformation of less than 1 Å. However, there were 19 sites with relatively large displacements of 2–6 Å upon phosphorylation, all of them belonged to tyrosine sites. Overall, phosphorylation of Tyr and Ser in protein structural regions led to the larger structural changes while phosphorylation of Thr produced very small changes in local structure on average (p -value of 1.3×10^{-26}) (Fig. 2b). Similar to the functional effect, the results on structural changes were not affected by the phosphosite clustering (Table S4).

Distinct phosphorylation sites may mediate crosstalk between signaling pathways

Protein phosphosites may provide points of crosstalk between signaling pathways through a single “junction” protein. The shared “junction” protein may have a *single* or *multiple* sites phosphorylated by the same or different kinases from the same or different

pathways. In one crosstalk scenario, multiple phosphorylation may lead to separation of signaling events in time or space by inhibiting or activating the target protein and consequently potentially inhibiting one pathway and activating another (site pairs with double arrows of different color; Fig. 1). In these cases, one might expect a certain functional heterogeneity between phosphorylation sites used for regulation [38(33 + 5)% and 4(4 + 0)% phosphosites regulated by multiple kinases and single kinase, respectively; Fig. 1]. In another scenario, multiple phosphorylation may support the integration and convergence of different pathways resulting in amplification, reduction or termination of the signal (the majority of sites, red and blue labels/arrows). In such cases, the functional homogeneity is expected between the sites.

We analyzed the principles of mediation of pathway crosstalk through phosphorylation. First we found that proteins containing sites phosphorylated by only one kinase participated in a smaller number of pathways (2.4 pathways on average) than proteins phosphorylated by multiple kinases (5.3 pathways, p -value of $\ll 0.01$). While the number of pathways increased with the diversity of kinases phosphorylating activating and dual sites, for inhibitory sites, no such association was observed (Fig. 2c). No correlation was found between the phosphoprotein length and number of pathways it controlled.

Different patterns of multiple phosphorylation

Furthermore, we hypothesized that the integration or separation of pathways provided by the shared protein and multiple phosphorylation sites might depend on their intra-molecular distance in space and along the sequence. Overall, the distance between two phosphorylation sites in protein three-dimensional structures correlated very well with the distance along the sequence (Pearson correlation coefficient ρ of 0.63) and almost no long-range spatial contacts between phosphosites separated by large sequence spans were detected. Consistent with the previous observation [19,20], we found that sites phosphorylated by the same kinase were closer in sequences than those phosphorylated by different kinases (p -value of $\ll 0.01$; Fig. S2). Importantly, we observed that two sites, which were located closer to each other in sequence, usually participated in regulation of the same pathway (p -value of 2.7×10^{-17} ; Fig. S2), although this effect was supported by KEGG and PID (Pathway Interaction Database) but not by Reactome pathway databases (Table S3).

Next we analyzed phosphosites with respect to the functional effect (activating, inhibitory or dual) of phosphorylation on the target protein. Sites having the opposite, antagonistic functional effect

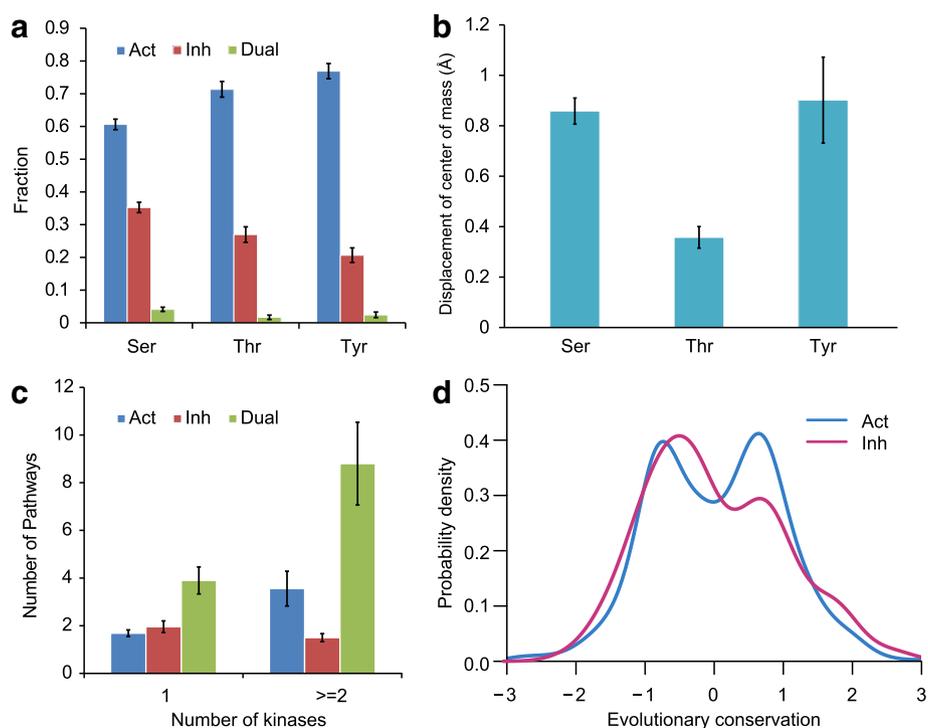


Fig. 2. Properties of all phosphorylation sites. (a) Fraction of phosphorylation sites with a given functional consequence (Act, activating, 1068 sites; Inh, inhibitory, 489 sites; Dual, dual phosphosites, 52 sites) for different residue types plotted with the standard errors. (b) Mean value of displacement of side-chain center of mass upon phosphorylation; standard error bars are shown. (c) Relationship between the numbers of kinases and pathways regulated by phosphosites with different functions. (d) Evolutionary conservation of phosphosites. Positive conservation score corresponds to sites that are more evolutionary conserved compared to other sites in a protein family. Conservation distribution for activating sites can be described by two normal distributions. The distributions were smoothed by the Gaussian kernel density estimation.

(hetero-functional site pairs) showed a larger separation along the sequence (p -value of 2.0×10^{-28} ; Fig. 3a) and in space (p -value of 6.7×10^{-5}) compared to homo-functional site pairs (see Materials and Methods for definitions). After subdividing homo-functional site pairs further into different residue types, we found that activating pairs were less clustered than inhibitory pairs. Moreover, activating pTyr pairs were less clustered than activating Ser/Thr pairs (p -value of 2.3×10^{-7} ; Fig. 3b) although no such distinction was observed between inhibitory sites of different types of residues. This result helps us to understand the previous observation that majority of pSer/pThr sites occur in clusters in sequence [1,21,22] while only 19% of pTyr are within 1–4 amino acids from each other [19].

Inspection of hetero-functional sites showed that they tend to be phosphorylated by different kinases (p -value of $\ll 0.01$; Fig. 3c). Here “shared” kinases refer to those kinases that can phosphorylate both sites in a pair. Next we analyzed the pathways that could be regulated by phosphorylation of individual sites. If kinase A phosphorylates protein B (A \rightarrow B), we used a corresponding pathway that contained both these proteins with recorded relationship

A \rightarrow B. We found that hetero-functional phosphosite pairs regulated fewer, common (the same) pathways (0.6 pathways on average) compared to homo-functional site pairs (2.3 pathways, p -value of $\ll 0.01$; Fig. 3d).

Patterns of evolutionary conservation of phosphosites with different functions

Previous studies showed that protein sites, which could be potentially phosphorylated, were under stronger evolutionary constraints compared to non-phosphorylated surface residues [5,21]. Here we analyzed evolutionary conservation by further distinguishing phosphosites based on their functional outcomes (Fig. 2d). Namely, we aligned phosphoprotein sequences from our data set to domain families as described in Materials and Methods and calculated evolutionary conservation using the entropy-based measure based on sequence-weighted observed amino acid frequencies for each column in the alignment. Unlike inhibitory sites, activating sites had bimodal conservation distribution that could be well described by two Gaussian distributions according to Lilliefors test

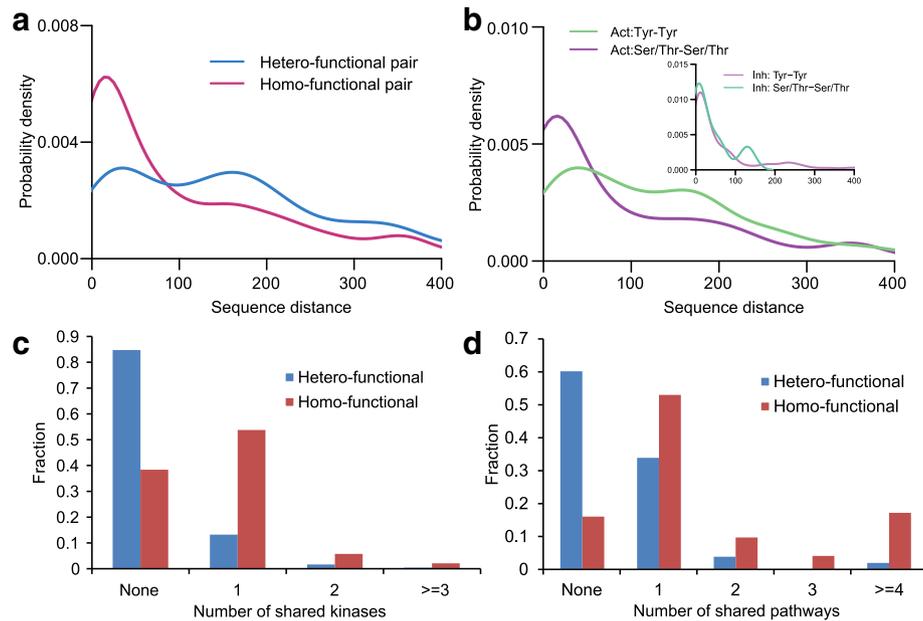


Fig. 3. Properties of multiple phosphorylation sites. (a) Probability density function of sequence distance between phosphorylation sites for hetero-functional (blue) and homo-functional (red) site pairs. (b) Probability density function of sequence distance between phosphorylation sites for activating (main figure) and inhibitory (inset) site pairs. Green, activating Tyr-Tyr pairs; purple, activating Ser/Thr-Ser/Thr pairs. (c) The histogram of the number of shared (same) kinases that can phosphorylate both sites for hetero-functional site (blue) and homo-functional site (red) pairs. (d) The histogram of the number of shared (same) pathways for hetero-functional site (blue) and homo-functional site (red) pairs. The distributions were smoothed by the Gaussian kernel density estimation.

(see Supplementary Materials). It comprised two distinct fractions of activating sites: one fraction included sites that were less conserved than the rest of the family (negative conservation values; Fig. 2d) while the second fraction contained evolutionarily conserved activating sites (Fig. S3b).

Structural and thermodynamic properties of regulatory phosphorylation sites

Phosphorylation can induce conformational changes and affect the stability and binding properties of proteins that, in turn, can be directly linked to the activity changes [3,23]. Multiple phosphorylation might further amplify this effect. We performed structural modeling as described in Materials and Methods and obtained 466 high-quality protein structural models containing 809 phosphosites, 161 proteins among them had several phosphorylation sites on one chain (altogether 582 phosphosite pairs in total). We repeated structural analysis using high-quality structural models obtained from the structural templates with more than 90% identity to the query phosphoprotein (537 sites from 304 proteins, “Model-90” set). Next we calculated changes in unfolding free energy ($\Delta\Delta G$) by attaching the phosphate group to each phosphorylation site in all possible pair combinations on one protein since the order of phosphorylation for these proteins is largely unknown.

We analyzed the effects of multiple phosphorylation events on protein stability in their relation to function using FoldX. Since phosphorylation might stabilize or destabilize protein state, we hypothesize here that sites with similar functional consequences of phosphorylation on a single target protein might have a similar effect on stability of a given protein state. In other words, if phosphorylation outcome is measured in terms of quantitative $\Delta\Delta G$ values, such phosphosites might have the same sign and similar amplitudes to their corresponding $\Delta\Delta G$ values. In support of this hypothesis, phosphosite pairs with the same activating or inhibitory functional consequences (activating-activating, inhibitory-inhibitory) showed more coherent behavior (measured in terms of $\Delta\Delta G$ values) compared to other site pairs. It was evident from measuring the relative differences between their $\Delta\Delta G$ values (Fig. S4a, p -value of 0.024). This effect was more pronounced for structural models from Model-90 set (p -value of 0.008) and was mostly attributed to heavily phosphorylated proteins with more than five sites (Table S2). Furthermore, we found a weak but significant linear correlation between $\Delta\Delta G$ values for sites with the same functional outcome (p -value of 0.035) that was further supported by Model-90 set (p -value of 8×10^{-5}). On the contrary, there was no correlation between $\Delta\Delta G$ values for sites with different functions (p -value of 0.5; Fig. S5). Interestingly, if two sites

shared the same kinase, the effect of their phosphorylation on protein stability was more similar (Fig. S4b, p -value of 0.008 and p -value of 0.02 for *Model-90* set).

Multiple phosphorylation of IRF3 (*interferon regulatory factor 3*) illustrates how phosphorylation sites located in different clusters might have differential effects on IRF3 stability and structure. IRF3 is expressed in cytoplasm as inactive monomer. Viral infection induces the phosphorylation of IRF3 that forms an active oligomer, enters the nucleus and activates the expression of interferon- α/β . Phosphorylation in Ser/Thr sites may activate IRF3 by inducing conformational rearrangements so that the C-terminal segment switches from an autoinhibitory state to an active dimer. According to PhosphositePlus database (“Function” set), seven phosphorylated Ser and Thr residues that are important for the activation are located in two major clusters; sites of each cluster are close to each other in sequence and in space. The first cluster includes residues Ser385 and Ser386 that are partially exposed and might trigger or complement subsequent phosphorylation events. According to our analysis of IRF3 autoinhibitory structure (Protein Data Bank code 1QWT), phosphorylation of Ser385 and Ser386 does not produce large effects on stability and conformation of inactive monomer ($\Delta\Delta G = +0.2$ – 0.6 kcal/mol and center-of-mass displacement upon phosphorylation of 0.1–0.2 Å) and phosphorylation of these sites might be necessary but not sufficient for IRF3 activation. It was originally proposed and subsequently confirmed by different experimental studies [24] that phosphorylation of residues in another cluster (Ser396, Ser398, Ser402, Thr404 and Ser405) releases IRF3 autoinhibition and allows it to interact with the coactivator CBP/p300 to initiate the transcription. Consistent with the autoinhibitory model, we show that phosphorylation of all but one residue within the cluster destabilizes the autoinhibitory IRF3 structure, the largest destabilizing effect on conformation is observed for phosphorylation of Ser398 and Ser396 residues ($\Delta\Delta G = +2.4$ and $+1.4$ kcal/mol, respectively; center-of-mass displacement of 1.3 Å). The critical role of Ser396 was previously confirmed as the minimal phosphoacceptor residue required for the *in vivo* activation of IRF3 [25]. All abovementioned phosphosites belong to homo-functional sites with the coherent outcome of their phosphorylation on IRF3 function and stability.

Discussion

Sequential reactions have certain advantages over single-step signaling in providing additional regulatory checkpoints or proofreading steps [26]. Moreover, selection of signal through the successive

regulatory checkpoints and dichotomic search should be more fast and efficient. Indeed, regulatory reactions are highly non-linear, which explains the predominance of indirect consequences of their perturbation and difficulty of functional annotations of many phosphorylation sites [27].

Here we applied the systems approach and analyzed the functional consequences of hundreds of phosphorylation events on target proteins and downstream signaling by crosslinking the data on individual phosphorylation events with the quarrying of biological pathways. We found that multiple phosphorylation and, in a few cases, single phosphorylation events on the same protein can serve as molecular switches allowing the biological crosstalk between different redundant and alternative pathways. These phosphosites can serve as hubs in the signal flow through the phosphorylation networks and possess special properties. Namely, phosphosites having similar regulatory functions (homo-functional site pairs), sharing the same kinases and participating in regulation of similar pathways are more likely to cluster in sequence and space, and this tendency for clustering is more pronounced for pThr/pSer sites. Furthermore, phosphosites with similar downstream functional consequences and phosphosites regulated by the same kinase have comparable effects on protein stability. In turn, this might point to possible amplification of phosphorylation effects when multiple sites are phosphorylated that can lead to integration of activating or inhibitory signals. Contrariwise, phosphosites with antagonistic regulatory functions tend to be located farther apart in sequence and structure while being phosphorylated by different kinases. Such sites may provide the separation of pathways in time and space.

We observe that there are more activating signaling events associated with Tyr phosphorylation compared to inhibitory flows. While activating role of phosphorylation could be explained in some cases by the evolution of phosphorylation sites from negatively charged amino acids [28], an understanding of its molecular mechanisms could be attained through the “conformational selection” hypothesis [29,30]. Within the framework of this hypothesis, activation by phosphorylation may occur when phosphorylation selects the relevant discrete conformation out of the entire preexisting ensemble, shifting the equilibrium of the conformational ensemble [31] and leading to activation of downstream signaling by stabilizing the active or destabilizing the inactive states. Recently, it was shown that pathway relations involving activation were characterized by a high percentage of structured regions and low disorder content [32]. We hypothesize here that one of the reasons why tyrosine is strongly associated with activation and depleted in inhibitory reactions is that pTyr is usually located in structured regions [33,34] and its phosphorylation leads to a more specific response. Such fine-tuned specific

response is more suitable for activating events that are generally accompanied by precise conformational changes at both local and global levels [35]. The association of tyrosine phosphorylation with activating events is consistent with a pivotal role of tyrosine in oncogene signaling, which extensively deregulates tyrosine phosphorylation [17,36]. It was shown recently that phosphorylation in structured regions, especially tyrosine phosphorylation, is related to longer-lasting effects of phosphorylation [34] compared to dynamically regulated phosphorylation in disordered regions for cell cycle pathways. At the same time, reactions involving disorder-to-order transitions and binding through unstructured regions are usually enriched with pSer and pThr [33,37] and can be used for activating and inhibitory flows.

It has been a controversial topic in the literature regarding the functional neutrality and importance of phosphorylation sites. One might argue that many or even the majority of phosphorylation events do not have phenotypic consequence and might not be vital for the cell survival [27]. Different explanations can support these ideas including the necessity to retain a certain level of redundancy in the system to ensure its robustness. However, regulatory phosphorylation sites analyzed here might represent cases with critical functional properties and severe consequences of their deregulation. Our results show that phosphorylation sites that can potentially provide pathway crosstalk are under selective pressure as they maintain a series of different sequence, structural and thermodynamic properties consistent with their function. This is evident from their evolutionary conservation: activating phosphosites consist of two fractions of slowly and rapidly evolving sites. The former fraction is under stronger evolutionary constraints compared to the rest of protein family sites.

Overall, our study complements and provides the novel insights into the current understanding of design principles of signaling regulation. We argue here that signaling crosstalk does not require a modular network structure in principle and may be achieved at the level of a single protein molecule in a considerable number of cases. Future experimental studies on the individual phosphosites and their functional roles may complete or challenge this view of the phosphorylation networks and their control principles.

Materials and Methods

Linking individual phosphorylation events with their functions and pathways

To compile a comprehensive list of human phosphorylation sites with the functional information on the individual phosphosites in biological pathways, we first extracted data on the locations of phosphorylation sites in human proteins, and for each site, we found its *in vivo* responsible

kinase and/or its function (“activating” and/or “inhibitory”) using the PhosphoSitePlus [38] database. Phosphorylation sites were confirmed by at least two independent high-throughput experimental studies or one low-throughput experimental study. Next we eliminated phosphorylation sites for which the responsible kinases and the function were unknown. After removing redundancy and excluding proteins with sequence identity of more than 50% (the longest protein was retained in the cluster), we ended up with a set of phosphorylation sites with the known functions of phosphorylation events (“Function” set, 1609 sites on 715 proteins) and a set with *in vivo* data on the corresponding kinases (“Kinase” set, 2465 sites on 1057 proteins) (see Supplementary Materials and Fig. S1 for details). There is an overlap between Function and Kinase sets, which contains 860 sites on 454 proteins.

Next we derived phosphorylation and protein activation/inhibition events in human-specific pathways from the biological pathway databases: KEGG [39], Reactome [40] and PID [41]. These pathways include biomolecules and their specific experimentally identified relationships such as protein binding, activation, inhibition and post-translational modifications. In KEGG and PID, target protein–kinase relations may have “activation (active)” or “inhibition (inactive)” attributes assigned to them, therefore indicating how phosphorylation may be associated with activation or inhibition of a target protein. On the other hand, Reactome uses a reaction-centered ontology to represent cellular processes. Activity consequences of phosphorylation events are represented explicitly through downstream reactions. To infer activity changes from Reactome, we searched for reactions where a phosphate group was added to a protein. We then looked for reactions that are catalyzed by either the modified state or another state that is produced directly from the modified state—i.e., it is one-step downstream. We annotated a phosphorylation reaction as “activating” if phosphorylation in Reactome caused activation or prevented inhibition of the target protein or its downstream reactions. Similarly, a phosphorylation was defined as “inhibiting” if it caused inhibition or prevented activation of the target protein or its downstream reactions. The redundancy within the pathways was further removed as described in Supplementary Materials[‡].

To integrate the data on phosphorylation sites, responsible kinases and biochemical pathways, we mapped phosphorylation sites onto pathways via *in vivo* kinase–substrate relation. Namely, if kinase A phosphorylates protein B ($A \rightarrow B$) *in vivo* in PhosphoSitePlus, we use a corresponding pathway that contains both of these proteins with recorded relationship $A \rightarrow B$. Then, functional annotation (activating or inhibitory) of phosphorylation events in Reactome, KEGG and PID pathway databases was used in the pathway analysis in addition to functional annotation derived from PhosphoSitePlus. It should be noted that 81% of phosphosite functions obtained from the pathway databases coincide with the functional outcomes annotated in PhosphoSitePlus. As a result, we obtained 354 phosphorylation sites on 152 non-redundant proteins distributed among 210 non-redundant pathways (“Pathway” set; Fig. S1). Furthermore, 85 out of 152 proteins with assigned pathways had two or more phosphorylation sites. Since pathway databases may differ in terms of the extent of manual annotation, we repeated all pathway analyses separately for each pathway database, and the results are listed in Table S3.

To exclude a possible bias coming from proteins enriched with phosphorylated sites, we also repeated all analyses without heavily phosphorylated proteins. All results reported in the paper were found to be robust with respect to adding or removing heavily phosphorylated proteins (*p*-values were less than 0.05 even though the data set was smaller). In addition, we repeated all analyses by a more stringent verification of phosphosites by requiring at least two independent low-throughput studies to support a phosphosite location. All results reported in the paper were found to be robust with respect to this additional filter (see tables in Supplementary Materials). We also performed clustering of phosphorylation sites to remove potential redundancy at the phosphosite level for single phosphorylation site analyses (see Table S4 for details).

We further subdivided all phosphosites into three categories depending on their activating or inhibitory effects on the target protein: activating, inhibitory and sites with dual properties (sites that can function as both activating and inhibitory according to PhosphoSitePlus database). For multisite proteins, we defined so-called “hetero-functional” phosphosite pairs as pairs between activating and inhibitory, between activating and dual and between inhibitory and dual sites. All other site pairs were denoted as “homo-functional” site pairs.

Structural modeling and *in silico* phosphorylation

Structural templates for modeling were chosen from the Protein Data Bank [42] using the BLAST algorithm [43]. To eliminate low-quality models, we selected only those templates with sequence identities of more than 40% to the query protein and with more than 80% of template structures covered by the BLAST alignment. These thresholds were chosen according to a previous study showing that high-quality models can always be built for proteins with sequence identity higher than 40% [44]. To build homology models, we employed the NEST program from the Jackal package [45] with the option that optimized the configurations of loops and secondary structure regions. The loop prediction program *loopy* in the Jackal package was used to complement missing coordinate regions of some structural templates. Similarly, the refinement of side-chain conformations was performed by the side-chain program *scap*. As a result, we obtained 466 structural models containing 809 phosphosites from the union of Function and Kinase sets.

For the next step, we phosphorylated the modeled proteins *in silico* and calculated the stability changes upon single phosphorylation events. Namely, we built phosphorylated models by attaching the phosphate to those Ser, Thr and Tyr side chains known to be phosphorylated. To assess the effect of phosphorylation on protein stability, we used the FoldX method that performed among the best three methods in predicting the experimental changes in stability produced by amino acid substitutions [46]. The FoldX program [47] estimates protein stability using an empirical force field; it attaches a phosphate group to Ser/Thr/Tyr, optimizes the side-chain conformations of the phosphorylated residue and calculates the difference in unfolding free energy between the original and phosphorylated complexes ($\Delta\Delta G$): $\Delta\Delta G = G^P - \Delta G^U$. Here ΔG^P and ΔG^U are the unfolding free energies of the phosphorylated and unphosphorylated states, respectively. Positive

and negative $\Delta\Delta G$ values correspond to destabilizing and stabilizing effects of phosphorylation.

Evolutionary conservation

We aligned phosphoprotein sequences from our data set to domain families from the Conserved Domain Database [48] and calculated evolutionary conservation using the AL2CO program [49] with default parameters. Conservation analysis could not be performed for dual phosphosites since the number of dual sites that could be mapped on CDD domains was very small. The conservation score represents the entropy-based measure calculated from sequence-weighted observed amino acid frequencies. The score is normalized by subtracting the mean and dividing by the standard deviation of the score distribution for the whole alignment. Therefore, the conservation score of a given site is negative if the site is less conserved than the average conservation background of protein family and vice versa. Statistical analyses are performed as described in Supplementary Materials.

Acknowledgements

We thank Thomas Madej and Yuri Wolf for insightful discussions. This work was supported by the Intramural Research Program of the National Library of Medicine at the US National Institutes of Health. E.D. was supported by grant 5U41-HG006623-02 of the National Human Genome Research Institute, National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2014.11.001>.

Received 12 January 2014;

Received in revised form 30 October 2014;

Accepted 1 November 2014

Available online xxxx

Keywords:

protein phosphorylation;
multiple phosphorylation sites;
signaling pathway;
pathway crosstalk

Present address: H. Nishi, Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi 980-8579, Japan.

‡All data are provided via the ftp site <ftp://ftp.ncbi.nih.gov/pub/panch/Phospho/>.

References

- [1] Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, et al. Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;127:635–48.
- [2] Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell* 2000;103:211–25.
- [3] Johnson LN. The regulation of protein phosphorylation. *Biochem Soc Trans* 2009;37:627–41.
- [4] Pawson T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 2004;116:191–203.
- [5] Nishi H, Hashimoto K, Panchenko AR. Phosphorylation in protein–protein binding: effect on stability and function. *Structure* 2011;19:1807–15.
- [6] Ferreon JC, Lee CW, Arai M, Martinez-Yamout MA, Dyson HJ, Wright PE. Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. *Proc Natl Acad Sci U S A* 2009;106:6591–6.
- [7] Hoffmann I, Clarke PR, Marcote MJ, Karsenti E, Draetta G. Phosphorylation and activation of human cdc25-C by cdc2—cyclin B and its involvement in the self-amplification of MPF at mitosis. *EMBO J* 1993;12:53–63.
- [8] Salazar C, Hofer T. Multisite protein phosphorylation—from molecular mechanisms to kinetic models. *FEBS J* 2009;276:3177–98.
- [9] Cohen P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem Sci* 2000;25:596–601.
- [10] Patwardhan P, Miller WT. Processive phosphorylation: mechanism and biological importance. *Cell Signalling* 2007;19:2218–26.
- [11] Kholodenko BN, Hancock JF, Kolch W. Signalling ballet in space and time. *Nat Rev Mol Cell Biol* 2010;11:414–26.
- [12] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999;402:C47–52.
- [13] Breikreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, et al. A global protein kinase and phosphatase interaction network in yeast. *Science* 2010;328:1043–6.
- [14] Ashby WR. Requisite variety and its implications for the control of complex systems. *Cybernetica* 1958;1:83–99.
- [15] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature* 2001;411:41–2.
- [16] Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 2007;3:e59.
- [17] Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, et al. A map of human cancer signaling. *Mol Syst Biol* 2007;3:152.
- [18] Songyang Z, Carraway KL, Eck MJ, Harrison SC, Feldman RA, Mohammadi M, et al. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* 1995;373:536–9.
- [19] Schweiger R, Linial M. Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data. *Biol Direct* 2010;5:6.
- [20] Yachie N, Saito R, Sugahara J, Tomita M, Ishihama Y. *In silico* analysis of phosphoproteome data suggests a rich-get-richer process of phosphosite accumulation over evolution. *Mol Cell Proteomics* 2009;8:1061–71.
- [21] Boekhorst J, van Breukelen B, Heck A, Snel B. Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol* 2008;9:R144.
- [22] Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004;32:1037–49.
- [23] Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. *Chem Rev* 2001;101:2209–42.
- [24] Qin BY, Liu C, Lam SS, Srinath H, Delston R, Correia JJ, et al. Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. *Nat Struct Biol* 2003;10:913–21.
- [25] Servant MJ, Grandvaux N, tenOever BR, Duguay D, Lin R, Hiscott J. Identification of the minimal phosphoacceptor site required for *in vivo* activation of interferon regulatory factor 3 in response to virus and double-stranded RNA. *J Biol Chem* 2003;278:9441–7.
- [26] Ferrell JE, Machleder EM. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* 1998;280:895–8.
- [27] Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signaling* 2010;3:rs4.
- [28] Pearlman SM, Serber Z, Ferrell JE. A mechanism for the evolution of phosphorylation sites. *Cell* 2011;147:934–46.
- [29] Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. *Protein Eng* 1999;12:713–20.
- [30] Boehr DD, Wright PE. Biochemistry. How do proteins interact? *Science* 2008;320:1429–30.
- [31] Shen T, Zong C, Hamelberg D, McCammon JA, Wolynes PG. The folding energy landscape and phosphorylation: modeling the conformational switch of the NFAT regulatory domain. *FASEB J* 2005;19:1389–95.
- [32] Fong JH, Shoemaker BA, Panchenko AR. Intrinsic protein disorder in human pathways. *Mol Biosyst* 2012;8:320–6.
- [33] Nishi H, Fong JH, Chang C, Teichmann SA, Panchenko AR. Regulation of protein–protein binding by coupling between phosphorylation and intrinsic disorder: analysis of human protein complexes. *Mol Biosyst* 2013;9:1620–6.
- [34] Tyanova S, Cox J, Olsen J, Mann M, Frishman D. Phosphorylation variation during the cell cycle scales with structural propensities of proteins. *PLoS Comput Biol* 2013;9:e1002842.
- [35] Xin F, Radivojac P. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* 2012;28:2905–13.
- [36] Li L, Tibiche C, Fu C, Kaneko T, Moran MF, Schiller MR, et al. The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer. *Genome Res* 2012;22:1222–30.
- [37] Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 2002;323:573–84.
- [38] Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 2012;40:D261–70.
- [39] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012;40:D109–14.
- [40] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011;39:D691–7.

- [41] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res* 2009;37:D674–9.
- [42] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- [43] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- [44] Zhang Y. Protein structure prediction: when is it useful? *Curr Opin Struct Biol* 2009;19:145–55.
- [45] Petrey D, Xiang Z, Tang CL, Xie L, Gimpelev M, Mitros T, et al. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* 2003;53:430–5.
- [46] Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel* 2009; 22:553–60.
- [47] Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 2002;320:369–87.
- [48] Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 2013;41: D348–52.
- [49] Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 2001; 17:700–12.
- [50] Nishi H, Shaytan A, Panchenko AR. Physicochemical mechanisms of protein regulation by phosphorylation. *Front Genet* 2014;5:270.