

Intrinsically disordered proteins link alternative splicing and post-translational modifications to complex cell signaling and regulation

Jianhong Zhou¹, Suwen Zhao^{1,2*}, A. Keith Dunker^{3*}

1. iHuman Institute, ShanghaiTech University. 2F Building 6, 99 Haike Road, Pudong New District, Shanghai, 201210, China.

2. School of Life science and Technology, ShanghaiTech University

3. Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W 10th street, Suite 5000, Indianapolis, IN 46202, USA.

Corresponding authors: Suwen Zhao, Email: zhaosw@shanghaitech.edu.cn and A. Keith Dunker, Email: kedunker@iupui.edu

Abstract: Intrinsically disordered proteins and regions (IDPs and IDRs) lack well-defined tertiary structures, yet carry out various important cellular functions, especially those associated with cell signaling and regulation. In eukaryotes, IDPs and IDRs contain the preferred loci for both alternative splicing (AS) and many post-translational modifications (PTMs). Furthermore, AS and/or PTMs at these loci generally alter the signaling outcomes associated with these IDPs or IDRs, where the functional cooperation of these three features is named the IDP-AS-PTM toolkit. However, the prevalence of such functional modulations remains unknown. Also, the signal-altering mechanisms by which AS, and PTMs modulate function and the extent to which AS and PTMs collaborate in their signaling modulations have not been well defined for particular protein examples. Here we focus on three important signaling and regulatory IDR-containing protein families in humans, namely G-protein coupled receptors (GPCRs), which are transmembrane

This is the author's manuscript of the article published in final edited form as:

Zhou, J., Zhao, S., & Keith Dunker, A. (2018). Intrinsically disordered proteins link alternative splicing and post-translational modifications to complex cell signaling and regulation. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2018.03.028>

proteins, the nuclear factors of activated T-cells (NFATs), which are transcription factors (TFs), and the Src family kinases (SFKs), which are signaling enzymes. The goal here is to determine how AS and PTMs individually alter the outcomes of the signaling carried out by the various IDRs and to determine whether AS and PTMs work together to bring about differential cellular responses. We also present data indicating that a wide range of other signaling IDPs or IDRs undergo both AS- and PTM-based modifications, suggesting that they, too, likely take advantage of signal outcome modulations that result from collaboration between these two events. Hence, we propose that the widespread cooperation of IDPs, AS and/or PTMs provides a IDP-AS-PTM toolkit and substantially contributes to the vast complexity of eukaryotic cell signaling systems.

Keywords: *Intrinsic disorder, alternative splicing, post-translational modification, differential and context-dependent signaling, signaling modulation, and regulation*

Abbreviations: IDPs or IDRs, intrinsically disordered proteins or regions; AS, alternative splicing; PTMs, post-translational modifications; GPCRs, G-protein coupled receptors; NFATs, nuclear factor of activated T-cells; TFs, transcription factors; SFKs, Src family kinases; CD, circular dichroism; NMR, nuclear magnetic resonance; ICL3, intercellular loop 3; CaN, calcineurin; GRK, G-protein coupled receptor kinase; PKA/C, protein kinase A/C; CDK, cyclin-dependent kinase; DBD, DNA-binding domain.

Highlights

- We propose that IDPs work in concert with AS and PTMs to provide an IDP-AS-PTM toolkit for complex context-dependent cell signaling and regulation.

- Three intensely studied, highly divergent signaling protein families, namely G-protein coupled receptors (GPCRs), nuclear factor of activated T-cells (NFAT) transcription factors (TFs)

and Src family kinases (SFKs), all use this toolkit to increase context-dependent signaling complexity.

▪PubMed text mining shows the widespread occurrence of IDPs, AS and PTMs in a large number of proteins including those involved in developmental signaling pathways, which indicates a common maybe even universal use of the IDP-AS-PTM toolkit in eukaryotic multicellular signaling systems.

Introduction

Intrinsically disordered proteins and regions (IDPs and IDRs) have been found to be heavily involved in cell signaling and regulation, especially in eukaryotes[1-9]. Due mainly to their amino acid compositions[10-12], IDPs and IDRs lack stable tertiary structures under physiological conditions and exist instead in highly dynamic, interconverting, flexible conformations[13]. Their functions complement those of structured proteins and underlie cellular differentiation, transcription, cell cycle regulation, DNA condensation, cell division and many other crucial biological processes[3, 14].

Protein/nucleic acid/ligand binding sites are often located in IDPs or IDRs, and their flexibility enable single, short IDRs to change their backbone and side-chain conformations and thereby bind tightly with multiple, distinctly shaped binding partner surfaces[15, 16]. Thus IDPs and IDRs can mediate interactions with a large number of partners and thus function as hubs or as partners to structured hub proteins in signaling networks[17-19]. Furthermore, IDPs and IDRs contain numerous sites for post-translational modifications (PTMs) that can reversibly regulate IDP- or IDR-binding in a cellular context-dependent manner by adding binding sites for new partners and/or by inhibiting binding by partners that recognize the unmodified IDR, thus further expanding an IDR's already formidable binding repertoire. In many cases the PTM-induced changes lead to

initiation or inhibition of specific signaling pathways[20-23], or can be involved in diseases[24] or protein translocation[25]. Another way of expanding the functionality of IDPs and IDRs is through alternative splicing (AS) of pre-mRNA that codes for the same IDP or IDR, a post-transcriptional process that generates two or sometimes many more protein forms from a single gene[26]. These versatilities offered by the flexibility of IDPs or IDRs, with AS and PTMs adding further complexity, enable signaling and regulatory proteins to efficiently accomplish dynamic functions in response to changes in cellular environments.

Based on the observation that both AS and PTMs, especially multiple PTMs[27, 28], alter the functions of many IDPs[23, 26], we previously proposed that IDPs work cooperatively with AS and PTMs to provide a toolkit (namely IDP-AS-PTM), for signaling diversification[29, 30]. We also showed that this toolkit is used by many of the proteins that carry out the functions underlying multicellularity, functions such as cell-to-cell adhesion, intercellular communication, development pathway specification, development pathway regulation over space and time, and tissue- or cell-type-specific physiology[31]. An important feature of this proposal is that tissue- or cell-type-specificity has been shown for both AS[32, 33] and PTMs[34, 35], suggesting that the new biological function enabled by the IDP-AS-PTM toolkit provides context-dependent signaling. By this we mean that specific tissues or specific cell-types are able to “rewire” or “remodel” protein pathways and genetic networks depending on the local context, while still using the same sets of genes. This rewiring results from the tissue- or cell-type-specific AS and/or the tissue- or cell-type-specific PTMs. That is, both AS and PTMs have been shown to have the capability to alter interactions between proteins or between proteins and nucleic acids, and both undergo tissue- or cell-type specific alterations[16, 26, 32, 33, 36-38].

So far, a significant number of important protein families have been predicted to contain substantial amounts of intrinsic disorder. Some families have been analyzed in a large scale (genome- or proteome-wide), including transcription factors[39, 40], nuclear hormone receptors[41], membrane proteins[42, 43], histones[44], spliceosomal proteins[45, 46], ribosomal proteins[47], DNA/RNA binding proteins[48], and many enzymes[49]. Other proteins have been investigated on a smaller scale, but with detailed features (e.g. domain organizations, molecular recognition features, binding interface properties), such as serine-arginine rich proteins[50], scaffold proteins[51], autoinhibited proteins[52], cytoskeletal proteins[53]. In addition, many IDPs are associated with human diseases, including cancer, neurodegeneration, cardiovascular disease, amyloidosis, diabetes and many others[2, 54]. While more IDPs continue to be characterized thus expanding the biological functions known to be associated with these proteins[49, 55, 56], the frequency of occurrence of the IDP-AS-PTM toolkit mentioned above remains unknown, and how IDPs or IDRs, AS and/or PTMs jointly regulate these proteins is not well characterized for any particular IDP or IDR.

Here we examine whether the IDP-AS-PTM toolkit is used by three important signaling and regulatory IDP families in humans, all of which contain members involved in development, namely the G-protein coupled receptors (GPCRs, a membrane receptor protein family), the nuclear factors of activated T-cells (NFATs, a transcription factor (TF) family), and the proto-oncogene Src family kinases (SFKs, a regulatory enzyme family). First we map the sequence locations of sites associated with AS (annotated in UniProt) and PTM (annotated in UniProt and PhosphoSitePlus[57]), and then we determine whether their IDRs are enriched in these two regulatory events as compared to their structured regions. Next we study selected examples for each family to determine whether, and if so, how IDRs, AS and PTMs collaborate to regulate signaling diversity for these particular protein examples.

GPCRs are of particular interest because they are the largest known seven transmembrane (7TM) protein family that have a particularly high amount of predicted disorder as compared to other transmembrane proteins[43]. This a very large membrane receptor protein family with over 800 members in humans, with an enormous diversity of ligands from rhodopsin to peptides, and with involvement in an extremely wide range of biological processes, including blood pressure regulation, olfactory function, embryogenesis and nearly every other physiological process[58, 59].

TFs play a particularly central role in transcription regulation. Particular TF examples have been known to contain IDRs since the 1980s[60], nearly two decades before bioinformatics examinations showed the widespread existence of massive amounts of intrinsic disorder in the eukaryotic TFs[39, 40]. Among the many well studied IDR-containing TFs, the NFAT family was selected because it has widespread importance, including critical roles in T-cell function, inflammation, angiogenesis, myocardial development, skeletal muscle development, cancer metastasis, and many other biological processes; because its own regulation depends on two very important signaling systems, namely regulation by calcium levels with calmodulin as the calcium sensor and regulation by phosphorylation by various kinases and dephosphorylation by calcineurin [61], and because its massive IDRs have been characterized by both circular dichroism (CD) and nuclear magnetic resonance (NMR)[62].

SFKs are non-receptor tyrosine kinases, and they are key regulators in signal transduction. The first-discovered SFK member, Src, is identified as the most highly connected hub in the whole kinome. Members of this family are predicted to contain large IDRs at their N- or C-terminus or between folded domains, and these predicted IDRs are significantly overlapped by

regions of missing electron density from available SFK structures[63]. Also, SFKs help to regulate a number of important processes such as differentiation, proliferation, migration, and survival[64].

For these three particular protein families, and many other IDPs mentioned above, numerous publications show that their functional complexity is substantially enhanced by the combined use of IDPs or IDRs, AS and/or PTMs. These observations support our hypothesis that the IDP-AS-PTM toolkit is commonly used to provide a mechanism for sophisticated signaling processes and indeed may be essential for the emergence of complex multicellular organisms as we suggested previously[31].

Results

GPCRs and the IDP-AS-PTM toolkit

Co-occurrence of IDRs, AS and PTMs in GPCRs

The full-length GPCRs include the extracellular N-terminus, the intracellular C-terminus, seven transmembrane helices (TM1-7), three intracellular loops (ICL1-3) and three extracellular loops (ECL1-3) connecting the helices. While the overall TM topology is well conserved across GPCR members, the N-terminus, ICL3 and the C-terminus exhibit high variability in terms of length and amino acid composition[65], and are predicted to be the most disordered regions (green histogram in Fig. 1A, left). This result is consistent with previous predictions[65-67] and the experimental confirmation on a few GPCRs by CD and NMR[68-70]. In addition, these results agree with the necessary truncation of these IDRs for most GPCRs during crystallization. Indeed, for the >200 available crystal structures belonging to 44 GPCRs in Protein Data Bank (PDB), a majority of their N- and C-termini and ICL3 regions are truncated to achieve crystallization success.

In accordance with the distribution of IDRs over GPCR secondary structures, the occurrence of both AS and PTMs (purple line in Fig. 1A and stacked bar in Fig. 1B, respectively) more often localize within the N- and C-termini and ICL3 regions as compared to other regions. Experimental data provides evidence that GPCRs can be modified in multiple ways, and undergo tissue-specific AS (Fig.1C), suggesting those two regulatory phenomena may mutually alter specific receptor functions. Indeed, Among the 308 AS regions, 94 (~31%) of them contain known PTM sites that may be altered by AS (Fig.2A). More importantly, as shown in the disorder distribution in Fig. 2B, when predicted to be fully ordered (disorder percentage=0%), AS regions with PTMs show lower percentage of than those without known PTMs (18% and 41% respectively); in opposite, when predicted to be disordered, AS regions with PTMs generally show higher percentage than those of without PTMs (Fig. 2B). This result suggests that AS regions with PTMs are more likely to predicted to be disordered, indicating that IDRs provide the preferential locations for AS to modulate PTMs. This enrichment of AS and multiple PTMs within IDRs of GPCRs provide massive combinations of IDR, AS and/or PTMs that would differentially “encode” receptor functional diversity, including differential downstream cellular signaling.

Regulation of GPCR functions by IDR-localized AS and/or PTMs

Given the preference of IDRs, AS and PTMs within N-terminus, ICL3 and C-terminus of GPCRs, in the following we explore how these three most disordered regions work in concert with AS and/or PTMs to enhance receptor functional diversity in a cell- or tissue- specific manner.

In general, the disordered regions of GPCRs are important for receptor ligand binding, surface expression, trafficking and signaling, thus alteration of them by AS would substantially affect GPCR activities. For instance, AS-isoforms differ in the N-terminus of many GPCRs show decreased or abolished ligand-binding activity[71, 72]. In many cases, the functions of N-

terminus-associated AS-variants remain to be elucidated, and some are speculated to act as a dominant-negative mutant of the wild-type receptors[71]. Interestingly, some GPCRs have a very long disordered ICL3 regions, within which the AS events often occur. For instance D2 dopamine receptor (DRD2), with a long ICL3 having 148 residues, generates three isoforms, D2short, D2long (canonical) and D2longer, with D2short and D2longer having deletion of 29 residues and insertion of two residues in ICL3, respectively[73, 74]. As a result, D2short and D2long couple with different α subunit of inhibitory G-proteins (D2short coupled preferentially with $G\alpha_{i1}$, and D2long couples selectively with $G\alpha_{i2}$) and activate distinct signaling pathways accordingly[75]. The free fatty acid receptor 4 (FFAR4) generates an isoform with insertion of 16 residues within ICL3, leading the receptor towards an arrestin-biased pathway[76]. Another extreme case is CXCR3 isoform CXCR3Alt, which lacks the whole ICL3 region and fails to induce either $G\alpha_i$ activation or β -arrestin recruitment[77]. As ICL3 is important for G-protein or arrestin interaction and subsequent activation of intracellular events[78], other GPCRs with AS-isoforms differing within this disordered region [e.g. D3 dopamine receptor (34 residues deletion), Gastrin/cholecystokinin type B receptor (69 residues insertion), Histamine H3 receptor (80 residues deletion)] are likely to have different signaling pathways or especially biased signaling pathways, thus illustrating the essential roles of IDR-localized AS in selective signal transduction.

A large number of phosphorylation sites are localized within the IDRs of GPCRs (Fig.1B), and have been described to generate different combinational patterns linked to differential GPCR signaling. Such patterns, which are termed phosphorylation codes, were initially discovered in M3-muscarinic acetylcholine receptor[79] and β 2AR[80], and are found in disordered C-tails/ICL3 for most GPCRs[81]. Besides phosphorylation, other PTMs, such as palmitoylation[82], glycosylation[83], different patterns of ubiquitination (mono- or poly-ubiquitination)[84], also participate in biased signaling, suggesting the likely presence of an expanded version to the

phosphorylation code—a PTM code that specifies GPCR activities using different types of PTMs. The potential to regulate distinct signaling outcome by combinatorial PTMs is illustrated by studies on CXCR4, which undergoes not only sulfation and glycosylation, but also phosphorylation and ubiquitination (Fig.3). Thus, in the case of CXCR4, there is evidence that site-specific PTMs can result in different signaling results, supporting the GPCR PTM code hypothesis mentioned in reference[85].

AS further modulates PTM incidence in GPCRs

AS can alter multiple PTMs sites for many GPCRs (Fig1C and Fig.2), leading to another layer of functional complexity by combining IDR-Co-occurring with AS and PTMs. This agrees well with the functional influence of tissue-specific AS on IDPs in general[32, 33]. For example, among the six GPCRs that undergo palmitoylation (Fig.1C), four of them (AVPR2[86], EDNRB and EDNRA[87], OPRM1[88]) lose their palmitoylation sites induced by AS. AVPR2 isoform without the palmitoylation sites (C341 and C342) and phosphorylation sites (S362, S363 and S364) can adopt two different topologies[89] and mainly remain inside the cell and down-regulates the surface expression of canonical AVPR2 by formation of heterodimers[90]. EDNRB isoforms that lack the palmitoylation sites fail to activate G proteins[91]. Replacement of phosphorylation sites within disordered C-terminus of TXA2R isoforms lead to distinct combinations of kinase phosphorylation, and subsequent separated biological processes (desensitization or internalization)[92, 93]. Although in many cases the AS-driven replacement or addition of potential PTMs remain unknown (i.e. those three N-terminal AS-variants of CXCR4 in Fig.3), the reported examples presented above show that the synergistic collaborations of AS and PTMs leads to enhanced, context-dependent signaling complexity. That is, IDR-localized AS creates alternative PTM patterns leading to different downstream outcomes. Overall, we conclude

that diverse signaling carried out by GPCRs use the IDP-AS-PTM toolkit in multiple essential ways.

TFs and the IDP-AS-PTM toolkit

Co-occurrence of IDRs, AS and PTMs in human TFs in general

Large portions of eukaryotic TFs have been predicted to contain IDRs, typically covering two-thirds or more of the TF sequences, and these predictions are in good agreement with experimental data[39, 40, 94]. Furthermore, recent work suggests that, for TF families associated with development, there is a strong positive correlation between the amount of predicted disorder and the complexity of the organism[95].

In human TFs, their intrinsic disorder is not restricted to DBDs but includes assigned Pfam domains outside of the DBD and also not restricted to transactivation domains or protein-interaction domains. Significantly, predicted IDRs include massive regions that have not yet been assigned to a particular type of domain (green boxplot in Fig.4A, top axis).

In accordance with the distribution of disorder prediction, higher numbers of TFs have AS events within the unassigned regions (URs) as compared to those in DBD and NonDBD (Fig.4A, purple-border histogram, bottom axis). Likewise, significantly more PTM sites are located within the more disordered URs than those in DBD and NonDBD (Fig.4B). Furthermore, the presence of multiple types of PTMs in TFs is very common; among the 1345 TF members with PTM annotations, 62% of them have more than one type of PTMs (Fig.4C). Besides, 93 of those 1345 TFs (~7%) are documented to have two or more types of PTMs targeting the same residues (totally 202 sites), with lysine being the most frequent site for alternative PTMs (Fig.4D). In addition, 882 (~41%) of all the AS regions (totally 2128 regions) contain known PTMs (Fig.5A), and for the fully predicted to be ordered AS regions, those with PTMs are about two times less

than those without PTMs (Fig.5B). That is, most of AS regions with PTMs (94%) are predicted to be disordered (red histogram in Fig. 5B); in comparison, only ~70% of AS regions without known PTMs show predicted disorder (blue histogram in Fig.5B). Thus, like GPCRs, human TFs generally display co-occurrence of IDRs, AS, and PTMs that likely act synergistically to help modulate the complicated aspects of transcriptional regulation.

Transcriptional regulation of NFATs by IDR-localized AS and PTMs

Extensive evidence indicates that AS and PTMs of TFs commonly alter DNA-binding affinity/specificity or their interactions with cofactors in cell- or tissue-specific manner[28, 96-99]. Here we present data for one important subfamily—the nuclear factors of activated T-cells (NFATs), to show how specific combinations of IDR-localized AS and PTMs affect the detailed functions of NFATs. As key regulators in T-cell development and function, NFATs have five members sharing a similar DNA-binding domain. Among them, four members (NFATc1-4) are specifically regulated by calcineurin (CaN), a Ca^{2+} /calmodulin-dependent serine phosphatase that is involved in many important signaling pathways[100]. The distant member, NFAT5, is not calcium-related due to the lack of regulatory domain that contains CaN-binding motifs, and it is activated by osmotic stress instead. NFATs are of our particular interest because they have been confirmed by CD and NMR to contain extremely long disordered domains besides the well-defined DNA-binding domain[62]. These experimental data agree very well with the disorder prediction in Fig.6A.

These IDRs of NFATs are the main locations where multiple splicing sites occur (Fig.6B). NFATs have been suggested to undergo tissue-specific AS events that contribute to isoform-specific transcriptional abilities[101]. The ten splice isoforms of NFATc1 are presented in Fig.6B.

These variants differ in the length of their disordered N- and/or C-terminus, and all the other NFAT genes are able to generate multiple AS isoforms in similar manner[101].

The NFATc proteins undergo multiple PTMs, including phosphorylation of the serine-rich regions (SRR1&2) and Serine–Proline repeat motifs (SP1-3), and sumoylation, most of which localize within the disordered region (Fig.6C). The large numbers of phosphorylation sites are required for maintaining the NFATc molecules within the cytoplasm, whereas dephosphorylation of these sites by CaN promotes nuclear import and initiation of target gene expression. These regulatory serine sites are phosphorylated by different kinases, specifically by PKA, DYRK, CK1, or GSK3 in a hierarchical pattern, creating a complex regulation that may allow for distinctive activation profiles in different cell types[102]. The sumoylation of NFATs, which is cell-specific and AS-isoform-specific, was recently shown to repress the transcriptional activity and regulate its nuclear retention, providing a new regulatory mechanism for NFAT functions[103].

To give more structural details, the nuclear localization signals (NLS) contain clusters of positively charged residues interspersed with serines (Fig.6C). Four of the five positively charged residues are arginines. Having arginines rather than lysines is likely important here because arginines form much stronger interactions with phosphates than do lysines. That is, each arginine has two hydrogens that are well-placed to form two hydrogen bonds with two phosphate oxygens, and it has even been observed that two arginines can bind to a single phosphate with the concomitant formation of four hydrogen bonds[104]. Such hydrogen bonding between the phosphates and arginines would inactivate the NLS, while dephosphorylation by calcineurin would lead to activation and nuclear import. Interestingly, up to 13 different phosphates, most of which are rather distant in the sequence from the NLS, are involved in this inactivation, and the

collaboration among the phosphates leads to a sensitive, on-off switch-like behavior for the nuclear import of NFAT1c[105, 106].

AS alter PTM sites of NFATs

AS-induced PTM changes are observed (Fig.6C and Fig.6D), and related functional consequences have been reported in some cases. The final ~150 residues of C-terminal IDR contain a second transactivation domain which contain two strong sumoylation sites (K702, K914). Sumoylation leads NFATc1 to its subnuclear relocalization and enable NFATc1 to suppress expression of its selective target genes (e.g.IL-2)[107]. The isoforms without these two sumoylation sites (e.g. isoform A, isoform IA), but contains a weaker sumoylation site K349, does not have the selectivity, suggesting the collaboration of AS and PTM bring out fine-tuning of NFATc1 functions. The N-terminal-altered AS isoforms of NFATc1 lose the annotated R23 methylation, but the functions of methylated R23 remains unknown, so does the consequences of losing this PTM. Also, the functions of the phosphorylation sites within the disordered C-terminus remains unknown. It has been reported that the C-terminal-altered isoform of NFATc1, B- β , fails to interact with TNF- α (a target gene of NFATc1) to which the common NFATc1 isoform strongly binds[108]. In addition, two alternative N-terminal transactivation domain regions are generated by AS in NFATc2, resulting isoforms having differential roles in the control of cell proliferation and transformation[109]. It would be very interesting to test whether this different regulation by NFATc is related to the AS-altered PTM mentioned above.

Taken together, the prevalence of the AS sites, as well as the large numbers of PTM sites within the IDRs, suggest that the different NFAT molecules utilize the IDP-AS-PTM toolkit for functional diversification.

Src family kinases (SFKs) and the IDP-AS-PTM toolkit

SFKs act as important intermediaries regulating a variety of cellular activities, and increased activity or overexpression of SFKs can lead to constant activation, which is considered as its common mechanism of causing cancers and other diseases[110]. Src is the first identified proto-oncogene[111], and its gene product is the first protein discovered to have tyrosine kinase activity[112]. SFKs generally include nine members that are divided into two groups based on sequence identity: Src, Fyn, Yes, Fgr, and Yrk, forming Group A, Lck, Hck, Blk and Lyn in Group B.

Regulatory role of disordered N-terminus of SFKs

The N-terminal domains of SFKs, including SH4 and Unique, have been confirmed to be disordered by NMR[113, 114], which is in good agreement with our prediction (Fig.7A). The other important functional regions are the classic regulatory domains SH3 and SH2, the catalytic domain SH1, SH2–SH1 linker and a regulatory C-terminal tail (Fig.7B), which are all involved in kinase autoinhibitory functions[115]. Remarkably, recent studies have highlighted a more complex regulatory mechanism mediated by the disordered SH4 and Unique domains[116-118]. Specifically, Unique is not only able to bind to SH3 and lipids, but also interacts with other proteins, such as calmodulin, to regulate kinase activities[116]. The lipid binding of Unique may change the ligands or substrates accessible to the kinase, thereby possibly inducing distinct pathways. This new finding is further supported by a study demonstrating that disordered SH4 and Unique are directly involved in kinase regulation by using SH3 as a scaffold[117], and these three domains form a fuzzy complex[118]. These recent investigations on SH4 and Unique have significantly filled the gap between unrecognized functions of the N-terminus and the well-known regulatory roles involving in SH2 and SH3.

Multiple types of PTMs within the disordered N-terminus of SFKs dynamically control kinase anchoring to the membrane. Specifically, all SFKs share the SH4 myristoylation sites (G2), a PTM that is necessary but not sufficient for membrane anchoring. Multiple positively charged residues at SH4, and the unique lipid binding regions of Unique domain, also markedly contribute to lipid binding as well. More importantly, in the case of Src, the lipid interactions with SH4 and Unique are disrupted to different extents by different combinations of phosphorylation. Those can either be phosphorylation of SH4 S17 by PKA and/or phosphorylation of Unique T37 and/or S75 by cyclin-dependent kinase (CDK). The lipid binding is disturbed because the phosphorylated residues are negatively charged, which electrostatically repels the negatively charged internal lipids. In addition, except for Src and Blk, other SFK members all have a conserved SH4 cysteine that can be palmitoylated after the N-myristoylation[119]. This dual acylation is critical for kinase localization and trafficking, and also promotes the productive plasma membrane binding.

IDR-localized AS modify PTMs and binding interactions of SFKs

AS within the disordered N-terminus of SFKs can delete PTM and has strong effects on SFKs cellular locations, protein partner binding and signaling pathways. In the case of Lck, the N-terminus-associated-isoforms not only show tissue-specific distribution, but also are devoid of the binding sites for plasma membrane and for other proteins[120] (Fig.7C and Fig.7D). Specifically, alternative installation of exon 1 (which encodes the first 10 residues including G2 myristoylation site and C3/C5 palmitoylation sites) and 1-prime (which encode 35 residues including a CD4/8 binding motif) altered these PTMs and bind sites (Fig.7D), thus may impact subcellular location and binding interactions of Lck, respectively[120]. Type I Lck variants (without exon 1) are speculated to show modified subcellular location and biological functions, the specific alteration remains undefined though[114]. The other member, Hck, is reported to generate an AS-variant without the palmitoylation site at the SH4 domain, which is required for significant plasma

membrane association[121]. As a consequence, the variant of Hck is mainly bound to lysosome membrane rather than plasma membrane, and triggers distinct biological responses as compared to those of the more common AS isoform of Hck[122]. Similarly, altered N-terminus variants are observed in Lyn, generating LynB with 21 missing residues required for protein-binding missing[123]. The modified protein–protein interactions between the most common AS isoform of Lyn and LynB affect their ability to associate with their antigen receptors in mast cell signaling and response[124]. Taking the above studies together, AS and multiple PTMs within the IDRs of SFKs provide the IDP-AS-PTM toolkit for multilevel regulation of kinase membrane anchoring, protein–protein interactions.

Discussion

Following our initial characterization of the functions carried out by IDPs and IDRs[1], we discovered that phosphorylation[20] and AS[26] individually show strong preferential mapping to IDRs as compared to structured protein regions. Later we noticed that p53 has functionally important IDRs that undergo both multiple PTMs as well as nearby AS. This co-localization of PTMs and AS means that the two types of modification can in principle work together to produce enormous signaling complexity[29]. Note that isoforms from a single gene of p53 arise not only by means of AS but also by alternative start or stop sites or by alternative promoter usage[125], but these various mechanisms all lead to the same result—the creation of different-length protein isoforms from a single gene. Others independently discovered that the same combination of features, e.g., co-localized IDRs, AS, and PTMs, are important for regulating the activities of the protein B-cell lymphoma 2 (Bcl-2), a protein important for its complicated roles in controlling apoptosis[126]. Further support for these concepts is provided by the observed correlations between IDPs and the evolution of cell-type diversification, including the modulation of IDP function by AS and PTMs[127].

As we began to look for other proteins that cooperatively use IDPs, AS, and PTMs, we found that these three features were very seldom considered together: one laboratory would focus on PTMs (especially phosphorylation), another on AS, and very few would consider that AS and PTMs, especially clusters of PTMs, preferentially occur in IDRs. For example, in our text-mining experiment to determine whether various IDR-containing pathways and proteins might use IDP-AS-PTM toolkit (Supplement Table S1), this tendency to focus on either PTMs or AS is very evident in the data in Table S1, where the number of papers that consider both AS and PTMs (the 5th column) is far fewer in every case as compared to the number that consider PTMs or AS individually (the 3rd and 4th columns). Thus, it was necessary to choose a few important proteins and carefully search the literature to determine whether or not there is evidence that IDRs, AS, and PTMs collaborate to increase context-dependent signaling complexity. We also noticed that very few of the researchers realized that the AS and PTMs they were studying were colocalized in IDP regions.

As result of these investigations, herein we report clear evidence for co-localization of IDRs, AS and PTMs for three important signaling proteins, and furthermore we report clear evidence that these three features collaborate to enable highly complex, context-dependent signaling that is important for cellular differentiation. Furthermore, we have searched in PubMed for various IDR-containing pathways and proteins for their presence of PTMs or AS, and the results (Table S1) indicate that the use of IDP-AS-PTM toolkit is likely very widespread. What is needed now is further detailed study to determine if IDPs, PTMs and AS are indeed all co-localized in individual proteins for the various pathways and in the various proteins, and, if so, how do these three features collaborate to bring about the complex context-dependent signaling that underlies these various important biological processes. In addition, we collected many IDPs that utilize PTM codes (or other different names used by different researchers, Table S2), and point out that

18

because multiple PTMs and AS both map to these IDPs (Table S2), it is likely that AS contributes to and enhances the complexity of PTM codes. All of these observations support the importance of the proposed IDP-AS-PTM toolkit.

Note that these three protein families exemplify a wide range of signaling proteins, namely a membrane receptor family, a transcription factor family, and a kinase family. We did not know in advance if these three mechanistically divergent protein families used IDPs, AS, and PTMs to create an array of context-dependent signaling outcomes. The result that all three of these very different signaling protein families utilize the IDP-AS-PTM toolkit indicates that this toolkit is very widely used. Indeed, when we learned that the Nobel Prize was awarded for work on the signaling system underlying circadian rhythms[128], we tested whether two of the key proteins shown by the Nobel Laureates to underlie these rhythms, namely clock and period[129], also use the IDP-AS-PTM toolkit. We found that these key proteins do contain large amounts of predicted disorder (data not shown) which contain both multiple PTMs as well as co-localized AS (Table S1 and further work in progress).

Additional proteins related to development[31], proteins related to cancer[2], proteins related to induction of pluripotent stem cells[7], and proteins that were studied specifically to understand the roles of IDRs or flexible regions in their functions[4, 36-38, 130] all show evidence for their likely use of the IDP-AS-PTM toolkit (Table S1). One of these proteins, p53, also shows up multiple times in Table S2, suggesting that PTM codes should also be considered for all these proteins. These data should encourage molecular biologists focusing on cancer or stem cells to test for the use of the IDP-AS-PTM toolkit and also for AS-modulated PTM codes for both of these processes.

The inhibition of NFAT's NLS depends on up to 13 well separated phosphorylation events, and, furthermore, the collaboration among the multiple phosphorylation events leads to a sensitive, switch-like on-off regulation of NFAT's NLS that may also involve simultaneous inhibition of the nuclear export signal (NES) upon dephosphorylation[105]. These data were interpreted in terms of a "conformational switch" resulting from protein-structure based models of allostery[105, 106]. However, the multiple phosphorylation events are located in IDRs, not structured regions, so a classical "conformational switch" seems very unlikely. A potential alternative model is provided by the on-off regulation of Sic1-Cdc4 interaction, which also results from multiple well separated phosphorylation events of Sic1. Unlike the NFAT "conformational switch model" that depends on structure, the model explaining the on-off behavior of Sic1-Cdc4 interactions is based on a flexible IDP having multiple phosphate groups that bind to a single site, so that, at higher levels of phosphorylation, rebinding by one of the other phosphates is the most likely event upon dissociation of a currently bound phosphate. However, at low levels of phosphorylation, escape is more likely than rebinding. The resulting kinetic model shows switch-like on-off behavior as the phosphorylation levels are changed [131, 132]. Such a kinetic model might also explain how multiple phosphorylations regulate NFAT's NLS.

A key question concerns the origin of the IDP-AS-PTM toolkit, and especially the origin of its ability to carry out context-dependent signaling. Tissue- or cell-type-specific AS has been solidly connected to the rewiring of protein signaling pathways[32, 33]. Also, AS has been solidly connected to alterations of gene regulatory networks[26], but to our knowledge it is yet to be shown that tissue- or cell-type-specific AS is directly connected to gene regulation. As for PTMs, such events have been shown to be both tissue- or cell-type-specific[34, 35] and to be capable of rewiring both protein pathways[16] and genetic networks[36-38], but again experiments directly connecting tissue- or cell-type-specific PTMs with rewiring protein pathway or genetic networks

are currently lacking. Also, it is unclear how tissue- or cell-type-specific PTMs arise. Do these arise from tissue- or cell-type specific expression of the proteins responsible for the PTMs or do these arise from tissue- or cell-type specific AS that add or remove the site of modification? Probably both mechanisms are involved. Thus, we encourage the development of experiments to further test the IDP-AS-PTM hypothesis.

Since context-dependent signaling depends on the capacity of tissue- or cell-type-specific PTMs or AS to rewire or remodel protein pathways or gene regulatory networks, it is tempting to suggest that this toolkit originated with multicellularity and became more complex as organism complexity increased. Indeed, the frequency of AS events has been shown to increase as the organism complexity increases[133] and so far AS in many single cell eukaryotes appears to be rather limited. However, there is another alternative, namely that the IDP-AS-PTM toolkit originated in single cellular organisms. This conjecture rests on the observations that single cell eukaryotic organisms have an abundance of IDRs[134], that they have AS events[135] even if such events appear to be rare, at least in the single-cell eukaryotes studied so far, that they use PTMs for signaling[135], and that they are sometimes observed to exist in quite different cellular states[136-138]. Another alternative is that the single-cell organisms having different cellular forms use a simpler toolkit consisting of just IDPs and PTMs; indeed, phosphorylation has already been shown to be important for the development of different cell types for yeast[135], but it remains to be determined for this example whether these important phosphorylation events occur in structured or IDP domains. Thus, we encourage the investigation of whether the IDP-AS-PTM toolkit or a slightly simpler IDP-PTM toolkit plays important roles in the formation of, or maintenance of, the different cell types observed for some single-cell organisms. If the latter, did the IDP-PTM toolkit provide a stepping stone to the IDP-AS-PTM toolkit? We look forward to a wider recognition of the connections among IDPs, AS and PTMs and to experimental tests of

whether these three features do or do not collaborate to form an important developmental toolkit that contributed to the evolution of multicellular organisms.

Finally, here we propose an integration of the context-dependent signaling arising from IDP-AS-PTM with more standard views of cellular differentiation. Explanations for cell-type specialization and specification focus on gene regulatory networks[139], particularly with regard to the use of differential gene expression to regulate cell-type specification[140-143]). Attempts to explain the advantages of cell specialization are based mainly on multilevel selection theory, soma-germ cell line requirements, and what it means to be an “individual”[144-148]. In our view these widely discussed concepts likely account for the broad outlines of cellular differentiation, but we speculate that gene regulatory networks alone are simply too coarse-grained for successful multicellular life. According to this view, the signaling modulations provided by the IDP-AS-PTM toolkit lead to the fine-tuning of the cell-cell signaling interactions provided by differential gene expression, a fine-tuning brought about by modulating the signaling interactions of these very same differentially expressed genes. This fine-tuning is proposed to enable individual cells to more appropriately respond to the various signals received from their surroundings, thereby promoting the integration of the different cell types into a more successful multicellular organism.

Materials and Methods

Dataset construction

The sequences for human GPCRs, NFATs, and SFKs are collected from UniProt. Initially, a full list of GPCR members from all species were obtained from the UniProtKB/Swiss-Prot document named 7tmrlist.txt. Only human GPCRs were included by searching the single keyword “Human” in the species column. A sequence identity cutoff of 100% was used to remove

redundancy by using CD-Hit (with other parameters default), and a total of 822 GPCR sequences with sequence identity range from 1.5% to 98.7% (average: 20%, standard deviation:10.9%) were retrieved. The names of the five families of NFATs and nine families of SFKs are identified from [100] and [115], respectively. One study reported herein is to carry out the alignment of disorder predictions across the nine SFKs families using the human sequences for these comparisons. However, in the human SFKs, the Yes kinase family member is a pseudogene, so the chicken Yes sequence was used in place of the human pseudogene for these comparisons. The sequence identity of the five NFAT sequences ranges from 26% to 47%, with an average of 36% and standard deviation of 8%. The sequence identity of the nine SFK sequences ranges from 55% to 79%, with an average of 63% and standard deviation of 6.9%.

The initial set of human TFs sequences (including 1691 gene members belonging to ~70 families) are retrieved from AminoAcidTFDB database[149], and a sequence identity cutoff of 100% was used to remove redundancy. Among the retrieved 1568 sequences, 1392 of them can be assigned to 99 specific DNA-binding domains (DBDs) by Pfam. There are also Pfam domains outside the DBDs, which are called Non-DBDs. The segments that have not yet been assigned with any Pfam domains are named unassigned regions (URs).

Disorder prediction

PONDR FIT was used for all of the disorder predictions reported herein. This is a meta-predictor that uses a neural network to combine the normalized outputs of six different disorder predictors, namely PONDR VLXT, PONDR VSL2, PONDR VL3, IUPred, FoldIndex, and TopIDP. Overall, this meta-predictor outperforms all of the individual predictors for nearly every protein and by an average of about 11%[134]. This predictor can be accessed at the following URL: http://disorder.compbio.iupui.edu/meta_predictor.php#PONDR-FIT.

Identification of AS events and PTM sites

AS events were retrieved from UniProt[150] by using the keywords “Event=Alternative splicing” in the CC table of the downloaded flat format files. UniProt provides the information on how each spliced isoform differs from the canonical sequence and, more importantly, often includes the related literature suggesting the functional relevance or tissue distribution of AS-isoforms. In addition, “alternative spliced variants + protein name” were used to search for additional information on tissue-specific AS.

PTMs information was extracted from both UniProt and PhosphoSitePlus[57]. Annotations in UniProt based on sequence similarity were not included in the current study. Phosphorylation, acetylation, methylation and sulfation were obtained from “MOD_RES” in the features table (FT) of UniProt flat format file. Glycosylation and palmitoylation were retrieved by using “CARBOHYD” and “LIPID” keywords, respectively. Other PTMs sites, especially ubiquitination, were obtained in PhosphoSitePlus.

Acknowledgements This work was supported by the Science and Technology Commission of Shanghai Municipality grant 16ZR1448500 and ShanghaiTech University.

References

- [1] Dunker A.K., Brown C.J., Lawson J.D., Iakoucheva L.M., Obradovic Z. (2002). Intrinsic disorder and protein function. *Biochemistry*. 41, 6573-6582.
- [2] Iakoucheva L.M., Brown C.J., Lawson J.D., Obradovic Z., Dunker A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 323, 573-584.
- [3] Ward J.J., Sodhi J.S., McGuffin L.J., Buxton B.F., Jones D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*. 337, 635-645.
- [4] Galea C.A., Wang Y., Sivakolundu S.G., Kriwacki R.W. (2008). Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*. 47, 7598-7609.

- [5] Babu M.M., van der Lee R., de Groot N.S., Gsponer J. (2011). Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol.* 21, 432-440.
- [6] Tantos A., Han K.H., Tompa P. (2012). Intrinsic disorder in cell signaling and gene transcription. *Mol Cell Endocrinol.* 348, 457-465.
- [7] Xue B., Oldfield C.J., Van Y.Y., Dunker A.K., Uversky V.N. (2012). Protein intrinsic disorder and induced pluripotent stem cells. *Mol Biosyst.* 8, 134-150.
- [8] Yruela I. (2015). Plant development regulation: Overview and perspectives. *J Plant Physiol.* 182, 62-78.
- [9] Wright P.E., Dyson H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* 16, 18-29.
- [10] Xie Q., Arnold G.E., Romero P., Obradovic Z., Garner E., Dunker A.K. (1998). The Sequence Attribute Method for Determining Relationships Between Sequence and Protein Disorder. *Genome Inform Ser Workshop Genome Inform.* 9, 193-200.
- [11] Dunker A.K., Brown C.J., Obradovic Z. (2002). Identification and functions of usefully disordered proteins. *Adv Protein Chem.* 62, 25-49.
- [12] Uversky V.N., Gillespie J.R., Fink A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins.* 41, 415-427.
- [13] Varadi M., Kosol S., Lebrun P., Valentini E., Blackledge M., Dunker A.K., et al. (2014). pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.* 42, D326-335.
- [14] Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Uversky V.N., et al. (2007). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res.* 6, 1882-1898.
- [15] Oldfield C.J., Meng J., Yang J.Y., Yang M.Q., Uversky V.N., Dunker A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics.* 9 Suppl 1, S1.
- [16] Hsu W.L., Oldfield C.J., Xue B., Meng J., Huang F., Romero P., et al. (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* 22, 258-273.
- [17] Dunker A.K., Cortese M.S., Romero P., Iakoucheva L.M., Uversky V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129-5148.
- [18] Haynes C., Oldfield C.J., Ji F., Klitgord N., Cusick M.E., Radivojac P., et al. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol.* 2, e100.
- [19] Kim P.M., Sboner A., Xia Y., Gerstein M. (2008). The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol.* 4, 179.

- [20] Iakoucheva L.M., Radivojac P., Brown C.J., O'Connor T.R., Sikes J.G., Obradovic Z., et al. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32, 1037-1049.
- [21] Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Obradovic Z., et al. (2007). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res.* 6, 1917-1932.
- [22] Gao J., Xu D. (2012). Correlation between posttranslational modification and intrinsic disorder in protein. *Pac Symp Biocomput.* 94-103.
- [23] Pejaver V., Hsu W.L., Xin F., Dunker A.K., Uversky V.N., Radivojac P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* 23, 1077-1093.
- [24] Huang Q., Chang J., Cheung M.K., Nong W., Li L., Lee M.T., et al. (2014). Human proteins with target sites of multiple post-translational modification types are more prone to be involved in disease. *J Proteome Res.* 13, 2735-2748.
- [25] Ota M., Gonja H., Koike R., Fukuchi S. (2016). Multiple-Localization and Hub Proteins. *PLoS One.* 11, e0156455.
- [26] Romero P.R., Zaidi S., Fang Y.Y., Uversky V.N., Radivojac P., Oldfield C.J., et al. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A.* 103, 8390-8395.
- [27] Strahl B.D., Allis C.D. (2000). The language of covalent histone modifications. *Nature.* 403, 41-45.
- [28] Benayoun B.A., Veitia R.A. (2009). A post-translational modification code for transcription factors: sorting through a sea of signals. *Trends Cell Biol.* 19, 189-197.
- [29] Dunker A.K., Silman I., Uversky V.N., Sussman J.L. (2008). Function and structure of inherently disordered proteins. *Curr Opin Struct Biol.* 18, 756-764.
- [30] Niklas K.J., Bondos S.E., Dunker A.K., Newman S.A. (2015). Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front Cell Dev Biol.* 3, 8.
- [31] Dunker A.K., Bondos S.E., Huang F., Oldfield C.J. (2015). Intrinsically disordered proteins and multicellular organisms. *Semin Cell Dev Biol.* 37, 44-55.
- [32] Buljan M., Chalancon G., Eustermann S., Wagner G.P., Fuxreiter M., Bateman A., et al. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell.* 46, 871-883.
- [33] Ellis J.D., Barrios-Rodiles M., Colak R., Irimia M., Kim T., Calarco J.A., et al. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell.* 46, 884-892.
- [34] Abdel-Hafiz H.A., Horwitz K.B. (2014). Post-translational modifications of the progesterone receptors. *J Steroid Biochem Mol Biol.* 140, 80-89.

- [35] Hussain M.M., Bucher N.L., Faris B., Franzblau C., Zannis V.I. (1988). Tissue-specific posttranslational modification of rat apoE. Synthesis of sialated apoE forms by neonatal rat aortic smooth muscle cells. *J Lipid Res.* 29, 915-923.
- [36] Slupsky C.M., Gentile L.N., Donaldson L.W., Mackereth C.D., Seidel J.J., Graves B.J., et al. (1998). Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proc Natl Acad Sci U S A.* 95, 12129-12134.
- [37] Liu Y., Matthews K.S., Bondos S.E. (2008). Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* hox protein ultrabithorax. *J Biol Chem.* 283, 20874-20887.
- [38] Nelson M.L., Kang H.S., Lee G.M., Blaszczyk A.G., Lau D.K., McIntosh L.P., et al. (2010). Ras signaling requires dynamic properties of Ets1 for phosphorylation-enhanced binding to coactivator CBP. *Proc Natl Acad Sci U S A.* 107, 10026-10031.
- [39] Liu J., Perumal N.B., Oldfield C.J., Su E.W., Uversky V.N., Dunker A.K. (2006). Intrinsic disorder in transcription factors. *Biochemistry.* 45, 6873-6888.
- [40] Minezaki Y., Homma K., Kinjo A.R., Nishikawa K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol.* 359, 1137-1149.
- [41] Krasowski M.D., Reschly E.J., Ekins S. (2008). Intrinsic disorder in nuclear hormone receptors. *J Proteome Res.* 7, 4359-4372.
- [42] Xue B., Li L., Meroueh S.O., Uversky V.N., Dunker A.K. (2009). Analysis of structured and intrinsically disordered regions of transmembrane proteins. *Mol Biosyst.* 5, 1688-1702.
- [43] Minezaki Y., Homma K., Nishikawa K. (2007). Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *J Mol Biol.* 368, 902-913.
- [44] Peng Z., Mizianty M.J., Xue B., Kurgan L., Uversky V.N. (2012). More than just tails: intrinsic disorder in histone proteins. *Mol Biosyst.* 8, 1886-1901.
- [45] Korneta I., Bujnicki J.M. (2012). Intrinsic disorder in the human spliceosomal proteome. *PLoS Comput Biol.* 8, e1002641.
- [46] Coelho Ribeiro Mde L., Espinosa J., Islam S., Martinez O., Thanki J.J., Mazariegos S., et al. (2013). Malleable ribonucleoprotein machine: protein intrinsic disorder in the *Saccharomyces cerevisiae* spliceosome. *PeerJ.* 1, e2.
- [47] Peng Z., Oldfield C.J., Xue B., Mizianty M.J., Dunker A.K., Kurgan L., et al. (2014). A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell Mol Life Sci.* 71, 1477-1504.
- [48] Wang C., Uversky V.N., Kurgan L. (2016). Disordered nucleome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics.* 16, 1486-1498.

- [49] DeForte S., Uversky V.N. (2017). Not an exception to the rule: the functional significance of intrinsically disordered protein regions in enzymes. *Mol Biosyst.* 13, 463-469.
- [50] Haynes C., Iakoucheva L.M. (2006). Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* 34, 305-312.
- [51] Cortese M.S., Uversky V.N., Dunker A.K. (2008). Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol.* 98, 85-106.
- [52] Trudeau T., Nassar R., Cumberworth A., Wong E.T., Woollard G., Gsponer J. (2013). Structure and intrinsic disorder in protein autoinhibition. *Structure.* 21, 332-341.
- [53] Guharoy M., Szabo B., Contreras Martos S., Kosol S., Tompa P. (2013). Intrinsic structural disorder in cytoskeletal proteins. *Cytoskeleton (Hoboken).* 70, 550-571.
- [54] Uversky V.N., Oldfield C.J., Midic U., Xie H., Xue B., Vucetic S., et al. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics.* 10 Suppl 1, S7.
- [55] van der Lee R., Buljan M., Lang B., Weatheritt R.J., Daughdrill G.W., Dunker A.K., et al. (2014). Classification of intrinsically disordered regions and proteins. *Chem Rev.* 114, 6589-6631.
- [56] Piovesan D., Tabaro F., Micetic I., Necci M., Quaglia F., Oldfield C.J., et al. (2017). DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* 45, D1123-D1124.
- [57] 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.
- [58] Katritch V., Cherezov V., Stevens R.C. (2013). Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol.* 53, 531-556.
- [59] Venkatakrisnan A.J., Deupi X., Lebon G., Tate C.G., Schertler G.F., Babu M.M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature.* 494, 185-194.
- [60] Sigler P.B. (1988). Transcriptional activation. Acid blobs and negative noodles. *Nature.* 333, 210-212.
- [61] Pan M.G., Xiong Y., Chen F. (2013). NFAT gene family in inflammation and cancer. *Curr Mol Med.* 13, 543-554.
- [62] Park S., Uesugi M., Verdine G.L. (2000). A second calcineurin binding site on the NFAT regulatory domain. *Proc Natl Acad Sci U S A.* 97, 7130-7135.
- [63] Kathiriya J.J., Pathak R.R., Clayman E., Xue B., Uversky V.N., Dave V. (2014). Presence and utility of intrinsically disordered regions in kinases. *Mol Biosyst.* 10, 2876-2888.
- [64] Espada J., Martin-Perez J. (2017). An Update on Src Family of Nonreceptor Tyrosine Kinases Biology. *Int Rev Cell Mol Biol.* 331, 83-122.
- [65] Jaakola V.P., Prilusky J., Sussman J.L., Goldman A. (2005). G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng Des Sel.* 18, 103-110.

- [66] Tovo-Rodrigues L., Roux A., Hutz M.H., Rohde L.A., Woods A.S. (2014). Functional characterization of G-protein-coupled receptors: a bioinformatics approach. *Neuroscience*. 277, 764-779.
- [67] Venkatakrisnan A.J., Flock T., Prado D.E., Oates M.E., Gough J., Madan Babu M. (2014). Structured and disordered facets of the GPCR fold. *Curr Opin Struct Biol*. 27, 129-137.
- [68] Veldkamp C.T., Seibert C., Peterson F.C., De la Cruz N.B., Haugner J.C., 3rd, Basnet H., et al. (2008). Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal*. 1, ra4.
- [69] Chen A.S., Kim Y.M., Gayen S., Huang Q., Raida M., Kang C. (2011). NMR structural study of the intracellular loop 3 of the serotonin 5-HT(1A) receptor and its interaction with calmodulin. *Biochim Biophys Acta*. 1808, 2224-2232.
- [70] Seebahn A., Dinkel H., Mohrluder J., Hartmann R., Vogel N., Becker C.M., et al. (2011). Structural characterization of intracellular C-terminal domains of group III metabotropic glutamate receptors. *FEBS Lett*. 585, 511-516.
- [71] Markovic D., Grammatopoulos D.K. (2009). Focus on the splicing of secretin GPCRs transmembrane-domain 7. *Trends Biochem Sci*. 34, 443-452.
- [72] Oladosu F.A., Maixner W., Nackley A.G. (2015). Alternative Splicing of G Protein-Coupled Receptors: Relevance to Pain Management. *Mayo Clin Proc*. 90, 1135-1151.
- [73] Dal Toso R., Sommer B., Ewert M., Herb A., Pritchett D.B., Bach A., et al. (1989). The dopamine D2 receptor: two molecular forms generated by alternative splicing. *EMBO J*. 8, 4025-4034.
- [74] Seeman P., Nam D., Ulpian C., Liu I.S., Tallerico T. (2000). New dopamine receptor, D2(Longer), with unique TG splice site, in human brain. *Brain Res Mol Brain Res*. 76, 132-141.
- [75] Guiramand J., Montmayeur J.P., Ceraline J., Bhatia M., Borrelli E. (1995). Alternative splicing of the dopamine D2 receptor directs specificity of coupling to G-proteins. *J Biol Chem*. 270, 7354-7358.
- [76] Watson S.J., Brown A.J., Holliday N.D. (2012). Differential signaling by splice variants of the human free fatty acid receptor GPR120. *Mol Pharmacol*. 81, 631-642.
- [77] Berchiche Y.A., Sakmar T.P. (2016). CXC Chemokine Receptor 3 Alternative Splice Variants Selectively Activate Different Signaling Pathways. *Mol Pharmacol*. 90, 483-495.
- [78] Rosenbaum D.M., Rasmussen S.G., Kobilka B.K. (2009). The structure and function of G-protein-coupled receptors. *Nature*. 459, 356-363.
- [79] Tobin A.B., Butcher A.J., Kong K.C. (2008). Location, location, location...site-specific GPCR phosphorylation offers a mechanism for cell-type-specific signalling. *Trends Pharmacol Sci*. 29, 413-420.
- [80] Butcher A.J., Prihandoko R., Kong K.C., McWilliams P., Edwards J.M., Bottrill A., et al. (2011). Differential G-protein-coupled receptor phosphorylation provides evidence for a signaling bar code. *J Biol Chem*. 286, 11506-11518.

- [81] Zhou X.E., He Y., de Waal P.W., Gao X., Kang Y., Van Eps N., et al. (2017). Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell*. 170, 457-469 e413.
- [82] Zheng H., Loh H.H., Law P.Y. (2013). Posttranslation modification of G protein-coupled receptor in relationship to biased agonism. *Methods Enzymol*. 522, 391-408.
- [83] Marada S., Navarro G., Truong A., Stewart D.P., Arensdorf A.M., Nachtergaele S., et al. (2015). Functional Divergence in the Role of N-Linked Glycosylation in Smoothed Signaling. *PLoS Genet*. 11, e1005473.
- [84] Skieterska K., Rondou P., Van Craenenbroeck K. (2017). Regulation of G Protein-Coupled Receptors by Ubiquitination. *Int J Mol Sci*. 18.
- [85] Prabakaran S., Lippens G., Steen H., Gunawardena J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med*. 4, 565-583.
- [86] Sadeghi H.M., Innamorati G., Dagarag M., Birnbaumer M. (1997). Palmitoylation of the V2 vasopressin receptor. *Mol Pharmacol*. 52, 21-29.
- [87] Stannard C., Lehenkari P., Godovac-Zimmermann J. (2003). Functional diversity of endothelin pathways in human lung fibroblasts may be based on structural diversity of the endothelin receptors. *Biochemistry*. 42, 13909-13918.
- [88] Zheng H., Pearsall E.A., Hurst D.P., Zhang Y., Chu J., Zhou Y., et al. (2012). Palmitoylation and membrane cholesterol stabilize mu-opioid receptor homodimerization and G protein coupling. *BMC Cell Biol*. 13, 6.
- [89] Gonzalez A., Borquez M., Trigo C.A., Brenet M., Sarmiento J.M., Figueroa C.D., et al. (2011). The splice variant of the V2 vasopressin receptor adopts alternative topologies. *Biochemistry*. 50, 4981-4986.
- [90] Sarmiento J.M., Anazco C.C., Campos D.M., Prado G.N., Navarro J., Gonzalez C.B. (2004). Novel down-regulatory mechanism of the surface expression of the vasopressin V2 receptor by an alternative splice receptor variant. *J Biol Chem*. 279, 47017-47023.
- [91] Okamoto Y., Ninomiya H., Tanioka M., Sakamoto A., Miwa S., Masaki T. (1997). Palmitoylation of human endothelinB. Its critical role in G protein coupling and a differential requirement for the cytoplasmic tail by G protein subtypes. *J Biol Chem*. 272, 21589-21596.
- [92] Reid H.M., Kinsella B.T. (2003). The alpha, but not the beta, isoform of the human thromboxane A2 receptor is a target for nitric oxide-mediated desensitization. Independent modulation of Tp alpha signaling by nitric oxide and prostacyclin. *J Biol Chem*. 278, 51190-51202.
- [93] Kelley-Hickie L.P., O'Keeffe M.B., Reid H.M., Kinsella B.T. (2007). Homologous desensitization of signalling by the alpha (alpha) isoform of the human thromboxane A2 receptor: a specific role for nitric oxide signalling. *Biochim Biophys Acta*. 1773, 970-989.
- [94] Guo X., Bulyk M.L., Hartemink A.J. (2012). Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac Symp Biocomput*. 104-115.

- [95] Yruela I., Oldfield C.J., Niklas K.J., Dunker A.K. (2017). Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity. *Genome Biol Evol.* 9, 1248-1265.
- [96] Taneri B., Snyder B., Novoradovsky A., Gaasterland T. (2004). Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. *Genome Biol.* 5, R75.
- [97] Jackson Behan K., Fair J., Singh S., Bogwitz M., Perry T., Grubor V., et al. (2005). Alternative splicing removes an Ets interaction domain from Lozenge during *Drosophila* eye development. *Dev Genes Evol.* 215, 423-435.
- [98] Tootle T.L., Rebay I. (2005). Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays.* 27, 285-298.
- [99] Meek D.W., Anderson C.W. (2009). Posttranslational modification of p53: cooperative integrators of function. *Cold Spring Harb Perspect Biol.* 1, a000950.
- [100] Hogan P.G., Chen L., Nardone J., Rao A. (2003). Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* 17, 2205-2232.
- [101] Vihma H., Pruunsild P., Timmusk T. (2008). Alternative splicing and expression of human and mouse NFAT genes. *Genomics.* 92, 279-291.
- [102] Sheridan C.M., Heist E.K., Beals C.R., Crabtree G.R., Gardner P. (2002). Protein kinase A negatively modulates the nuclear accumulation of NF-ATc1 by priming for subsequent phosphorylation by glycogen synthase kinase-3. *J Biol Chem.* 277, 48664-48676.
- [103] Vihma H., Timmusk T. (2017). Sumoylation regulates the transcriptional activity of different human NFAT isoforms in neurons. *Neurosci Lett.* 653, 302-307.
- [104] Johnson L.N., O'Reilly M. (1996). Control by phosphorylation. *Curr Opin Struct Biol.* 6, 762-769.
- [105] Okamura H., Aramburu J., Garcia-Rodriguez C., Viola J.P., Raghavan A., Tahliliani M., et al. (2000). Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. *Mol Cell.* 6, 539-550.
- [106] Salazar C., Hofer T. (2003). Allosteric regulation of the transcription factor NFAT1 by multiple phosphorylation sites: a mathematical analysis. *J Mol Biol.* 327, 31-45.
- [107] Nayak A., Glockner-Pagel J., Vaeth M., Schumann J.E., Buttmann M., Bopp T., et al. (2009). Sumoylation of the transcription factor NFATc1 leads to its subnuclear relocalization and interleukin-2 repression by histone deacetylase. *J Biol Chem.* 284, 10935-10946.
- [108] Park J., Takeuchi A., Sharma S. (1996). Characterization of a new isoform of the NFAT (nuclear factor of activated T cells) gene family member NFATc. *J Biol Chem.* 271, 20914-20921.
- [109] Lucena P.I., Faget D.V., Pachulec E., Robaina M.C., Klumb C.E., Robbs B.K., et al. (2016). NFAT2 Isoforms Differentially Regulate Gene Expression, Cell Death, and Transformation through Alternative N-Terminal Domains. *Mol Cell Biol.* 36, 119-131.

- [110] Thomas S.M., Brugge J.S. (1997). Cellular functions regulated by Src family kinases. *Annu Rev Cell Dev Biol.* 13, 513-609.
- [111] Oppermann H., Levinson A.D., Varmus H.E., Levintow L., Bishop J.M. (1979). Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). *Proc Natl Acad Sci U S A.* 76, 1804-1808.
- [112] Hunter T., Sefton B.M. (1980). Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A.* 77, 1311-1315.
- [113] Perez Y., Gairi M., Pons M., Bernado P. (2009). Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *J Mol Biol.* 391, 136-148.
- [114] Kim P.W., Sun Z.Y., Blacklow S.C., Wagner G., Eck M.J. (2003). A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science.* 301, 1725-1728.
- [115] Roskoski R., Jr. (2004). Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun.* 324, 1155-1164.
- [116] Perez Y., Maffei M., Igea A., Amata I., Gairi M., Nebreda A.R., et al. (2013). Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci Rep.* 3, 1295.
- [117] Maffei M., Arbesu M., Le Roux A.L., Amata I., Roche S., Pons M. (2015). The SH3 Domain Acts as a Scaffold for the N-Terminal Intrinsically Disordered Regions of c-Src. *Structure.* 23, 893-902.
- [118] Arbesu M., Maffei M., Cordeiro T.N., Teixeira J.M., Perez Y., Bernado P., et al. (2017). The Unique Domain Forms a Fuzzy Intramolecular Complex in Src Family Kinases. *Structure.* 25, 630-640 e634.
- [119] Liang X., Lu Y., Wilkes M., Neubert T.A., Resh M.D. (2004). The N-terminal SH4 region of the Src family kinase Fyn is modified by methylation and heterogeneous fatty acylation: role in membrane targeting, cell adhesion, and spreading. *J Biol Chem.* 279, 8133-8139.
- [120] Rouer E., Benarous R. (1992). Alternative splicing in the human lck gene leads to the deletion of exon 1' and results in a new type II lck transcript. *Oncogene.* 7, 2535-2538.
- [121] Lock P., Ralph S., Stanley E., Boulet I., Ramsay R., Dunn A.R. (1991). Two isoforms of murine hck, generated by utilization of alternative translational initiation codons, exhibit different patterns of subcellular localization. *Mol Cell Biol.* 11, 4363-4370.
- [122] Carreno S., Caron E., Cougoule C., Emorine L.J., Maridonneau-Parini I. (2002). p59Hck isoform induces F-actin reorganization to form protrusions of the plasma membrane in a Cdc42- and Rac-dependent manner. *J Biol Chem.* 277, 21007-21016.
- [123] Yamanashi Y., Miyasaka M., Takeuchi M., Ilic D., Mizuguchi J., Yamamoto T. (1991). Differential responses of p56lyn and p53lyn, products of alternatively spliced lyn mRNA, on stimulation of B-cell antigen receptor. *Cell Regul.* 2, 979-987.

- [124] Alvarez-Errico D., Yamashita Y., Suzuki R., Odom S., Furumoto Y., Yamashita T., et al. (2010). Functional analysis of Lyn kinase A and B isoforms reveals redundant and distinct roles in Fc epsilon RI-dependent mast cell activation. *J Immunol.* 184, 5000-5008.
- [125] Khoury M.P., Bourdon J.C. (2011). p53 Isoforms: An Intracellular Microprocessor? *Genes Cancer.* 2, 453-465.
- [126] Rautureau G.J., Day C.L., Hinds M.G. (2010). Intrinsically disordered proteins in bcl-2 regulated apoptosis. *Int J Mol Sci.* 11, 1808-1824.
- [127] Niklas K.J., Dunker A.K., Yruela I. (2018). The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J Exp Bot.* 69, 1437-1446.
- [128] Burki T. (2017). Nobel Prize awarded for discoveries in circadian rhythm. *Lancet.* 390, e25.
- [129] Price J.L., Dembinska M.E., Young M.W., Rosbash M. (1995). Suppression of PERIOD protein abundance and circadian cycling by the *Drosophila* clock mutation timeless. *EMBO J.* 14, 4044-4049.
- [130] Jie J., Lohr F., Barbar E. (2017). Dynein Binding of Competitive Regulators Dynactin and NudE Involves Novel Interplay between Phosphorylation Site and Disordered Spliced Linkers. *Structure.* 25, 421-433.
- [131] Mittag T., Orlicky S., Choy W.Y., Tang X., Lin H., Sicheri F., et al. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A.* 105, 17772-17777.
- [132] Klein P., Pawson T., Tyers M. (2003). Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. *Curr Biol.* 13, 1669-1678.
- [133] Chen L., Bush S.J., Tovar-Corona J.M., Castillo-Morales A., Urrutia A.O. (2014). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol.* 31, 1402-1413.
- [134] Xue B., Dunker A.K., Uversky V.N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn.* 30, 137-149.
- [135] Hossain M.A., Rodriguez C.M., Johnson T.L. (2011). Key features of the two-intron *Saccharomyces cerevisiae* gene *SUS1* contribute to its alternative splicing. *Nucleic Acids Res.* 39, 8612-8627.
- [136] Herskowitz I. (1989). A regulatory hierarchy for cell specialization in yeast. *Nature.* 342, 749-757.
- [137] Niklas K.J. (2014). The evolutionary-developmental origins of multicellularity. *Am J Bot.* 101, 6-25.
- [138] Duran-Nebreda S., Bonforti A., Montanez R., Valverde S., Sole R. (2016). Emergence of proto-organisms from bistable stochastic differentiation and adhesion. *J R Soc Interface.* 13.
- [139] Britten R.J., Davidson E.H. (1969). Gene regulation for higher cells: a theory. *Science.* 165, 349-357.

- [140] Carroll S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 134, 25-36.
- [141] Erwin D.H., Davidson E.H. (2009). The evolution of hierarchical gene regulatory networks. *Nat Rev Genet*. 10, 141-148.
- [142] Levine M., Cattoglio C., Tjian R. (2014). Looping back to leap forward: transcription enters a new era. *Cell*. 157, 13-25.
- [143] Arenas-Mena C. (2017). The origins of developmental gene regulation. *Evol Dev*. 19, 96-107.
- [144] Labarbera M. (1988). *The Evolution of Individuality*. Leo W. Buss. Princeton University Press, Princeton, NJ, 1988. xvi, 203 pp., illus. \$40; paper, \$12.95. *Science*. 240, 1073-1074.
- [145] John Damuth I.L.H. (1988). Alternative formulations of multilevel selection. *Biology and Philosophy*. 3, 407-430.
- [146] Michod R.E. (1997). Evolution of the individual. *Am Nat*. 150 Suppl 1, S5-S21.
- [147] Bonner J. (1998). The origins of multicellularity. *Intergrative Biology: Issues, News, and Reviews*. 1, 27-36.
- [148] Folse H.J., 3rd, Roughgarden J. (2010). What is an individual organism? A multilevel selection perspective. *Q Rev Biol*. 85, 447-472.
- [149] Zhang H.M., Liu T., Liu C.J., Song S., Zhang X., Liu W., et al. (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res*. 43, D76-81.
- [150] UniProt C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res*. 43, D204-212.
- [151] Xue B., Dunbrack R.L., Williams R.W., Dunker A.K., Uversky V.N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 1804, 996-1010.
- [152] Seibert C., Veldkamp C.T., Peterson F.C., Chait B.T., Volkman B.F., Sakmar T.P. (2008). Sequential tyrosine sulfation of CXCR4 by tyrosylprotein sulfotransferases. *Biochemistry*. 47, 11251-11262.
- [153] Chabot D.J., Chen H., Dimitrov D.S., Broder C.C. (2000). N-linked glycosylation of CXCR4 masks coreceptor function for CCR5-dependent human immunodeficiency virus type 1 isolates. *J Virol*. 74, 4404-4413.
- [154] Ahr B., Denizot M., Robert-Hebmann V., Brelot A., Biard-Piechaczyk M. (2005). Identification of the cytoplasmic domains of CXCR4 involved in Jak2 and STAT3 phosphorylation. *J Biol Chem*. 280, 6692-6700.
- [155] Busillo J.M., Armando S., Sengupta R., Meucci O., Bouvier M., Benovic J.L. (2010). Site-specific phosphorylation of CXCR4 is dynamically regulated by multiple kinases and results in differential modulation of CXCR4 signaling. *J Biol Chem*. 285, 7805-7817.

[156] Marchese A., Raiborg C., Santini F., Keen J.H., Stenmark H., Benovic J.L. (2003). The E3 ubiquitin ligase AIP4 mediates ubiquitination and sorting of the G protein-coupled receptor CXCR4. *Dev Cell.* 5, 709-722.

[157] Sheftic S.R., Page R., Peti W. (2016). Investigating the human Calcineurin Interaction Network using the piLxVP SLiM. *Sci Rep.* 6, 38920.

ACCEPTED MANUSCRIPT

Figure Legends

Figure 1 Abundance of alternative splicing (AS) and post-translational modifications (PTMs) within IDRs of human GPCRs. (A) Distribution of average percentage of predicted disordered residues (score ≥ 0.5 predicted by PONDR-FIT predictor [151], green histogram) and AS regions (marked purple line) within each region. (B) Distribution of PTMs (stacked bar, multiple colors). (C) AS and PTMs may cooperate to further enhance GPCR signaling complexity. Asterisks (*) indicate GPCRs with reported tissue/region-specific AS.

Figure 2. AS-modulated PTMs are mostly localized within IDRs of GPCRs. (A) Fraction of all GPCR AS regions with and without known PTMs. (B) Distribution of disorder prediction for AS regions with and without known PTMs (red and blue histogram, respectively). X-axis is the range of predicted disorder. Y-axis is the percentage of AS regions within each bin.

Figure 3. IDR-localized AS and PTMs regulate differential CXCR4 signaling. Predicted disorder by PONDR-FIT is shown as heat map (lower left), where red and blue indicate predicted disorder and order, respectively. A crystal structure of the structured regions (28–303 residues, PDB id: 3OE9) is represented in blue ribbon. AS generate three tissue-specific isoforms by replacing the first five residues at the disordered N-terminus and strongly affect receptor functions. Multiple PTM combinations lead to different signaling: sulfation of Y7, Y12, and Y21 regulates receptor-ligand binding and dimerization[152], and glycosylation of N11 plays a role in masking the coreceptor functional activity[153]; Phosphorylation of Y157 is required for activation of the G β -independent JAK2/STAT3 pathway[154]. C-terminal PTM combinations have been reported to associate with three different biological processes. Specifically, GRK6 phosphorylation (S339), and possibly GRK2 phosphorylation (two residues from S346-S348 and S351-S352), lead to receptor-arrestin3 binding, G-protein uncoupling and subsequent receptor desensitization. In contrast, GRK3 (at the same regions as GRK2, but probably different residues), and GRK6 (S330 and S339) phosphorylation result in arrestin2 recruitment and subsequent ERK1/2 activation[155]. Additionally, PKC and GRK6 phosphorylation (S324 or S325, S330 respectively) initiate degradation modulated by ubiquitination of K327, K331, and K333[155, 156].

Figure 4 Both AS and PTMs of human transcription factors (TFs) are significantly abundant within IDRs, and multiple PTMs are very common. (A) Distribution of percentage of predicted disorder (Green boxplot) and the number of TFs with annotated AS regions (purple-border

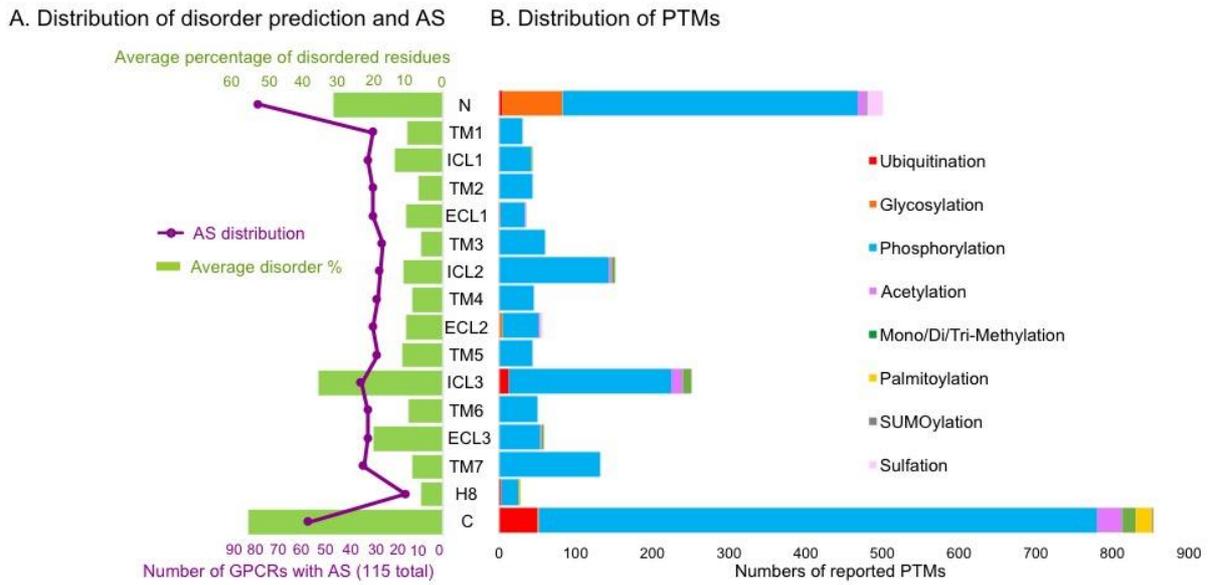
histogram) for three different sets of regions, including Pfam DNA-binding domains (DBD), Pfam Non-DBD domains (NonDBD) and unassigned regions (URs). Note that each set of region contains more outliers than it indicates in the graph, and the default value 1.5 was used to define the outliers. The median value for each set is indicated as a thick line. Statistical significance between sets was assessed by Wilcoxon rank sum test, and p values are indicated. (B) Distribution of PTM sites. Different types of PTMs are indicated in distinct colors. (C) Percentage of TFs with 1, 2, 3, 4, ≥ 5 types of PTMs. (D) Multiple PTMs targeting the same residue.

Figure 5. AS-modulated PTMs are mostly localized within IDRs of TFs. (A) Fraction of all TF AS regions with and without known PTMs. (B) Distribution of disorder prediction for AS regions with and without known PTMs (red and blue histogram, respectively). X-axis is the range of predicted disorder. Y-axis is the percentage of AS regions within each bin.

Figure 6 NFATs and the IDP-AS-PTM toolkit. (A) Disorder prediction of the five members of NFATs (same heat map as Fig.2). (B) The splice variants of NFATc1. The missing regions are indicated in dash line and replaced segments are presented in green boxes. (C) Multiple PTMs from PhosphoSitePlus and available publications mainly localize within the IDRs of NFATc1. PTMs are indicated by different colors. The important IDR-localized functional regions are: two calcineurin-binding motifs (PxIxIT and LxVP, x indicates any residues), nuclear localization signal (NLS), three serine-proline-rich repeat motifs (SP1–3) and two serine-rich regions (SRR1&2). The structure of the second LxVP short motif is from the NFATc1–calcineurin binding complex (384-390 residues, PDB id: 5SVE)[157].

Figure 7 IDR-localized AS regions modulate PTMs and binding interactions of SFKs. (A) Disorder prediction of the nine members of SFKs by PONDR-FIT (same heap map as in Fig.2). (B) Domain architecture of SFKs: SH4 (yellow), Unique (red), SH3 (blue), SH2 (green), SH1 (tan). Linkers are colored in red line. (C) Gene structures and N-terminal alternative variants of Lck gene. Exons are shown in colored boxes and introns are shown as lines. (D) Multiple functional PTMs of the disordered N-terminal region of Lck.

Figure 1



C AS may further modulate PTMs

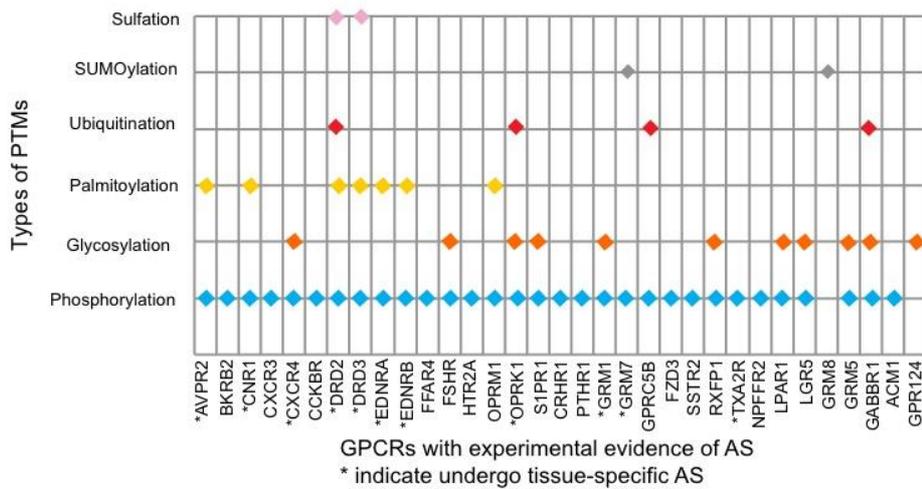


Figure 2.

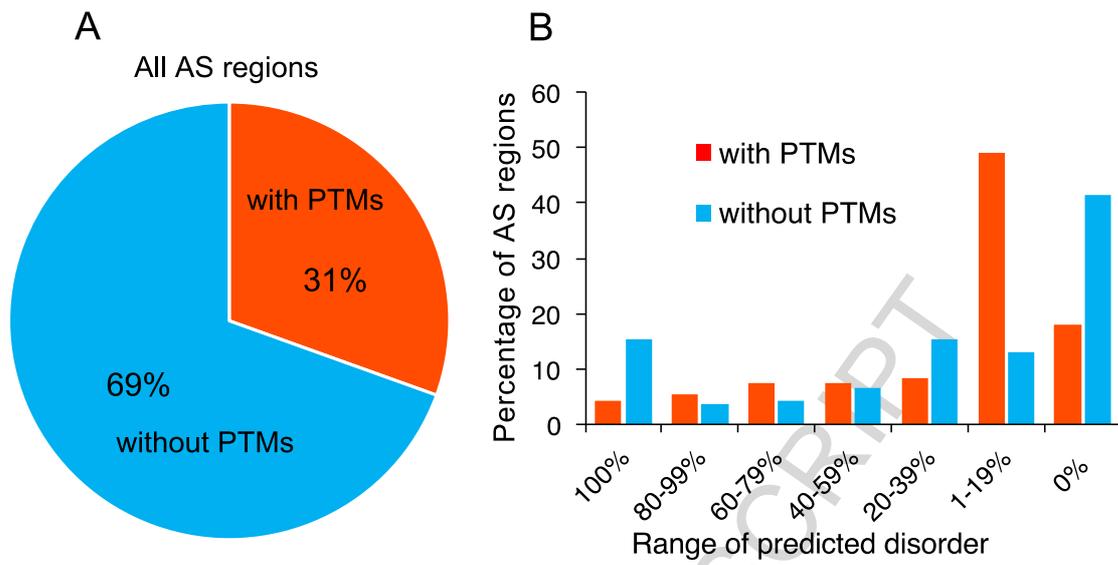


Figure 3

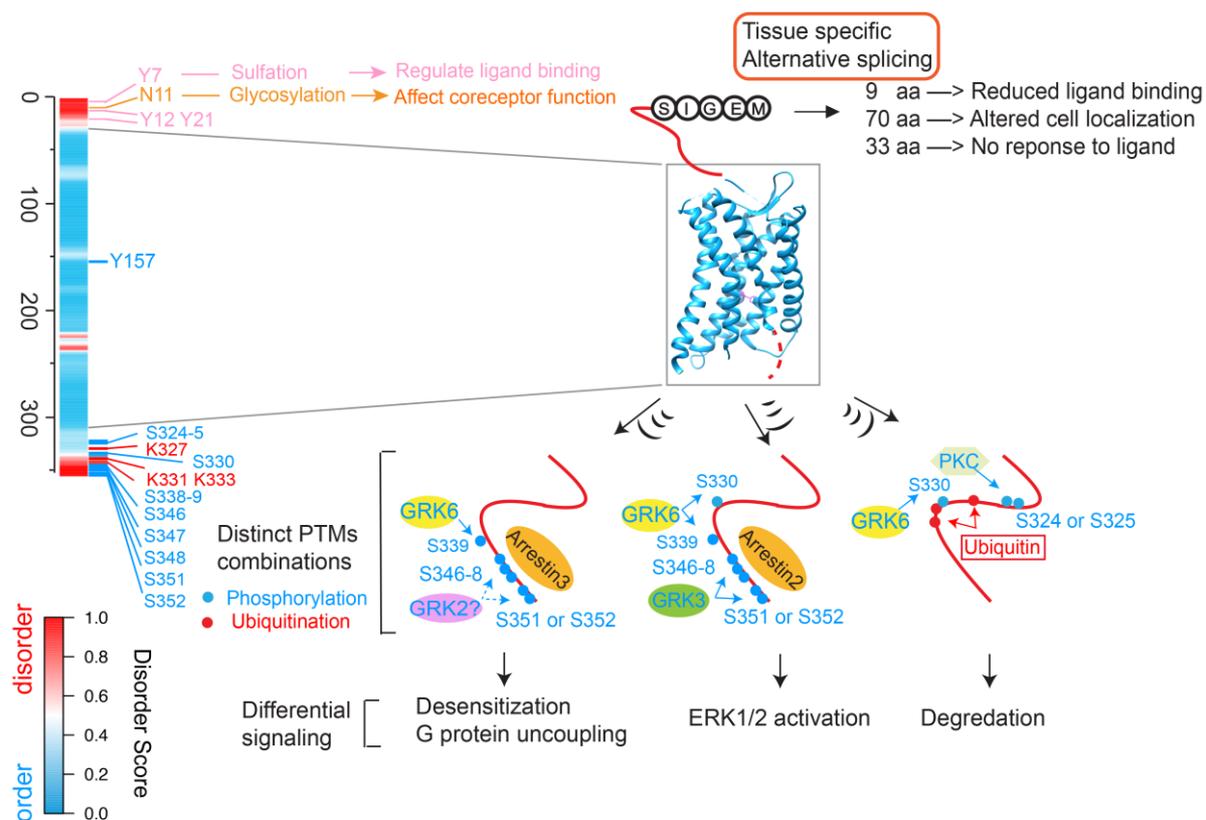


Figure 4

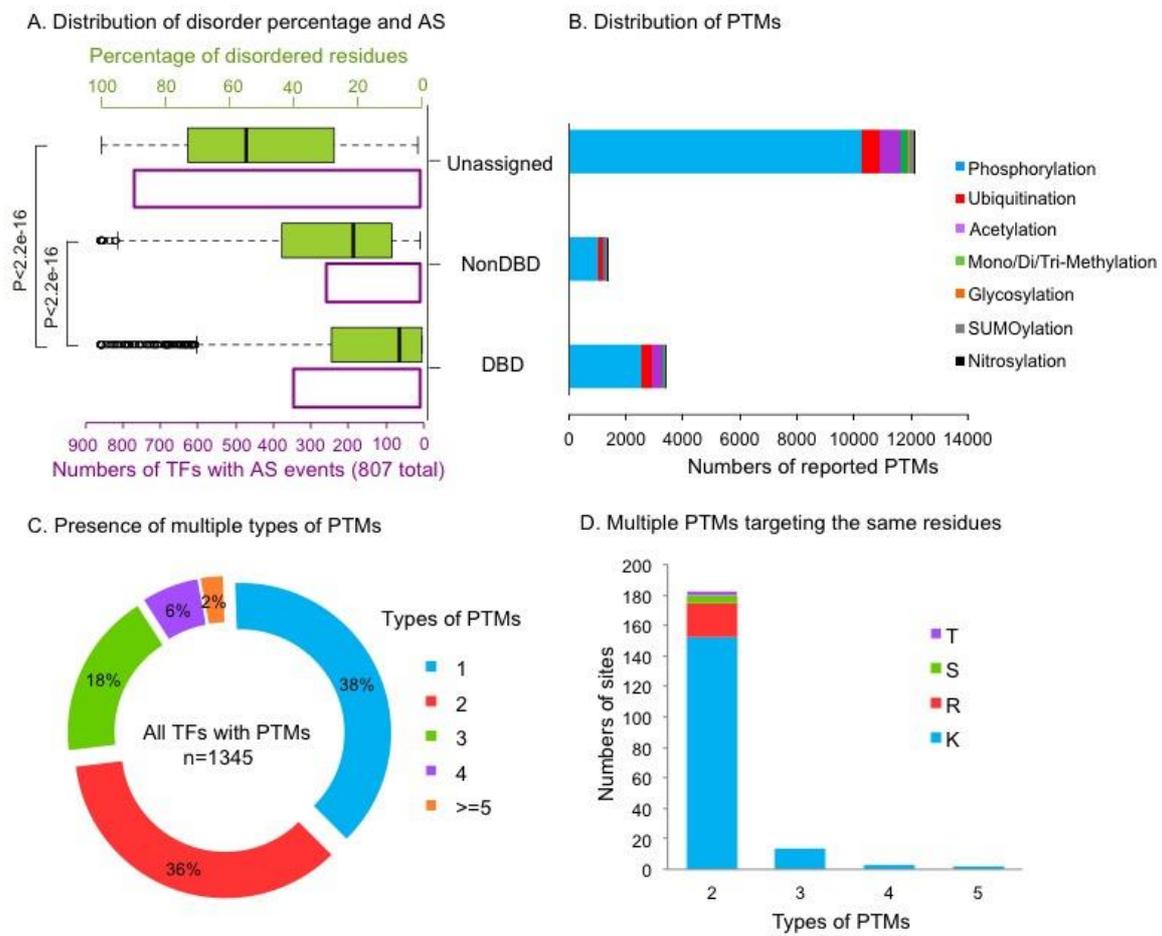


Figure 5

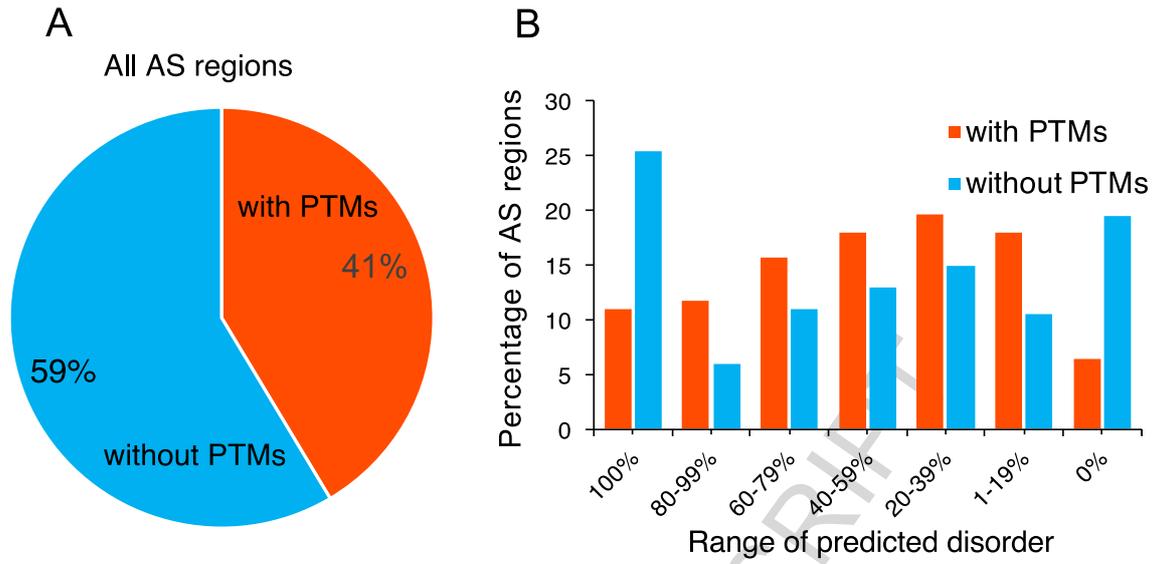
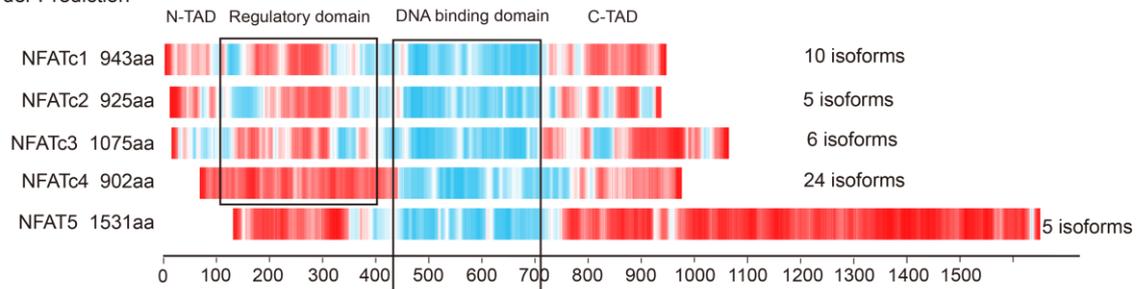
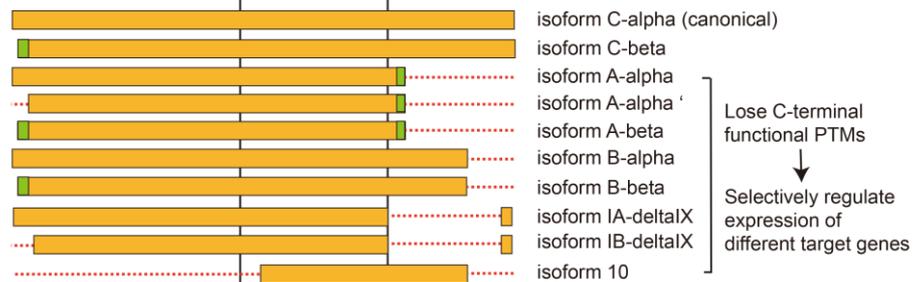


Figure 6

A. Disorder Prediction



B. Splice Variants of NFATc1



C. Multiple IDR-localized PTMs of NFATc1

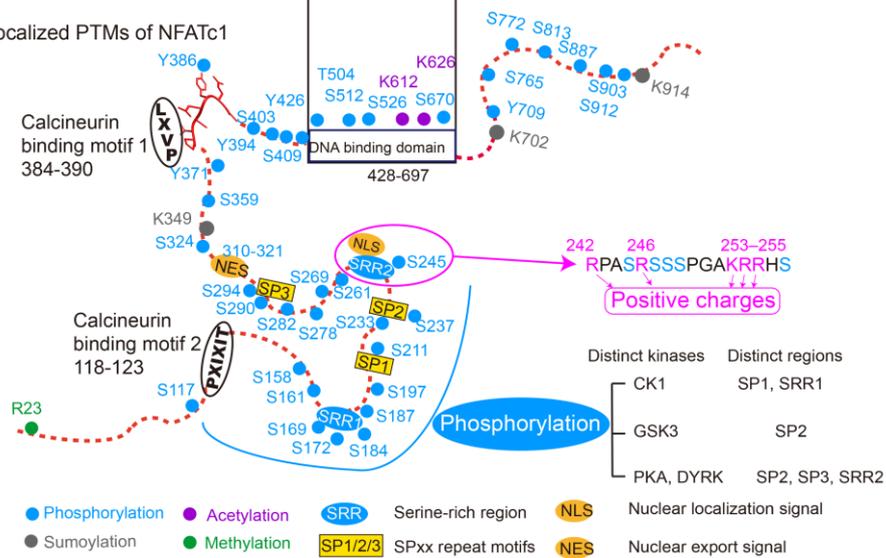
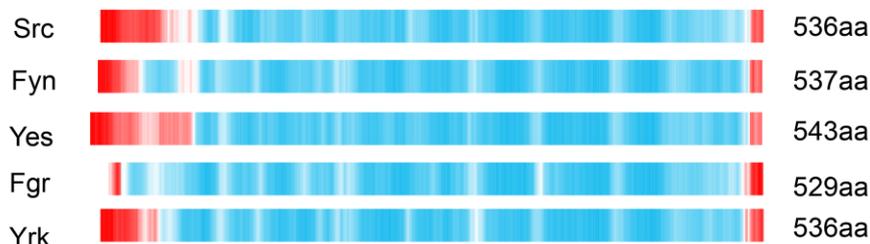


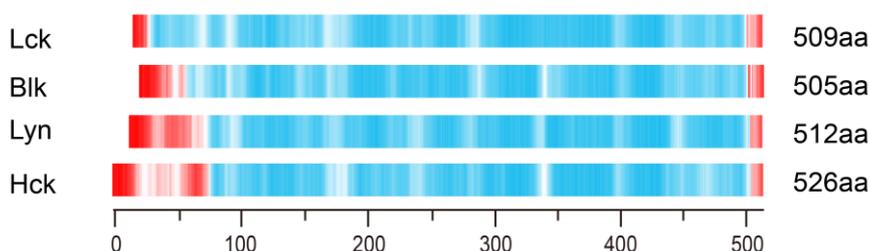
Figure 7

A. Disorder prediction

Group A



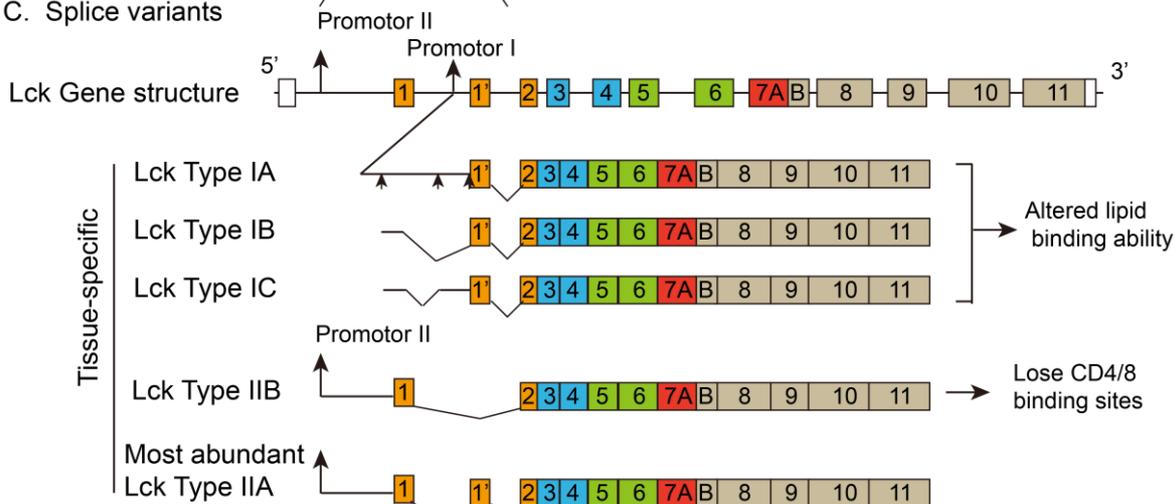
Group B



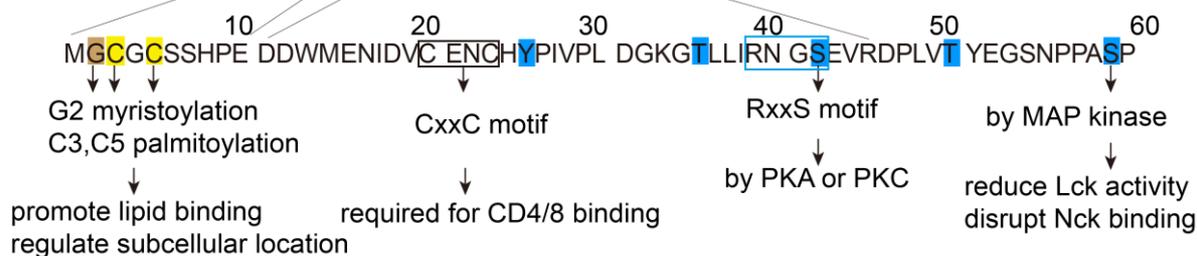
B. Domains

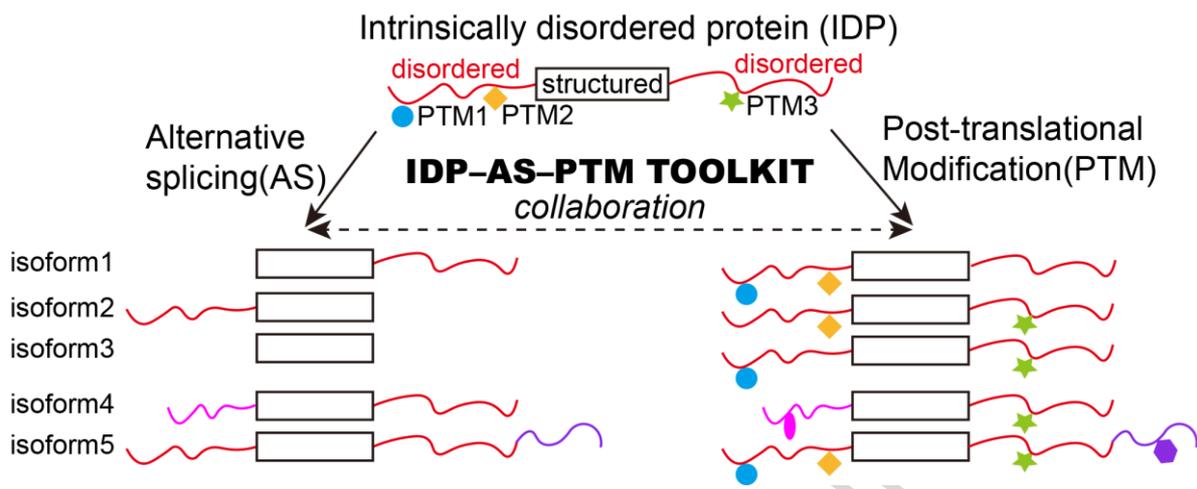


C. Splice variants



D. N-terminal PTMs & binding site of Lck





Graphical abstract