Version submitted 27.07.2019
Final version available online in Archives of Biochemistry and Biophysics.
https://doi.org/10.1016/j.abb.2019.108161

Integrating disorder in globular multidomain proteins: fuzzy sensors and the role of SH3 domains.

Miguel Arbesú[1], Miquel Pons*

Biomolecular NMR laboratory. Department of Inorganic and Organic Chemistry. University of Barcelona. Baldiri Reixac, 10-12, 08028 Barcelona, Spain.

[1] Present address. Department of NMR-supported Structural Biology. Leibniz-Forschungsinstitut für Molekulare Pharmakologie. Robert-Rössle-strasse, 10, 13125 Berlin, Germany

**Corresponding author: mpons@ub.edu*

Highlights

- Disordered regions can act as sensors of the cell environment.
- Intramolecular fuzzy complexes in Src family kinases couple disordered and SH3 domains
- Nearly 50% of disordered regions bound to SH3 domains in Uniprot database are tails
- Linkers and tails connected to SH3 domains are enriched in SH3 binding motifs
- A widespread role of SH3 domains coupling disordered and folded domains is suggested

Abstract

Intrinsically disordered proteins represent about one third of eukaryotic proteins. An additional third correspond to proteins containing folded domains as well as large intrinsically disordered regions (IDR). While IDRs may represent functionally autonomous domains, in some instances it has become clear that they provide a new layer of regulation for the activity displayed by the folded domains. The sensitivity of the conformational ensembles defining the properties of IDR to small changes in the cellular environment and the capacity to modulate this response through post-translational modifications makes IDR ideal sensors enabling continuous, integrative responses to complex cellular inputs. Folded domains (FD), on the other hand, are ideal effectors, e.g. by catalyzing enzymatic reactions or participating in binary on/off switches. In this perspective review we discuss the possible role of intramolecular fuzzy complexes to integrate the very different dynamic scales of IDR and FD, inspired on the recent observations of such dynamic complexes in Src family kinases, and we explore the possible general role of the SH3 domains connecting IDRs and FD.

1. Introduction

More than 70% of eukaryotic proteins are formed by multiple structurally defined domains [1]. Intrinsically disordered regions (IDR), another defining characteristic of eukaryotic proteomes, do not adopt a single structure (or a narrow range of related conformations) but exist as dynamic conformational ensembles. Some of the IDR may fold upon binding to other proteins but some remain highly flexible even in their bound state. Proteins in which the disordered region extends to most of the molecule are referred to as Intrinsically Disordered Proteins (IDPs) [2-5].

IDPs and IDRs are much more abundant in eukaryotes than in prokaryotes or archaea, consistent with their prominent role in high-level regulation [6-7]. The special properties of IDR include i) the capacity to interact with multiple binding partners while still being able to be highly specific, ii) the variable exposure of multiple interactors (e.g. short linear motifs) modulated by conformational fluctuations, iii) the sensitivity of the populations of individual conformations from large ensembles to environmental changes, and iv) the capacity to modulate the conformational ensembles by post-translational modifications or alternative splicing. Entropy plays a key role in the properties of IDRs [8]. Intuitively, binding by highly flexible IDRs is expected to involve a potentially large entropic penalty if a rigid complex is formed. Surprisingly, comparisons of thermodynamic data of binary protein complexes involving IDPs or ordered proteins found that ΔG^{o} of IDP complexes were only 2.5 kcal mol^{-1} less stable than those involving folded proteins [9]. The entropic loss can be alleviated by forming multiple weak contacts, which may not be acting simultaneously but rapidly exchanging. Thus the bound form, operatively defined as the two interacting partners remaining at a short distance and with a preferred relative orientation, may be characterized as an ensemble, the members of which engage in alternative local contacts. An illustrative example is the binding of Sic1 and Cdc4 in which the fraction of bound form increases sharply when the number of randomly phosphorylated serine and threonine residues exceeds a threshold, showing that the individual specific interactions are exchangeable [10]. Electrostatic interactions play an important role: being long range they can act globally affecting the size, shape and amplitude of the fluctuations of the conformational ensemble, as well as driving together the interacting partners. An example of an electrostatically driven picomolar interaction between two IDPs that remain disordered in the bound state has been reported recently [11].

Tompa and Fuxreiter [12] introduced the concept of fuzzy complexes that maintain structural ambiguity upon protein-protein interactions. The FuzDB database contains experimentally observed fuzzy interactions with functional implications [13].

The large conformational fluctuations of IDRs contrast with the detailed structural requirements of functions typically associated with folded domains, such as enzymatic catalysis requiring a detailed positioning of the intervening atoms in the active center, or the precise shape complementarity in rigid-body molecular recognition. Of course, dynamics is also present in folded domains but the time scale or the relative populations of the exchanging states are orders of magnitude different from those of IDRs.

The regulatory roles of IDRs exploit their unique capacity to respond and integrate complex cellular inputs and to provide rheostat-like responses. However, the communication between the disordered sensors and the folded actuators must bridge the gap between two highly different structural and dynamic regimes.

An example of a fuzzy intramolecular complex between the disordered N-terminal region of human c-Src and its neighbor SH3 domain was recently described [14]. The focus of this review is on the interactions between folded and disordered regions that are part of the same protein, with a special look on the possible relevance of SH3 domains in the transduction of information between disordered and globular domains.

2. Disorder in Multidomain Proteins.

Disordered regions are distributed unequally around (and inside) the folded domains and in the protein termini [15]. Disordered flexible linkers (DFL) can be predicted from protein sequences using DFL predict [16]. The analysis of the whole human proteome reveals that about 10% of proteins have more than 30% of their residues as part of DFLs. By comparing the distribution of DFLs inside and between domains, it was concluded that a large number of those highly flexible

regions (with an average length of 25 residues) link structural elements within globular domains. Intradomain DFLs form a subclass of loop regions displaying the characteristics of IDRs. The functional connection between ordered regions and disordered loops belonging to the same domain is widely accepted, but the active role of disordered linkers between folded domains is less appreciated [17]. The relevance of disordered tails, extending from folded domains toward the protein termini has been highlighted in a review [18]. The abundance and variety of disordered tails clearly point to a variety of functional roles. The functional repertoire of intrinsically disordered protein tails in Uversky's review [18] is dominated by interaction-based functions. An important class is formed by disordered cis-acting inhibitory sequence elements forming an inhibitory module located in the same polypeptide chain of the functional domain [19]. Trudeau et al. [20] showed that autoinhibited proteins are enriched in intrinsic disorder. They found in two-thirds of the cases studied that inhibition was modulated by interactions with a binding partner, around one-third by phosphorylation and about 10% by proteolysis.

Comparison of inhibitory modules across members of different families of autoinhibited proteins showed a very broad spectrum of disorder. For example, giant protein kinases have inhibitory modules that bind and inhibit the kinase domain by adopting a helical conformation but can be sequestered by binding to calmodulin or S100, causing kinase activation [21]. The inhibitory modules were found to range from nearly ordered to 80% disordered, suggesting that the level of residual structure in the inhibitory module may represent an evolutionary tool to fine-tune the balance between active and inactive states.

In many of the examples mentioned in the article by Trudeau et al., activation is associated to structural changes associated to helix-to-coil transitions "melting" a helical segment that interacted and and inhibited a functional domain. In some cases, like Vav1, melting of the helical segment is fast and followed by phosphorylation, which is the activating event, although the rate of phosphorylation depends on the population of the state with a melted autoinhibitory helix [22].

Regulatory mechanisms at the individual protein level may be broadly described as "switches", basically providing an on/off response, or "rheostats" giving a gradual response to external stimuli. The autoinhibition model is still based on the binary switch concept, even if the population of the "on" and "off" states can be changed in a continuous way.

3. Coping with the effective volume differences between interacting IDRs and folded domains.

An obvious major difference between folded and unfolded domains is the volume that can be, transiently or permanently, occupied by the peptide chain. The N-terminal disordered region of human c-Src has an experimentally determined radius of gyration of around 3 nm, which roughly corresponds to a sphere of 113 nm^3. It forms a fuzzy complex with the SH3 domain, which has a volume of circa 6.5 nm^3, meaning that the multiple interacting sites are located in about 5% of the volume sampled by the non-interacting disordered region (Figure 1). Thus, the size of the conformational ensemble sampled by the disordered regions is expected to be important in the energetics of intramolecular fuzzy complexes.

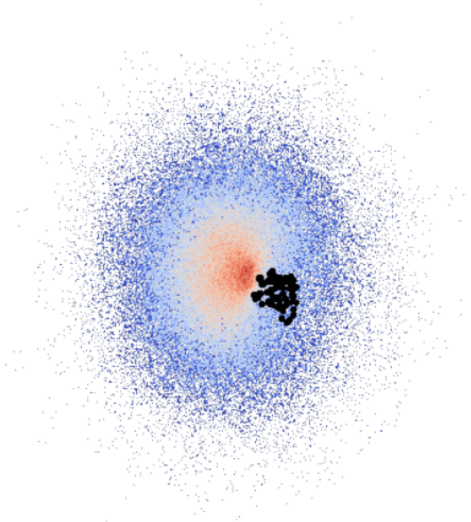


Figure 1. The effective volume of folded and disordered domain. Dots represent two dimensional projections of the positions of α carbons in a random coil ensemble of the disordered N-terminal region of human c-Src bound to the natural SH3 domain, represented in black. (Figure one-column wide)

The hydrodynamic radius of chemically denatured proteins scales with the number of residues following a power law [23]

$$R_h = R_0 N^{\nu}$$

with $R_0 = 2.2\,A$ and ν = 0.57. For IDPs, in the absence of urea or guanidinium chloride, Marsh and Forman-Kay [24] found a similar power low with ν = 0.509 and the value of R_0 depending on the fraction of proline residues f_{pro} and the net charge $|Q|$, according to

$$R_0 = \left(1.24\, f_{pro} + 0.904\right)\left(0.00759\,|Q| + 0.963\right)2.49$$

If the fraction of proline and charge are set to zero, R_0 = 2.16.
Proline residues are abundant in IDPs. Their singular conformational properties result from the Cδ atom of the aliphatic side chain being linked to the backbone nitrogen, introducing steric bulk and restricting the range of allowed ϕ dihedral angle values [25]. The polyproline II (PPII) conformation found in the collagen helix PPII, is abundant in disordered proteins. Although non-proline residues can also adopt PPII conformations, proline-rich segments show a higher propensity for this extended conformation. The PPII propensities measured by the Hilser group [26] correlate with the R_h following a power law, in which the exponent depends on the average PPII propensity along the chain

$$R_h = 2.16\, N^{0.503 - 0.11 \ln\,(1 - f_{PPII})}$$

The correlation between R_h and f_{PPII} was experimentally demonstrated in a set of 22 IDPs [27].
In addition to its role in modulating the overall size of IDR ensembles, PPII conformations are recognized by SH3 domains, and canonical motifs recognized by SH3 domains adopt a PPII conformation in their bound form [25].

Other sequence related parameter affecting the conformational ensemble of IDPs are the frequency and sequence distribution of oppositely charged residues, described by the fraction of charged residues (FCR) = ($f_+ + f_-$) and the net charge per residue (NCPR) = $| f_+ - f_- |$, where f_+ and f_- denote the fraction of positively and negatively charged residues [28]. The overall charge asymmetry is defined as $\sigma = NCPR^2/FCR$.

Electrostatic repulsion in strongly charged polyelectrolytes leads to swollen coils. In contrast, proteins with low NCPR tend to form collapsed quasi-spherical globules, even when the sequence is depleted of hydrophobic residues, as in the case of IDPs. The polymeric nature of protein chains is important and the properties of individual isolated residues or in short peptides may not reflect the properties observed in IDPs [29].
The conformational properties of an IDP are determined by the balance between chain-chain, chain-solvent and solvent-solvent interactions integrated over an effective length (“blobs”) in which the net interaction energy exceeds kT. Typically the length of the blobs is between 5 and 7 residues.
In diluted solutions, the global balance in a good solvent favors the repulsion between blobs favoring an expanded coil, while in a poor solvent; blobs tend to collapse into spherical globules. The transition is sharp for long polymers. The exponent of the power law linking R_h or related magnitudes with the number of residues is a measure of the goodness of the solvent, i.e. the balance between solvent-chain and chain-chain interactions.
R_h of peptide chains formed by polyglycine [30] or polyglutamine [31] chains scale with ν = 0.33 suggesting that water is actually acting as a poor solvent for the peptide backbone and polar, non charged side chains, favoring collapsed globules.
At high peptide concentrations, intermolecular interactions can compete with the intramolecular collapsed state, giving rise to aggregation [32] or liquid phase separation [33,34].

4. The effect of tethered folded domains in disordered regions

Mittal et al. [35] have compared computational simulations of the conformational ensembles of IDRs as autonomous units, with those of the same IDRs tethered as C-terminal tails to folded domains, or as linkers connecting two folded domains. The folded domains were a SH3 domain from the tyrosine kinase adaptor protein NCK1 and the WW domain from peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (PIN1). The choice of the folded domain was determined by their small size and their distinct charge distribution: the WW domain is uniformly basic over its entire surface while half of the surface of the SH3 domain is negatively charged, while the other half is neutral/mildly basic.
The IDR were selected from the regions R1, R2 and R3 of the Das-Pappu diagram of states (Figure 2), where the majority of linkers and tails in the SwissPfam database are located. They are characterized by low values |NCPR| ≤ 0.35 and differ by the total fraction of charged residues.

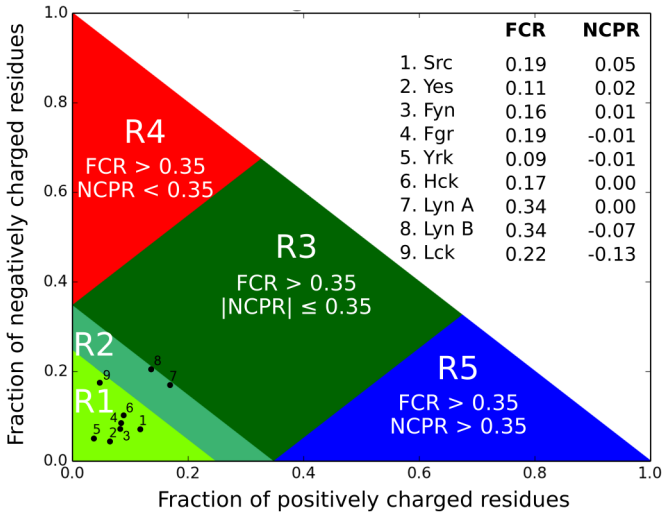


Figure 2. Das-Pappu diagram of states. A large majority of linkers and tails are located in R1, R2 and R3 regions. The localization of the disordered regions (SH4 + Unique domains) of SFKs are indicated.
(Figure one-column wide)

IDR in the R1 region (FCR < 0.25) form collapsed globules and are little affected by the tethered domains. IDRs in the R2 region (0.25 < FCR < 0.35) or R3 region (FCR > 0.35) that were not forming collapsed globules showed the largest context sensitivity.
The simulations of Mittal et al. of the effect of tethered folded domains explicitly avoided natural functional interactions between the folded and disordered units that probably exist in natural proteins, thus providing a null model of the sequence determinants in the disordered domains that affect their context dependency.

5. A regulatory intramolecular fuzzy complex between the disordered and SH3 domains of c-Src.

The N-terminal regions of Src family kinases (SFK) are intrinsically disordered and are directly bound to SH3 domains. The functional relevance of this disordered tail is suggested by the presence of multiple phosphorylation sites [36] and alternatively spliced forms in some of the members of the family [37]. Comparison of the regions preceding the SH3 domain in various SFKs shows a striking diversity, that lead to name this region as the “Unique domain”. This is in contrast with the high conservation of the folded domains, specially the SH2 and kinase domains.

Mutation of a small number of residues in the Unique domain of c-Src results in a 50% decrease in the invasive capacity of c-Src-dependent colorectal cancer cells [14]. Preliminary data [38] show remarkable changes in the whole cell phosphoproteome patterns, suggesting that the intrinsically disordered region is actually affecting c-Src specificity rather than its activation.
In contrast to the case of the Giant Protein Kinases mentioned in section 2, the level of disorder is similarly high across the SFKs, suggesting that the functional modulation induced by the intrinsically disordered regions is not linked to induced folding by the Unique domains upon binding either to other regions of the same protein or external activators.
The sequences of the disordered regions of SFK fall in the R1 and R2 regions of the Das-Pappu diagram (Figure 2). Analysis of the small angle X-ray scattering of the disordered region of human c-Src bound to its native neighbor SH3 domain show the distribution of radius of gyrations in an ensemble of conformations reproducing the observed scattering had two maxima, the first one corresponding to the maximum observed in the isolated disordered region and a second one indicating a more compact, yet still highly disordered ensemble [14]. NMR data confirmed that the region remained disordered in the presence of the SH3 domain although consistent transient contacts between the Unique and SH3 domains were observed using paramagnetic relaxation enhancement. Interestingly, the SH3 regions in contact with the disordered domains correspond to the flexible loops suggesting that the SH3 domain share characteristics of folded domains and disordered regions.
Paramagnetic relaxation enhancements within the disordered regions were analyzed as ΔPRE, to emphasize departures from a random coil model [39]. The ΔPRE pattern confirms transient contacts within the Unique domain that are retained in the isolated and SH3 bound forms. These results are consistent with a fuzzy complex between the Unique and SH3 domain. The N-terminal SH4 region, which acts also as a lipid-binding region anchoring c-Src to lipid membranes, actively participates in the stabilization of the fuzzy complex. The fuzzy complex involving the SH4, Unique and SH3 domains, which we refer to as Src N-terminal Regulatory Element (SNRE) is retained in the membrane anchored form of the myristoylated protein [40].

6. Conserved interactions between disordered and SH3 domains in SFKs.

The analysis of the fuzzy complex formed by the SH4, Unique and SH3 domains of c-Src provides hints on the relevance of connecting ordered and disordered domains, as well as some indication of a possible general role of the SH3 domain as a specialized interface between the folded and the highly dynamic regions.
The SFK contain two subfamilies, of which c-Src and Lyn are representative examples. Similarly to Src, the Unique domain of Lyn also forms an intramolecular fuzzy complex with the SH3 domain [41]. Lyn exists in two alternatively spliced isoforms differing exclusively in a short region of the Unique domain. Interestingly, the alternatively spliced region is directly interacting with the RT loop of the SH3 domain. The region of the Unique domain that is common to the two isoforms preferentially interacts with the nSrc loop. Thus, the two isoforms present natural alternative fuzzy complexes, highlighting the regulatory role of the intramolecular fuzzy complexes in SFKs.

Limited sequence variability under evolutionary pressure at a particular site is an indicator of the functional importance of this site. When two sites interact, sequence variations can be correlated. This can be understood considering that mutations in the first site that could impair the interaction may be rescued by a compensating mutation in the second site. Thus, while the two individual single mutants at each site may have a low probability to be selected, the double mutant is energetically “well fit”, and preserved by evolution. Coevolving residues may be part of the same protein or involve the interface between complexes. The interaction may be a direct structural contact or other functionally relevant event, including folding. Analysis of coevolution extracted from the alignment of natural sequences has been successfully used to predict 3D structures or to identify residues across protein interfaces [42-44]. A high-throughput analysis of coevolution outperformed experimental methods to detect protein-protein interactions such as yeast two hybrids or affinity purification mass spectrometry [45].

The analysis of interactions with disordered regions using coevolution methods is more problematic as disordered protein regions present a higher variability than folded structures [46].

In addition, fuzzy interactions most often involve multiple conformations that may weaken the coevolution signal. However, coevolution analysis of large sequence sets can identify multiple conformations and, therefore, is not restricted to rigid contacts [47].

Coevolutionary analysis critically depends on a) the reliability of the sequence alignments and, b) the amount of sequence information available, determined by the number of aligned sequences with respect to the average length of the sequence(s) analyzed. Toth-Petroczy et al. [48] reported an optimized method for the alignment of disordered sequences. Their coevolution analysis of the disordered human proteome suggests that 42% of the regions may adopt secondary structures under some condition and that 50% may have 3D contacts. The abundance of evolutionary contacts in disordered regions reinforces the view of their functional importance and tight evolutionary selection. On the other hand, it challenges the naïf binary divide of the protein world into ordered and disordered proteins in favor of a continuous view of the dynamics of protein regions and their interactions. Evolutionary couplings detect functional interactions and may provide clues on the mechanisms by which sequence information encodes higher order information and the way this information is converted into function.

In a recent study, Pancsa et al. [49] analyzed intermolecular co-evolutionary couplings involving a folded domain in one of the partners and an intrinsically disordered region in the other. The evolutionary couplings detected frequently involved multiple contacts in long regions with high propensity to adopt transient α-helical structures, while short linear motifs (SLiMs), that are known to mediate many interactions in IDPs were not detected, probably because they are diluted by the large variability of neighbor sequences.

A coevolution analysis identified intramolecular interactions between disordered regions and SH3 domains in SFKs [14]. In this case the conserved SH3 domain facilitated the alignment of the disordered regions. Also, the variety of homologues and orthologues among SFKs of different species sharing the same architecture provided enough sequence information to perform the analysis (Figure 3). The observed intramolecular evolutionary couplings matched experimentally detected interactions, although no induced folding could be observed. This example stresses the fact that evolutionary couplings reflect functional interactions that are evolutionary preserved, even if they do not have a rigid structure correlate, as in the case of fuzzy complexes.

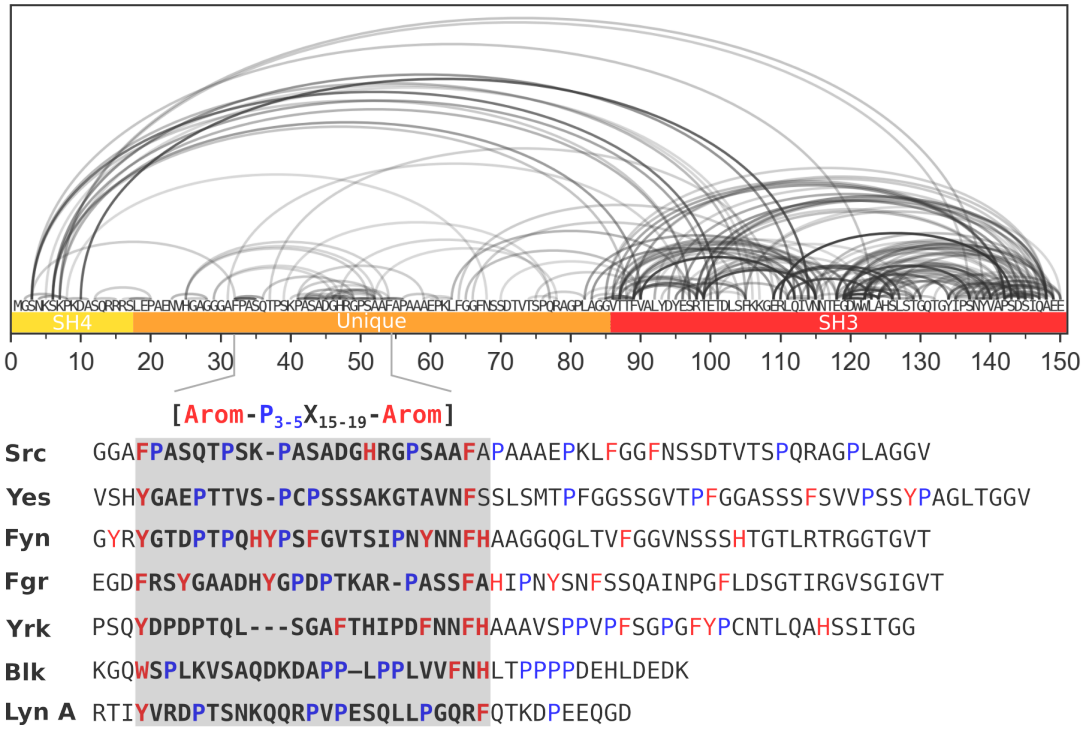


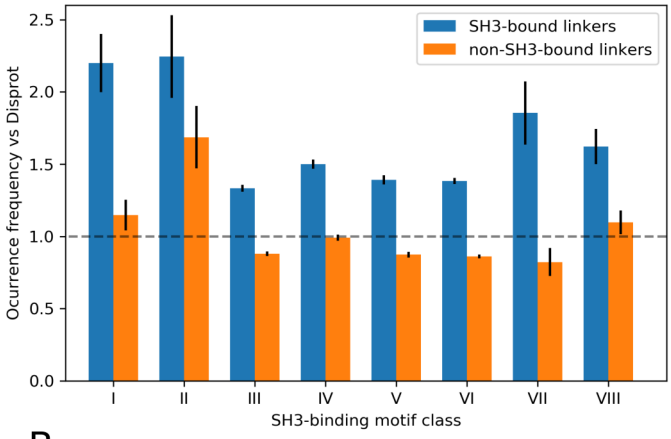
Figure 3. Coevolution of the disordered and SH3 regions of SFK. Conserved aromatic residues within the intrinsically disordered Unique domain bracket a proline rich region. (Figure two columns wide)

In addition to coevolution signals between the disordered and SH3 domains, additional couplings are observed within the Unique domain. The most intense correspond to a region defined by a conserved pattern of aromatic residues separated by 19-23 residues, including 3-5 prolines, observed in seven of the SFK family members (Figure 3).
IDP sequences are typically depleted of hydrophobic residues, including aromatic residues. However, aromatic residues are unusually abundant in the Unique domain of SFKs and may have functional significance in the context of the intramolecular fuzzy complexes. Experimental validation of the role of phenylalanine residues in the Unique domain of Src was obtained by NMR using point mutants in which the individual phenylalanines were changed to alanine [14].
Prolines represent close to 20% of the residues separating the pair of conserved aromatic residues. This enrichment is significant even for an IDR. The average abundance of prolines in IDPs is around 7% and for globular proteins around 4%. Thus, the observed pattern can be defined by the length of the inter-aromatic segment and its proline-rich character.
Two other SFK members show a similar pattern formed by a pair of aromatic residues separated by a proline-rich sequence, but the separation between the aromatic residues is larger (Lck:25 residues) or smaller (Hck:14 residues).
As a reference, the abundance of proline residues outside the conserved pattern in the disordered N-terminal regions (including the SH4 and Unique domains) of nine SFKs is 8,4%, close to the average value for an IDP. Thus, the enrichment of proline residues in the region of the Unique domain flanked by conserved aromatic residues is likely to be functionally relevant for the formation of a fuzzy complex with SH3 domains. Fuzzy complexes result from multivalency and alternative, nearly isoenergetic contacts. The presence of multiple proline-rich short elements in close proximity may stabilize the fuzzy complex in SFK. Indeed, the splice variant of Lyn lacking one of the conserved aromatic groups and two prolines does not interact with the RT loop of the SH3 domain, although the interaction with the n-Src loop is retained [38]. Thus, sequence variations introduced by alternative splicing generate alternative fuzzy complexes, although we are still far from understanding the sequence rules.

7. SH3, linkers and tails.

SH3 is an abundant and versatile protein-binding domain [50]. The consensus binding sequence is proline rich and the two canonical binding motifs: RxxPxxP (class I) and PxxPxR (class II) bind in opposite directions. A recent unbiased analysis of peptides selected by phage display for binding to 115 SH3 domains identified 154 specificity profiles, about half of which are non-canonical, i.e. do not correspond to the PxxP canonical sequence [51]. Many, although not all, of the selected sequences contain proline and charged residues and are compatible with IDR forming sequences.
The relevance of intramolecular interactions involving SH3 domains can be estimated by analyzing the disordered regions located between domains or between the domains and the termini. 1104 non-redundant, reviewed containing at least one domain of the Pfam SH3 families were found in Uniprot. 2464 tails and interdomain regions were identified, of which 1505 are directly adjacent to a SH3 domain while 959 are not directly attached to SH3 domains.
The average IUPred score [52] of the segments is 0.477 ± 0.173 and 0.572 ± 0.183 for the linkers attached or not to SH3 domains, respectively. Thus the two sets of linkers show similar degrees of disorder.

A



B

Type	Class	Pattern
Canonical	I	RxxPxxP
	II	PxxPxR
Non canonical	III	RxxxxxP
	IV	PxxxxR
	V	RxxPxxx
	VI	xxxPxR
	VII	PxKP
	VIII	PxxPxxP

Figure 4. Linker and tail regions in SH3-containing proteins in Uniprot. (A) Relative frequency of the eight classes of SH3 binding motifs observed in the unbiased analysis of Teyra et al [51]. Linkers and tails (longer that 25-residues) were separated between those directly connected to a SH3 domain or other domains. The number of occurrences of each motif, normalized by the total number of residues in the set, was divided by the frequency observed in the Disprot database. Linkers and tails that are directly connected to SH3 domains show a significantly higher frequency of SH3 binding motifs than the linkers or tails co-occurring with SH3 domains but flanking other domains, which show around the same frequency as the Disprot database. Error bars were estimated from repeated searches of the same motifs in randomly generated sequences to account for the different motif lengths. (B) SH3 binding motifs.

The motifs defined by Teyra et al. are enriched in the linkers connected to SH3 domains as compared to sequences in the Disprot data base [53]. In contrast, SH3 binding motifs are not significantly enriched in linkers connected to other domains, even if the protein contains a SH3 domain.
Interestingly enrichments were observed for canonical and non-canonical motifs, supporting the generality of the non-canonical motifs identified by Teyra et al. These authors identified additional non-canonical motifs that were classified as class IX (“atypical”) indicating that the range of sequences recognized by SH3 domains is wider that usually assumed. The diversity of motifs included in class IX prevents an analysis of their abundance in SH3 bound linkers, but suggests that the abundance of SH3 binding motifs mediating intramolecular complexes is even higher than the one estimated in figure 4. On the other hand, they found that one third (38 of 115) of the SH3 domains they studied showed multiple binding motifs, indicative of a remarkable promiscuity.

The enrichment in SH3 binding motifs in linkers connected with SH3 domains suggests a widespread role of intramolecular interactions between disordered domains and SH3 domains, generalizing the observation of a intramolecular fuzzy complex between the disordered regions and the SH3 domains of Src Family Kinases.

Close to one half (46,4%) of the segments attached to SH3 domains include the protein termini (correspond to disordered tails). In 10% of these tails there are clusters of SH3-binding motifs (defined as containing at least two motifs of classes I,II, VI and VII, separated by less than 30 residues. An illustrative example is the SH3 domain containing protein 19 (Uniprot Q5HYK7) [54], that has 4 SH3 domains and its N-terminal tail contains 13 SH3 binding motifs (3 class I; 4 class VII; 6 class VIII). Another example is the Melanoma inhibitory protein 2 (Uniprot Q96PC5) [55] that has a single SH3 domain at its N-terminus and a very long tail terminated with 8 SH3 binding motifs (2 class I, 3 class II and 3 class VIII) in the last 109 residues. The intermediate long region includes coiled coil regions and an additional class I SH3 binding motif.

The FuzDB [13] contains several entries of intermolecular interactions involving fuzzy complexes with a SH3 domain. One example is the binding of non-structural protein 5A (Uniprot Q9YKI6) to a variety of SH3 domains of Src family kinases (Fyn, Lyn, Lck, HcK) and adaptor proteins Grb2 and Bin1. The interaction is mediated by canonical and non-canonical motifs. A second example is the binding of c-Myc to Bin1 SH3 domain through at least two sites forming a dynamic complex and shifting the population of the conformational ensemble samples by Myc.
A third example, connecting intra- and intermolecular interactions is that of the linker between the first and second SH3 domains of the adaptor protein Nck2 (Uniprot: O43639) that weakly interacts with the latter domain through a non-canonical basic motifs interacting with the negatively charged SH3 domain. At high concentrations Nck phase separates. This is an interesting example of the electrostatic-driven interactions between folded and disordered domains leading to formation of liquid phases [56].

Although SH3 domains have been extensively used as archetypal models of folded domains, a large portion of their sequence is formed by loops. When comparing the conservation of residues between nine SFKs (Src, Yes, Fyn, Fgr, Lck, Lyn, Hck, Blk, and Frk) aligned using Muscle [57], 56% of the positions in the kinase domain are strictly conserved in at least eight of the sequences, but only 36% of those in the SH3 domain. As expected, conservation is lower in the RT and n-Src loops of the SH3 domain, which are also the most perturbed regions by the presence of the disordered domains in Src and Lyn.
This suggests the intriguing hypothesis that SH3 domains, known for their affinity to polyproline regions, may be regarded, more broadly, as receptors for intrinsically disordered regions, playing the role of interfaces between the intrinsically disordered sensors and the globular actuators.
Figure 5 suggests a putative role in the soft signaling control of SFKs by the disordered regions, connected through the SH3 domain to the other globular domains. The SH3 domain is known to participate also in the switch between the open and closed states of Src and the linker between the SH2 and kinase domain adopts a polyproline conformation contributing to the stabilization of the Src closed state. The abundance of binding motifs in linkers connected to SH3 domains suggests that the IDR-SH3 interactions may represent a widespread mechanism for intramolecular regulation by IDR. The presence of multiple motifs in many of the linkers examined would be compatible with the abundance of intramolecular fuzzy complexes nucleated by SH3 domains also outside the SFKs.

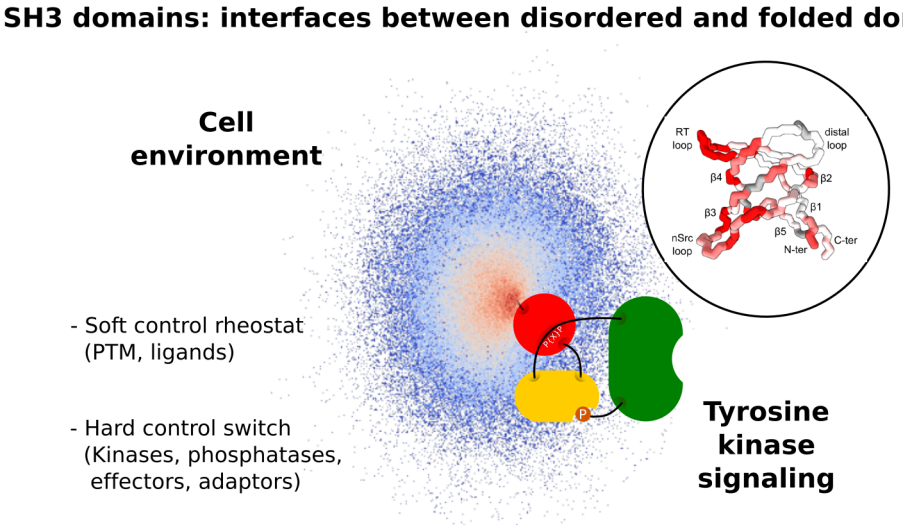


Figure 5. SFKs have at least two levels of regulation: a soft control exerted through the N-terminal disordered region and the classical hard control exemplified by the on-off switching mediated by the interaction of a phosphorylated tyrosine in the C-tail with the SH2 domain. We suggest that the SH3 domain plays a key role in the two modes, by contributing to the stability of the closed, inactive form of the globular domains, and by transducing the signals received by the Unique and SH4 domains to the kinase domain. The abundance of SH3 binding motifs (canonical and non-canonical) in the SH3 bound linkers of other proteins supports the general role of SH3 as an adaptor domain between disordered and folded domains. (Figure 1.5 column wide)

8. Concluding remarks: Signaling and information channels

The evolution of pTyr signaling has been interpreted as the emergence of a new cellular communication technology, some 600 million years ago, just prior to the emergence of multicellular animals. Wendell Lim and colleagues [58] have argued on a model in which components of the elements responsible for the writing (tyrosine kinases), erasing (tyrosine phosphatases) and reading (SH2 domains) of the signals had slowly evolved independently until the three components became interlinked. The highly efficient communication channel unleashed a quantum jump in the evolution of complexity, which the authors compare to the effects of linking laser and fiber optic technologies in human communication networks. In their insightful essay the authors ask whether the pTyr communication channel system is saturated or there is still available encoding potential for further evolution (natural or synthetic).
In our view a fourth module, not surprisingly often found together with the kinase and SH2 domains, is formed by the intrinsically disordered regions coupled to SH3 domains. This complementary module implements a large bandwidth channel, enabling the incorporation of complex cellular environmental clues and their conversion into proper signaling responses, thus expanding and fully exploiting the phosphotyrosine signal encoding potential.

Acknowledgements

This work was supported by the Fundació Marató TV3 (20132830/31) and the Spanish Agencia Estatal de Investigación (BIO2016-78006R), co-financed with EU structural funds (FEDER).

References

[1] A. G.Murzin, S. E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247 (1995) 536–540. https://doi.org/10.1006/jmbi.1995.0159

[2] P.E. Wright, H.J. Dyson, Intrinsically unstructures proteins: re-assessing the structure-function paradigm. J. Mol. Biol. 293 (1999) 321-331. https://doi.org/10.1006/jmbi.1999.3110.

[3] V. Uversky, Unusual biophysics of intrinsically disordered proteins. Biochim. Biophys Acta 1834 (2013) 932-951. https://doi.org/10.1016/j.bbapap.2012.12.008

[4] M. Fuzreiter, Fuzziness in protein interactions- A historical perspective. J. Mol. Biol. 430 (2018) 2278-2287. https://doi.org/10.1016/j.jmb.2018.02.015

[5] V. Uversky, Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics, Front. Physics 7(2019) 10 https://doi.org/10.3389/fphy.2019.00010.

[6] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradovic, Intrinsics disorder and protein function, Biochemistry 41 (2002) 6573-6582. https://doi.org/10.1021/bi012159+

[7] J. Zhou, S. Zhao, A.K. Dunker, Intrinsically disordered proteins link alternatively splicing and post-translational modifications to complex cell signalling and regulation, J. Mol. Biol. 430 (2018) 2342-2359. https://doi.org/10.1016/j.jmb.2018.03.028

[8] T. Flock, R. J. Weatheritt, N. S. Latysheva, M. M. Madan Babu, Controlling entropy to tune the functions of intrinsically disordered regions. Curr. Opin. Struct. Biol. 26 (2014) 62–72. https://doi.org/10.1016/j.sbi.2014.0 5.007

[9] K.Teilum, J. G. Olsen, B. B. Kragelund, Globular and disordered— the non-identical twins in protein-protein interactions, Front. Mol. Biosci. 2 (2015) 40. https://doi.org/10.3389/fmolb.2015.00040

[10] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, et al. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure 18 (2010) 494–506. https://doi.org/10.1016/j.str.20 10.01.020

[11] A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature 555 (2018) 61–66. https://doi.org/10.1038/nature25762

[12] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, Trends Biochem. Sci. 33 (2008) 2–8. https://doi.org/10.1016/j.tibs.2007.10.003

[13] M. Miskei, C. Antal, M. Fuxreiter, FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher order assemblies. Nucleic Acid Res. 45 (2017), D228-D234. https://doi.org/10.1093/nar/gkw1019

[14] M. Arbesú, M. Maffei, T.N. Cordeiro, J.M.C. Teixeira, Y. Pérez, P. Bernadó, S. Roche, M. Pons, The Unique domain forms a fuzzy intramolecular complex in Src family kinases, Structure 25 (2017) 630-640. https://doi.org/10.1016/j.str.2017.02.011

[15] M.Y. Lobanov, E.I. Furletova, N.S. Bogatyreva, M.A, Roytberg, O.V. Galzitskaya, Library of disordered patterns in 3D protein structures. PLoS Comput. Biol. 6 (2010) e1000958. https://doi.org/10.1371/journal.pcbi.1000958

[16] F. Meng, L. Kurgan, DFLPred: high-throughput prediction of disordered flexible linker regions in protein sequences. Bioinformatics 32 (2016) i341-i350. https://doi.org/10.1093/bioinformatics/btw280

[17] J.H. Han, S. Batey, A.A. Nickson, S.A.Teichmann, J. Clarke, The folding and evolution of multidomain proteins, Nat Rev Mol Cell Biol. 8 (2007) 319-30 https://doi.org/10.1038/nrm2144.

[18] V.N. Uversky, The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini, FEBS Letters 587 (2013) 1891-1901. https://doi.org/10.1016/j.febslet.2013.04.042

[19] M.A. Pufall, B.J. Graves, Autoinhibitory domains: modular effectors of cellular regulation. Annu. Rev. Cell Dev. Biol. 18 (2002) 421–462. https://doi.org/10.1146/annurev.cellbio.18.031502.133614

[20] T. Trudeau, R. Nassar, A. Cumberworth, E.T.C. Wong, G. Woollard, J. Gsponer, Structure and intrinsic disorder in protein autoinhibition, Structure 21 (2013) 332–341. https://doi.org/10.1016/j.str.2012.12.013

[21] B. Kobe, J. Heierhorst, S.C. Feil, M.W. Parker, G.M. Benian, K.R. Weiss, B.E. Kemp, Giant protein kinases: domain interactions and structural basis of autoregulation, EMBO J. 15 (1996) 6810-6821.

[22] P. Li, I.R. Martins, G.K. Amarasinghe, M.K. Rosen, Internal dynamics control activation and activity of the autoinhibited Vav DH domain. Nat Struct. Mol. Biol. 15 (2008) 613-618. https://doi.org/10.1038/nsmb.1428
[23] D.K, Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson. J.A. Jones, L.J. Smith, Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 38 (1999) 16424–16431. https://doi.org/10.1021/bi991765q
[24] J.A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, Biophys J. 98 (2010) 2383-90. https://doi.org/10.1016/j.bpj.2010.02.006.
[25] A. Rath, A. R. Davidson, C. M. Deber, The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition, Biopolymers 80 (2005) 179–185. https://doi.org/10.1002/bip.20227
[26] W.A, Elam, T.P. Schrank, A.J. Campagnolo, V.J. Hilser, Evolutionary conservation of the polyproline II con- formation surrounding intrinsically disordered phosphorylation sites, Protein Sci. 22 (2013) 405–417. https://doi.org/10.1002/pro.2217.
[27] M.E. Tomasso, M.J.Tarver, D. Devarajan, S.T. Whitten, Hydrodynamic radii of intrinsically disordered proteins determined from experimental polyproline II propensities. PLoS Comput Biol.12 (2016) e1004686. https://doi.org/10.1371/journal.pcbi.1004686.
[28] R.K. Das, R.V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, Proc. Natl. Acad. Sci. USA 110 (2013) 13392-13397. https://doi.org/0.1073/pnas.1304749110.
[29] R.V. Pappu, X. Wang, A. Vitalis, S. L. Crick, A polymer physics perspective on driving forces and mechanisms for protein aggregation, Arch. Biochem. Biophys. 469 (2008) 132–141. https://doi.org/10.1016/j.abb.2007.08.033
[30] D.P.Teufel, C.M. Johnson, J.K. Lum, H. Neuweiler, Backbone-driven collapse in unfolded protein chains, J. Mol. Biol. 409 (2011) 250–262. https://doi.org/10.1016/j.jmb.2011.03.066
[31] S.L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R.V. Pappu, Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions Proc. Natl. Acad. Sci. USA 103 (2006) 16764–16769. https://doi.org/10.1073 pnas.0608175103.
[32] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, A natively unfolded yeast prion monomer adoptsan ensemble of collapsed and rapidly fluctuating structures. Proc. Natl. Acad. Sci. USA 104 (2007) 2649–2654. https://doi.org/10.1073 pnas.0611503104
[33] C.P. Brangwynne, T.J. Mitchison, A.A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc Natl Acad Sci U S A 108 (2011) 4334-4339. https://doi.org/10.1073/pnas.1017150108.
[34] E. Gomes, J. Shorter, The molecular language of membraneless organelles, J. Biol. Chem. 294 (2019) 7115-7127. https://doi.org/10.1074/jbc.TM118.001192
[35] A. Mittal, S.S. Holehouse, M.C. Cohan, R.V. Pappu, Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins, J. Mol. Biol. 430 (2018) 2403-2421. https://doi.org/10.1016/j.jmb.2018.05.012
[36] I. Amata, M. Maffei, M. Pons, Phosphorylation of unique domains of Src family kinases, Frontiers in Genetics 5 (2014)181. https://doi.org/10.3389/fgene.2014.00181.
[37] T.Yi, J.B. Bolen, J.N. Ihle, Hematopoietic cells express two forms of lyn kinase differing by 21 amino acids in the amino terminus. Mol. Cell. Biol. 11 (1991) 2391–2398. https://doi.org/10.1128/mcb.11.5.2391
[38] M. Maffei. PhD thesis. U. of Barcelona. 2014 http://hdl.handle.net/2445/61812.
[39] J.M.C. Teixeira, S.P. Skinner, M. Arbesú, A. Breeze, M. Pons, Farseer-NMR: automatic treatment, analysis and plotting of large, multi-variable NMR data, J. Biomol. NMR. 71 (2018) 1-9. https://doi.org/10.1007/s10858-018-0182-5
[40] A.L. Le Roux, I.-L.Mohammad, B. Mateos, M. Arbesú, M. Gairí, F.Ali Khan, J.M.C. Teixeira, M. Pons, A myristoyl binding site in the SH3 domain modulates c-Src membrane anchoring iScience,12 (2019)194-203 https://doi.org/10.1016/j.isci.2019.01.010
[41] J.M.C. Teixeira, H. Fuentes, S. Bielskute, M. Gairi, S. Zĕrko, W. Koźmiński, M. Pons, The Two Isoforms of Lyn Display Different Intramolecular Fuzzy Complexes with the SH3 Domain, Molecules 23 (2018) 2731 https://doi.org/10.3390/molecules23112731.
[42] T.A. Hopf, C.P. Scharfe, J.P. Rodrigues, A.G. Green, O. Kohlbacher, C. Sander, A.M. Bonvin, D.S. Marks, Sequence co-evolution gives 3D contacts and structures of protein complexes. eLife 3 (2014) e03430 https://doi.org/10.7554/eLife.03430.

[43] S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife 3 (2014) e02030. https://doi.org/10.7554/eLife.02030
[44] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D.S. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 108 (2011) E1293–E1301. https://doi.org/10.1073/pnas.1111471108
[45] Q. Cong, I. Anishchenko, S. Ovchinnnikov, D. Baker, Protein interaction networks reveled by proteome coevolution. Science 365 (2019) 185-189. https://doi.org/10.1126/science.aaw6718.
[46] C.J. Brown, A.K. Johnson, G.W. Daughdrill, Comparing models of evolution for ordered and disordered proteins. Mol Biol Evol. 27 (2010) 609–621. https://doi.org/10.1093/molbev/msp277
[47] F. Morcos, B. Jana, T. Hwa, J.N. Onuchic. Coevolutionary signals across protein lineages help capture multiple protein conformations. Proc Natl Acad Sci U S A. 110 (2013) 20533–20538. https://doi.org/10.1073/pnas.1315625110.
[48] A. Toth-Petroczy, P. Palmedo, J. Ingraham, T.A. Hopf, B. Berger, C. Sander, D.S.Marks, Structured states of disordered proteins from genomic sequences Cell 167 (2016)158–170. https://doi.org/10.1016/j.cell.2016.09.010
[49] R. Pancsa, F. Zsolyomi, P. Tompa, Co-evolution of intrinsically disordered proteins with folded partners witnessed by evolutionary couplings Int. J. Mol. Sci. 19 (2018) E3315, https://doi.org/10.3390/ijms19113315.
[50] B.J. Mayer, SH3 domains: complexity in moderation, J. Cell Science 114 (2001)1253-1263;
[51] J. Teyra, H. Huang, S. Jain, X. Guan, A. Dong, Y. Liu, W. Tempel, J. Min, Y. Tong, P.M. Kim, G.D. Bder, S. S. Sidhu, Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities, Structure 25 (2017) 1598–1610. https://doi.org/10.1016/j.str.2017.07.017
[52] Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics. 21 (2005) 3433-3434. https://doi.org/10.1093/bioinformatics/bti541
[53] M. Sickmeier, J. A. Hamilton,T. LeGall, V. Vacic, M.S. Cortese, A.Tantos, B. Szabo,P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, A. K. Dunker, DisProt: the database of disordered proteins, Nucleic Acids Res. 35 (2007) D786–D793. https://doi.org/10.1093/nar/gkl893
[54] J.W. Yam, D.Y. C.W. Jin So, L.C. Chan, Identification and characterization of EBP, a novel EEN binding protein that inhibits Ras signaling and is recruited into the nucleus by the MLL-EEN fusion protein, Blood 103 (2004) 1445-1453. https://doi.org/10.1182/blood-2003-07-2452
[55] J.L. Pitman, D.J. Bonnet, L.K. Curtiss, N. Gekakis, Reduced cholesterol and triglycerides in mice with a mutation in Mia2, a liver protein that localizes to ER exit sites, J. Lipid Res. 52 (2011) 1775-1786. https://doi.org/10.1194/jlr.M017277
[56] S. Banjade, Q. Wu, A. Mittal, W.B. Peeples, R.V. Pappu, M. K. Rosen, Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. Proc Natl Acad Sci U S A. 112 (2015) E6426-35. https://doi.org/10.1073/pnas.1508778112.
[57] R.C. Edgar MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792-1797. https://doi.org/10.1093/nar/gkh340
[58] W.A. Lim, T. Pawson, Phosphotyrosine signaling: evolving a new cellular communication system. Cell 142 (2010) 661-667. https://doi.org/10.1016/j.cell.2010.08.023.