Title:

Evidence for conserved fuzzy complexes involving a preorganized Unique domain in the Src family of kinases.

Authors:

Miguel Arbesú[1], Mariano Maffei[1a], Tiago N. Cordeiro[2,3], João M.C. Teixeira[1], Yolanda Pérez[1c], Pau Bernadó[2] Serge Roche[4], and Miquel Pons[1*].

[1] BioNMR Laboratory, Inorganic and Organic Chemistry Department, Universitat de Barcelona, Baldiri Reixac 10-12, 08028 Barcelona, Spain.

[2] Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Université Montpellier 1 and 2, 34092 Montpellier, France.

[3] Instituto de Tecnologia Química e Biológica António Xavier, ITQB NOVA, Av. da República, 2780-157 Oeiras, Portugal

[4] CNRS UMR5237, University of Montpellier, CRBM, 1919 route de Mende, 34000 Montpellier, France.

[a] Present address: European Molecular Biology Laboratory (EMBL), Mouse Biology Unit, Via Ramarini 32, 00015 Monterotondo, Italy.

[b] Present address: Institut de Química Avançada de Catalunya. CSIC. Jordi Girona, 18-26 08034 Barcelona. Spain.

* Corresponding author: mpons@ub.edu

Keywords:

Intrinsically disordered proteins, Fuzzy complex, Protein scaffold, Preorganization, c-Src, Src Family Kinases, SH4, SH3, Unique domain.

SUMMARY

The N-terminal regulatory region of c-Src including the SH4, Unique and SH3 domains adopts a compact, yet highly dynamic, structure that can be described as an intramolecular fuzzy complex. Most of the long-range interactions within the Unique domain are also observed in constructs lacking the structured SH3, indicating a considerable degree of preorganization of the disordered Unique domain. Here we report that members of the Src family of kinases (SFK) share well-conserved sequence features involving aromatic residues in their Unique domains. This observation contrasts with the supposed lack of sequence homology implied by the name of these domains and suggests that the other members of SFK also have a regulatory region involving their Unique domains. We argue that the Unique domain of each SFK is sensitive to specific input signals, encoded by each specific sequence, but the entire family shares a common mechanism for connecting the disordered and structured domains.

INTRODUCTION

The Src Family of non-receptor tyrosine Kinases (SFK) is formed by at least nine members (Src, Fyn, Yes, Yrk, Fgr, Hck, Lyn, Blk, and Lck) implicated in cell signaling pathways related to cell growth, migration, invasion and survival (Thomas and Brugge, 1997; Boggon and Eck, 2004). High levels of Src activity are associated to poor prognosis in colorectal, prostate and breast cancers (Sirvent et al., 2012; Hynes 2000; Yeatman, 2004). The closest family members to Src are Yes and Fyn, which perform overlapping functions and in some cases, can compensate for the loss of the other kinases (Stein et al., 1994). Yes is also implicated in cancer (Dubois et al., 2015).

SFKs share a common domain architecture with three folded domains (SH1, which contains the catalytic center; SH2 and SH3 that are regulatory domains) and a N-terminal intrinsically disordered region (IDR) that includes the SH4 and Unique domains (UD) (Fig. 1). While the folded domains are highly homologous across the whole SFK, the disordered domains of the various SFK have different lengths and sequences. The term "Unique" refers specifically to this diversity.

The structure and functions of the SH3, SH2 and kinase domains have been extensively studied. The interaction of the SH2 domain with a phosphorylated tyrosine located close to the C-terminus, together with the interaction of the SH3 domain with the SH1 domain and the SH2-SH1 connecting segment contribute to maintain SFKs in a closed, inactive, basal state (Xu et al., 1999; Brábek et al. 2002). Dephosphorylation of the tyrosine located close to the C-terminus, followed by additional phosphorylation of an alternative tyrosine site in the SH1 domain lead to the conversion into an active form.

The disordered regions are not observed in the crystal structures and their function remains obscure, except for the lipid-anchoring role of its SH4 domain (Sigal et al., 1994). An exception is the Unique domain of Lck, directly implicated in the interaction with co-receptors via a zinc complex (Kim et al., 2003). The Unique domains of Fyn, Lyn (Luciano et al. 2001) and Src (Hossain et al. 2013) mediate the induction of apoptosis. The association between Src and NMDA receptors is lost in the presence of a peptide with the 40-49 sequence of the UD (Liu et al. 2008). Mutations in the Unique domain of Src cause strong phenotypes in the maturation of *Xenopus laevis* oocytes (Pérez et al., 2013) while phosphorylation events play important regulatory roles in SFKs (reviewed in Amata et al. 2014). However, the functional connection between the disordered region and the folded domains is not yet understood.

The UD and SH4 regions of Src are constrained around the SH3 domain but retain a high flexibility (Maffei et al.,

2015). Thus, the SH3 domain forms the scaffold of a fuzzy intramolecular complex (Tompa and Fuxreiter, 2008). Here we present a detailed NMR and Small-angle X-ray scattering (SAXS) study of the internal contacts occurring in the 1-150 region of human Src that include the SH4, UD and SH3 domains (from now USH3). The set of mutated and/or truncated variants studied is summarized in Figure 1.

We have found that key long-range interactions within the IDR of USH3 are maintained in the absence of the SH3 domain, suggesting that Src UD and SH4 are preorganized. Moreover, mutations affecting some of the residues involved in those interactions display strong phenotypes in cancer cells. This finding sheds light on the connection between ordered and disordered domains of Src. Finally, we show that, in spite of the large sequence diversity among the UD of SFKs, some features, highlighted by our NMR results, are conserved in other SFK members, suggesting that preorganization and scaffolding of the disordered regions around the SH3 domains may be present in other SFKs and could represent a general mechanism to incorporate the environment sensing capacity of intrinsically disordered regions (Wright and Dyson, 2014) as part of regulatory mechanisms.

RESULTS

USH3 adopts a compact fuzzy structure around the SH3 domain.

SAXS provides a sensitive method to monitor the degree of compactness of a protein. This is most conveniently represented using Kratky plots ($I(s)s^2$ vs s, where $I(s)$ is the scattering intensity at a given momentum transfer $s=4\cdot\pi\cdot\sin(\theta/\lambda)$ value. The Kratky plot of the SAXS curves of USH3 (Fig. 2A) presents the typical shape expected for a protein with a globular part and a disordered tail, but could not be simulated by a pool of 10.000 structures generated by Flexible Meccano (Ozenne et al., 2012) for the 1-88 and 148-150 regions attached to a rigid core with the structure of human SH3. However, the curves were well reproduced by a sub-ensemble of structures selected using the EOM algorithm (Bernado et al., 2007). The SAXS scattering profile and the residual deviations between the experimental and calculated curves, showing the goodness of fit, is given in Figure S1.

The distribution of radius of gyration (Fig. 2B) in the reference pool has a broad maximum at around 30 Å. In contrast, the selected subensemble shows a multimodal distribution with a main peak at *circa* 20 Å that contains ~30% of the structures and a smaller peak centered at 30 Å, confirming the compact structure of USH3. The compact structure reflects conformational restrictions that could be functionally important for the regulation of Src through its N-terminal region. The large population of compact structures was not observed in the SAXS data from a construct lacking the SH3 domain (Fig. 2C-D), highlighting the scaffolding role of the SH3 domain.

To analyze the origin of the compact structures and the possible persistence of intramolecular interactions in the

disordered region not detectable by SAXS we used Paramagnetic Relaxation Enhancement (PRE) NMR experiments.

Mapping SH4 and UD contact sites in SH3 with using PREs.

Compaction can be defined by intramolecular contacts. PRE NMR experiments report on the distance between NMR observable nuclei and a paramagnetic group, in this case a MTSL ((1-oxy-2,2,5,5-tetra-methyl-D-pyrroline-3-methyl)-methanethiosulfonate) nitroxide radical, covalently attached to a cysteine residue introduced by site-directed mutagenesis. The relaxation induced by the unpaired electron extends to long distances and, due to the $<r^{-6}_{electron\text{-}nucleus}>$ dependency, close approximation of amide protons to the radical center results in substantial peak broadening even if the contact is transient (Clore and Iwahara, 2009). PRE is measured as the ratio of peak intensities in the paramagnetic protein and a diamagnetic sample produced by reducing the nitroxide group with ascorbic acid: PRE = (I_{para}/I_{dia}).

Figs.3 A-C show, in a color scale, the PRE induced by MTSL in the SH3 domain of USH3 at positions 1, 27 and 59, respectively. The complete profiles are given in Fig. S2. The most perturbed regions, independently of the MTSL position, included residues 96 and 98-102 in the RT loop; residues 114, 116, 120 and 121 of strands β2 and β3 and the connecting nSrc loop; and residues 132-133 in strand β4. The distal loop was only weakly affected. Additional regions, including the entire β4 strand and the following 3_{10} helix as well as the β3 strand and the entire nSrc loop were affected when the MTSL was located on residues 27 and, specially, 59. Thus, various probes well separated along the disordered region of USH3 cause similar effects in the RT and nSrc loops of the SH3 domain.

In order to dissect the contributions from the SH4 and UD to the interactions between the unfolded and folded regions, we prepared two truncated USH3 variants (USH3 Δ10 and USH3 Δ20). The former retains the last residues of the SH4 domain, including three arginine residues (R14, R15, R16), but lacks the three lysine residues (K5, K7 and K9). USH3 Δ20 lacks the entire SH4 domain.

Fig. 3D-E show the PRE effects induced by the paramagnetic probe at position 27 on USH3 Δ10 and USH3 Δ20, respectively, while Fig. 3B shows the profile of full length USH3. The truncated USH3 retained the most intense interactions with the RT and nSrc loop regions but showed reduced PRE on other regions. Thus, although the SH4 domain contributes to the interactions with the SH3 domain, the UD and SH3 domains interact also in the absence of the SH4 domain.

Probing SH4 interactions with the SH3 domain using multiple Chemical Shift Perturbations.

While PREs efficiently detect even weak and transient interactions, chemical shift perturbations (CSP) are sensitive to the local environment and could provide a more detailed mapping than PREs of the interaction between the SH4 and SH3 domains. Fig. 4A presents combined CSP values observed in the SH3 domain of the truncated USH3 variants compared with native USH3. The most significant variations were observed in the RT loop with smaller perturbations

126 in the distal and nSrc loops. In addition to the truncated USH3 variants we included a full length USH3 variant with
127 residues K5 and S6, highlighted in previous NMR studies (Maffei et al., 2015), replaced by alanine.
128 Figs. 4B and S3 show 2D plots, showing the changes in ^{1}H and ^{15}N shifts of representative RT and nSrc loop residues in
129 USH3 variants, taking as a reference the chemical shifts of the isolated SH3 domain. The use of this reference
130 emphasizes the interactions already present in native USH3 and how are they changed by perturbations in the SH4
131 domain.
132 The chemical shifts in the nSrc loop of all USH3 variants were similar, but distinct from those of the isolated SH3
133 domain, indicating that the nSrc loop is affected by the UD but is only weakly sensitive to variations in the SH4 region.
134 In contrast, the RT loop residues were significantly affected by changes in the SH4 domain. The Δ10 and K5A S6A
135 USH3 variants showed similar chemical shifts but distinct from those of Δ20 USH3. Interestingly, Δ10 and Δ20
136 deletions caused shifts in opposite directions with respect to native USH3.
137 Table 1 presents the integrated $^{1}H,^{15}N$ CSP with respect to isolated SH3 of residues belonging to specific SH3 regions
138 in native USH3 and the percentage change of these values in USH3 Δ20, USH3 Δ10, and USH3 K5A S6A variants.
139 The central region of the SH3 domain showed 15% larger CSP in the K5A S6A and Δ10 variants than in native USH3.
140 On the other hand, USH3 Δ20 showed ~8% smaller perturbations than native USH3. The same trend was even more
141 apparent when the analysis was restricted to the RT loop. In contrast, the nSrc loop region showed a decrease in CSP, as
142 compared with native USH3, that was similar for the three SH4 variants.
143 The interactions between UD and SH3 domains were retained when the entire SH4 was deleted. In fact, UD residues
144 that report on the interaction with the SH3 domain (T37, A55 and E60) had similar chemical shifts in USH3 WT and
145 USH3 Δ20, and therefore sense similar environments (Fig. S5).
146 Of notice is the opposite effect of deleting only the first ten residues or the entire SH4 domain. The CSP with respect to
147 the isolated SH3 domain may be interpreted in terms of an equilibrium between a “closed state”, in which specific
148 regions of SH3 are interacting with the intrinsically disordered region and an “open state”, in which these interactions
149 do not occur and, therefore, has the chemical shifts of the isolated SH3 domain. The larger CSP observed when the
150 initial SH4 residues are deleted or modified suggests that these residues are inhibiting the closed state in native USH3,
151 probably by participating in competitive interactions outside the SH3 domain, i.e. within the disordered region itself.
152 This observation is in line with the growing understanding of the role of conformational ensembles in allostery.
153 (Motlagh et al. 2014). In order to characterize long-range interactions within the disordered region of USH3, we
154 measured PRE of the amide protons of SH4 and UD residues.
155
156 **ΔPRE detected long-range interactions in the SH4 and UD of USH3.**
157 Since the spin-label and the affected residues are part of the same flexible chain, we calculated the deviations between

158 the experimental PRE and those predicted from a random-coil model ($\Delta PRE = [I_{para}/I_{dia}]_{exp} - [I_{para}/I_{dia}]_{rc}$).
159 A smoother representation is obtained by calculating a running average of the ΔPRE from neighbor residues with a
160 Gaussian weighting function. In order to emphasize the most affected regions, values above a threshold of one standard
161 deviation were represented in the form of heat maps (Fig 5A). Fig. 5B (upper rows) compares the ΔPRE heat maps
162 generated by labels in residues 1, 27 and 59 of USH3.
163 The MTSL in the SH4 domain showed long-range interactions with residues in the region from 55 to 67 of the UD.
164 Residues 51, 53, 55 and 60-67 form the Unique Lipid Binding Region (ULBR, see Fig. 1) (Pérez et al., 2013). The
165 paramagnetic probe in position 59 approaches residues 15-25, including part of the SH4 domain, and 30-35, as
166 previously described (Pérez et al., 2009). Consistently, the MTSL probe in residue 27, caused significant PRE in the
167 region 49-67, which includes position 59 and the ULBR, as well as in residues 35-45. The complementary long-range
168 interactions observed between the regions around position 27 and 59 and the latter with the SH4 domain are mutually
169 self-consistent and confirm the adoption of compact conformations in the disordered domains of USH3.
170
171 **Long-range interactions inside the disordered region are conserved in the absence of the SH3 domain.**
172 We next asked the question: is the interaction of the SH4 and UD with the SH3 domain causing the observed
173 compaction of the disordered region, or is the compaction a property of the Src N-terminal region facilitating the
174 interaction with the SH3 domain? To answer this question we compared the PREs in USH3 and in a previously
175 described variant containing SH4 and UD but lacking the SH3 domain (Pérez et al., 2009). The ΔPRE heat maps of the
176 SH4-UD variants are presented in Fig. 5B (lower rows). The observed long-range interactions in the absence and in the
177 presence of the SH3 domain are summarized in Figs.5C-D, respectively.
178 The regions contacting residue 59 in USH3 and SH4-UD were almost the same in USH3 and SH4-UD, and some of the
179 strongest contacts from residue 27 (with residues 42-43) were also retained. The SH4 domain showed no interactions
180 with the UD in the absence of the SH3 domain. However, in its presence, showed a number of contacts, the strongest
181 with the 55-67 region. In the presence of the SH3 domain the PREs from position 27 on residues 15-16 decreased and
182 new contacts with the ULBR (residues 65-66) were observed.
183 The similar long-range interactions observed in USH3 and SH4-UD suggest that the intrinsically disordered region of
184 Src is preorganized even in the absence of the SH3 domain, a feature that would alleviate the entropic cost of restricting
185 the flexible peptide chain to remain in the proximity of the SH3 domain. The new contacts between SH4 and the UD
186 observed in the presence of the SH3 domain confirm its scaffolding role.
187
188 **Conserved sequence features in the Unique domains of SFKs.**
189 The sequences of the UD of the various SFK members show very low similarity but we reasoned that they might

display common features enabling functional interactions with their neighboring, well conserved, SH3 domains. In order to identify sequence patterns encoding for the observed conformational preferences, we used MUSCLE (Edgar, 2004) to align the sequences of the UD of Src in various species and of the human forms of Src, Yes, Fyn and Fgr, the closest SFK members (Fig. 6A). Fgr (Feline Gardner-Rasheed Sarcoma) is related to chicken Yrk (Yes related kinase) (Tronick et al., 1985).

The presence of aromatic residues in similar positions and the distribution of proline residues with respect to these aromatic residues are the most striking features.

The UD of all SFK contain an unusually high number of hydrophobic residues for an IDR (Dunker et al., 2001). The disordered region of human Src contains 9 hydrophobic residues (4 Phe, 3 Leu, 2 Val) most of them conserved in other species. Of these, the aromatic residues corresponding to positions 32 and 54 in human Src are highly conserved. Position 32 is a phenylalanine in Src but a tyrosine in the other SFKs. Interestingly, some of these tyrosine residues are known to be phosphorylated (Amata et al., 2014) and therefore putative modulation sites of these interactions. The region between the conserved aromatic residues (F32 and F54 in c-Src) contains four proline residues in Src, Fyn and Fgr. Yes contains only three but includes P41 that is conserved in the four compared SFKs. The other SFK have between 3 and 5 proline residues between a pair of conserved aromatic residues.

Another conserved feature is related to the characteristic pattern associated to the functionally important ULBR (Pérez et al., 2013) in Src, with two phenylalanine residues (F64 and F67) separated by two glycine residues. The pattern $\Phi_1$xx Φ_2, where Φ_1 is phenylalanine or tyrosine, x is a turn promoting residue (G,S,N) and Φ_2 is an aromatic or a hydrophobic residue, is found in Fyn (^{50}YNNF^{53}, ^{64}FGGV^{67}), Yes (^{54}FSSL^{57}) and Fgr (^{53}YSNF^{57}) in the Src subfamily but is not present in Hck or Lyn from the other SFK subfamily.

Chemical Shift Perturbations caused by phenylalanine mutations in USH3.

In order to investigate the role of the phenylalanine residues in the UD of Src, we individually mutated each of them to alanine in USH3. The CSP with respect to native USH3 are shown in Fig. 7. The F#A mutants cause CSP in residues distant along the sequence of the UD and SH4 domains with respect to native USH3, confirming long-range interactions of the aromatic residues within the disordered region. The F32A mutation causes large CSP from residues 21 to 49 and F54A perturbs residues from 25 to 66. Mutants F64A and F67A, in the ULBR, also cause long-range CSP from residues 55 to 88, the latter already in the SH3 domain. These effects are likely caused by an overall reorientation of the connection between the Unique and SH3 domains. Additionally, F67A affects other residues in the UD (R48, G29, G30) and the SH4 domain (R14, R15, R16 and S17) and F64A affects K5 and also S17, confirming a mutual interaction between the ULBR and the SH4 domain.

Focusing on the SH3 domain, the F32A and F54 mutations induced shifts in the same regions, although the F32A

effects are stronger: the distal and nSrc loops and strands β2, β3 and β4. In contrast, the F64A and F67A mutations caused CSP mainly in the RT loop. A direct interaction between the ULBR and the RT loop is consistent with the correlated evolution observed between these regions in viral Src.

Phenylalanine residues contribute to UD compaction also in the absence of SH3.

To test the role of the conserved phenylalanine residues in the compaction of the disordered region in the absence of SH3, we introduced phenylalanine mutations in the SH4-UD construct. We compared ΔPRE profiles of variants F32A, F54A, F67A and a previously studied triple mutant in which F64, as well as L63 and G65, had been mutated to alanine. We placed the paramagnetic probe at position 27 or 59, choosing the one closer to the F#A mutation site while observing the changes in distant regions. Using the native variants with the spin label in the same position as a reference (Pérez et al., 2009) the average ΔPRE values in the region 30-50 (for MTSL in position 59) or 50-70 (for the spin label in position 27) decreased between 34% and 52% in the F#A variants (Fig. S7). Mutations of F32 (52%) and F54 (43%) have the largest effect, confirming the important role of these residues in the compaction of the UD.

Proline residues may alleviate the entropic cost of compacting a disordered domain.

Local dynamics could minimize the entropic cost of compacting the disordered region of Src. Correlated motions of peptide segments retain local flexibility while avoiding large-scale motions of the entire protein, preserving long-range interactions (Baxa et al., 2014). The abundant proline residues in IDPs may extend correlated motions beyond the nearest neighbor residues. We hypothesized that correlated motions could result in concerted departures from the average distance to a nearby paramagnetic center that could result in alternating maxima and minima in the observed PRE. This was indeed observed in the ΔPRE profiles of the disordered region of USH3 or the SH4-UD variant, with the MTSL probe attached at position 27 or 59 (Fig. 8A-B).

The separation of consecutive maxima has a narrow distribution with a peak every 5-6 residues. Nine of the maxima are found in residues situated next to proline residues, and all of the prolines have a ΔPRE maximum or minimum in the preceding or following residues. The ΔPRE extremes not directly associated to prolines include aromatic residues (H47, and the F64-F67 pair) or the 15-16 region of the SH4 domain involved in long-range interactions. Interestingly, these features are located in the longer interproline gaps and therefore, the pattern is more regular than the distribution of proline residues in the Src SH4 and UD. The observed ΔPRE oscillations are similar in the presence and in the absence of the SH3 domain. For the initial 50 residues the positions of the maxima are the same. However, in the region 51-70 the maximum ΔPRE values are observed for residues preceding proline in USH3 but for the residues after proline in the absence of the SH3 domain. The similar pattern of maxima and minima in the presence and in the absence of SH3 indicates that this is an intrinsic property of the disordered region and is consistent with preorganization of the SH4 and

Unique domains.

Phosphorylation of S17 affects the compaction of the disordered region of Src.

Phoshorylation of the UD od SFK is an important regulatory element (Amata et al. 2014). Phosphorylation of S17 of Src, mediated by PKA (cAMP-dependent protein kinase), is a well-characterized process although the mechanism by which this phosphorylation changes Src signaling is not well understood (Obara et al. 2004). Treatment with PKA and ATP of SH4-UD leads to complete and exclusive phosphorylation of S17.

Figure 8C compares the ΔPRE profiles of SH4-UD and its S17-phosphorylated forms with MTSL in position 27. The phosphorylated form shows PRE values closer to those predicted for a random coil in the 15-20 and 35-45 regions suggesting that preorganization of the region close to the interface between the SH4 and UD is disturbed by phosphorylation of S17. Interestingly, the ΔPRE become slightly larger in other regions more distant from the phosphorylation site, suggesting that interactions within the disordered regions are mutually connected in fuzzy complexes.

Human cancer cells expressing full length Src with mutations in the UD display reduced invasiveness.

The SH3 domain is in direct contact with the kinase domain in the autoinhibited form of Src (Xu et a. 1999) and even its open state is substantially compact in solution (Bernadó et al. 2008). This raises the question of the physiological role of the observed interactions in USH3 in the context of the entire protein. Figure 9A shows that the regions of the SH3 domain that are approached by residue 59 of the UD are also accessible in the full-length protein, even in its closed form.

Although the effect of mutations in the UD of human Src had been previously determined in an heterologous system (maturation of *Xenopus laevis* oocytes, Pérez et al., 2013), we tested the effect of UD mutation in SW620 cells. These cells have been isolated from a metastatic ganglion of a Dukes' type C colorectal carcinoma where Src is highly deregulated. However, they express a low level of endogenous SFK and therefore they exhibit moderate invasive activity. Previous studies (Leroy et al. 2009) demonstrated that the over-expression of the wild type form of human full-length Src strongly increased cell invasiveness.

Cell invasion assays were carried out using SW620 stable cell lines overexpressing either full-length Src WT or a mutated Src form with the three residues ^{63}LFG^{65} mutated to alanine (AAA mutant). As a negative control we used cells transformed with an empty vector (mock). Results are shown in B. Cells over-expressing the wild type form of c-Src displayed enhanced cell invasiveness (>30 fold higher) in comparison with control cells (mock). Conversely, the invasion ability of cells expressing the AAA mutant was reduced more than 50% with respect to wild type. The expression levels of the two full-length Src variants were very similar in each of the cell lines (Fig. S8). These results

demonstrate that the UD can play an important regulatory role in colorectal cancer cells.

Multiple sequence alignment identifies co-evolving residues linking the SH4, UD and SH3 domains

When comparing evolutionarily related proteins, correlated variations in pairs of residues are usually indicative of a functional interaction that is preserved by evolution. We used the GREMLIN software (Kamisetty et al., 2013) to systematically search for coevolving regions in an ensemble of related sequences sharing the SH4-UD-SH3 architecture. Residues in the SH4 and SH3 domains show co-evolution. Co-evolution is also detected inside the region between F32 and F54 of the UD and between the SH4 and UD (Fig. 6B). We had previously reported the correlation between variations in the ULBR and the RT loop in a series of viral Src variants (Maffei et al., 2015). The co-evolution patterns are in agreement with the structural proximity of the affected regions detected by PRE and support the compact structure of the intrinsically disordered domains of Src around the SH3 scaffold and its physiological relevance in the entire SFK family.

DISCUSSION

The largest sequence diversity among SFKs is at the UD. In spite of its intrinsically disordered character, the N-terminal region of Src is compacted (Fig. 2) around the SH3 domain, as previously suggested (Maffei et al., 2015). We have now found that many of the long-range interactions defining a compact form of the disordered domains of Src are preserved also in the absence of the SH3 domain (Fig. 5) and, therefore, we show that compaction is an intrinsic property of this disordered region. Mutation of the phenylalanine residues of the UD results in a lost of long-range interactions. Residues F32 and F54 are specially important in this respect and, interestingly, the corresponding positions are occupied by aromatic residues in all SKFs. A second feature, also related to aromatic residues, is the pattern $\Phi_1xx\Phi_2$, where Φ_1 is phenylalanine or tyrosine, x is a turn promoting residue (G,S,N) and Φ_2 is an aromatic or a hydrophobic residue, which includes F64 and F67 in Src, and is conserved in Yes, Fyn, Fgr and Yrk. This pattern is at the core of the ULBR that, when mutated, causes strong phenotypes in Src dependent processes (Pérez et al., 2013).

Long range contacts associated to a compact N-terminal regulatory region in SFKs are supported by sequence coevolution analysis (Fig. 6), highlighting the interaction between the SH4 and SH3 domains, as well as contacts involving the UD, in the region flanked by the conserved F32 and F54 residues, and the ULBR, containing F64 and F67 in Src.

The disordered region of Src forms multiple contacts with the SH3 domain, but the RT loop is the region most affected by perturbations in the SH4 domain (Fig. 4 and Table 1) or the ULBR (Fig. 7). Interestingly, the perturbations caused by deleting the initial ten residues of the SH4 domain or the entire domain go in opposite directions (Fig. 4). The interactions between the SH4 and ULBR are only observed in the presence of the SH3 domain (Fig. 5), confirming the

scaffolding role of the SH3 domain. The SAXS data in the absence of the SH3 domain does not show a detectable population of forms with a low radius of gyration, in spite of the fact that most of the compacting interactions detected by NMR in USH3 are also detected in the constructs lacking the folded domain. This apparent discrepancy between SAXS and NMR views of random coil behavior has been previously discussed (Kohn et al. 2004). Short distances dominate PRE data and individual contacts are detected independently from each other. In contrast, SAXS data reflect the global electron distribution. The observation of a substantial population of compact structures in the presence of the folded domain suggests that the scaffolding effect of the SH3 domain may originate from enhancing the cooperativity of interactions within the disordered region as well as simultaneous multiple interactions with the SH3 domain itself.

In conclusion, we have shown that the SH4, UD and SH3 domains of the human tyrosine kinase Src are intimately linked in an intramolecular fuzzy complex: the disordered domains are maintained in a compact, yet dynamic, form that can be described as a "constrained cloud", even in the absence of the SH3 domain. The term "fuzzy complex" is generally used for inter-molecular interactions involving disordered region but not leading to folding. This is a special case of intra-molecular fuzziness. This preorganization, favored by conserved patterns of aromatic and proline residues, may compensate for the entropic cost of the formation of the fuzzy complex around the SH3 scaffold. Aromatic-based interactions are often involved in higher-order fuzzy protein assemblies (Wu and Fuxreiter, 2016). The collapse of disordered proteins is of entropic nature, which is dominated by the content of polyproline II more than the proline abundance (Cheng et al. 2010). Proline, with a cyclic structure that locks its φ dihedral angle is generally considered to increase protein rigidity. However, in the context of disordered proteins, the capacity of proline to break regular secondary structures increases the flexibility of the region where proline is located and favor more compact ensembles. On the other hand, local correlated motions permit retaining long-range intramolecular interaction without a high entropic cost. Thus, the combination of proline and aromatic residues seems well suited to favor a compact state in the UD of Src which is preorganized to allow multiple weak contacts with the SH3 domain.

Some of the Src UD sequence features associated to this model are present in other UDs, suggesting a general mechanism by which the inherent characteristic of intrinsically disordered proteins can be incorporated to introduce new functions to a common module, in this case the core multidomain structure formed by the SH3, SH2 and SH1 domains of SFKs. The dramatic decrease of the invasive capacity of cancer cells in response to a localized mutation in the UD highlights the functional relevance of the disordered region of Src under physiological conditions. A complete characterization of the signaling pathways specifically affected by changes in the disordered regions will be reported elsewhere. The association of structural ambiguity with function is one of the indicators of fuzzy complexes (Miskei et al. 2016).

We hypothesize that fuzzy complexes incorporating the specific UD of SFKs form the basis of regulatory mechanisms fine-tuning the specific response of each kinase to environmental cues, sensed by the corresponding UD, for example as

phosphorylation events, interactions with lipids or calmodulin binding in the case of Src, the later regulated by calcium. Interestingly, the UD mediates the association of Src with a calcium channel in neurons and is cleaved in a calpain-mediated process when the calcium levels are increased. Phosphorylation of the UD may affect the fuzzy complexes directly, as shown here in the case of S17, or indirectly by changing the interactions of the UD with lipids or other proteins.

The SH3 domain is known to be part of the classical SFK regulation mechanism toggling between active and inactive states. Fuzzy intramolecular complexes nucleated by the SH3 domain of the various SFKs could mediate the coupling of a very diverse set of sensors (the specific UDs) to a highly conserved set of actuators (the conserved SH1 domains). The IDR regions could thus endow Src and the other SFK with the plasticity to integrate complex signaling responses. Our analysis of the fuzzy interactions in Src show how functional information can be encoded in the sequences and, hopefully, brings us one step closer to the goal of deciphering the IDR functional codes.

EXPERIMENTAL PROCEDURES

Cloning and mutagenesis

The USH3 protein variants, including the Δ10 and Δ20 truncated forms and the various single residue mutants, as well as the isolated SH3 were expressed as His_6-GST fusion proteins in a pETM30 vector with a TEV cleavage site after GST. The N-terminus resulting from cloning has the sequence GAMA. In the truncated forms the last alanine coincides with the native A11 or A21. In full length USH3 and SH3 the first native residues (G2 and V86 respectively) are located after alanine. SH4-UD constructs were cloned in a pET-14b plasmid along with a C-terminal Streptag® (SAWSHPQFEK) for purification purposes. Single residue mutations were introduced using the QuickChange site-directed mutagenesis kit (Stratagene). Detailed protocols on sample expression and purification are given in the SI.

SAXS and USH3 ensemble optimization

The reference ensemble of 10.000 structures was computed by attaching the disordered N- and C-terminal extensions, generated using Flexible-Meccano (Ozenne et al., 2012), to the human Src SH3 structure (PDB: 4HXJ). The N-terminal ensemble was generated using local dihedral angle φ/ψ propensities previously shown to reproduce the experimental residual dipolar coupling profiles of SH4-UD (Pérez et al., 2009). A sub-ensemble of 50 structures whose average theoretical scattering profiles reproduce the experimental curves (in the momentum transfer range of $0.0224 < s < 0.5$Å^{-1}) was selected using the ensemble optimization method (EOM) (Bernado et al., 2007). Detailed information on SAXS experiments and modeling can be found in the SI.

NMR PRE experiments

A 16-fold excess of MTSL was immediately added to DTT-free samples, just after elution from PD-10 columns, and left reacting overnight at 4ºC in the dark. Excess MTSL was removed using PD-10 columns. Reference diamagnetic spectra were obtained after adding 5-fold excess of ascorbic acid. Further details on the NMR experiments are given in the SI.

PRE analysis

Simulated random coil PRE values were calculated using Flexible Meccano 1.1 (Ozenne et al., 2012) from a pool of 50,000 conformers using the average line width in the ^{1}H dimension of the diamagnetic spectrum of each protein. ΔPREs were calculated as the difference between simulated and experimental data for each residue. Smoothed plots were obtained by convolution with a normalized 1D Gaussian kernel of window size = 7 and one standard deviation. Heat maps were generated using one standard deviation of the total ΔPREs as a lower threshold and the largest ΔPRE of all datasets as a common maximum in the gray scale.

Invasion assays

Full-length human Src was subcloned in pMX-pS-CESAR vector. Mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene). SW620 cells (American Type Culture Collection) were grown, infected and selected as described in Sirvent et al., 2007. Src expressers were isolated by fluorescence-activated cell sorting and maintained in culture with Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum in a humidified incubator at 37 °C and 5% CO_2. Cell invasion assays were performed in Boyden chambers (BD Bioscience) using 100,000 cells and in presence of 0.1 mL of Matrigel (0.5 mg/mL; BD Bioscience). After 32 hrs of invasion (37°C), cells were resuspended in PBS with Calcein for 30 minutes and counted from optical microscope images. Cell counts are the mean of three independent experiments.

Coevolution analysis

We used the GREMLIN software (Kamisetty et al. 2013) to systematically search for co-evolving regions with potential structural significance. Briefly, the HHblits algorithm (Remmert et al., 2011) searches a pre-clustered Uniprot database using a Hiden Markov Model to generate an ensemble of related sequences that are then used for Multiple Sequence Alignment. Sequences having >75% gaps are filtered out. This methodology enriches the search with orthologous sequences sharing domain architecture (Hegyi and Gerstein, 2001). Thus, we took profit of the SH4-Unique-SH3 architecture to bypass the low homology between Unique domains and used the 1-150 Src sequence as input. Although GREMLIN is typically used to obtain structural restrains for modeling 3D structures, our results show its potential

applicability to study also interacting points in fuzzy complexes between folded and disordered domains.

The search returned 151 c-Src USH3-related sequences that included human and non-human SFKs. The complete protein list is given in supplementary figure S6.

AUTHOR CONTRIBUTIONS

MP conceived the project; MP and MA designed the experiments; MA, MM and YP produced samples and performed NMR experiments; MA and JMCT processed and analyzed the data; TNC performed structural modeling and measured and analyzed SAXS data together with MA and PB; MM and SR performed and analyzed functional experiments. MP and MA wrote the manuscript, with contributions from other authors.

ACKNOWLEDGEMENTS

We thank N. Oliver for preparing the K5A S6A mutant. This work was supported by the Fundació Marató TV3 and the Spanish MINECO (BIO2013-45993R and BIO2016-78006R), co-financed with EU structural funds. We acknowledge the use of the LRB NMR of the University of Barcelona. The authors declare no competing financial interests.

REFERENCES

Amata, I., Maffei, M., Pons, M., 2014. Phosphorylation of unique domains of Src family kinases. Front. Genet. 5, 181.

Baxa, M.C., Haddadian, E.J., Jumper, J.M., Freed, K.F., Sosnick, T.R., 2014. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. Proc. Natl. Acad. Sci. U. S. A. 111, 15396–401.

Bernado, P., Mylonas, E., Petoukhov, M. V, Blackledge, M., Svergun, D.I., 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. 129, 5656–5664.

Bernadó, P., Pérez, Y., Svergun, D.I., and Pons, M. 2008. Structural characterization of the active and inactive states of Src kinase in solution by small- angle X-ray scattering. J.Mol.Biol. 376, 492–505.

Boggon, T.J., Eck, M.J., 2004. Structure and regulation of Src family kinases. Oncogene 23, 7918–7927.

Cheng, S., Cetinkaya, M., Gräter, F., 2010. How sequence determines elasticity of disordered proteins. Bioph. J. 99,

3863-3869.

Clore, G.M., Iwahara, J., 2009. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem. Rev. 109, 4108–39.

Dubois, F., Leroy, C. Simon, V. Bénistant, C.Roche S. 2015. YES oncogenic activity is specified by its SH4 domain and regulates RAS/MAPK signaling in colon carcinoma cells. Am J Cancer Res 5, 1972-87.

Dunker, A. K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z., 2001. Intrinsically disordered protein. J. Mol. Graph. Model. 19, 26–59.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Hegyi, H., Gerstein, M., 2001. Annotation Transfer for Genomics : Measuring Functional Divergence in Multi-Domain Proteins Annotation Transfer for Genomics : Measuring Functional Divergence in Multi-Domain Proteins. Genome Res. 1632–1640.

Hossain, M. I., Roulston, C.L., Kamaruddin, M.A., Chu, P.W., Ng, D.C.H., Dusting, G.J., Bjorge, J.D., Williamson, N.A., Fujita, D.J., Cheung, S.N., Chan, T.O., Hill, A.F., Cheng, H-C., 2013. A truncated fragment of Src protein kinase generated by calpain-mediated cleavage is a mediator of neuronal death in excitotoxicity. J. Biol. Chem., 288, 9696-9709.

Kamisetty, H., Ovchinnikov, S., Baker, D., 2013. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. 110, 15674–15679.

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.

Kohn, J.E., Millett, I.S., Jacob, J., Zagrovic, B., Dillon, T.M., Cingel, N., Dothager, R.S., Seifert, S., Thiyagarajan, P., Sosnick, T.R., Hasan, M.Z., Pande, V.S., Ruczinski, I., Doniach, S., Plaxco, K.W., 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc. Natl. Acad. Sci. 101, 12491-12496.

Leroy, C., Fialin, C., Sirvent, A., Simon, V., Urbach, S., Poncet, J., Robert, B., Jouin, P., Roche, S. 2009. Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates SRC invasive activity in advanced colon carcinoma cells. Cancer Res., 69, 2279-2286.

Liu, X.J, Gingrich, J.R., Vargas-Caballero, M., Dong, Y.N.Sengar, A., Beggs, S., Wang, S-H, Ding, H.K., Franklanda, P. W., Salter, M.W., 2008. Treatment of inflammatory and neuropathic pain by uncoupling Src from the NMDA receptor complex. Nature Medicine, 14, 1325-1332.

Luciano, F., Ricci, J.-E., and Auberger, P., 2001. Cleavage of Fyn and Lyn in their N-terminal unique regions during induction of apoptosis: a new mechanism for Src kinase regulation. *Oncogene* 20, 4935–4941.

Maffei, M., Arbesú, 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.

Miskei, M., Antal, C., Fuxreiter, M., 2016. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. Nucl. Acids Res.

Motlagh, H.N., Wrabl, J.O., Li, J., Hilser, V.J., 2014. The ensemble nature of allostery. Nature, 508, 331-339.

Obara, Y., Labudda, K., Dillon, T. J. & Stork, P. J. S. 2004. PKA phosphorylation of Src mediates Rap1 activation in NGF and cAMP signaling in PC12 cells. J. Cell Sci. 117, 6085–6094.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M.R., Segard, S., Bernadó, P., Charavay, C., Blackledge, M., 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28, 1463–70.

Pérez, Y., Gairí, M., Pons, M., Bernadó, 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.

Pérez, Y., Maffei, M., Igea, A., Amata, I., Gairí, M., Nebreda, A.R., Bernadó, P., Pons, M., 2013. Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. Sci. Rep. 3, 1295.

Remmert, M., Biegert, A., Hauser, A., Söding, J., 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods 9, 173–175.

Sigal, C.T., Zhou, W., Buser, C. a, McLaughlin, S., Resh, M.D., 1994. Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. Proc. Natl. Acad. Sci. 91, 12253–12257.

Sirvent, A., Boureux, A., Simon, V., Leroy, C., Roche, S., 2007. The tyrosine kinase Abl is required for Src-transforming activity in mouse fibroblasts and human breast cancer cells. Oncogene 26, 7313–7323.

Stein, P.L., Vogel, H., Soriano, P., 1994. Combined deficiencies of Src, Fyn, and Yes tyrosine kinases in mutant mice.

Genes Dev. 8, 1999–2007.

Tompa, P., Fuxreiter, M., 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. 33, 2–8.

Tronick, S.R., Popescu, N.C., Cheah, M.S., Swan, D.C., Amsbaugh, S.C., Lengel, C.R., DiPaolo, J. a, Robbins, K.C., 1985. Isolation and chromosomal localization of the human fgr protooncogene, a distinct member of the tyrosine kinase gene family. Proc. Natl. Acad. Sci. U. S. A. 82, 6595–6599.

Wright, P.E., Dyson, H.J., 2014. Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18–29.

Wu, H. And Fuxreiter, M., 2016. The structure and dynamics of higher-order assemblies: amlyloids, signalosomes, and granules. Cell, 165, 1055-1066.

Xu, W., Doshi, A., Lei, M., Eck, M. J. and Harrison, S. C. (1999) 'Crystal structures of c-Src reveal features of its autoinhibitory mechanism', *Molecular Cell*, 3(5), pp. 629–638.

FIGURE LEGENDS

Fig. 1. Structure of Src. The figure shows the domain structure of full length Src and the constructs used in this study. The constructs are represented schematically, to show the domains included. Multiple mutated variants were studied for several constructs and are listed in the right column. (*) The F64A mutant of the SH4-UD construct contained the additional L63A and G65A mutations. The sequence of residues 1-150 of wild-type Src is shown in the bottom with the location of secondary structure elements and loops on the SH3 domain highlighted. For PRE experiments, the native alanine residues were mutated to cysteines. The N-terminus of the constructs with a SH3 domain has the additional sequence GAMA resulting from TEV cleavage. The C-terminus from the SH4-UD constructs has a C-terminal Streptag® (SAWSHPQFEK). The alanine residues mutated to cysteine to introduce the MTSL tag are marked in green. The phenylalanine residues are marked in blue. The 3D structure of the folded part (i.e. excluding the SH4 and Unique domains) of the closed form of Src is shown in Figure 9 with the domains represented with the same colors.

Fig. 2. Compaction of the disordered domains of Src. Small Angle X-Ray Scattering (SAXS) of USH3 (A-B) and SH4-UD (C-D). A and C) Experimental Kratky plot (dots) overlapped with the curves calculated from the random coil pool (black lines) and the best fit ensemble (red for USH3, blue for SH4-UD). B and D) Frequency distributions of radii of gyration in the reference ensembles (grey bars) and the optimized sub-ensembles reproducing the experimental SAXS data (red for USH3, blue for SH4-UD). See also Fig. S1 and Table S1.

Fig. 3. Paramagnetic Relaxation Enhancement (PRE) observed in the SH3 domain of USH3 variants. PRE values measured in the SH3 domain of various constructs are mapped on the structure of the SH3 domain (PDB:4HXJ) with a color code scale indicated next to the PRE profile shown in the center of the figure. This profile corresponds to full length USH3 with the tag in position 27. The experimental profiles of the other constructs are given in Fig S2. (top) full length USH3 (A-C) with the paramagnetic tag introduced in residue 1 (A), 27 (B), or 59 (C); (bottom) N-terminal truncated by 10 (D) or 20 residues (E) with the tag in position 27.

Fig. 4. Chemical Shift Perturbations (CSP) in SH3 domain residues induced by changes in the SH4 domain. A) CSP with respect to native USH3 induced by truncation of the initial 10 (Δ10) or 20 (Δ20) residues. The threshold for statistical significance (gray line) was defined as the mean value of the lowest decile of CSP values in the set plus 5 standard deviations. (B) 2D plots of the ^{1}H and ^{15}N chemical shift differences of NH signals from individual residues in USH3 variants, with respect to the isolated SH3 domain. The gray cross marks the coordinate origin, corresponding to the reference (isolated SH3) and indicates the relative axis scale, representing 0.01 ppm in each direction of both axis; the gray circle corresponds to the chemical shift change observed in native USH3; pentagons correspond to Δ20 USH3; squares to Δ10 USH3; triangles to the K5A S6A mutant. Residues R98 and E100 belong to the RT-loop. Residues V114 and N116 to the nSrc loop. See also Figs. S3 and S4.

Fig. 5. ΔPRE values observed in the disordered regions of USH3 and the SH4-UD variant lacking the SH3 domain. A) Top: Experimental PRE values (bars) and the predicted random coil model (red line). Center: Differences between experimental and predicted values (ΔPRE), presented as single residue values (circles) or smoothed curves using a running Gaussian weighting function (black line), residues too close to the paramagnetic tag (in this example in position 59) are not assigned. Bottom: Smoothed ΔPRE values above a threshold are represented as heat maps. B) Compared pairs of ΔPRE heat maps for variants with (top) and without the SH3 domain (bottom). Residues 27 and 59 are in the UD. The label in the SH4 region was in position 1 for USH3 but in position 2 for SH4-UD due to different cloning strategies. C,D) Summary of the observed long range contacts in the absence (C) and in the presence of the SH3 domain (D). Wider lines indicate stronger relaxation effects. In panel D blue lines mark the PRE contacts with the SH3 domain shown in the center (PDB: 4HXJ). Raw data are available in Fig. S5.

Fig. 6. Sequence alignments and co-evolution. A) Sequence alignments, using MUSCLE (Edgar, 2004), of the UD of various human Src Family Kinase members (top) and of homologues of Src in various species (bottom). The most

559 conserved phenylalanine residues are indicated in orange. The conserved ΦxxΦ patterns (see text) are also indicated.
560 Proline residues are highlighted in green and histidines in blue**.** B) Co-evolution: the lines connect pairs of residues
561 identified as co-evolving. The width of the lines indicates the significance of the connections. The search returned 151
562 Src USH3-related sequences (Fig. S6). The co-evolution analysis was done using GREMLIN (Kamisetty et al., 2013).
563
564 **Fig. 7. CSP induced by phenylalanine to alanine USH3 mutations.** Plots of CSP of individual F#A mutants with
565 respect to native USH3: F32A (A), F54A (B), F64A (C), and F67A (D). Perturbations on the UD and SH4 domains (left
566 panels) were measured at 278K. CSPs in the SH3 domain (right panels) were measured at 298K. The threshold (gray
567 line) was defined as in Figure 4.
568
569 **Fig. 8. Oscillatory ΔPRE patterns**. ΔPRE patterns showing oscillations in the disordered regions of constructs lacking
570 (A) or including (B) the SH3 domain, with the spin label in residue 27 (blue and purple) or residue 59 (red and green).
571 Circles represent ΔPRE calculated for individual residues and the continuous lines are smoothed by averaging the
572 values from neighbor residues with a Gaussian filter. See also Fig. S7. (C) Effect of the phosphorylation of S17 of the
573 SH4-UD construct with the spin label in position 27. The ΔPRE patterns of the phosphorylated form (orange) and non-
574 phosphorylated (black) forms are compared. The phosphorylation site is marked by the red pointed line.
575
576 **Figure 9. Full length Src and the effect of the Unique domain in the invasive capacity of cancer cells**. A) Two
577 views of the X-ray structure of human Src, showing only the SH3, SH2 and SH1 domains in its closed form (PDB:
578 2SRC, Xu *et al.*, 1999). The domains are represented with the same color code as in Figure 1, except for the SH3
579 domain which is represented in white but with the sites affected by the approximation of residue 59 given as shades of
580 red to show that most of the SH3 residues interacting with the disordered domain are accessible even in the closed form.
581 The orientation and color code of the SH3 domain in the ribbon representation are directly comparable to those of Fig.
582 3. The views of the ribbon and surface representations are rotated 60ºC. The linker between the SH2 and SH1 domains,
583 which is interacting with the SH3 domain is shown in dark blue. The Unique domain, missing in the crystal structure,
584 would be connected to the N-terminus of the SH3 domain. B). Invasion assay. Micrographs of SW620 human
585 colorectal cancer cells stably transfected with empty vector (Mock), wild type full length Src (WT) and full length Src
586 with residues ^{63}LFG^{65} mutated to alanine (AAA). The cell counts of the invading cells are shown with red bars over
587 each micrograhs, with a common scale shown in the right. Red bars represent the mean of 3 independent experiments;
588 the error bar corresponds to one standard deviation. See also Fig. S8.

589 **TABLES**

590 **Table 1: Integrated CSP of SH4 variants of USH3 with respect to isolated SH3**

Region	CSP (ppm)	% change from native form		
	Native	K5AS6A	Δ10	Δ20
Central[a]	0.6771	+15.0%	+15.0%	-7.7%
RT-loop[b]	0.2161	+22.8%	+17.9%	-26.5%
nSrc loop[c]	0.1975	-6.3%	-10.8%	-10.8%

591 a) Residues 95-143; b) Residues 94-107; c) Residues 111-120