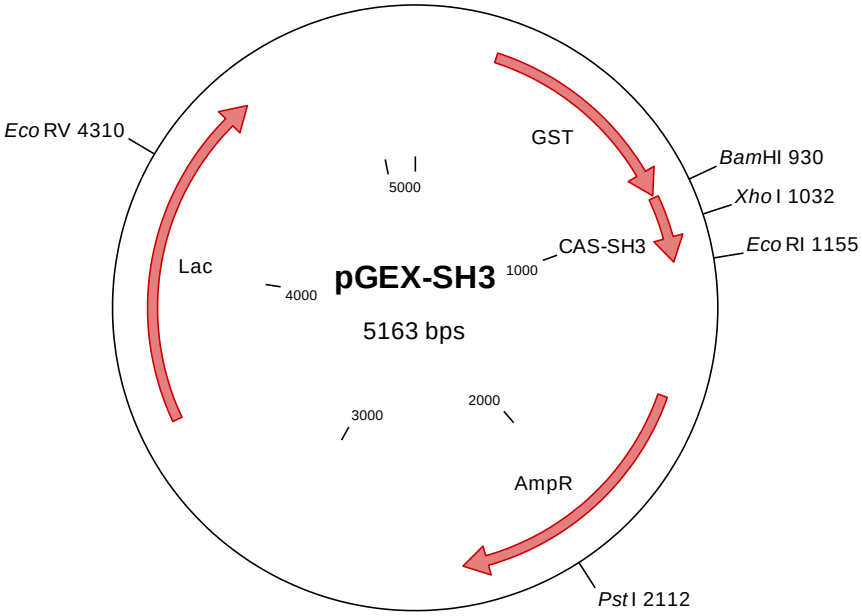
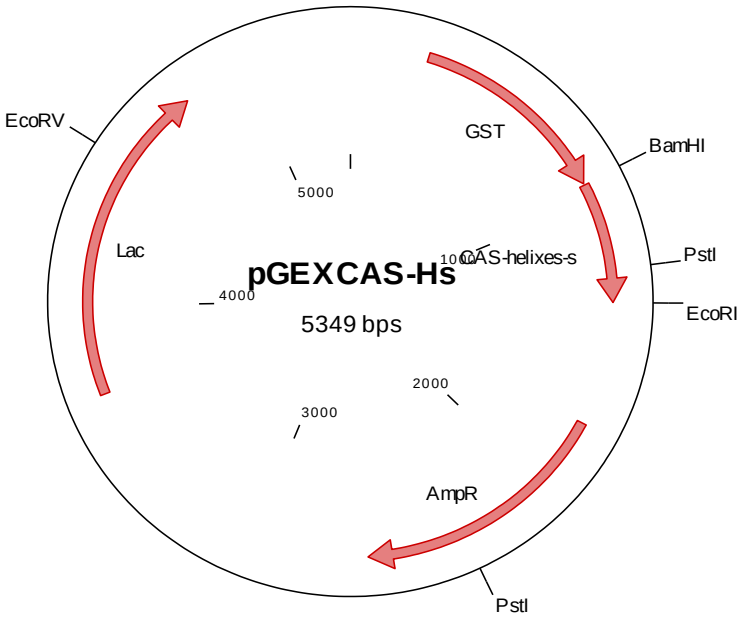
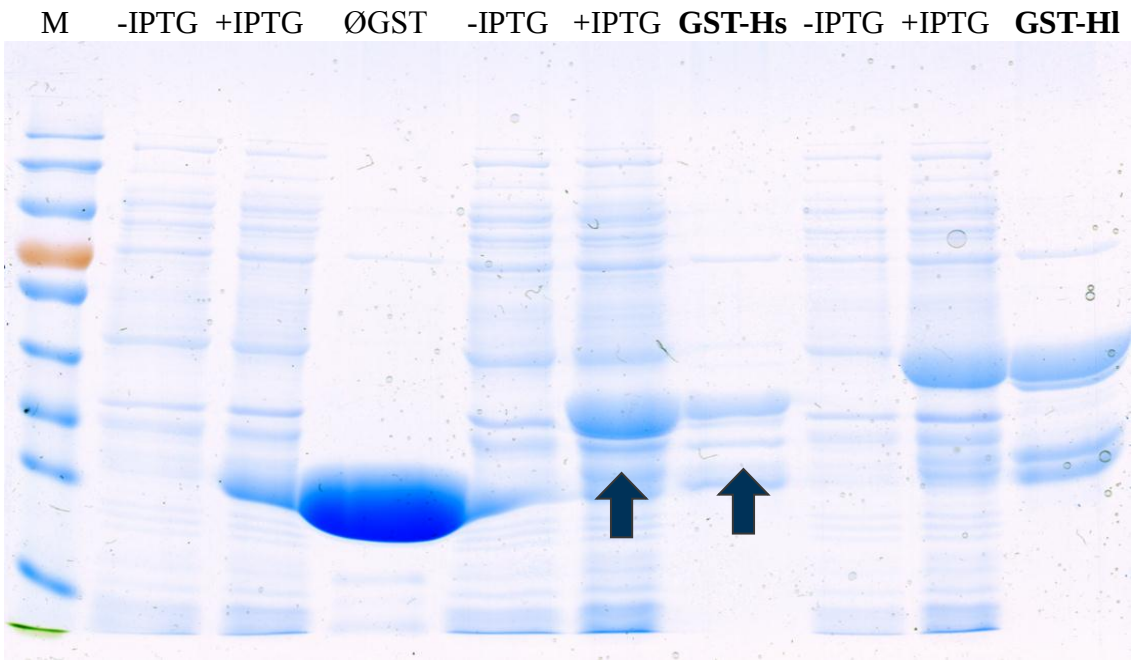


Supporting data S5.1 Expression plasmid pGEX-2T-CAS-SH3_WT was used for expression of GST fused CAS SH3 domain in bacteria *E. coli* and subsequent purification. The plasmid was created by Dr. Rösel.

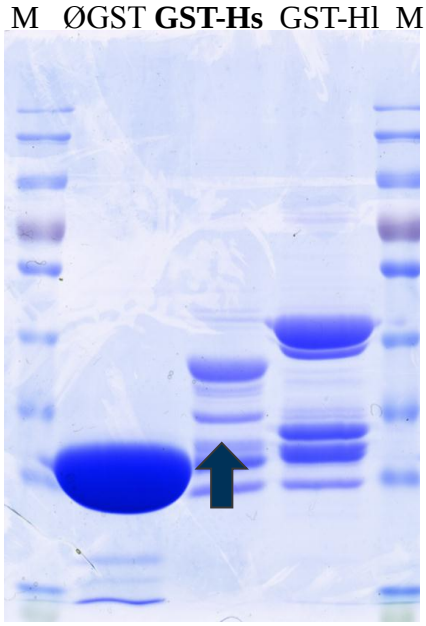


Supporting data S5.2 A schematic map of bacterial expression plasmid pGEX-CAS-Hs showing restriction sites and domain structure. Size of vector fragment is 4938bp and size of insert is 411bp. Restriction enzymes used for plasmids were BamH1 and EcoR1.

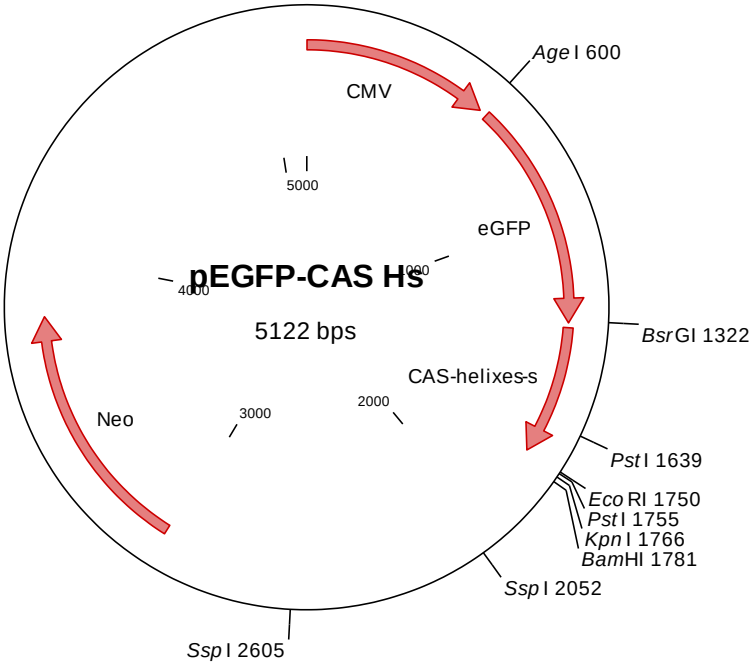
A)



B)

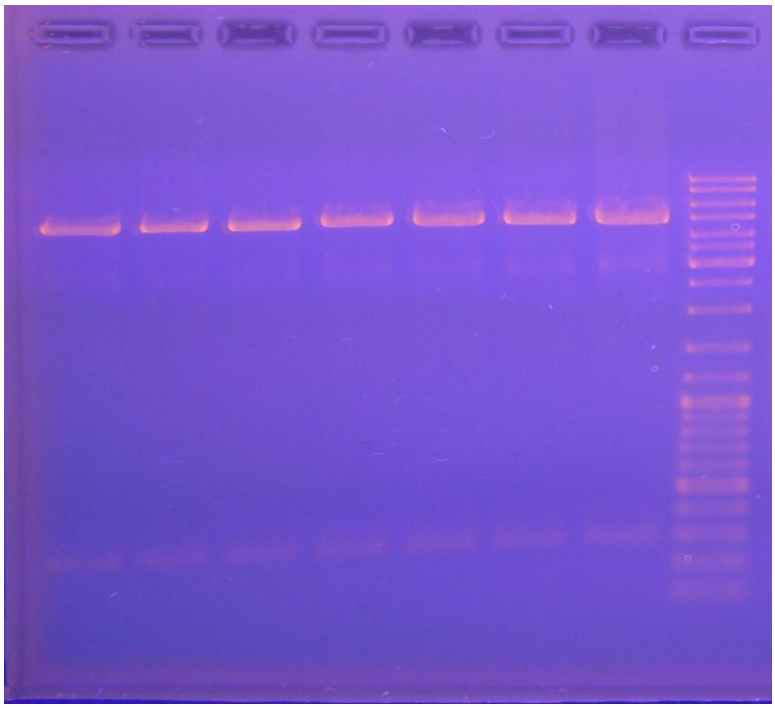


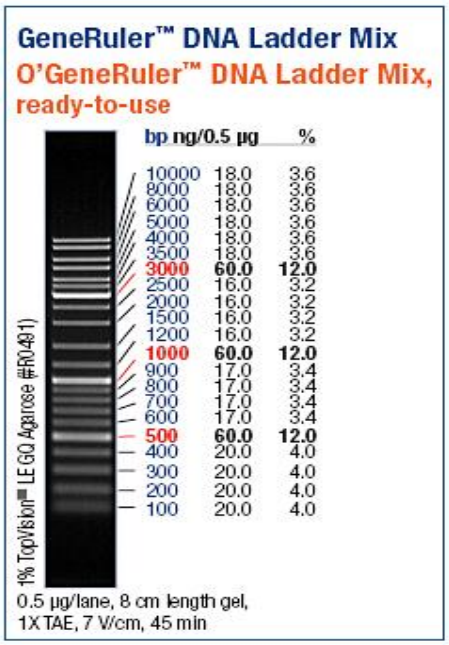
Supporting data S5.3 Difficulties of GST-Hs purification. A) The first gel shows that expression and purification of three recombinant proteins domains: ØGST („empty“ GST), GST-Hs (Helix short, C - terminal 133 amino acid region of CAS), GST-Hl (Helix long, 181aa C-terminal region of CAS). The first of three lines is always a sample of bacterial cell culture (transformed with the concrete pGEX expression plasmid) before addition of IPTG. Second line represent sample of bacteria culture four hours after IPTG induction. The third line is showing purified GST recombinant protein. Even the expression in bacterial cells is high, the purification efficiency is low (especially for GST-Hs, arrows). B) Optimizing the purification process of CAS CCH domain did not help to improve the purification of Short variant of CAS CCH domain (GST-Hs, arrow).



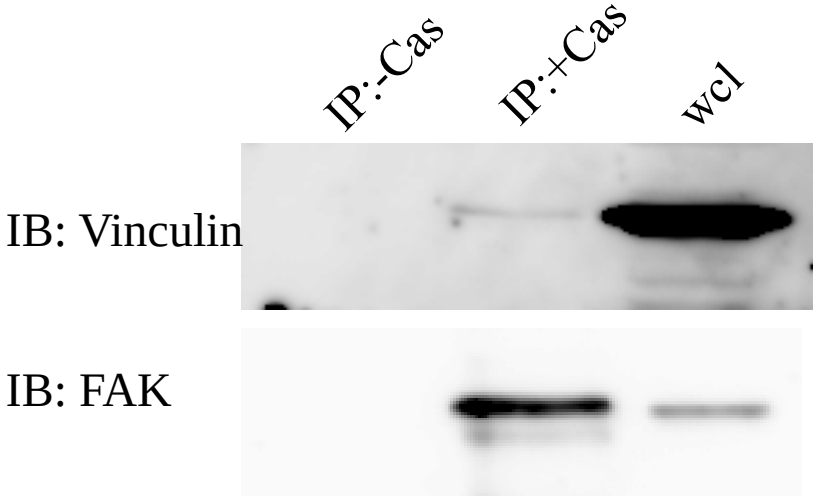
Supporting data S5.4 A schematic map of bacterial expression plasmid pEGFP-CAS-Hs showing restriction sites and domain structure. Size of vector fragment is 4711bp and size of insert is 411bp. Restriction enzymes used for plasmid preparation were BglI1 and EcoR1.

GFP-CAS-Hs / GFP-CAS-Hl M





Supporting data S5.5 Verification of GFP-CCH constructs by restriction digestion. Seven colonies (three of GFP-Hs and four of GFP-Hl were picked of kanamycine selective plates that were used for plasmid isolation). Plasmids were digested with BglII and EcoRI restriction endonucleases. The restriction mixture was run by agarose electrophoresis and the fragments were visualized with the use of ethidium bromide and UV. All seven colonies were proved to be convenient for further experiments according to expected fragmets sizes (4835bp, 287bp for eGFP-CAS-Hs and 4979bp and 287 for eGFP-CAS-Hl). DNA ladder #SM0331 (Fermentas) was used.



Supporting data S5.6 Interaction of CAS with vinculin is much weaker when compared to FAK. Endogenous p130Cas was immunoprecipitated from Hela cell lysates (P130Cas (C-20), sc-860) as described in Methods (4.4.4) and co-precipitated vinculin, top panel (Vinculin (N-19), sc-7649) and co-precipitated FAK, bottom panel (FAK (C-20)) was detected by immunobloting. Precipitate without antibody was used as negative control (IP: -Cas). Whole cell lysate (wcl) was used as positive control.

Abi-1 (human)	Q8IZP0		
Abi-2 (human)	Q9NYB9		
Abl (human)	P00519	PLCG2 (human)	P16885
Arg (human)	P42684	PRMT2 (human)	P55345
ARHGAP12 (human)	Q8IWW6	SAMSN1 (human)	Q9NSI8
ARHGEF6 (human)	Q15052	SASH1 (human)	O94885
Btk (human)	Q06187	SH3D19 (human)	Q5HYK7
Cas-L (human)	Q14511	SH3GL1 (human)	Q99961
CASKIN1 (human)	Q8WXD9	SH3PXD2B (human)	A1X283
CASKIN2 (human)	Q8WXE0	SH3RF1 (human)	Q7Z6J0
cortactin (human)	Q14247	SLAP-130 (human)	O15117
Crk (human)	P46108	SLY (human)	O75995
CrkL (human)	P46109	SPTAN1 (human)	Q13813
Eps8 (human)	Q12929	TEC (human)	P42680
Fgr (human)	P09769	Tks5 (human)	Q5TCZ1
FNBP1 (human)	Q96RU3	TXK (human)	P42681
Fyn (human)	P06241	VAV1 (human)	P15498
GADS (human)	O75791	Yes (human)	P07947
Grb2 (human)	P62993	ABI3 (mouse)	Q8BYZ1
Hck (human)	P08631	CIN85 (mouse)	Q8R550
ITK (human)	Q08881	ephexin-1 (mouse)	Q923H2
ITSN2 (human)	Q9NZM3	Nck1 (mouse)	Q9Z279
MLK4 (human)	Q5TCX8	P130Cas (mouse)	Q61140
Nck2 (human)	O43639	PSD-95 (mouse)	Q62108
PACSIN1 (human)	Q9BY11	SORBS1 iso5 (mouse)	Q62417-5
PACSIN3 (human)	Q9UKS6	Src (mouse)	P05480
PLCG1 (human)	P19174	Src iso2 (mouse)	NP_001020566

Supporting data S5.7 UNIPROT codes of proteins with tyrosine phosphorylation within SH3 domain. Outcome of PSP database to October, 2011.

Y-p	protein (organism)	UniProt	Y-p	protein (organism)	UniProt
Y455-p	Abi-1 (human)	Q8IZP0	Y428-p	Abi-1 (mouse)	Q8CBW3
Y457-p	Abi-1 (human)	Q8IZP0	Y430-p	Abi-1 (mouse)	Q8CBW3
Y460-p	Abi-2 (human)	Q9NYB9	Y393-p	Abi-2 (mouse)	P62484
Y462-p	Abi-2 (human)	Q9NYB9	Y395-p	Abi-2 (mouse)	P62484
Y93-p	Abl (human)	P00519	Y93-p	Abl (mouse)	P00520
Y223-p	Btk (human)	Q06187	Y223-p	Btk (mouse)	P35991
Y225-p	Btk (human)	Q06187	Y225-p	Btk (mouse)	P35991
Y86-p	Fgr (human)	P09769	Y74-p	Fgr (mouse)	P14234
Y91-p	Fyn (human)	P06241	Y91-p	Fyn (mouse)	P39688
Y180-p	ITK (human)	Q08881	Y186-p	ITK (mouse)	Q03526
Y47-p	MLK4 (human)	Q5TCX8	Y33-p	MLK4 (mouse)	Q8VDG6
Y372-p	PACSIN3 (human)	Q9UKS6	Y372-p	PACSIN3 (rat)	Q5I2Z0
Y811-p	PLCG2 (human)	P16885	Y811-p	PLCG2 (mouse)	Q8CIH5
Y818-p	PLCG2 (human)	P16885	Y818-p	PLCG2 (mouse)	Q8CIH5
Y315-p	SH3GL1 (human)	Q99961	Y315-p	SH3GL1 (mouse)	Q62419
Y189-p	SLY (human)	O75995	Y189-p	SLY (mouse)	Q8K352
Y791-p	VAV1 (human)	P15498	Y791-p	VAV1 (mouse)	P27870
Y826-p	VAV1 (human)	P15498	Y826-p	VAV1 (mouse)	P27870
Y100-p	Yes (human)	P07947	Y98-p	Yes (mouse)	Q04736
Y432-p	PSD-95 (mouse)	Q62108	Y432-p	PSD-95 (rat)	P31016

Supporting data S5.8 List of phosphorylated tyrosines that occured in SH3 domains of two orthologue organisms.

Y-p	protein (organism)	UniProt	Y-p	protein (organism)	UniProt
Y296-p	CASKIN1 (human)	Q8WXD9	Y296-p	CASKIN1 (mouse)	Q6P9K8
Y239-p	Crk (human)	P46108	Y239-p	Crk (mouse)	Q64010
Y251-p	Crk (human)	P46108	Y251-p	Crk (mouse)	Q64010
Y132-p	CrkL (human)	P46109	Y132-p	CrkL (mouse)	P47941
Y251-p	CrkL (human)	P46109	Y251-p	CrkL (mouse)	P47941
Y50-p	Nck2 (human)	O43639	Y50-p	Nck2 (mouse)	O55033
Y179-p	SAMSN1 (human)	Q9NSI8	Y179-p	SAMSN1 (mouse)	P57725
Y570-p	SASH1 (human)	O94885	Y563-p	SASH1 (mouse)	P59808

Y-p	protein (organism)	UniProt
Y296-p	CASKIN1 (rat)	Q8VHK2
Y240-p	Crk (chicken)	Q04929
Y251-p	Crk (rat)	Q63768
Y132-p	CrkL (rat)	Q5U2U2
Y251-p	CrkL (rat)	Q5U2U2
Y50-p	Nck2 (rat)	NP_001101686
Y171-p	SAMSN1 (rat)	Q8VI91
Y567-p	SASH1 (rat)	XP_001069000

Supporting data S5.9 List of phosphorylated tyrosines that occured in SH3 domains of three orthologues

Q9BY11	P06241	P16885	Q8N157	P42681
Q6FIA3	P09769	Q05BB3	Q5TCX8	Q9UKS6
P08631	Q8TEJ3	Q5JY90	O15034	P07947
O94868	A6NI59	Q13690	Q99962	Q5HYK7
O75791	Q08881	Q1JPZ3	Q8IW46b	
P42684	Q8WV41	Q99961	Q05D26	
O60498b	Q13813	Q14247	O43586	
Q9H3Y6	Q9UJU6	P27986-4	Q59FK4	

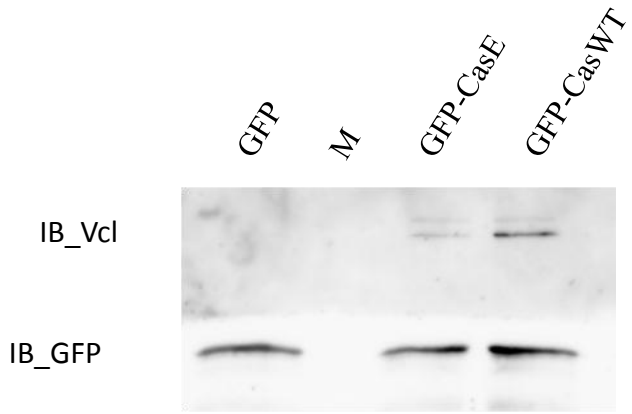
Supporting data S5.10 UNIPROT codes of proteins with ALYDY/ALYDF motif within SH3 domain. Based on SMART database.

Q05D26/64-110	Q59F39/222-268	Q5TZC3/391-438	B4DEM2/274-320
A2VCM5/275-305	B4DXL9/326-372	D3DQR0/370-417	D3DPL7/83-130
O94868/573-621	B7Z6W7/83-113	Q9H803/411-457	Q96IN1/64-110
B4E358/182-224	Q9UJU6/377-423	Q8N707/461-503	Q53HG7/498-543
Q13813/973-1018	B4DUI2/369-411	Q4FD35/1057-1103	B4DDD6/353-399
B3KRK4/182-227	P07947/97-144	Q14247/498-543	B4DQ92/397-442
Q5TCX8/44-94	P08631/84-130	Q9P0L4/56-95	B4DDU5/307-353
Q9BY11/391-438	P09769/83-130	Q5TGY6/83-130	B4DGT1/973-1018
B4DDP6/329-375	P78453/83-130	B4DUF9/283-329	D3DUH1/97-144
B4DKZ4/315-361	B4DTV8/973-1018	Q96H99/461-506	C9K004/345-390
A8MUA4/182-227	P12931/90-137	A8K1I1/62-108	P51451/64-110
P16885/775-821	Q9UNF0/432-479	O43586/365-410	D3DUL3/775-821
B4E3H3/642-688	Q9UKS6/369-416	Q6FIA3/391-438	B2RA70/97-144
B2RAW1/220-266	A8K4G3/64-110	Q8N157/1057-1103	Q5JY90/220-266
Q5TZC2/349-396	Q08881/177-223	A6NG51/973-1018	Q06187/220-266
B4DFJ9/391-415	B4DK38/437-485	P06241/88-135	Q59FH4/305-351

Supporting data S5.11 UNIPROT codes of proteins with ALYDY motif within SH3 domain. Based on Pfam. SH3 domain amino-acid-regions are shown next to the UniProt codes.

C9JNJ4/262-311	A2RQD7/319-365	P00519/67-113
B7UEF2/77-123	B3KY23/491-536	O75886/208-253
P42684/113-159	C9JDW6/733-778	Q13690/116-162
Q13691/67-113	Q8TEJ3/262-311	B7Z8F8/251-296
D1MPS6/77-123	Q6FGM0/312-357	Q8WV41/6-53
Q9BVL7/43-88	Q7Z6J0/202-251	Q14020/77-123
B7Z7M7/249-292	B7Z8J2/211-256	Q9UQD4/297-342
B7Z8I3/164-209	A2RQD6/406-452	Q59FK4/104-150
D3DVD8/983-1028	B5MEB6/92-138	Q6FHA6/277-322
A9UF07/727-773	B7UEF5/92-138	Q13848/65-111
D3DPA1/208-253	A1Z199/64-97	Q9BWP4/42-87
P42681/88-134	A3RL30/74-107	Q96RF0/6-53
B7UEF3/113-159	Q13915/40-86	A9UF02/581-627
B7UEF4/92-138	B0ZRR1/103-149	O75791/277-322
P02549/983-1028	B1NM17/6-53	B3KS04/300-345
Q99961/312-357	Q5HYK7/736-781	Q6FI14/277-322
B7Z8E3/205-250		

Supporting data S5.12 UNIPROT codes of proteins with ALYDF motif within SH3 domain. Based on Pfam. SH3 domain amino-acid-regions are shown next to the UniProt codes.



Supporting data S6.1 Interaction between CAS and vinculin is decreased but not totally diminished when <u>full-length</u> Y12E variant of CAS is used. MDA cells were transfected with empty GFP vector (negative control), full-length CAS fused with GFP and a Y12E variant CAS fused with GFP. The phosphomimiking mutation of CAS decreased its interaction with vinculin. By contrast, pull-down experiment with isolated Y12E CAS SH3 domain showed almost no interaction with vinculin (Figure 5.11). This suggests that some other part of CAS protein interact with vinculin. Later, we reported that CAS CCH domain bind vinculin. This experiment was not repeated.

SH3 domain tyrosine phosphorylation – sites, role and evolution

Zuzana Tatárová[1], Jan Brábek[1], Daniel Rösel[1] and Marian Novotný[1]

1 Department of Cell Biology, Faculty of Science, Charles University in Prague

Funding:

This research was supported by grants of the Grant Agency of the Czech Republic P305/10/0205, League against Cancer Prague and in part by Czech Ministry of Education Youth and Sport MSM0021620858). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

SH3 domains are eukaryotic protein domains that participate in a plethora of cellular processes including signal transduction, proliferation, and cellular movement. Several studies indicate that tyrosine phosphorylation could play a significant role in the regulation of SH3 domains.

Results

To explore the incidence of the tyrosine phosphorylation within SH3 domains we queried the PhosphoSite Plus database of phosphorylation sites. Over 100 tyrosine phosphorylations occurring on 20 different SH3 domain positions were identified. The tyrosine corresponding to c-Src Tyr-90 was by far the most frequently identified SH3 domain phosphorylation site. A comparison of sequences around this tyrosine led to delineation of a preferred sequence motif AL**Y**D(Y/F). This motif is present in about 15% of human SH3 domains and is structurally well conserved. We further observed that tyrosine phosphorylation is more abundant than serine or threonine phosphorylation within SH3 domains and other adaptor domains, such as SH2 or WW domains. Tyrosine phosphorylation could represent an important regulatory mechanism of adaptor domains.

Conclusions

While tyrosine phosphorylation typically promotes signaling protein interactions via SH2 or PTB domains, its role in SH3 domains is the opposite - it blocks or prevents interactions. The regulatory function of tyrosine phosphorylation is most likely achieved by the phosphate moiety and its charge interfering with binding of polyproline helices of SH3 domain interacting partners.

Introduction

The SH3 domain is one of the most well characterized protein interaction modules. SH3 domain-mediated signaling is involved in all basic cellular processes as well as in many pathological conditions, including malignant transformation (reviewed in [1]).

SH3-mediated signaling processes are mostly driven by the recognition of polyproline-II helices by SH3 domain structures [2]. The SH3 domain ligand- binding surface plays a key role in intramolecular and intermolecular interactions [3]. It contains three hydrophobic pockets, each containing a cluster of conserved amino acid residues. Mutational analysis of SH3 domains identified key residues necessary for interactions with ligands. For example, the essential residues for Src SH3 domain ligand binding are Y90, N135 and Y136 in the first pocket, Y92, W118 and P133 in the second pocket, and D99 and Y131 in the third pocket (numbering based on chicken c-Src)[4]; [5].

Protein phosphorylation is one of the most fundamental regulatory events in eukaryotic cells [6]. The importance of reversible tyrosine phosphorylation in the regulation of essential cellular functions is underscored by the fact that tyrosine kinases comprise the largest group of oncoproteins [7].

Tyrosine phosphorylation is a relatively recent evolutionary innovation, having emerged approximately 600 million years ago, just prior to emergence of the first multicellular organisms [8]. The full phosphotyrosine signaling system including “writer” (kinase), “eraser” (phosphate) and “reader” (SH2 domain) is present in choanoflagellate *Monosiga brevicollis* [9]; [10], but elements of the system already appear in other unicellular organisms, such as *Acanthamoeba* [11].

During the past two decades, tyrosine phosphorylation within SH3 domains of several signaling proteins was discovered [12]; [13]; [14]. In some cases, mutational analyses were performed to

determine the functional importance of a particular phosphorylated tyrosine. Results of these studies brought substantial evidence for a significant role of phosphorylation on the well conserved tyrosines within SH3 domain hydrophobic pockets in regulating the binding capacity of the SH3 domain and intramolecular regulation of signaling proteins. This mechanism of regulation seems to be used in various cellular processes and we hypothesize that it could be universally applicable to regulate signal transduction pathways mediated by proteins containing SH3 domains.

We surveyed available data from phosphoproteomic and structural studies to explore the abundance and variability of SH3 domain tyrosine phosphorylation sites, and to identify SH3 domain phosphorylation motifs. We also analyzed structural conservation of the AL**Y**D(Y/F) motif - the most frequently phosphorylated SH3 domain motif. Our results further support recent experimental observations that tyrosine phosphorylation within SH3 domains plays a critical role in the regulation of their function.

Results and Discussion

SURVEY OF SH3 DOMAIN PHOSPHORYLATION

SH3 domains are common protein interaction modules. Over 16000 SH3 domains in more than 12500 different proteins are described in the SMART database (October 2011). More than 97 % of those occur in eukaryotic proteins. A growing body of experimental evidence indicates that tyrosine phosphorylation plays a significant role in regulation of many SH3 domains (Table 1).

Tyrosine phosphorylation of SH3 domains has an unorthodox effect on protein function. Tyrosine phosphorylation is perhaps best known for its role in facilitating protein-protein interactions through the recognition of phosphotyrosine by a protein with a SH2 or PTB domain [15]. This usually leads to signal propagation. In contrast, the tyrosine phosphorylation of SH3 domains prevents or reduces the affinity of protein-protein interactions (Table 1). This can cause a switch in cell behavior, as in the case of chronic myeloid leukemia cells where phosphorylation of the SH3 domain of c-Abl enhances transformation potential [16], [13].

We queried the PhosphoSite Plus database for all phosphorylations within SH3 domains. At the time of the survey (October 2011), 188 distinct phosphorylation sites in 127 different SH3 domains were described in the database (File S1). Of these, 106 were tyrosine phosphorylations which were further analyzed.

SH3 domain sequences were aligned to determine the abundance of tyrosine phosphorylations at individual positions within the domain (Figure 1). To avoid redundancy, we included only one of the orthologue and paralogue (isoforms) sequences with identical phosphorylation pattern in the alignment. Fifty-two unique SMART-based SH3 domain sequences were aligned (File S2). A total of 36 protein domains were phosphorylated at one tyrosine site, 15 of them at two tyrosine sites and one (PLCγ2) on three sites.

We also analyzed the conservation of phosphorylation amongst orthologue sequences. We

found 20 phosphorylated tyrosines that occurred in SH3 domains of two ortologues (Table S1) and eight phosphorylated tyrosines that were present in SH3 domains of three orthologues (Table S2). Our previous experimental data have further confirmed the phosphorylation of Tyr 12 in human, rat and mouse p130Cas [14]. Although available phosphorylation data are incomplete, we can conclude that a significant proportion of phosphosites (64 out of 106 in our survey) is present in more than one organism, further supporting the importance of SH3 domain tyrosine phosphorylation.

To unify the numbering of positions, we used the protein amino acid positions in the alignment in Figure 1 as our reference. The alignment showed that most tyrosine phosphorylations were detected at positions 7 and 66 (Table 2). These positions correspond to Y90 and Y131 in chicken Src SH3 domain localized, respectively, in the first and the third surface hydrophobic pockets. Therefore, both of these tyrosines are involved in ligand binding [17].

ANALYSIS OF ABUNDANT PHOSPHORYLATION SITES

To further analyze the sequence surrounding the two most phospho-enriched positions of Tyr 7 and 66, we created sequence logos (Figure 2, WebLogo [18]). The resulting consensus logos show an absence of strong amino acid conservation around Tyr 66 (with the exception of Pro on the position +2), while sequence around phosphorylated Tyr 7 is more conserved. Alanine at the position -2, leucine at position -1, and aspartate at the position +1 from Tyr 7, all show very strong conservation. The position +2 is predominantly occupied by amino acids with an aromatic ring – tyrosine and phenylalanine. Thus AL**Y**D(Y/F) is the most favorable motif for tyrosine phosphorylation in the SH3 domain. Since the sequence around Tyr 66 was not as well conserved and there were fewer observations of tyrosine phosphorylation on this site, further analysis was concentrated on the Tyr 7 site.

Of 304 human SH3 domains in the SMART database, the AL**Y**DY motif around Tyr 7 appears in 21 domains and the AL**Y**DF motif appears in 15 SH3 domains (File S3). Of those 36 sequences, 12 are known to be phosphorylated at Tyr 7 according to PhospositePlus. There are many domain definition programs available and they differ significantly in a number of predicted proteins with a particular domain. We therefore also evaluated the number of human SH3 domains with AL**Y**D(Y/F) sequence with an independent domain definition program – Pfam [19]. There are 750 human SH3_1 domains in the Pfam database. Among those, there are 113 sequences with either AL**Y**DY (64) or AL**Y**DF (49) motif (File S4 and S5). Results from the two independent domain definition systems roughly agree on the estimate that 12-15% of human SH3 domains possess AL**Y**D(Y/F) sequence motif that can potentially be phosphorylated.

Although it is very unlikely that all these motifs will get phosphorylated, we expect more experimental evidence on the significance of Tyr 7 phosphorylation in the near future due to an ever increasing amount of phosphosite data.

The high conservation of the sequence around Tyr 7 suggests that it could be phosphorylated by a specific group of kinases. We used GPS [20] and PhosphoMotifFinder [21] database/software to predict kinases responsible for the phosphorylation of Tyr 7 within the AL**Y**D(Y/F) motif. Both programs indicated Src-family kinases as the likely kinases for this site. GPS further suggested FAK, Btk, PDGF and Abl as potential kinases. AL**Y**DY motif of Itk was shown to by autophosphorylated by Itk itself [22]. This suggests that the phosphorylation of Tyr 7 is not mediated by a particular kinase in a specific cellular compartment. Rather, different kinase families working in different compartments of the cell may regulate SH3 domains through phosphorylation of the AL**Y**D(Y/F) motif.

STRUCTURAL CONSERVATION OF THE AL**Y**D(Y/F) MOTIF

Structural alignment was employed to further evaluate the AL**Y**D(Y/F) motif. There are 104 known 3D structures of SH3 domains with an AL**Y**D(Y/F) motif, representing 16 different proteins. A representative structure was selected for each protein and structurally aligned to c-Src structure (1FMK). Results of the structural alignment are shown in Table 2. The AL**Y**D(Y/F) motif is a part of the loop that connects the first and second strands in the structure. The loop folds in the conformation that is similar to the structure of two interacting strands in β-sheet (Figure S1). The conformation of the loop is (in the case of 1FMK) stabilized by three hydrogen bonds in-between the main chain atoms of amino acids within this loop: two hydrogen bonds between Tyr 9 and Phe 24 and by a hydrogen bond between Ala 5 and Gly 27. It is also stabilized by a hydrogen bond between the main chain atoms of Leu 6 and Tyr 71 that lies in the loop connecting strands four and five (numbering based on alignment in Figure1).

The AL**Y**D(Y/F) motif is structurally well conserved. The root mean square distance (RMSD) for the C-alpha atoms in the motif was found to be typically less than half of the average RMSD for the whole SH3 domain (Table 3). Figure 3 further shows that even side chain conformations of the residues in this motif are very well conserved. The structural similarity holds true even for the AL**Y**D(Y/F) motif in proteins without experimentally verified phosphorylation in SH3 domains.

The RMSD values of the AL**Y**D(Y/F) motif did not correspond to sequence identity of aligned structures. Even the structures with a rather low sequence identity to c-Src had better RMSD values to c-Src than closely related proteins from Src kinase family. The best structural match of AL**Y**D(Y/F) motif of human c-Src was found in SH3 domain of myosin IB from *Acantamoeba castellanii* (2DRM, 32 % seq. identity, 0.18 Å RMSD).

Acanthameboa castellanii belongs to Amebozoa, sister group to Opisthokonta (fungi and animals). Interestingly, according to a gene discovery study, *Acantamoeba castellanii* does contain basic elements of phosphotyrosine signaling pathway, including animal tyrosine kinase

families, tyrosine phosphatases and proteins with SH2 domains [11].

A very good structural match was also found between AL**Y**D(Y/F) motif of human c-Src and those of two SH3 domains from *Saccharomyces cerevisie.* AL**Y**D(Y/F) motif occurs in four out of 29 *S. cerevisie* SH3 domains. However, animal tyrosine kinases have not been detected in yeast [15]. Nevertheless, this does not mean that phosphorylation on tyrosines does not occur in yeast. For example, kinase Swe1 inhibits the activity of Cdc28 by phosphorylation of its Tyr 19 [23].

The strong structural conservation of the AL**Y**D(Y/F) motif in Amebozoa and Opisthokonta could indicate that this mode of regulation is not a recent invention, but appeared before Amebozoa and Opisthokonta segregated.

TYROSINE PHOSPHORYLATION IS ENRICHED IN OTHER DOCKING DOMAINS

We observed that tyrosine phosphorylations represent an unusually high proportion (68 %) of all phosphorylations in SH3 domains,. We thus wanted to find out whether a prevalence of tyrosine phosphorylations is unique to SH3 domain or could be observed in other adaptor domains. We chose SH2, PH, PDZ, WW, PTB, EH, PX for further analysis [24]. Using PhosphositePlus database we searched for phosphorylation sites within these domains separately. We used only human proteins to avoid redundancy. For each domain we counted the ratio of tyrosine-phosphorylated sites to all of phosphorylations (Table 4). The statistics showed that there are 13324 (21,4%) human phosphotyrosine sites, 11618 (18,6 %) human phosphothreonine sites and human 37410 (60%) phosphoserine sites in the PhosphoSitePlus database The tyrosine phosphorylation was overrepresented (in comparison to the database statistics) in five (out of seven) selected adaptor domains. There are three domains (SH2, WW and EH), where more than 50% of all documented phosphorylations are tyrosine

phosphorylations. However, only a very few phosphorylations of EH have been observed. This suggests that tyrosine phosphorylation could also be an important regulatory mechanism for other adaptor domains. However, in human protein evolution tyrosine loss is strongly favored, most notably in protein subsets that are not known to be tyrosine phosphorylated (Tan-25). Thus the higher proportion of tyrosines in adaptor domains is in agreement with their higher tyrosine phosphorylation. Morover, the trend for enrichment of tyrosine phosphorylation in adaptor domains is maintained even after a correction to tyrosine content (Table 4).

An other possible explanation of tyrosine phosphorylations enrichment in adaptor domains could be provided by the work of Fabian et al. which showed that while phosphorylation of serine residue had no impact on the structure of non-phosphorylated tau peptide, phosphorylation of the tyrosine results in considerable conformational changes [25].

In this study, we showed that tyrosine phosphorylation has been detected in a number of SH3 domains. The most phosphorylations have been detected at the position in the SH3 domain that is responsible for substrate binding. The experimental evidence shows that this tyrosine phosphorylation interferes with binding of SH3 domain to its interacting partners. We also showed that tyrosine phosphorylations occur frequently in other adaptor domains and could therefore represent an important regulatory mechanism of these domains.

MATERIALS AND METHODS

Phosphorylation search and evaluation

All tyrosine phosphorylation sites in SH3 domains were identified in the PhosphoSite Plus database [26], curated and currently one of the most comprehensive databases of posttranslational modifications. For each hit from the PhosphoSite Plus, the occurrence of phosphorylation site in SH3 domain was carefully validated using the SMART (Simple Modular Architecture Research Tool) domain identification program [27]. Hits from PhosphoSite Plus that were not part of SMART-defined SH3 domains were not included in subsequent analyses. In case of doubt, the Uniprot annotation team was contacted for consultation, which led once to update of domain definition of a particular entry in the Uniprot database [28] .

PhosphoSitePlus was also used to find tyrosine phosphorylation sites in other adaptor proteins. To avoid redundancy, we used only human proteins to calculate the ratio of tyrosine to all phosphorylations.

The occurrence of serines, threonines and tyrosines was calculated for the set of all human proteins as defined by Uniprot and compared to occurrence of these amino acids in the sets of all human SH3, SH2, PH, PDZ, PTB, EH,PX, WW domains as defined by Pfam [19].

Normalized relative phosphotyrosine enrichment was calculated as the ratio of tyrosine phosphorylations to the number of tyrosines in adaptor domains to the ratio of tyrosine phosphorylations to the number of tyrosines for all human proteins.

Motif definition and Motif searches

The SH3 domains with identified tyrosine phosphorylations were aligned using ClustalW [29]. The alignment was further used to describe sequence motifs around two most frequently

phosphorylated positions using WebLogo [18].

The AL**Y**D(Y/F) motif, identified around most frequently phosphorylated position 7, was used to estimate abundance of tyrosine phosphorylation in SH3 domains of human proteome.

Simple text search was used to locate AL**Y**D(Y/F) motif in all SH3 domains in the SMART and Pfam databases [27]; [19]. The Clustal W [29] was used to align sequences with identified AL**Y**D(Y/F) motif and sequences with AL**Y**D(Y/F) motif around Tyr 7 were selected.

GPS 2.1 (Group-based Prediction System) Online service [20] and PhosphoMotifFinder [21] were used to identify kinases that could phosphorylate tyrosines in ALYD(Y/F) motif.

Structural analysis

The PDB [30] was used to find all AL**Y**D(Y/F) motifs in SH3 domains with known 3D structures. One representative structure for each SH3 domain with more than one experimentally solved structure was selected. All selected 3D structures of SH3 domains with AL**Y**D(Y/F) motif were aligned to a reference SH3 structure (1FMK; [5], a high-resolution structure of human Src protein, using LSQMAN program [31]. The RMSD (Root mean square distance) for the whole SH3 domain and the described motifs were calculated and compared. ClustalW at the EBI webpage was used to calculate sequence identities between aligned structures [32]. PyMol was employed to visualize the results.

Supporting Information

Figure S1. β-sheet-like structure of a loop with ALYDY motif.

File S1. UNIPROT codes of proteins with SH3 domain phosphorylation.

File S2. UNIPROT codes of proteins with tyrosine phosphorylation within SH3 domain.

File S3. UNIPROT codes of proteins with ALYDY/ALYDF motif within SH3 domain. Based on SMART database.

File S4. UNIPROT codes of proteins with ALYDY motif within SH3 domain. Based on Pfam.

File S5. UNIPROT codes of proteins with ALYDF motif within SH3 domain. Based on Pfam.

Table S1. List of phosphorylated tyrosines that occured in SH3 domains of two orthologues

Table S2. List of phosphorylated tyrosines that occured in SH3 domains of three orthologues

Supporting Information Legends

Figure S1

The AL**Y**DY motif is located in the loop that connects first and second β-strand in Src SH3 domain (1FMK). The loop conformation is stabilized by three hydrogen bonds in-between loop residues and by a hydrogen bond between Leu 6 and Tyr 71 (orange).

Acknowledgements

We thank Prof. S.K. Hanks for valuable comments. We thank Peter Hornbeck, director of PhosphoSite Plus, for providing us recent statistics about deposited data. We thank Njainday P. Jobe for proofreading of the manuscript. We thank reviewers for useful comments.

References

1. Mayer BJ (2001) SH3 domains: complexity in moderation. J Cell Sci 114: 1253-1263.

2. Yu H, Chen JK, Feng S, Dalgarno DC, Brauer AW, et al. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains. Cell 76: 933-945.

3. Kaneko T, Li L, Li SS (2008) The SH3 domain--a family of versatile peptide- and protein-recognition module. Front Biosci 13: 4938-4952.

4. Erpel T, Superti-Furga G, Courtneidge SA (1995) Mutational analysis of the Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. EMBO J 14: 963-975.

5. Xu W, Harrison SC, Eck MJ (1997) Three-dimensional structure of the tyrosine kinase c-Src. Nature 385: 595-602.

6. Hunter T (2009) Tyrosine phosphorylation: thirty years and counting. Curr Opin Cell Biol 21: 140-146.

7. Rodrigues GA, Park M (1994) Oncogenic activation of tyrosine kinases. Curr Opin Genet Dev 4: 15-24.

8. King N, Hittinger CT, Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301: 361-363.

9. Lim WA, Pawson T (2010) Phosphotyrosine signaling: evolving a new cellular communication system. Cell 142: 661-667.

10. Manning G, Young SL, Miller WT, Zhai Y (2008) The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. Proc Natl Acad Sci U S A 105: 9674-9679.

11. Anderson IJ, Watkins RF, Samuelson J, Spencer DF, Majoros WH, et al. (2005) Gene

discovery in the Acanthamoeba castellanii genome. Protist 156: 203-214.

12. Park H, Wahl MI, Afar DE, Turck CW, Rawlings DJ, et al. (1996) Regulation of Btk function by a major autophosphorylation site within the SH3 domain. Immunity 4: 515-525.

13. Meyn MA, Wilson MB, Abdi FA, Fahey N, Schiavone AP, et al. (2006) Src family kinases phosphorylate the Bcr-Abl SH3-SH2 region and modulate Bcr-Abl transforming activity. J Biol Chem 281: 30907-30916.

14. Janostiak R, Tolde O, Bruhová Z, Novotny M, Hanks SK, et al. (2011) Tyrosine phosphorylation within the SH3 domain regulates CAS subcellular localization, cell migration, and invasiveness. Mol Biol Cell 22: 4256-4267.

15. Hunter T, Plowman GD (1997) The protein kinases of budding yeast: six score and more. Trends Biochem Sci 22: 18-22.

16. Chen S, O'Reilly LP, Smithgall TE, Engen JR (2008) Tyrosine phosphorylation in the SH3 domain disrupts negative regulatory interactions within the c-Abl kinase core. J Mol Biol 383: 414-423.

17. Agrawal V, Kishan KV (2002) Promiscuous binding nature of SH3 domains to their target proteins. Protein Pept Lett 9: 185-193.

18. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188-1190.

19. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res 38: D211-222.

20. Xue Y, Ren J, Gao X, Jin C, Wen L, et al. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics 7: 1598-1608.

21. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, et al. (2009) Human Protein Reference Database--2009 update. Nucleic Acids Res 37: D767-772.

22. Joseph RE, Fulton DB, Andreotti AH (2007) Mechanism and functional significance of Itk autophosphorylation. J Mol Biol 373: 1281-1292.

23. Booher RN, Deshaies RJ, Kirschner MW (1993) Properties of Saccharomyces cerevisiae wee1 and its differential regulation of p34CDC28 in response to G1 and G2 cyclins. EMBO J 12: 3417-3426.

24. Cesareni G, Gimona M, Sudol M, Yaffe M (2005) Modular Protein Domains. Wiley-VCH. 524p.

25. Fabian H, Otvos L, Szendrei GI, Lang E, Mantsch HH (1994) Tyrosine- versus serine-phosphorylation leads to conformational changes in a synthetic tau peptide. J Biomol Struct Dyn 12: 573-579.

26. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics 4: 1551-1561.

27. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95: 5857-5864.

28. Consortium U (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40: D71-75.

29. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

30. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, et al. (2011) The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res 39: D392-401.

31. Kleywegt GJ (1996) Use of non-crystallographic symmetry in protein structure refinement.

Acta Crystallogr D Biol Crystallogr 52: 842-857.

32. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

33. Sato M, Maruoka M, Yokota N, Kuwano M, Matsui A, et al. (2011) Identification and functional analysis of a new phosphorylation site (Y398) in the SH3 domain of Abi-1. FEBS Lett 585: 834-840.

34. Morrogh LM, Hinshelwood S, Costello P, Cory GO, Kinnon C (1999) The SH3 domain of Bruton's tyrosine kinase displays altered ligand binding properties when auto-phosphorylated in vitro. Eur J Immunol 29: 2269-2279.

35. Sriram G, Reichman C, Tunceroglu A, Kaushal N, Saleh T, et al. (2011) Phosphorylation of Crk on tyrosine 251 in the RT loop of the SH3C domain promotes Abl kinase transactivation. Oncogene 30: 4645-4655.

36. Li S, Couvillon AD, Brasher BB, Van Etten RA (2001) Tyrosine phosphorylation of Grb2 by Bcr/Abl and epidermal growth factor receptor: a novel regulatory mechanism for tyrosine kinase signaling. EMBO J 20: 6793-6804.

37. Wilcox HM, Berg LJ (2003) Itk phosphorylation sites are required for functional activity in primary T cells. J Biol Chem 278: 37112-37121.

38. Wu Y, Spencer SD, Lasky LA (1998) Tyrosine phosphorylation regulates the SH3-mediated binding of the Wiskott-Aldrich syndrome protein to PSTPIP, a cytoskeletal-associated protein. J Biol Chem 273: 5765-5770.

39. Wu X, Gan B, Yoo Y, Guan JL (2005) FAK-mediated src phosphorylation of endophilin A2 inhibits endocytosis of MT1-MMP and promotes ECM degradation. Dev Cell 9: 185-196.

40. Sylvester M, Kliche S, Lange S, Geithner S, Klemm C, et al. (2010) Adhesion and degranulation promoting adapter protein (ADAP) is a central hub for phosphotyrosine-

mediated interactions in T cells. PLoS One 5: e11708.

41. Fernow I, Tomasovic A, Siehoff-Icking A, Tikkanen R (2009) Cbl-associated protein is tyrosine phosphorylated by c-Abl and c-Src kinases. BMC Cell Biol 10: 80.

42. Broome MA, Hunter T (1996) Requirement for c-Src catalytic activity and the SH3 domain in platelet-derived growth factor BB and epidermal growth factor mitogenic signaling. J Biol Chem 271: 16798-16806.

43. Kashiwakura J, Suzuki N, Takeno M, Itoh S, Oku T, et al. (2002) Evidence of autophosphorylation in Txk: Y91 is an autophosphorylation site. Biol Pharm Bull 25: 718-721.

44. Lazer G, Pe'er L, Farago M, Machida K, Mayer BJ, et al. (2010) Tyrosine residues at the carboxyl terminus of Vav1 play an important role in regulation of its biological activity. J Biol Chem 285: 23075-23085.

45. Luo W, Slebos RJ, Hill S, Li M, Brábek J, et al. (2008) Global impact of oncogenic Src on a phosphotyrosine proteome. J Proteome Res 7: 3447-3460.

Figure legends:

Figure 1. Multiple sequence alignment of tyrosine-phosphorylated SH3 domains. Phosphosite Plus database was searched for tyrosine phosphorylation within SH3 domains. The identified SH3 domains were aligned using ClustalW. Human sequences are shown except those depicted with _**m**, which come form mouse. All the sequences were obtained using SMART server. Names of proteins according to UniProt database are situated on the left including the domain range if there are more then one SH3 domains within a protein. Alignment is numbered at the top. Phosphorylated tyrosines are highlighted in red. Orthologous and paralogous sequences with identical phosphosites are not included.

Figure 2. Weblogo of aligned segments of SH3 domains phosphorylated at Y7 and Y66 sites. Sequence logos were created using WebLogo (http://weblogo.berkeley.edu/logo.cgi) from 9aa long multiple sequence alignments (Fig. 1) with Y7 (A) and Y66 (B) in central position. The numbers of sequences for each WebLogo are indicated in upper left corner.

Figure 3. Structural alignment of ALYD(Y/F) motifs.

Structures of human Src (green), human Abl (orange), yeast Pex13 (cyan) and Acanthamoeba Myosin Ib (grey) SH3 domains were aligned using LSQMAN. The whole SH3 domains were aligned. For the sake of clarity only the AL**Y**D(Y/F) motif is shown. The first amino acid of the motif (Ala) is hidden behind the plane of the figure. The phosphorylation on Tyr 7 (Y7) in the AL**Y**D(Y/F) motif was documented in mouse Src [45] and in human Abl [13]. The figure was created using PyMol.

Table 1. Summary of effects caused by mutation or phosphorylation at tyrosine sites in SH3 domains

Protein	non-phospho-rylatable mutation	phospho-mimicking mutation	phospho-tyrosine	corresponding position in the alignment	Effect of mutation/phosphorylation	Ref.
Abi-1			Y398p	Y7	reduces binding to Abl	[33]
Abl	Y89F			Y7	decreases Bcr-Abl-mediated transformation of TF-1 myeloid cells to cytokine independence	[13]
			Y89p	Y7	decreases interaction of SH3 domain with binding partners both in *cis* and in *trans*	[16]
Btk	Y223F			Y7	blocks Btk autophosphorylation and potentiates the transforming activity of Btk in fibroblasts	[12]
			Y223p	Y7	disrupts the interaction with WASP	[34]
Crk			Y251p	Y17	induces Abl kinase transactivation	[35]
Grb2			Y209p	Y71	reduces binding to Sos	[36]
Itk	Y180F			Y7	plays positive role in Itk signaling	[37]
p130CAS	Y12F			Y7	decreases invasiveness in Src-transformed cells	[14]
		Y12E		Y7	decreases interaction of SH3 domain with FAK and PTP-PEST	[14]
			Y12p	Y7	decreases interaction of SH3 domain with FAK	[14]
PST-PIP		Y367E		Y7	decreases interaction with WASP	[38]
Endophilin		Y315E		Y7	decreases interaction of SH3 domain with Dynamin	[39]
ADAP			Y559p	Y66	positively affects interaction with Nck protein	[40]
CAP	Y623F			Y7	results in partial nuclear localization of CAP protein	[41]
Src	Y90A, Y92A			Y7, Y9	disrupts the interaction with Sam68 and PI3K-p85α	[4]
	Y133F,Y138F			Y66,Y71	inhibit PDGF and EGF mitogenic signaling	[42]
Txk			Y91p	Y7	contributes to upregulated IFN-g gene transcription	[43]
Vav1	Y826F			Y55	reduces binding to CSK	[44]

Table 2. Position-based phosphotyrosine abundance within SH3 domain.

Phospho-Y position within SH3	2	3	**7**	9	11	17	30	31	35	42	49	50	51	54	55	56	59	64	**66**	71
Number of phoshorylations detected per site	2	2	**24**	6	1	6	2	1	2	1	1	1	1	1	3	1	1	1	**11**	1

Phosphotyrosine position within SH3 refers to the position in alignment in Figure 1.

Table 3. Structural alignment of all SH3 domains with AL**Y**D(Y/F) motif with known 3D structure.

Structure	Protein	Organism	Motiv	Ex. Method	Seq. Identity	Average RMSD [Å]	RMSD (5-9) [Å]
1FMK	Src	human	AL**Y**DY	x-ray	100	0	0
2hda	Yes	human	AL**Y**DY	x-ray	76	1.02	0.39
3cqt	Fyn	chicken	AL**Y**DY	x-ray	71	1.04	0.22
4hck	Hck	human	AL**Y**DY	NMR	51	1.29	0.35
1aww	Btk	human	AL**Y**DY	NMR	42	1.45	0.81
1bbz	Abl	human	AL**Y**DF	x-ray	37	1.1	0.47
2d0n	Grb2 - related protein	mouse	AL**Y**DF	x-ray	33	1.04	0.4
1u06	Spectrin	chicken	AL**Y**DY	x-ray	33	1.1	0.37
2drm	Myosin	*A. castellanii*	AL**Y**DY	x-ray	32	1	0.18
2ed0	Abl2	human	AL**Y**DY	NMR	32	1.17	0.72
2v1r	Peroxin-13	*S. cerevisiae*	AL**Y**DF	x-ray	30	1.09	0.36
1x2q	Stam2	human	AL**Y**DF	NMR	30	1.19	0.38
3i5s	Pi3K	human	AL**Y**DY	x-ray	30	1	0.45
1yn8	Nap-1 binding protein	*S. cerevisiae*	AL**Y**DF	x-ray	28	1.07	0.46
1x69	Cortactin	human	AL**Y**DY	NMR	28	1.18	0.43
2yuq	Itk	human	AL**Y**DY	NMR	28	1.31	0.55

Structures were aligned to SH3 domain of human Src protein (1FMK). Two parameters are measured for each structural alignment – root mean square distance of the whole SH3 domains (Average RMSD) and of the AL**Y**D(Y/F) motif (RMSD (5-9)). Sequence identity to human 1FMK SH3 domain was calculated using ClustalW.

Table 4. Ratios of phosphotyrosine to all phosphorylations within selected adaptor domains

Domain	Number of pY to all phosphorylations	Ratio of pY to all phosphorylations [%]	Number of tyrosines (%)	Normalized relative pY enrichment
Human proteome	13324/62352	21,4	2,6	1,00
SH3	59/87	67,8 %	3,8	2,17
SH2	72/104	69,2 %	5,3	1,59
PH	75/159	47,2 %	3,7	1,55
PDZ	13/73	17,8 %	1,2	1,80
WW	11/14	78,6 %	7,2	1,33
PTB	0/1	-	1,5	
EH	2/2	100%	2,0	6,08
PX	18/37	48,6 %	3,7	1,60

The number of phosphorylations within selected domains was analyzed by PhosphoSite Plus and the ratio of phosphotyrosine sites (pY) to all sites was calculated. The complete human proteome from the Uniprot database was chosen to calculate the number of tyrosines among human proteins. The complete sets of human proteins containing adaptor domains was selected using the Pfam database.

Normalized relative enrichment of tyrosine phosphorylation shows ratio of percentage of pY to percentage of pY in human proteome normalized to number of tyrosines.

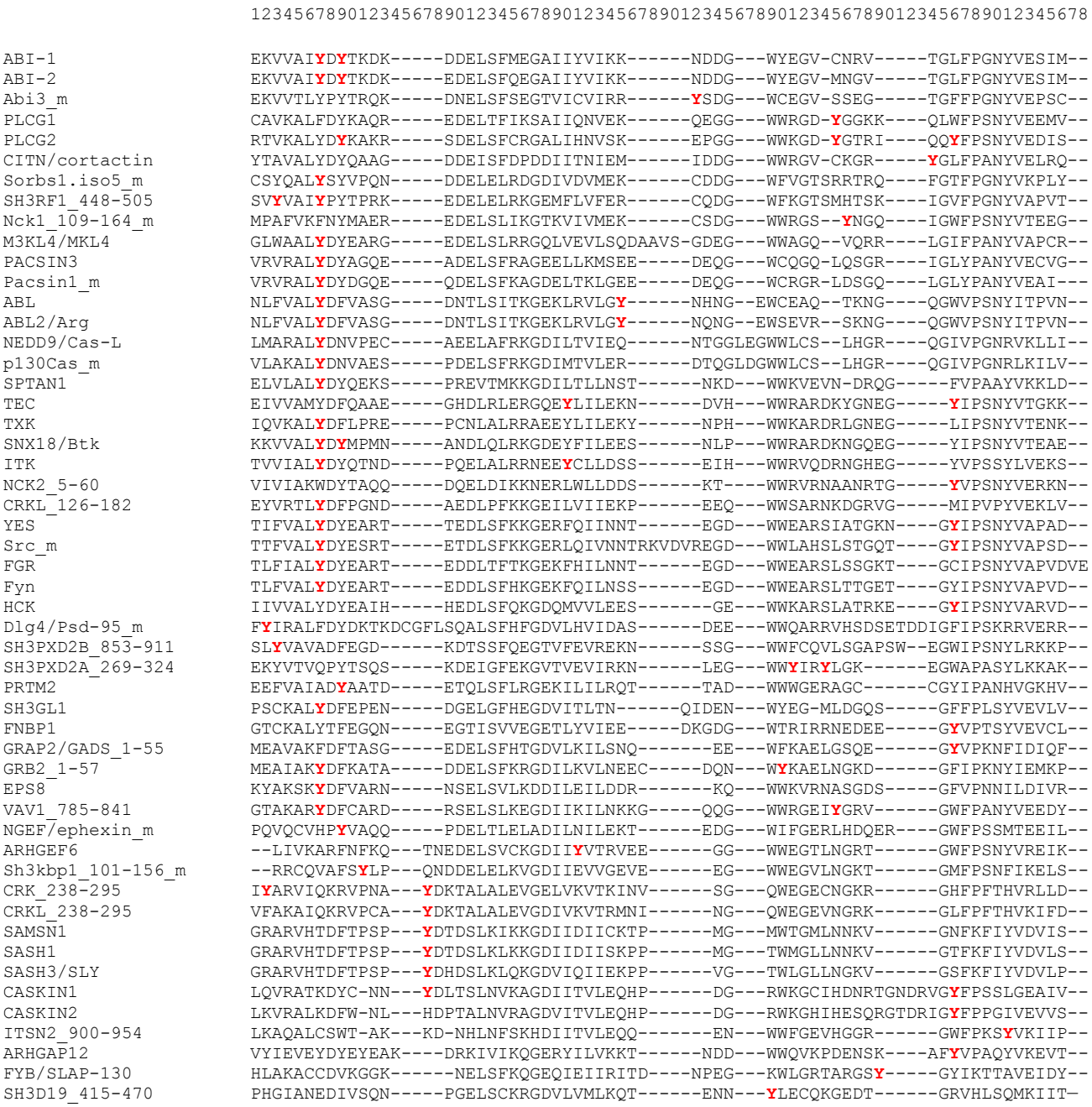
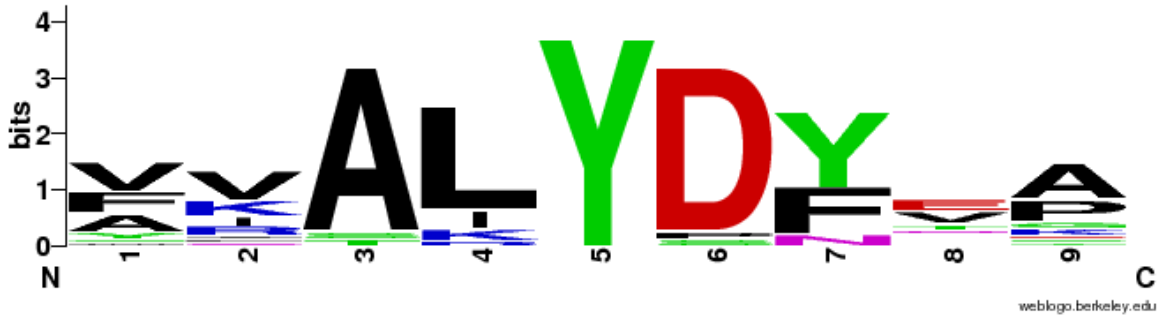


Figure 1. Multiple sequence alignment of tyrosine-phosphorylated SH3 domains.

(A) 24



(B) 11

Figure 2. Weblogo of aligned segments of SH3 domains phosphorylated at Y7 and Y66 sites.

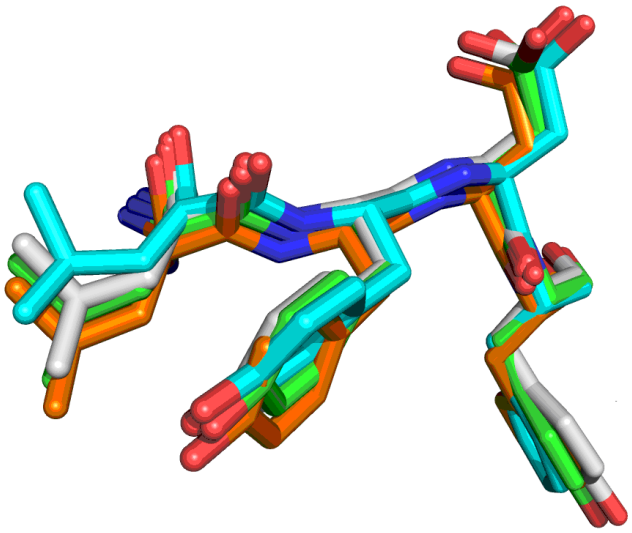


Figure 3. Structural alignment of ALYD(Y/F) motifs.

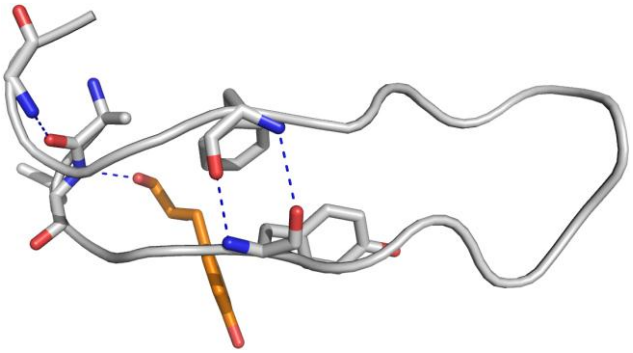


Figure S1. β-sheet-like structure of a loop with ALYDY motif.