Supporting Information

Aitio et al. 10.1073/pnas.1010243107

SI Materials and Methods

Cloning, Expression, and Purification of IRTKS SH3 Domain and $EspF_U$ $R47_5$. The gene encoding the SH3 domain (residues 339–402) of human IRTKS protein (UniProt - Swiss-Prof Protein database ID Q9UHR4) was cloned to pET15b vector (Novagen) into the NdeI and XhoI sites. Production of ^{15}N and ^{13}C, ^{15}N-labeled IRTKS SH3 was carried out by transforming plasmids into BL21(DE3) cells. Cells were grown in M9 minimal media, supplemented with 1 g/L $^{15}NH_4Cl$ and 2 g/L ^{13}C-D-glucose as the sole nitrogen and carbon sources, respectively. Protein production was induced with 1 mM isopropyl-β-D-thiogalactopyranoside when the OD_{600} of the cell culture reached 0.6. Cells were further incubated at 37 °C for 5 h and collected by centrifugation. Unlabeled protein was produced similarly, except LB was used as a culture medium and incubation time after induction was 3 h. For purification of IRTKS SH3, cells were disrupted with sonication and the resulting supernatant was clarified by centrifugation at 30,000 × *g*. Clarified supernatant was applied to the 7-mL Ni-NTA column (GE Healthcare) according to the manufacturer's instructions. Eluted proteins were extensively dialyzed against PBS. His-Tag from IRTKS SH3 was removed by thrombin digestion. Digestion mixture was applied to the Ni-NTA column. The flow-trough, containing cleaved IRTKS SH3, was concentrated by means of a Vivaspin2 (SartoriusStedim) concentrator, and samples were separately applied into the Superdex30 (16/60) gel filtration column (GE Healthcare). Buffer used in gel filtration contained 20 mM sodium phosphate (pH 7) and 50 mM NaCl (NMR buffer). All chromatographic purifications were performed by using the ÄKTA Purifier FLPC purification system (GE Healthcare). The purity of the both proteins was analyzed by SDS/PAGE and MALDI-TOF mass spectra, which showed that neither protein contained degradation products or other protein impurities.

The cloning, expression, and purification procedures of the gene-encoding residues 268–314 of *E. coli* O157:H7 $EspF_U$-like protein [the fifth repeat (UniProt Q8 × 2D5), called $R47_5$] will be described elsewhere.

The synthetic peptides EHIPPAPN and PAPTPPVQ corresponding to residues 25–32 and 34–41 of $EspF_U$ $R47_5$ were obtained from GenScript USA, Inc.

NMR Spectroscopy. All NMR experiments were carried out at 25 °C on a Varian Unity INOVA 800-MHz spectrometer or a Varian Unity INOVA 600-MHz spectrometer, equipped with a $^{15}N/^{13}C/^{1}H$ triple-resonance cold probe and an actively shielded z-axis gradient system. A set of well-established triple-resonance experiments [e.g., 3D iHNCACB, CBCA(CO)NH, HBHA(CO)NH, CC(CO)NH, HCC(CO)NH, HCCH-COSY] was used for the assignment of backbone and aliphatic side chain resonances of $^{15}N,^{13}C$-labeled IRTKS SH3 (1, 2). Aromatic resonances were assigned using (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, and aromatic HCCH-COSY experiments (1). For the assignment of ^{15}N, ^{13}C-labeled $EspF_U$ $R47_5$, a similar set of triple-resonance experiments as in the case of IRTKS SH3 was used, but the assignment of main chain resonances was supplemented with a set of HA-detected experiments (3). The SH3 domain interproton distance restraints were determined from 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra (1). A mixing time of 100 ms was used in all NOESY experiments.

Initial characterization of IRTKS SH3:$EspF_U$ $R47_5$ interaction was carried out by adding unlabeled $R47_5$ into a sample containing ^{15}N, ^{13}C-labeled SH3. The titration series was monitored by detecting the amide correlations using the ^{15}N-HSQC experiment (4). IRTKS SH3:$EspF_U$ $R47_5$ molar ratios at each data point were 1:0, 1:0.25, 1:0.75, and 1:1.

All spectra were processed using the VNMRJ 2.1 revision B software package (Varian Associates) and analyzed with Sparky 3.110 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Structure Calculations. For the structure calculation of IRTKS SH3:$EspF_U$ R47 complex, the peptide sequence was connected to the C terminus of the SH3 domain sequence through a set of weightless noninteracting dummy atoms. Peaks were picked manually from ^{15}N- and ^{13}C-edited NOESY spectra. The peak lists, together with the chemical shift assignments, were used as input for the iterative NOE assignment and structure calculations. We generated 200 conformers in each of the seven cycles of the combined automated NOESY and structure calculation algorithm. The final 20 structures were energy-minimized, using CYANA-derived NOE restraints, in AMBER 8 (5). One thousand iterations with the standard AMBER force field and generalized Born implicit solvent model were performed. Quality of structure was analyzed with PROCHECK-NMR (6), indicating that 90.2% and 9.8% of the residues are in the most favored and additionally allowed regions, respectively, of the Ramachandran plot (Table S1).

Peptide Array. Peptides of interest were synthesized by the Peptide and Protein Laboratory (Haartman Institute, University of Helsinki) as peptide-cellulose conjugates using Multipep (INTAVIS Bioanalytical Instruments) according to the manufacturer's protocol for SPOT synthesis (7) and printed in parallel arrays on glass slides using SlideSpotter equipment (INTAVIS Bioanalytical Instruments). The array was immersed for 2 h in blocking solution [PBST (137 mM NaCl, 2.7 mM KCl, 100 mM Na_2HPO_4, 2 mM KH_2PO_4, 0.05% Tween-20), supplemented with 5% (wt/vol) nonfat dry milk] and prewashed three times in PBST. Biotinylated IRTKS SH3 protein used for probing the arrayed peptides was expressed as GST fusion protein, including a 123-residue biotinylation acceptor domain from *Propionibacterium shermanii* transcarboxylase between the GST and SH3 moieties as described previously (8). Biotinylated IRTKS SH3 was diluted with PBST supplemented with 1.5% (wt/vol) BSA to a final concentration of 1 μM and incubated with the array for 2 h at room temperature, followed by three 5-min washes with PBST and one with PBS. The binding was detected with infrared fluorescent IRDye 680-conjugated streptavidin (LI-COR Biosciences) and quantified with an Odyssey Infrared Imaging System (LI-COR Biosciences).

1. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34:93–158.
2. Tossavainen H, Permi P (2004) Optimized pathway selection in intraresidual triple-resonance experiments. *J Magn Reson* 170:244–251.
3. Mäntylahti S, Aitio O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR* 47: 171–181.
4. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663–10665.
5. Case DA, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
6. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486.

7. Frank R (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—Principles and applications. *J Immunol Methods* 267:13–26.
8. Heikkinen LS, et al. (2008) Avian and 1918 Spanish influenza a virus NS1 proteins bind to Crk/CrkL Src homology 3 domains to activate host cell signaling. *J Biol Chem* 283: 5719–5727.

Type III secretion signal sequence	1	MINNVSSLFPTVNRNITAVYKKSSFSVSPQKITLNPVKISSPFSPSS	47
	48	SSISATTLFRAPNAHSASFHRQSTAESSLHQQ	79
Repeat 1 ($R47_1$)	80	LPNVRQRLIQHLAEHGIKPARSMAEHIPPAPNWPAPPPPVQNEQSRP	126
Repeat 2 ($R47_2$)	127	LPDVAQRLVQHLAEHGIQPARNMAEHIPPAPNWPAPPLPVQNEQSRP	173
Repeat 3 ($R47_3$)	174	LPDVAQRLVQHLAEHGIQPARSMAEHIPPAPNWPAPPPPVQNEQSRP	220
Repeat 4 ($R47_4$)	221	LPDVAQRLMQHLAEHGIQPARNMAEHIPPAPNWPAPTPPVQNEQSRP	267
Repeat 5 ($R47_5$)	268	LPDVAQRLMQHLAEHGIQPARNMAEHIPPAPNWPAPTPPVQNEQSRP	314
Repeat 6 ($R18_6$)	315	LPDVAQRLMQHLAEHGINTSKRS	337

Fig. S1. Amino acid sequence of $EspF_U$. The first 79 residues correspond to the type III secretion signal sequence, which is followed by five almost identical 47-residue repeats referred to as $R47_x$, where x stands for the corresponding repeat number. The sixth repeat ($R18_6$) contains only 18 conserved N-terminal residues of repeats 1–5, followed by 6 additional residues (332–337). The fifth repeat, $R47_5$, used in this study is highlighted in yellow. The sequence identity of the repeats is high, and only a few (<6) amino acid differences are found among them (1, 2).

1. Sallee NA, et al. (2008) The pathogen protein EspF(U) hijacks actin polymerization using mimicry and multivalency. *Nature* 454:1005–1008.
2. Cheng HC, Skehan BM, Campellone KG, Leong JM, Rosen MK (2008) Structural mechanism of WASP activation by the enterohaemorrhagic E. coli effector EspF(U). *Nature* 454: 1009–1013.

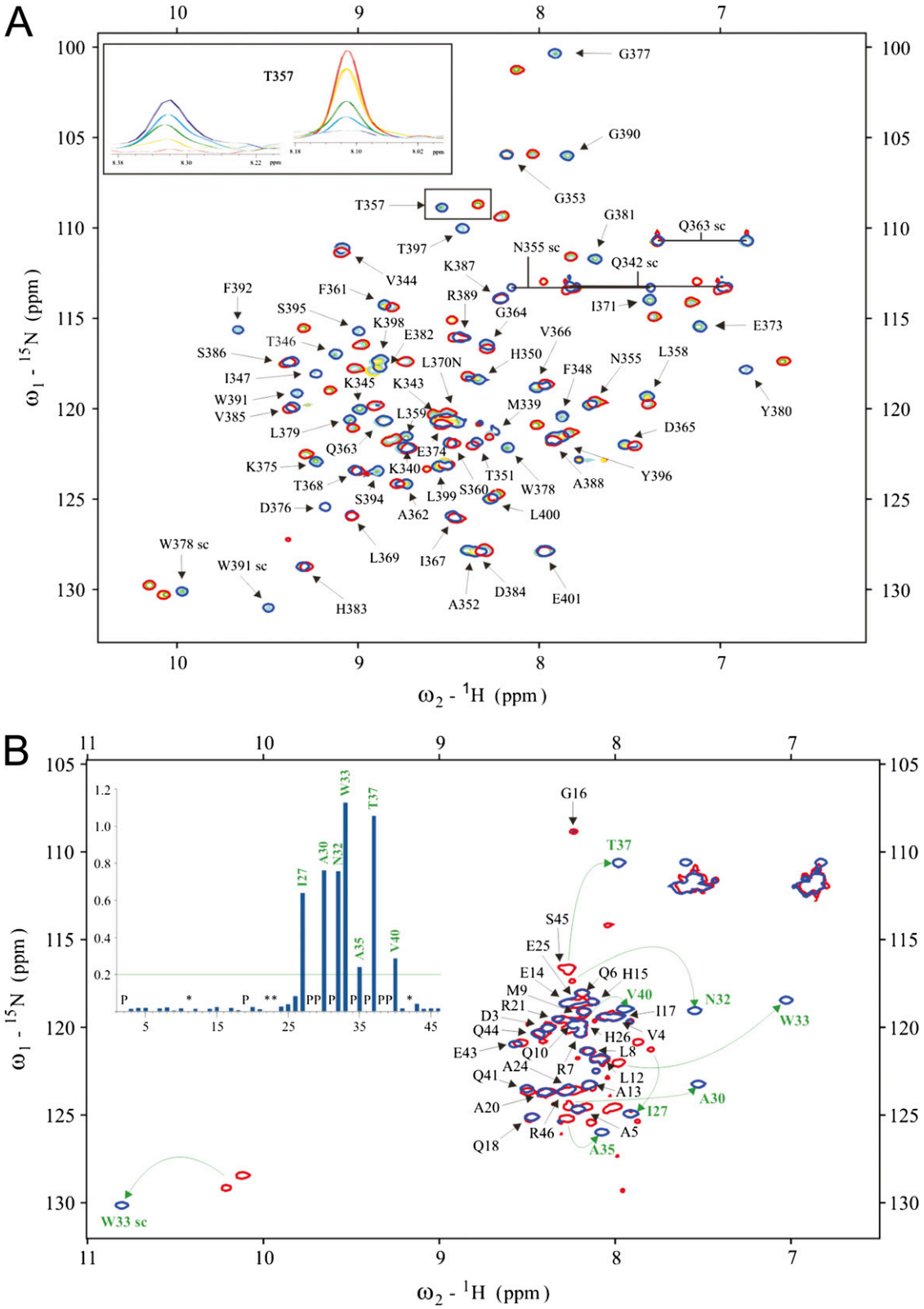

Fig. S2. (*A*) ^{15}N-HSQC spectra showing the chemical shift perturbations of the ^{1}H-^{15}N correlations measured for ^{15}N-labeled IRTKS SH3 on addition of unlabeled R47$_5$. The assignment of resonances corresponds to the bound state with IRTKS SH3:EspF$_U$ R47$_5$, with a 1:1 molar ratio. Color coding for IRTKS SH3: EspF$_U$ R47$_5$ molar ratios: 1:0 (red), 1:0.25 (orange), 1:0.5 (green), 1:0.75 (cyan), and 1:1 (blue). (*Inset*) Overlaid 1D traces along the ^{1}H axis of T357 ^{1}H-^{15}N correlation during titration. Resonances corresponding to free and bound states coexist and do not shift on addition of unlabeled EspF$_U$ R47$_5$, indicating tight binding (i.e., slow exchange in the NMR time scale). (*B*) Comparison of ^{15}N-HSQC spectra of free and bound ^{15}N-labeled EspF$_U$ R47$_5$. The free form of ^{15}N-labeled EspF$_U$ R47$_5$ is shown in red, and ^{15}N-labeled EspF$_U$ R47$_5$ with 1:1 molar ratio of unlabeled IRTKS SH3 is shown in blue. Assignment corresponds to the bound state of EspF$_U$ R47$_5$. Resonances exhibiting the largest changes in chemical shift are shown in green, and a curved arrow connects the resonances

Legend continued on following page

corresponding to the free and bound states. (*Inset*) Histogram of normalized chemical shift perturbations according to $\delta\Delta = [(\Delta H^N)^2 + (0.17\Delta N)^2]^{\frac{1}{2}}$ between the free and bound $EspF_U$ $R47_5$. Proline residues are marked with a P, and unassigned residues are marked with an asterisk. The free $EspF_U$ $R47_5$ exists in several conformations, as evidenced, for example, by two correlations corresponding to side chain NH of W33.

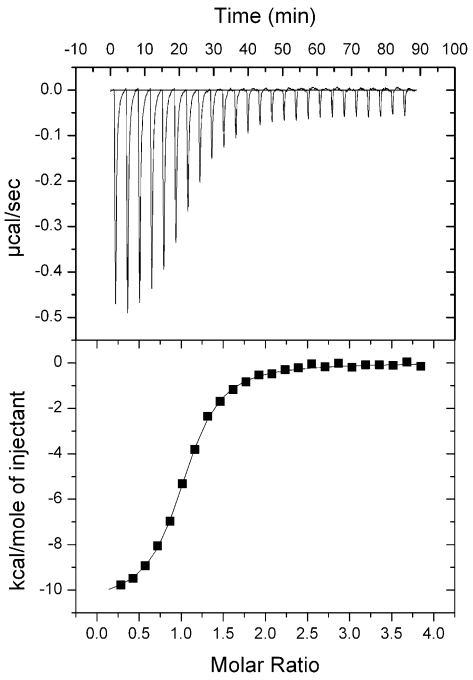

Fig. S3. Calorimetric (ITC) titration of IRTKS SH3 with $R47_5$. The thermodynamic profile of the interaction was obtained by nonlinear least-square fitting of data using a single-site binding model.

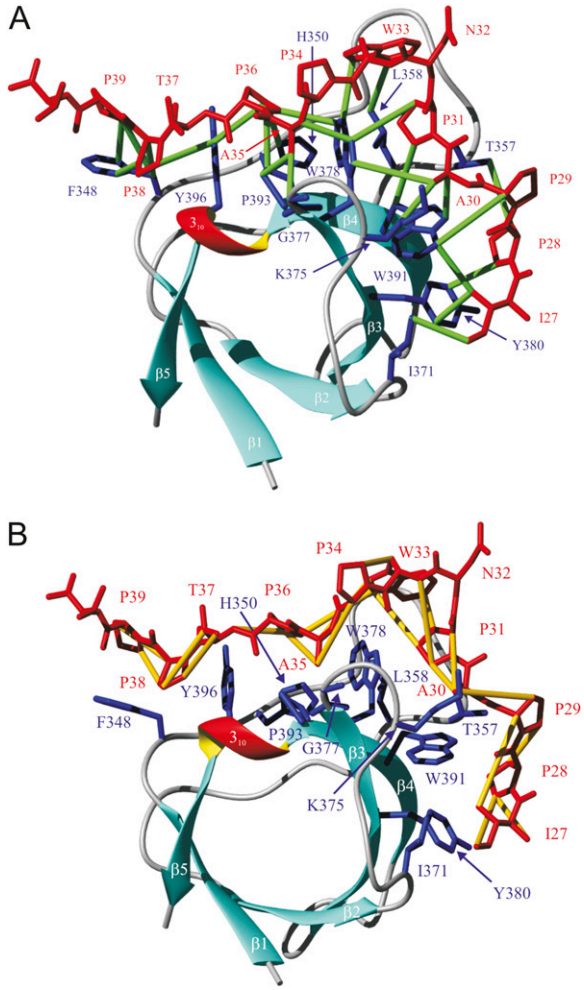

Fig. S4. Selected intermolecular (*A*; green tubes) and intra-$EspF_U$ $R47_5$ (*B*; yellow tubes) NOEs used for structure determination of the complex.

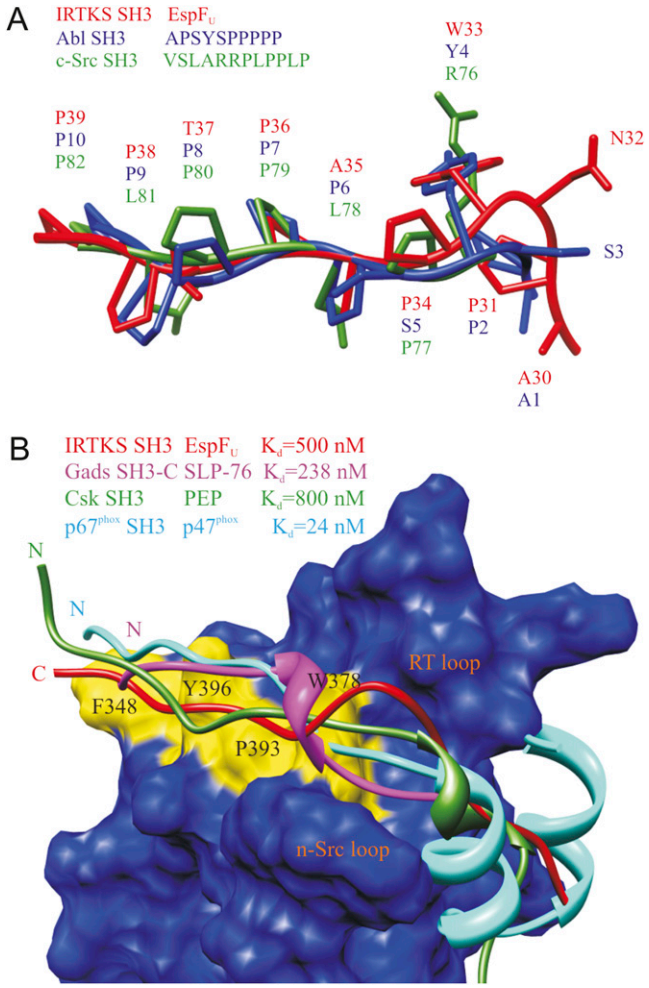

Fig. S5. (*A*) Comparison of IRTKS SH3-bound EspF$_U$ with other PPII helical structures. The C-terminal PPII helix of EspF$_U$ P$_{39}$PTPAP$_{34}$ (red) is superimposed with corresponding residues of c-Src SH3-bound VSL12 peptide VSLARRPLPPLP (green), representing the typical class I RxxPxxP motif (PDB ID code 1QWF) (1), and with Abl SH3-bound APSYSPPPPP (blue; PDB ID code 1BBZ) (2). Only the relevant residues in the peptides are shown. A comparison shows that W$_{33}$ of EspF$_U$ and Y$_8$ of Abl SH3-bound peptide occupy the same position in the specificity pocket as R$_{76}$ in VSL12 peptide and that hydrophobic residues are accommodated in the specificity pocket in similar conformation in IRTKS SH3 and Abl SH3. (*B*) Comparison of IRTKS SH3 EspF$_U$ complex with other high-affinity complexes between SH3 domains and natural ligands. SH3 domains of IRTKS bound to EspF$_U$, Gads SH3-C bound to SLP-76 peptide (PDB ID code 1OEB) (3), Csk SH3 bound to PEP peptide (PDB ID code 1JEG) (4), and p67^{phox} SH3 bound to p47^{phox} peptide (PDB ID code 1K4U) (5) are superimposed. The surface of IRTKS SH3 is shown in blue, and the yellow surface with amino acid labeling corresponds to residues forming the classic PxxP-binding site. Bound peptides are shown as a ribbon presentation, and coloring corresponds to given K$_d$ values. The location of n-Src and RT loops is also indicated. A comparison shows that SH3 peptide interactions are mediated through similar docking surfaces involving the conventional PxxP-binding site and an extended surface area in the proximity of the specificity pocket and that the interactions are stabilized through secondary structure elements on the peptide. Despite the binding to a similar area, the peptide-SH3 domain contacts differ considerably. Only EspF$_U$ has a class I binding orientation, and it makes no polar contacts with the specificity pocket. SLP-76 peptide lacks the PxxP motif and contains a nontypical PxxRxxKP instead.

1. Feng S, Kasahara C, Rickles RJ, Schreiber SL (1995) Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc Natl Acad Sci USA* 92:12408–12415.
2. Pisabarro MT, Serrano L, Wilmanns M (1998) Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: Implications for SH3-ligand interactions. *J Mol Biol* 281:513–521.
3. Harkiolaki M, et al. (2003) Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76. *EMBO J* 22:2571–2582.
4. Ghose R, Shekhtman A, Goger MJ, Ji H, Cowburn D (2001) A novel, specific interaction involving the Csk SH3 domain and its natural ligand. *Nat Struct Biol* 8:998–1004.
5. Kami K, Takeya R, Sumimoto H, Kohda D (2002) Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p. *EMBO J* 21:4268–4276.

Table S1. Structural statistics for IRTKS SH3:EspFU R475 complex

Distance constraints	
Total	1,526
Short range, |i–j|< = 1	703
Medium range, 1<|i–j|<5	154
Long range, |i–j| ≥ 5	669
No. constraints per residue	18.2
Intermolecular	145
Structure statistics	
Average AMBER energy (kcal/mol)	−2,382.9 ± 17.1
Violations	
Distance constraints violations >0.3 Å	—
Deviations from idealized geometry	
Bond lengths, Å	0.0099 ± 0.0001
Bond angles, °	2.316 ± 0.023
Average RMSD from mean coordinates, Å	
Backbone	0.32 ± 0.08*
Heavy	0.62 ± 0.08*
Ramachandran plot, %†	90.2/9.8/0/0

*RMSD values are shown for the 20 calculated complex structures (IRTKS SH3 343–400 and $EspF_U$ 27–40).
†Residues in most favored/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot.

Table S2. Main chain dihedral angles ϕ and ψ for EspFU R475 peptide in complex with IRTKS SH3

Residue	ϕ	ψ
ILE 27	−68.7	132.7
PRO 28	−69.7	158.7
PRO 29	−64.1	144.8
ALA 30	−69.7	154.8
PRO 31	−68.6	174.1
ASN 32	−86.7	2.0
TRP 33	−132.8	153.5
PRO 34	−71.4	152.9
ALA 35	−68.6	143.9
PRO 36	−73.5	−179.3
THR 37	−134.3	155.1
PRO 38	−67.6	163.1
PRO 39	−64.5	149.7