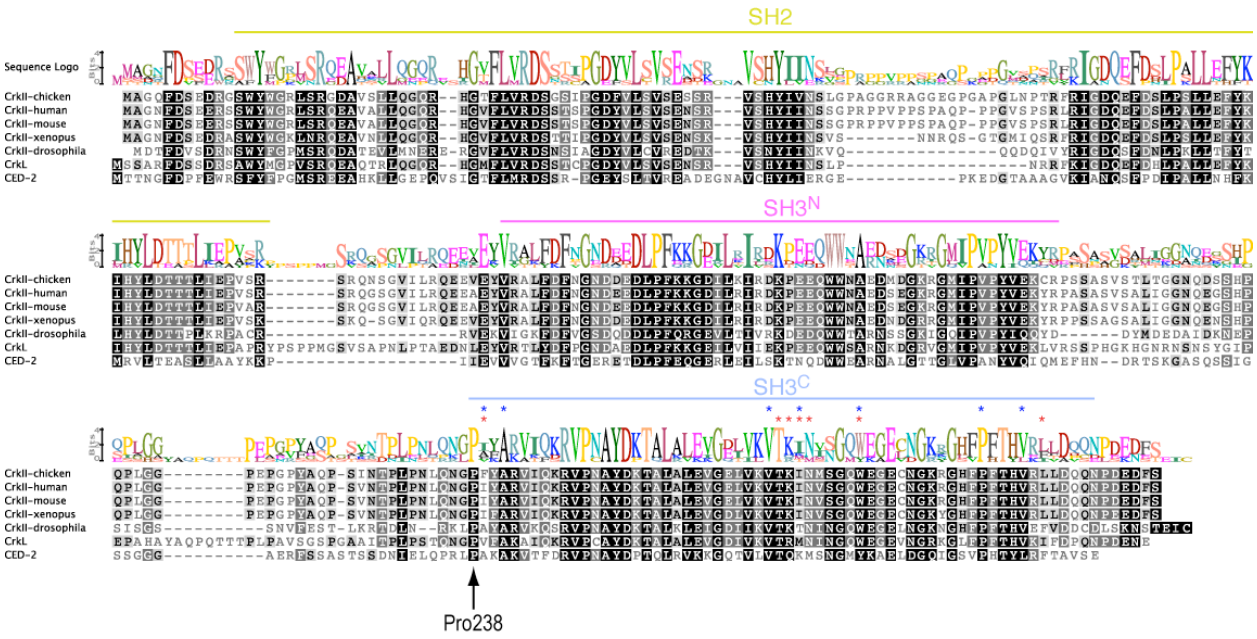
Supplementary Information

Structural basis for regulation of the Crk signaling protein by a proline switch

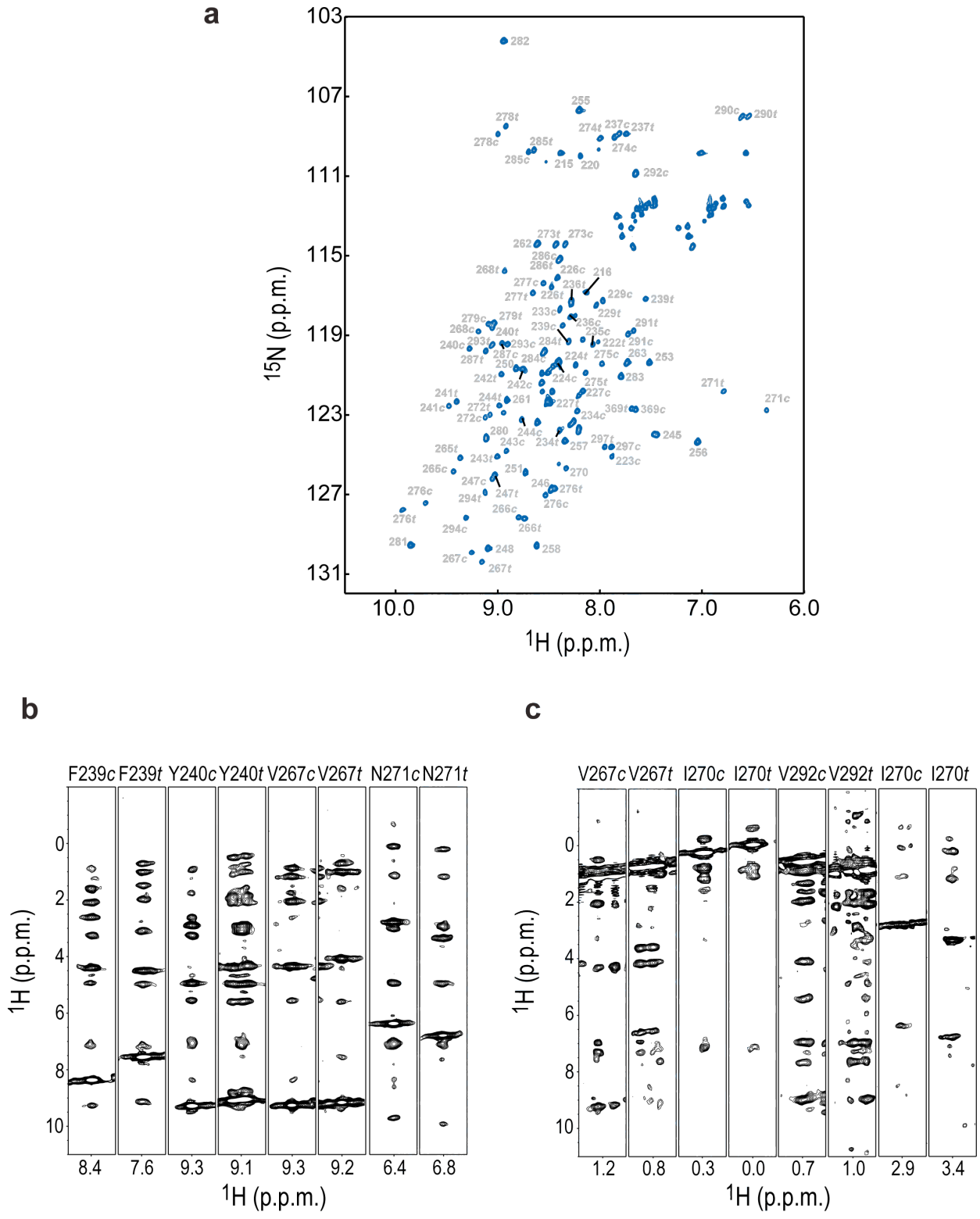
Paramita Sarkar[1], Tamjeed Saleh[1], Shiou-Ru Tzeng[1], Raymond B. Birge[2] & Charalampos G. Kalodimos[1,3,4*]

[1]Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, NJ 08854.
[2]Department of Biochemistry & Molecular Biology, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, NJ 07103. [3]BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, NJ 08854. [4]Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854.

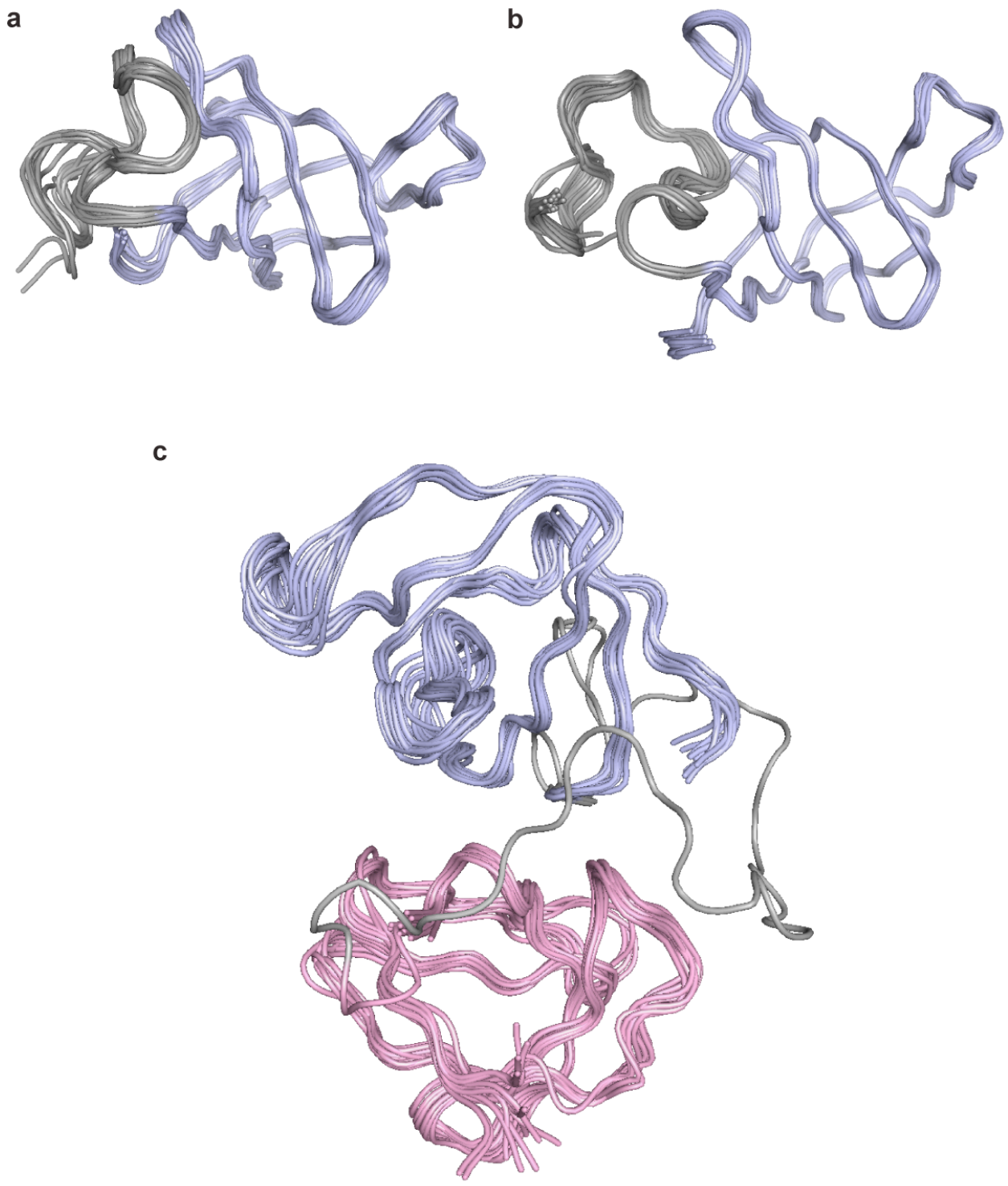
Supplementary Results



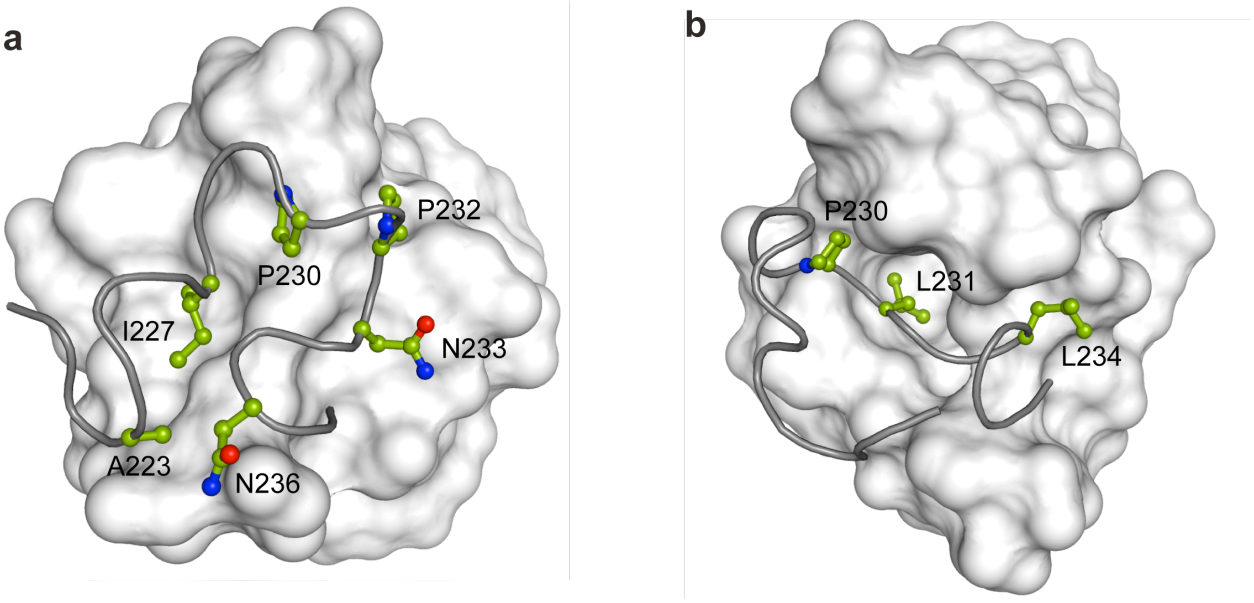
Supplementary Fig. 1. Sequence alignment of Crk-family proteins. The SH domain boundaries are indicated. Residues in $SH3^{C}$ that interact with the linker in the *trans* and *cis* conformations are indicated with a red and blue asterisk, respectively.



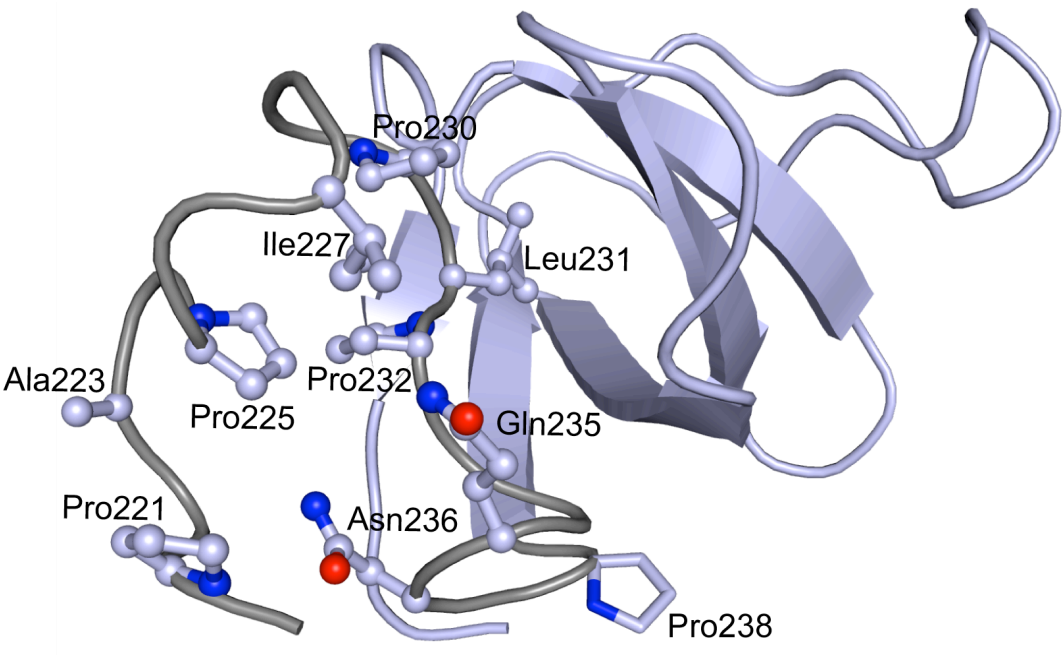
Supplementary Fig. 2. (a) ^{1}H-^{15}N HSQC spectrum of Crk I-SH3^{C}. Assignment is included. **(b)** Strips from ^{1}H-^{1}H planes extracted from a 3D ^{15}N-NOESY-HSQC spectrum. **(c)** Strips from ^{1}H-^{1}H planes extracted from a 3D ^{13}C-NOESY-HSQC spectrum.



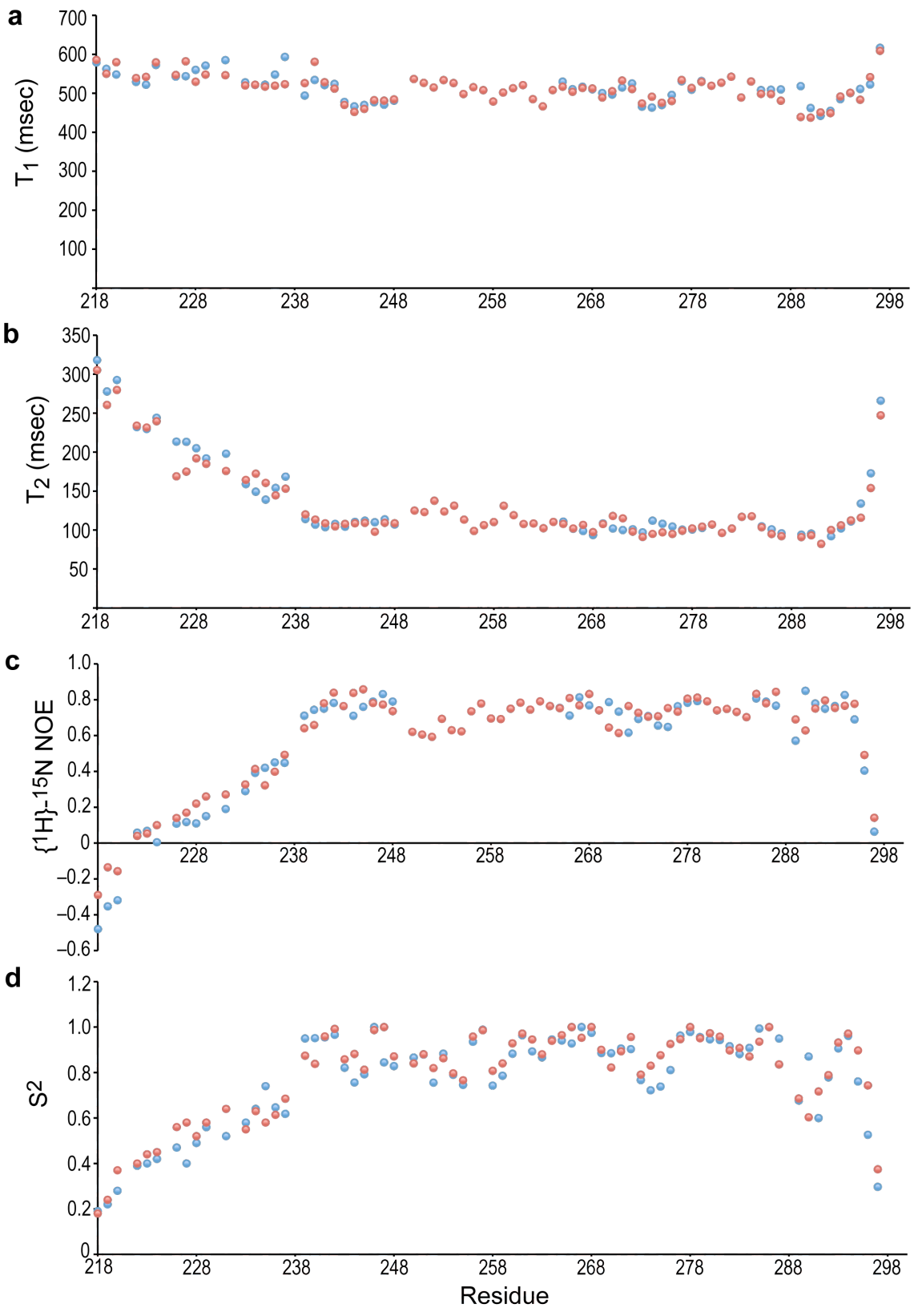
Supplementary Fig. 3. Overlay of the 20 lowest-energy conformers of **(a)** *trans* Crk l-SH3^{C}, **(b)** *cis* Crk l-SH3^{C} and **(c)** Crk^{SLS}. The SH3^{N} and SH3^{C} domains are in blue and pink, respectively, and the linker in grey. In (c) the entire linker is displayed only in one conformer as there is very poor overlap because of its intrinsic flexibility.



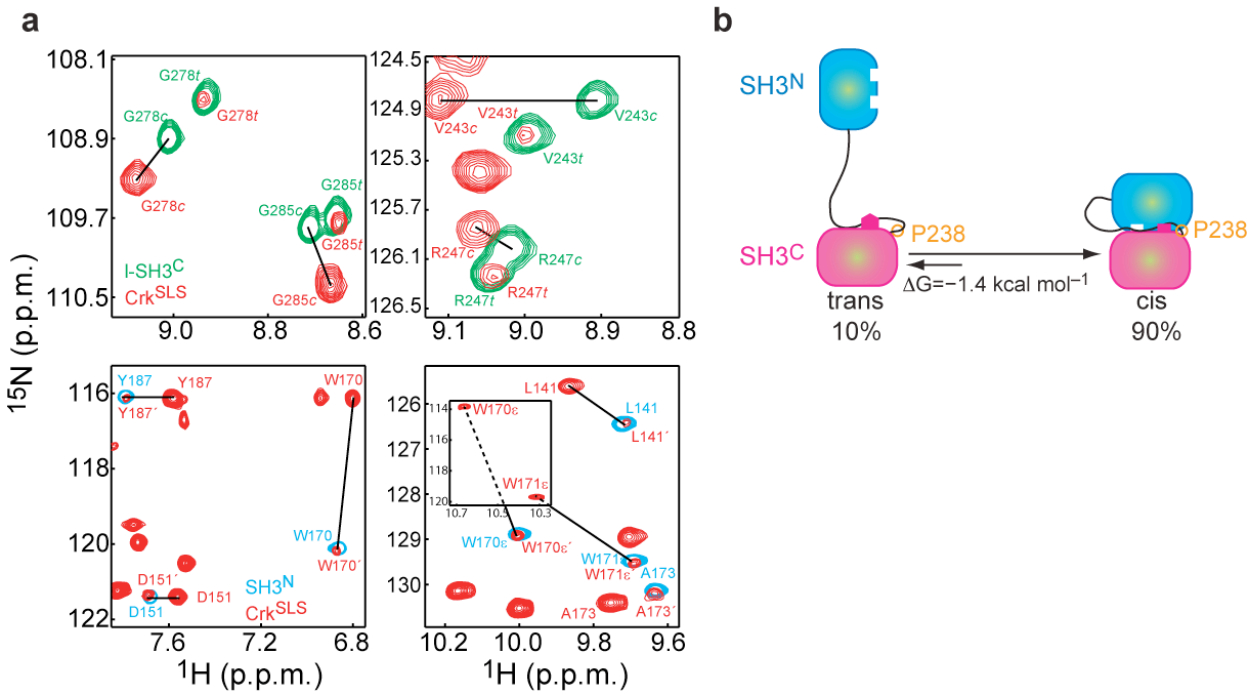
Supplementary Fig. 4. Structure of Crk l-SH3^{C} in the **(a)** *trans* and **(b)** *cis* form. The SH3^{C} domain is shown as a solvent-accessible surface and the linker as tube cartoon. Linker residues that interact with SH3^{C} are shown as sticks and are labeled.



Supplementary Fig. 5. Detailed view of the interactions within the linker in the *cis* form of l-SH3^{C}.

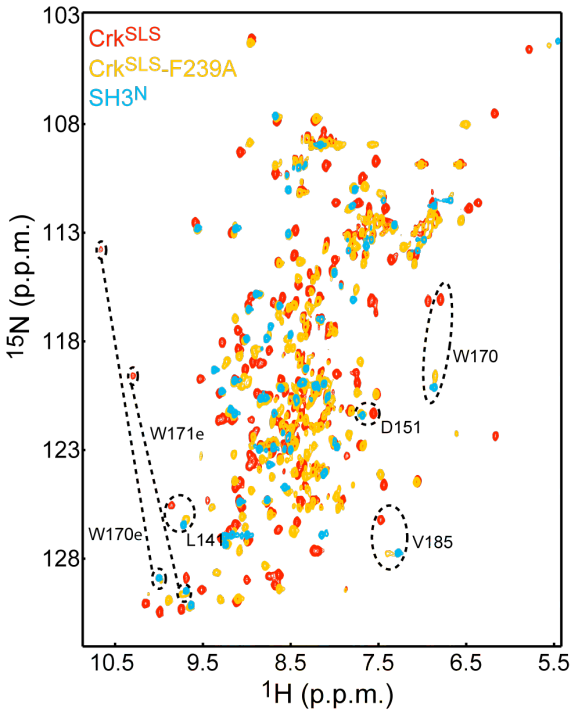


Supplementary Fig. 6. Relaxation rates (T_1, T_2 and $\{^1H\}$-^{15}N-NOE) and N-H order parameters (S^2) for Crk I-SH3^C in the *trans* (blue) and *cis* (red) forms as a function of residue number.



Supplementary Fig. 7. (a) Overlaid ^{1}H-^{15}N HSQC spectra of isolated SH3^{N} (cyan), l-SH3^{C} (green), and Crk^{SLS} (red) indicating the presence of a minor (~10%) conformation of Crk^{SLS} in an open, uninhibited conformation[1]. In this conformation the canonical binding site in SH3^{N} is completely accessible and the Gly237–Pro238 prolyl bond adopts only the *trans* conformation. SH3^{N} residues in the minor conformation are primed. **(b)** Model of the equilibrium of conformational states of Crk^{SLS} (ref.[1]).

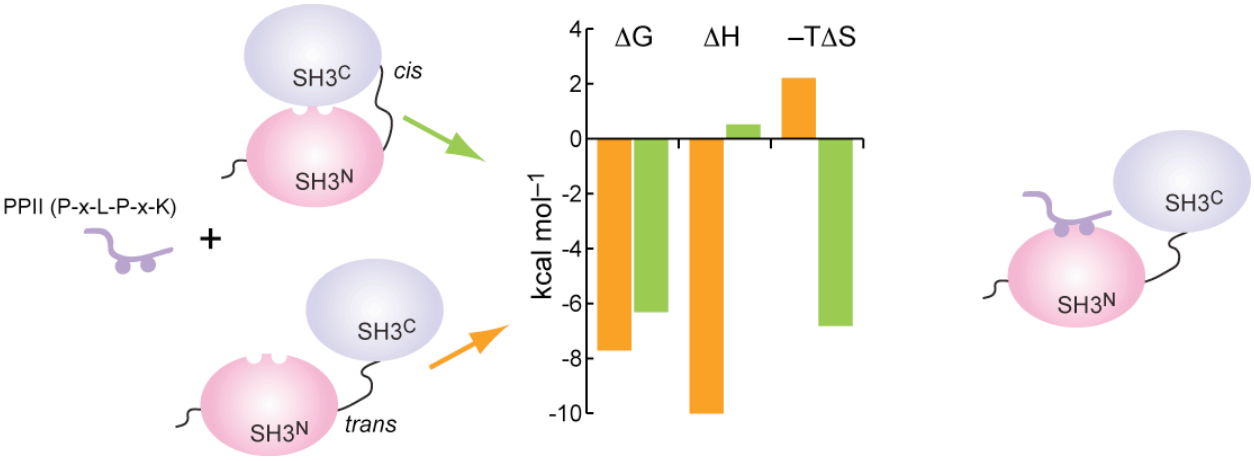
a



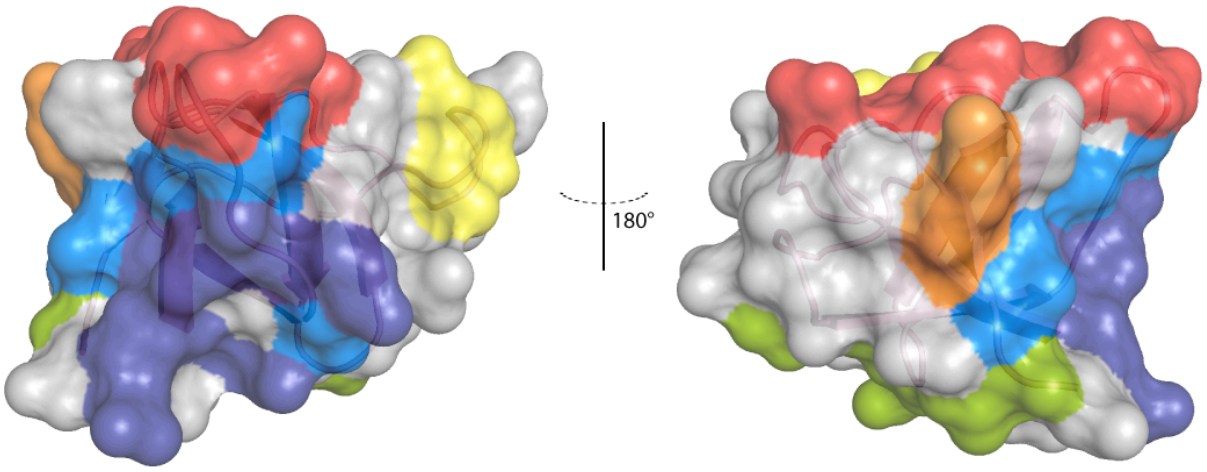
b

Protein	Open (%)	Closed (%)	ΔG (kcal mol⁻¹)
WT-Crk^{SLS}	10	90	-1.4
Crk^{SLS}-P232A	40	60	-0.3
Crk^{SLS}-F239A	92	8	1.5
Crk^{SLS}-P238A	>95	<5	>1.9
Crk^{SLS}-L231G	>95	<5	>1.9

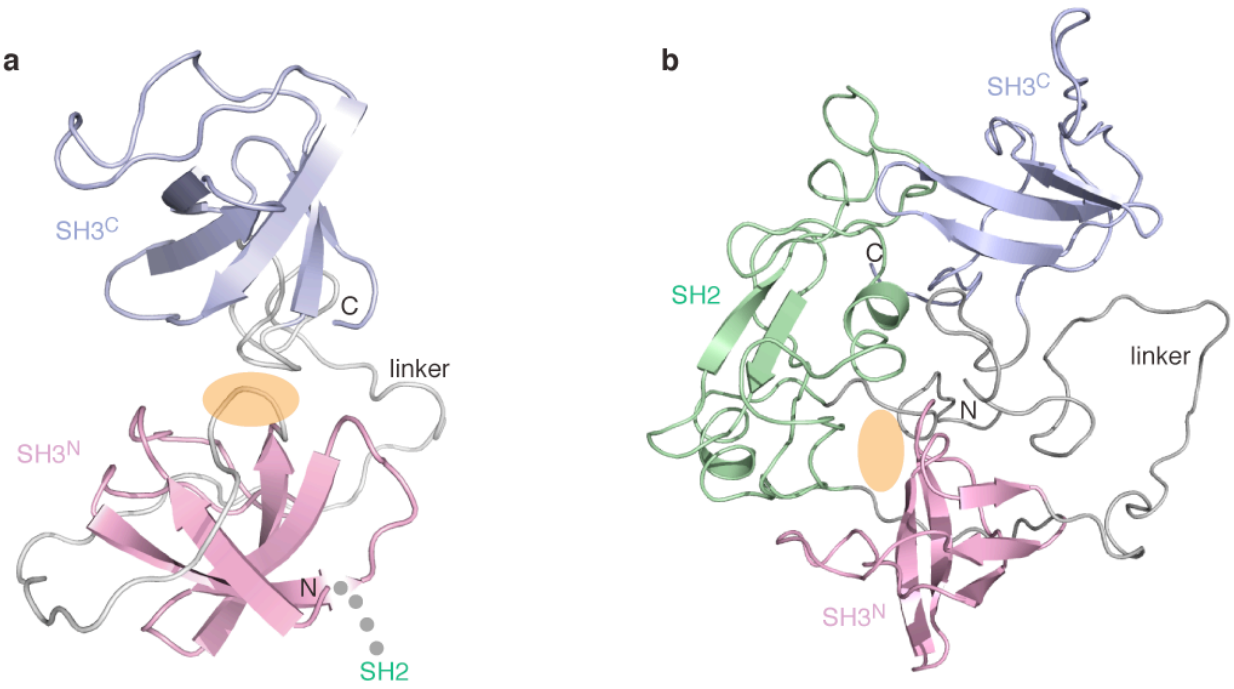
Supplementary Fig. 8. (a) Overlaid ^{1}H-^{15}N HSQC spectra of Crk^{SLS} (red), Crk^{SLS}-F239A (orange), and SH3^{N} (cyan). The F239A substitution results in all resonances of the SH3^{N} domain in Crk^{SLS} shifting to the corresponding chemical shifts of the isolated domain (characteristic shifts are indicated in the figure). **(b)** Percentage of population of Crk^{SLS} molecules adopting the closed and open conformation as assessed by integrating well resolved resonances of the two conformations by NMR.



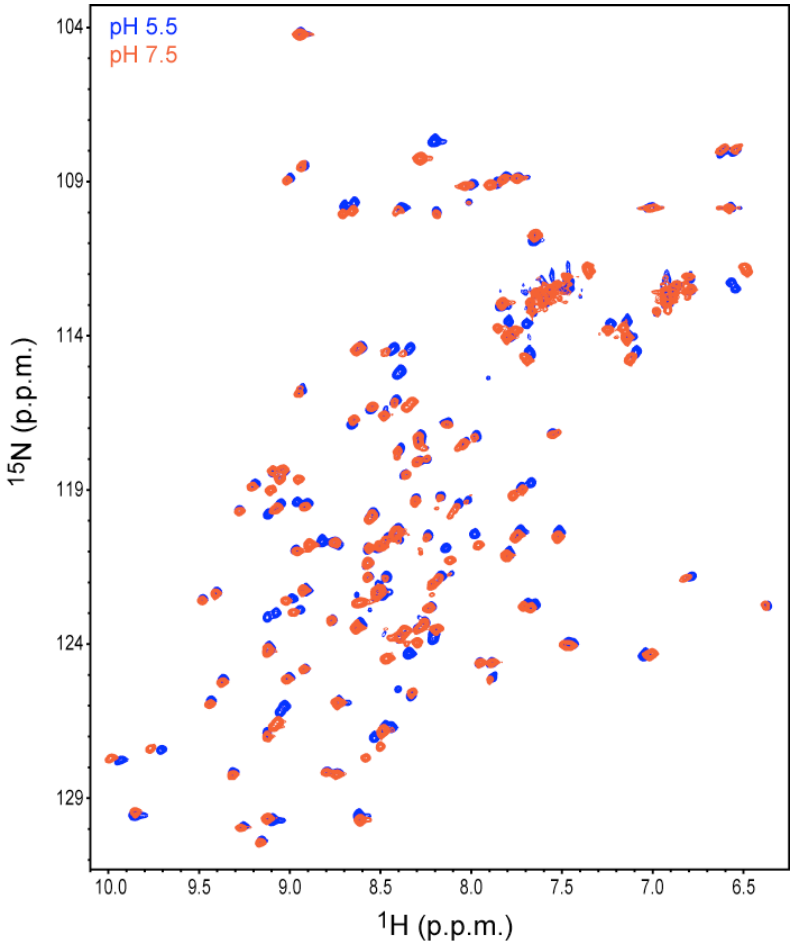
Supplementary Fig. 9. Effect of the autoinhibitory Crk conformation on its interaction with Abl PPII region. Energetics for the interaction of the Crk-binding P-x-L-P-x-K motif of Abl kinase with WT-Crk^{SLS} (closed) and Crk^{SLS}-F239A (open) determined by isothermal titration calorimetry (ITC). Binding to the closed conformation is suppressed by a factor of ~10 (ref.[1]).



Supplementary Fig. 10. SH3^{C} uses novel binding surfaces. Known surfaces used by SH3 domains to interact with non-canonical ligands are colored. Binding surfaces are colored as follows: PPII ligands, Itk SH2[2] and ubiquitin-like domain (Ubl)[3], red; PINCH-1 LIM4[4], orange; SAP SH2[5], yellow; myosin 7 intrapeptide[6], green; Crk intrapeptide linker, light blue (this work); Crk SH3^{N} domain, dark blue (this work).



Supplementary Fig. 11. Structures of **(a)** chicken Crk (the Crk^{SLS} fragment, this work) and **(b)** human full-length Crk[7]. NMR analysis of the chicken full-length Crk and Crk^{SLS} show that the structure of Crk^{SLS} represents the structure of the tandem SH3 domains in the context of the full-length Crk. In both chicken and human Crk proteins the PPII-binding site at $SH3^N$ (indicated by the orange ellipsoid) is autoinhibited. However, different structural elements appear to be involved in blocking the site in the two species. In chicken Crk, it is the $SH3^C$ domain that directly blocks the PPII-binding site whereas in human Crk it is part of the linker in concert with the SH2 domain. It is likely that sequence variation in crucial positions (e.g. Phe239; see Supplementary Figure 1) gives rise to alternative structures. Thus, it appears that the two proteins use different structural strategies to achieve the same goal, that is, autoinhibition of the $SH3^N$ domain.



Supplementary Fig. 12. Overlaid ^{1}H-^{15}N HSQC spectra of Crk I-SH3^{C} recorded in pH 5.5 (blue) and in pH 7.5 (green). The data show that Pro238 isomerization is not sensitive to the pH range 5.5-7.5. Similarly, the relative population of the *cis* and *trans* isomers is pH independent within this pH range.

Supplementary Methods

Protein preparation. Isotopically labeled samples for NMR studies were prepared by growing the cells in M9 media supplemented with 1 g l^{-1} of $^{15}NH_4Cl$ and 2 g l^{-1} of $^{13}C_6$-glucose. Cultures for all constructs were grown at 37 °C, and protein synthesis was induced by addition on 0.5 mM IPTG at OD_{600} ~0.4. The cytosolic fraction was separated from the membrane fraction by centrifugation at 70,000 × g. The lysate was loaded on glutathione Sepharose fast-flow 4% agarose resin (GE) pre-equilibrated with phosphate-buffered saline (PBS), 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4 (pH 7.3). The GST-fused protein was eluted with 10 mM reduced glutathione in 50 mM Tris-HCl (pH 8.0) and was cleaved overnight by the addition of PreScission protease. For the final purification step, the sample was concentrated and applied to a Superdex 75 size exclusion column (GE). For NMR studies, the samples were dialyzed in NMR buffer (50 mM KPi [pH 5.5], 140 mM NaCl and 1 mM DTT) and concentrated using Amicon cell units (Millipore). All fragments are monomeric in solution at concentrations used for the NMR studies (typically 0.6-1.0 mM), as indicated by gel filtration and light scattering. Protein concentration was determined spectrophotometrically at 280 nm using an extinction coefficient of 41035, 24075, and 9970 M^{-1} cm^{-1} for full-length Crk, Crk^{SLS} and L-$SH3^C$, respectively.

Residual Dipolar Coupling (RDC) Measurements. Alignment of the proteins for RDC measurements was achieved using poly(ethylene glycol)/alcohol mixtures[8]. A 5% C12E5/hexanol (molar ratio=0.96) mixture was prepared. C12E5 was used to a final concentration of 5% (w/w) in 90% H2O:10% D2O solution. The pH was adjusted using sodium hydroxide. The amount of hexanol was added dropwise, while vigorously shaking, to a final molar ratio C12E5:hexanol of 0.96. Air bubbles were removed by centrifugation at 5,000 × g for few minutes. The HDO quadrupolar deuterium splitting was checked to confirm the presence of the crystalline phase (a splitting of ~25 Hz was observed). For the measurement of RDCs in the protein, 250 µl of the C12E5: hexanol stock solution was added into 50 µl of protein in buffer. ^{15}N-HSQC (IPAP) and HNCO based experiments were used to measure one-bond N-H and CαC´ RDCs[9]. The alignment tensor was determined as described[10].

Relaxation measurement and analysis. Three relaxation parameters were measured for all backbone amides of Crk^{SLS}: the ^{1}H-^{15}N NOE, the longitudinal relaxation rate R_1 and the transverse relaxation rate R_2. ^{15}N R_1 values were measured from 2D spectra recorded with relaxation delays 100, 200, 300, 400, 500, 600, 700, 800, 1000, 1200 and 1400 ms; ^{15}N R_2 values were measured from 2D spectra recorded with relaxation delays 7.7, 23.2, 31.0, 46.6, 54.3, 62.1, 77.6, 85.4, 93.2, 100, 107.8 and 115.5 ms. Data sets were acquired as 256 x 1,024 complex points in the t1

x t2 time-domain dimensions. Data points were fitted as a function of the length of the parametric relaxation delay to two-parameter decay curves of the form $I(t)=I_{0e}^{-Rt}$, where I is the intensity of the magnetization. ^{1}H-^{15}N NOE data were obtained by recording, in an interleaved manner, one spectrum with a 4-s recycle delay followed by a 4-s saturation and another spectrum with no saturation and a 8-s recycle delay.

ModelFree (http://biochemistry.hs.columbia.edu/labs/palmer/software/modelfree.html) was used to optimize the fit of both internal dynamics and global tumbling parameters using the model-free approach[11]. Initial estimates for the rotational diffusion tensor were obtained from the ratio of longitudinal and transverse relaxation rates[12]. Residues with ^{1}H-^{15}N NOE value less than 0.65 or with R_1 or R_2 values exceeding one standard deviation from the mean and residues experiencing R_{ex} contributions were excluded from the fitting[13]. Overall correlation times and rotational diffusion tensors for isotropic, axially symmetric, and fully anisotropic models were estimated from the R_2/R_1 ratios of the remaining backbone amide groups and the structures of Crk proteins using the programs quadric_diffusion (http://www.palmer.hs.columbia.edu/software/quadric.html) and TENSOR[14]. The most appropriate diffusion tensor was selected by a comparison of χ^2 goodness-of-fit parameters and using F-statistical analysis. Using these rotational diffusion tensors, backbone relaxation data were fit to the five standard Lipari-Szabo model-free formalism models[15]. The fitted dynamics parameters for each model are as follows: model 1, order parameter (S^2); model 2, S^2 and internal correlation time (τ_e); model 3, S^2 and R_{ex}; model 4, S^2, τ_e, and R_{ex}; model 5, order parameters for two time scales (S_f^2 and S_s^2) and τ_e for the slower time scale. Parameters of the model-free formalism were optimized for each residue individually, and the best parameter set identified by model selection. All parameters, including the diffusion tensor, were then optimized. This process was repeated until the solution converged. The quality of the fits between the experimental data and each model were calculated as χ^2 statistics, and the different models were then compared to each other using F statistics. For both the *cis* and *trans* forms of L-SH3^{C} (τ_C~6.2 ns) an axially symmetric diffusion tensor was optimal, with $D_{\parallel}/D_{\perp}$~1.23 and ~1.25, respectively.

Calorimetry. Calorimetric titrations of Abl-PxxP (aa 50-535) or a PPII peptide (N-DNSPPPALPPKKRQSAPSC) with Crk^{SLS}, Crk^{SLS}-F239A and SH3^{N} were performed on an iTC200 microcalorimeter (Microcal). Protein samples were extensively dialyzed against the ITC buffer containing 50 mM KPi (pH 5.5 or 7.0), 140 mM NaCl, and 1 mM TCEP. All solutions were filtered using membrane filters (pore size, 0.45 mm) and thoroughly degassed for 20 min by gentle stirring under vacuum. The sample cell was filled with a 50 μM solution of protein, and the injection

syringe was filled with 0.5 μM titrating PPII peptide. The ligand solution was prepared by dissolving peptide in the flowthrough of the last buffer exchange. Each titration typically consisted of a preliminary injection followed by 15-20 subsequent injections. Data for the preliminary injection, which are affected by diffusion of the solution from and into the injection syringe during the initial equilibration period, were discarded. Binding isotherms were generated by plotting heats of reaction normalized by the modes of injectant versus the ratio of total injectant to total protein per injection. The data were fitted using Origin 7.0 (Microcal).

References

1. Sarkar, P., Reichman, C., Saleh, T., Birge, R. & Kalodimos, C. Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Mol Cell* **25**, 413-26 (2007).

2. Severin, A., Joseph, R., Boyken, S., Fulton, D. & Andreotti, A. Proline isomerization preorganizes the Itk SH2 domain for binding to the Itk SH3 domain. *J Mol Biol* **387**, 726-43 (2009).

3. Trempe, J.-F. et al. SH3 domains from a subset of BAR proteins define a Ubl-binding domain and implicate parkin in synaptic ubiquitination. *Mol Cell* **36**, 1034-47 (2009).

4. Vaynberg, J. et al. Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. *Mol Cell* **17**, 513-23 (2005).

5. Chan, B. et al. SAP couples Fyn to SLAM immune receptors. *Nat Cell Biol* **5**, 155-60 (2003).

6. Wang, Q. et al. The SH3 domain of a M7 interacts with its C-terminal proline-rich region. *Protein Sci* **16**, 189-96 (2007).

7. Kobashigawa, Y. et al. Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nat Struct Mol Biol* **14**, 503-10 (2007).

8. Ruckert, M. & Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* **122**, 7793-7797 (2000).

9. Yao, L., Ying, J. & Bax, A. Improved accuracy of 15N-1H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J Biomol NMR* **43**, 161-70 (2009).

10. Clore, G., Gronenborn, A. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* **133**, 216-21 (1998).

11. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546-4559 (1982).

12. Tjandra, N., Feller, S., Pastor, R. & Bax, A. Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *J Am Chem Soc* **117**, 12562-12566 (1995).

13. Hwang, P., Skrynnikov, N. & Kay, L. Domain orientation in beta-cyclodextrin-loaded maltose binding protein: diffusion anisotropy measurements confirm the results of a dipolar coupling study. *J Biomol NMR* **20**, 83-8 (2001).

14. Dosset, P., Hus, J., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **16**, 23-8 (2000).

15. Mandel, A., Akke, M. & Palmer, A. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* **246**, 144-63 (1995).

Supplementary Table 1 NMR and refinement statistics for Crk structures

	l-SH3^{C} *trans*	l-SH3^{C} *cis*	Crk^{SLS}
NMR distance and dihedral constraints			
Distance constraints			
Total NOE	1523	1557	3094
Intra-residue	328	333	760
Inter-residue			
Sequential ($\|i-j\|=1$)	375	380	752
Medium-range ($\|i-j\|<4$)	187	194	368
Long-range ($\|i-j\|>5$)	633	650	1214
SH3^{N}-SH3^{C} interdomain			42
Hydrogen bonds	22	22	46
Total dihedral angle restraints			
ϕ	52	48	92
ψ	52	49	88
Total RDCs			
$^{1}D_{HN}$			90
Q (%)			18
Structure statistics			
Violations (mean and s.d.)			
Distance constraints (Å)	0.06±0.008	0.08±0.014	0.07±0.010
Dihedral angle constraints (º)	2.15±0.241	3.03±0.369	4.20±0.542
Max. dihedral angle violation (º)	3.2	4.1	6.0
Max. distance constraint violation (Å)	0.31	0.42	0.45
Deviations from idealized geometry			
Bond lengths (Å)	0.012	0.013	0.013
Bond angles (º)	1.2	1.5	1.5
Impropers (º)	2.3	2.2	2.4
Average pairwise r.m.s. deviation** (Å)			
Heavy	0.7	0.5	1.2
Backbone	0.4	0.3	0.7

** Pairwise r.m.s. deviation was calculated among 20 refined structures.