Structure, Volume *24*

Supplemental Information

Expanding the PP2A Interactome by Defining a B56-Specific SLiM

Xinru Wang, Rakhi Bajaj, Mathieu Bollen, Wolfgang Peti, and Rebecca Page

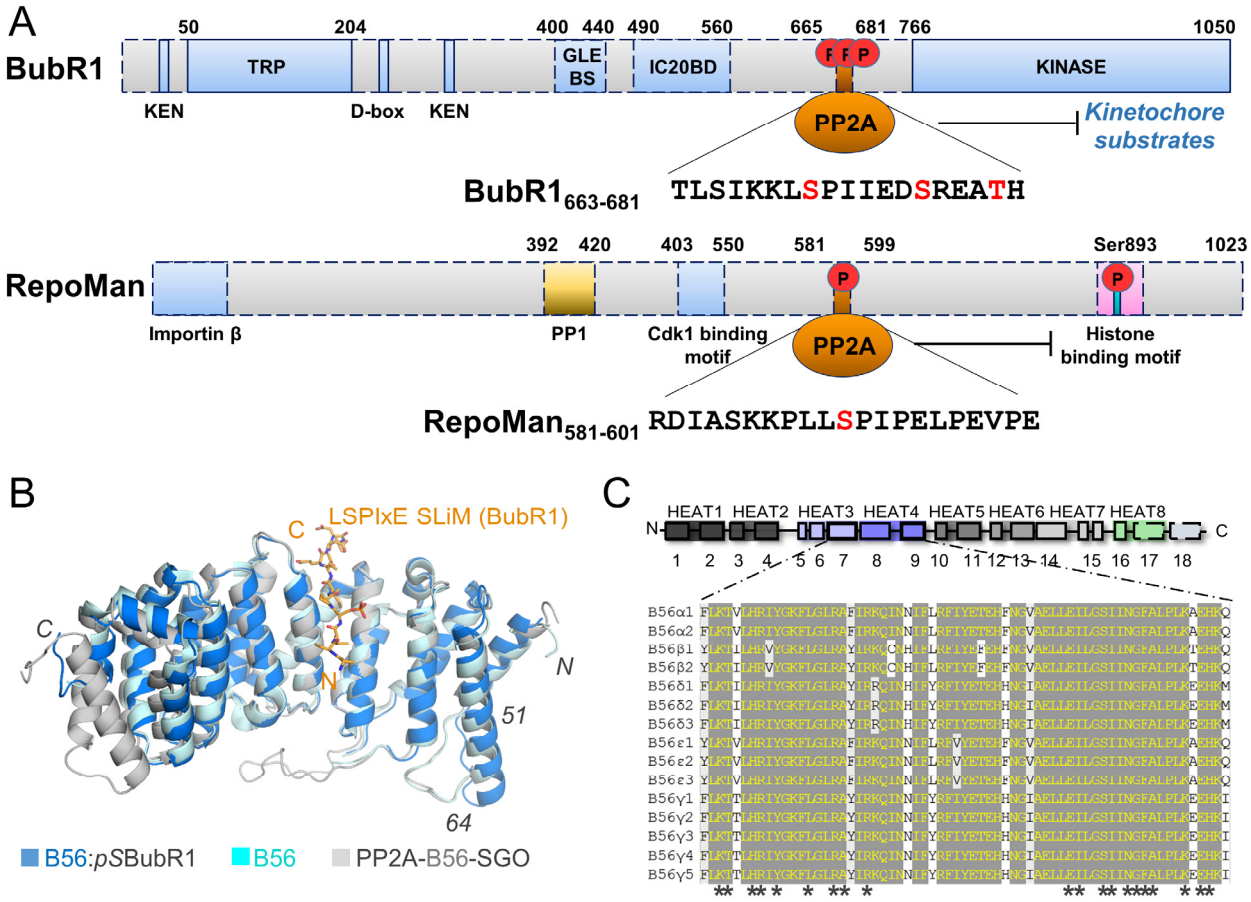

Figure S1. B56 targeting proteins and the overlay of distinct B56 complexes. Related to Figure 1.
(A) Domain structure of BubR1 and RepoMan; known protein interaction domains are indicated. Dotted lines indicate intrinsically disordered regions (IDRs); solid lines indicate folded domains. PP2A:B56 interaction sequence indicated.
(B) Overlay of B56 from the B56:*pS*BubR1 complex (blue) with free B56 (PDBID 2JAK; cyan) and B56 from the PP2A-B56-SGO complex (PDBID 3FGA; grey). Residue number 51-64 indicate the location of the loop that adopts a distinct conformation in the B56:*pS*RepoMan and B56:*pS*BubR1 complexes due to crystal contacts.
(C) Sequence alignment of the residues from B56 isoforms (homo sapiens) with the residues that constitute the LSPIxE binding pocket indicted by a '*'.

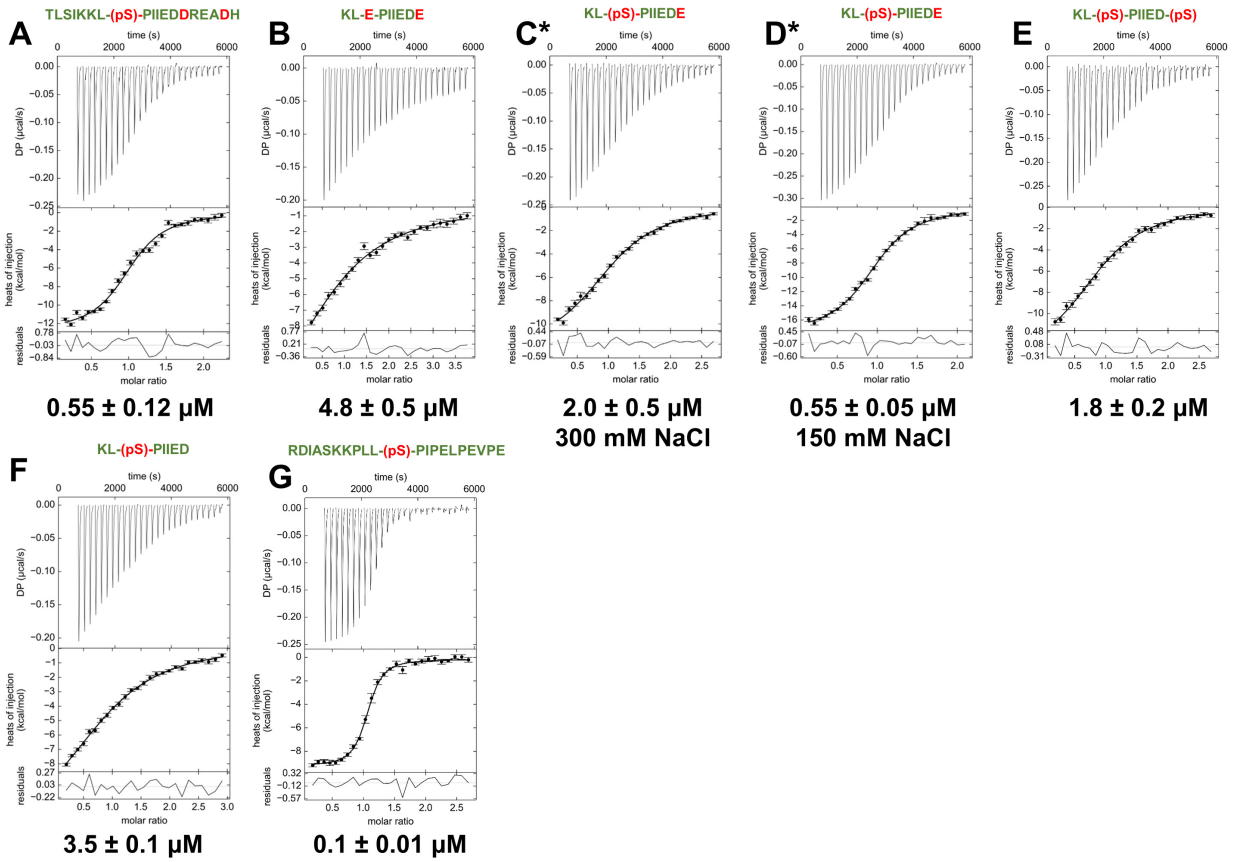

Figure S2. Isothermal titration calorimetry of LSPIxE peptides. **Related to Figure 1.**
(A) B56γ$_{12\text{-}380}$:BubR1 ^{663}TLSIKKL(pS)PIIED*D*REA*D*H^{681}
(B) B56γ$_{12\text{-}380}$:BubR1 ^{668}KLEPIIEDE^{676}
(C) B56γ$_{12\text{-}380}$:BubR1 ^{668}KL-(pS)-PIIEDE^{676}
(D) B56γ$_{12\text{-}380}$:BubR1 ^{668}KL-(pS)-PIIEDE^{676} , *low salt.*
(E) B56γ$_{12\text{-}380}$:BubR1 ^{668}KL-(pS)-PIIED(pS)^{676}
(F) B56γ$_{12\text{-}380}$:BubR1 ^{668}KL-(pS)-PIIED^{675}
(G) B56γ$_{12\text{-}380}$:RepoMan ^{581}RDIASKKPLL(pS)PIPELPEVPE^{601}.
All experiments performed in ITC buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 0.5 mM TCEP), with the exception of (D), which was performed in a low salt buffer (50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP). All measurements represent an n = 2 or 3.

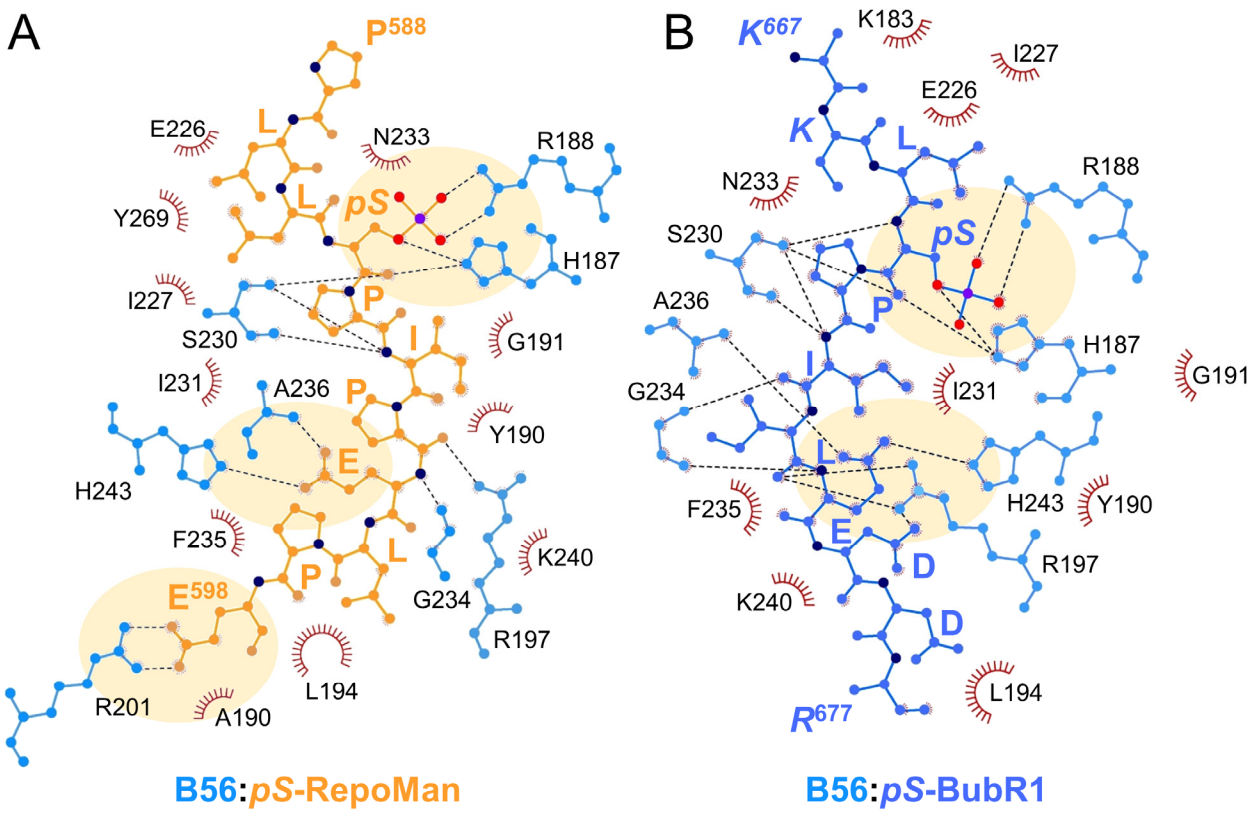

Figure S3. Ligplot figures of the hydrophobic and electrostatic/polar interactions between LSPIxE peptides and B56. Related to Figure 2.

(A) Ligplot diagram of the hydrophobic (red semi-circles) and electrostatic/polar interactions (dashed lines) between B56 (blue) and RepoMan (orange). B56 residues are labeled in black while RepoMan residues are in orange (^{588}PLL-*pS*-PIPELPE^{598}). Residues in key electrostatic interactions are highlighted by an orange circle.

(B) Ligplot diagram of the hydrophobic (red semi-circles) and electrostatic/polar interactions (dashed lines) between B56 (blue) and BubR1 (lavender). B56 residues are labeled in black while BubR1 residues are in blue (^{567}*KK*L-*pS*-PILEDD*R*^{677}; sidechains were not modeled for italicized residue labels due to a lack of density). Residues in key electrostatic interactions are highlighted by an orange circle.

Table S1: Isothermal titration calorimetry (ITC) measurements of PP2A-B56 with BuBR1 and RepoMan. Related to Figure 1.

B56	Peptides	K_D (μM)	ΔH (kcal/mol)	TΔS (kcal/Mol)
12-380	BUBR1 - ^{668}KLEPIIEDE^{676}	4.8 ± 0.5	-15.5 ± 0.8	-8.2 ± 0.7
12-380	BUBR1 - ^{668}KL-(pS)-PIIEDE^{676}	2.0 ± 0.5	-14.7 ± 1.6	-7.0 ± 1.8
12-380*	BUBR1 - ^{668}KL-(pS)-PIIEDE^{676}	0.55 ± 0.05	-17.6 ± 0.1	-9.1 ± 0.1
12-380	BUBR1 - ^{668}KL-(pS)-PIIED(pS)^{676}	1.8 ± 0.2	-13.8 ± 0.9	-6.0 ± 1.0
12-380	BUBR1 - ^{668}KL-(pS)-PIIED^{675}	3.5 ± 0.1	-14.5 ± 1.6	-7.1 ± 1.7
12-380	BUBR1 - ^{663}TLSIKKL(pS)PIIEDDREADH^{681}	0.55 ± 0.12	-12.9 ± 0.1	-4.4 ± 0.1
12-380	RepoMan - ^{581}RDIASKKPLL(pS)PIPELPEVPE^{601}	0.1 ± 0.01	-10.0 ± 0.9	-0.5 ± 1.0

All reported measurements are performed with ITC buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 0.5 mM TCEP). Errors are from duplicate or triplicate measurements

*Low salt ITC buffer: 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP

Table S2: Residues that define the B56:BubR1 (LSPIxE) interface. Related to Figure 1.

B56°	**BSA^ (Å²)**	***pS*- BubR1***	**BSA (Å²)**	***pSpS*-BubR1***	**BSA (Å²)**	***pS*-RepoMan***	**BSA (Å²)**
Leu 194	63.0	**Ile 672**	147.9	**Ile 672**	147.1	**Ile 593**	148.9
His 187	60.2	**Glu 674**	142.6	**Leu 669**	136.2	**Leu 590**	131.7
Ser 230	54.4	**Leu 669**	135.1	**Glu674**	134.7	**Glu 595**	126.1
Gly 234	53.9	***p*Ser 670**	57.9	***p*Ser 670**	55.2	**Glu 598***	99.3
Arg 197	51.9	**Pro 671**	46.4	Asp 675	50.5	***p*Ser 591**	58.5
Glu 226	38.1	Arg 677	38.4	**Pro 671**	42.3	**Leu 589**	49.4
Arg 188	36.6	Asp 675	37.7	Ile 673	24.2	**Pro 592**	45.6
Asn 233	26.9	Lys 668	26.7	Lys 668	21.6	Leu 596	39.3
Lys 240	26.6	Ile 673	21.1			Pro 594	18.9
Ala 236	24.2	Asp 676	11.0			Pro 597	14.1
Tyr 190	16.4	Lys 667	6.1			Pro 588	5.3
Ile 231	14.9						
Tyr 269	14.5						
His 243	12.8						
Phe 235	11.9						
Lys 183	11.7						
Thr 184	11.2						
Gly 191	9.0						
Arg 201	8.8						
Gly 195	5.5						
Ile 227	2.9						

°Values of BSA for B56 residues obtained from the B56:pS-BubR1 complex
^BSA, buried solvent accessible surface area
*LSPI-E residues in bold; the RepoMan Glu residue (Glu 598) that makes an additional salt bridge with B56 is in red

Table S3: Human proteins with LSPIxE motifs. Related to Figure 3.

Protein*	Gene	UniProt	Localization	Motif
Mitotic checkpoint serine/threonine-protein kinase BUBR1	*BUB1B*	*O60566*	Cyto/Nucleus	*LSPIiE*
RepoMan	*CDCA2*	*Q69YH5*	Nucleus	*LSPIpE*
Very large A-kinase anchor protein (vIAKAP)	CRBG3	Q69YH5	n/a	LSPIyE
Synaptopodin	SYNPO	Q8N3V7	Cell Junction	LSPIkE
Oxidation Resistance Protein 1	OXR1	Q8N573	Mitochondria	LSPIrE
Inhibitor of ASPP protein	iASPP	Q8WUf5	Cyto/Nucleus	LSPItE
Remodeling and spacing factor	RSF1	Q96T23	Nucleus	LSPIpE
Zinc finger protein 541	ZN541	Q9H0D2	Nucleus	LSPIrE
Homeobox protein nkx-2.4 (NK-2 homolog D)	NKX24	Q9H2Z4	Nucleus	LSPIeE
Uncharacterized protein KIAA1107	K1107	Q9UPP5	n/a	LSPIyE

*proteins in *italics* contain confirmed LSPIxE SLiMs that bind directly to B56

Table S4: **Number of human sequences identified in the UniProt database using increasingly less restrictive LSPIxE motifs. Related to Figure 3.**

Search Motif	Hits	Filter*	Search Motif	Hits	Filter*
[L]SP[I]	161				
[L]SP[I]xE	13	10	[L][ST]P[I]xE	18	11
[LCVM]SP[I]xE	26	15	[LCVM][ST]P[I]xE	33	17
[LCVMIF]SP[I]xE	38	20	[LCVMIF][ST]P[I]xE	57	22
[LCVM]SP[ILV]xE	117	49	[LCVM][ST]P[ILV]xE	206	72
[LCVMIF]SP[ILV]xE	173	67	[LCVMIF][ST]P[ILV]xE	301	92
[LCVMIF]SP[ILVM]xE	186	75	[LCVMIF][ST]P[ILVM]xE	327	104

*filter indicates those motifs in which the ser/thr residue satisfies the following two criteria (a) it is likely to be phosphorylated and (b) it is predicted to be in a IDR

Table S5: Excel Table of all 104 proteins identified using the following search motif: [LCVMIF][ST]P[ILVM]xE. Related to Figure 3.

Supplemental Experimental Procedures

Cloning and expression. Human B56γ1 (B56γ_{12-380} and B56γ_{31-380}) was sub-cloned into pRP1b vector (Peti and Page, 2007). B56γ_{12-380} and B56γ_{31-380} were expressed in *E. coli* BL21 (DE3) (Agilent). Cells were grown in Luria Broth in the presence of selective antibiotics at 37°C to an OD_{600} of ~0.8, and expression was induced by the addition of 0.5 mM isopropyl β-D- thiogalactoside (IPTG). Induction proceeded for ~18-20 h at 18°C prior to harvesting by centrifugation at 6,000 ×*g*. Cell pellets were stored at -80°C until purification. All peptides were synthesized, HPLC purified (>95% purity) and analyzed (MS) by Bio-Synthesis (Lewisville, TX).

Protein Purification. B56γ cell pellets were resuspended in ice-cold lysis buffer (50 mM Tris pH 8.0, 0.5 M NaCl, 5 mM imidazole, 0.1% Triton X-100 containing EDTA-free protease inhibitor tablet [Roche]), lysed by high-pressure cell homogenization (Avestin C3 Emulsiflex) and centrifuged (35,000 ×*g*, 40 min, 4°C). The supernatant was loaded onto a HisTrap HP column (GE Healthcare) pre-equilibrated with Buffer A (50 mM Tris pH 8.0, 500 mM NaCl and 5 mM imidazole) and was eluted using a linear gradient of Buffer B (50 mM Tris pH 8.0, 500 mM NaCl, 500 mM imidazole). Fractions containing the protein were pooled and dialyzed overnight at 4 °C (50 mM Tris pH 8.0, 500 mM NaCl) with TEV protease to cleave the His_6-tag. The cleaved protein was incubated with Ni^{2+}-NTA beads (GEHealthcare) and the flow-through collected. The protein was concentrated and purified using size exclusion chromatography (SEC; Superdex 75 26/60 [GE Healthcare]) pre-equilibrated in ITC Buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 0.5 mM TCEP) or Crystallization Buffer (20 mM HEPES pH 7.8, 500 mM NaCl, 0.5 mM TCEP). Fractions were pooled, concentrated to designated concentration for experiments or stored at -80˚C.

Crystallization and structure determination. Pooled B56γ_{31-380} (hereafter, B56) in 20 mM HEPES pH 7.8, 500 mM NaCl, 0.5 mM TCEP was concentrated and combined with *pS*-BubR1, *pSpS*-BubR1 and *pS*-RepoMan dissolved in the same buffer at a 1:5 molar ratio to a final concentration of 10 mg/ml. Initial crystals of the complexes were identified in 0.1 M HEPES pH 7.75, 0.8 M LiCl and 8% PEG6K (B56:*pS*-BubR1) or 8% PEG8K (B56:*pSpS*-BubR1, B56:*pS*-RepoMan) using hanging drop vapor diffusion at RT. Layered crystals grew overnight. Microseeding was necessary to generate single crystals. Crystals were cryo-protected by a 30 sec soak in mother liquor with 30% glycerol and immediately flash frozen. Data for the three complexes were collected at SSRL beamline 12.2 at 100 K and a wavelength of 0.98 Å using a Pilatus 6M PAD detector. The data were processed using Xds (Kabsch, 2010), Aimless (Evans and Murshudov, 2013) and Truncate (French and Wilson, 1978). The structure of B56:*pS*BubR1 was solved by molecular replacement using Phaser (Zwart et al., 2008), using B56 (PDBID 2JAK) as the search model (Magnusdottir et al., 2009). A solution was obtained in space group $P2_12_12_1$; clear electron density for the BubR1 peptide was visible in the initial maps. The initial models of the complex were built without the peptide using AutoBuild, followed by iterative rounds of refinement in PHENIX and manual building using Coot (Emsley et al., 2010). The peptide coordinates were then added followed by iterative rounds of refinement in PHENIX and manual building using Coot. The remaining complexes were phased using Fourier synthesis and refined in similar fashion. Data collection and refinement details are provided in **Table 1**.

Isothermal titration calorimetry. SEC was performed to polish B56 and exchange into ITC Buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 0.5 mM TCEP). BubR1 (70-100 μM) or RepoMan (70-100 μM) peptides were titrated into B56γ_{12-380} (7-10 μM) using a VP-ITC micro-calorimeter at 25°C (Malvern). Data were analyzed using NITPIC, SEDPHAT and GUSSI (Scheuermann and Brautigam, 2015; Zhao et al., 2015).

Bioinformatics. Multiple methods were used to confirm that the identified motifs are present in IDPs. First, IUPRED was used to distinguish sequences likely folded (IUPRED scores ≤0.4) from those likely disordered (IUPRED score> 0.4) (Dosztanyi et al., 2005). The 126 sites identified to be present in IDPs using IUPRED were then subsequently examined using both DisEMBL (loops/coils definition) (Linding et al., 2003) and PONDR (using both the VL3-BA and the VLXT predictors) (Romero et al., 2001). This resulted in a total of 102 sites classified as disordered by all three prediction algorithms with 24 sites predicted to be disordered by only 1 or 2 algorithms. The latter sites were investigated for (a) their presence in the PDB, (b) the presence of distal homologs whose structures are known (FFAS) (Xu et al., 2014) and (c) the likelihood that they are present in transmembrane domains or in coiled-coils. This additional analysis resulted in 2 of the 24 sites classified as disordered. The remaining 22 sites were eliminated as their sequences were present in protein domains that are structured, structures had been determined or proteins classified as secreted.

The ConSurf server (using 150 unique B56 sequences with the lowest E values (Ashkenazy et al., 2010)) were used to calculate the conservation scores. Figures generated using PYMOL. The T-Coffee Multiple Sequence Alignment Server (Notredame et al., 2000) was used to generate the sequence alignment for B56 isoforms and splice variants. Multiple Align Show (www.bioinformatics.org/sms/multi_align.html) was used to enhance the output from T-Coffee.

Supplemental References

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol *347*, 827-839.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204-1214.

French, G.S., and Wilson, K.S. (1978). On the treatment of negative intensity observations. Acta Crystallogr A *34*, 517-525.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. Structure *11*, 1453-1459.

Magnusdottir, A., Stenmark, P., Flodin, S., Nyman, T., Kotenyova, T., Graslund, S., Ogg, D., and Nordlund, P. (2009). The structure of the PP2A regulatory subunit B56 gamma: the remaining piece of the PP2A jigsaw puzzle. Proteins *74*, 212-221.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Peti, W., and Page, R. (2007). Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. Protein Expr Purif *51*, 1-10.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins *42*, 38-48.

Scheuermann, T.H., and Brautigam, C.A. (2015). High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. Methods *76*, 87-98.

Xu, D., Jaroszewski, L., Li, Z., and Godzik, A. (2014). FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. Bioinformatics *30*, 660-667.

Zhao, H., Piszczek, G., and Schuck, P. (2015). SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. Methods *76*, 137-148.

Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., McKee, E., Moriarty, N.W., Read, R.J., Sacchettini, J.C.*, et al.* (2008). Automated structure solution with the PHENIX suite. Methods Mol Biol *426*, 419-435.