Novel Interactors of the SH2 Domain of the Signaling Adaptors CRK and CRKL Identified in Neuro2A Cells

*Caroline M. Dumas†, Anna M. Schmoker‡, Shannon R. Bennett‡, Amara S. Chittenden‡, Chelsea B. Darwin‡, Helena K. Gaffney‡, Hannah L. Lewis‡, Eliana Moskovitz‡, Jonah T. Rehak‡, Anna A. Renzi‡, Claire E. Rothfelder‡, Adam J. Slamin‡, Megan E. Tammaro‡, Leigh M. Sweet, & Bryan A. Ballif**

Department of Biology, University of Vermont, 109 Carrigan Drive, 120A Marsh Life Sciences, Burlington, VT 05405, USA

https://doi.org/10.33697/ajur.2022.068

Students: Caroline.Dumas@uvm.edu, Shannon.Bennett@uvm.edu, Amara.Chittenden@uvm.edu, Cheslea.Darwin@uvm.edu, Helena.Gaffney@vum.edu, Hannah.Lewis@uvm.edu, Eliana.Moskovitz@uvm.edu, Jonah.Rehak@uvm.edu, Anna.Renzi@uvm.edu, Claire.Rothfelder@uvm.edu, Adam.Slamin@uvm.edu, Megan.Tammaro@uvm.edu
*Mentor: bballif@uvm.edu**

† Co-first authors, equal contribution
‡Equal contribution

ABSTRACT
CT10 regulator of kinase (CRK) and CRK-like (CRKL) form a family of signaling adaptor proteins that serve important roles in the regulation of fundamental cellular processes, including cell motility and proliferation, in a variety of cell types. The Src Homology 2 (SH2) domain of CRK and CRKL interacts with proteins containing phosphorylated tyrosine-X-X-proline (pYXXP) motifs, facilitating complex formation during signaling events. A handful of CRK/CRKL-SH2-specific interactors have been identified to date, although *in silico* analyses suggest that many additional interactors remain to be found. To identify CRK/CRKL-SH2 interactors with potential involvement in neuronal development, we conducted a mass spectrometry-based proteomics screen using a neuronal cell line (Neuro2A, or N2A). This resulted in the identification of 132 (6 known and 126 novel) YXXP-containing CRK/CRKL-SH2 interactors, of which 77 were stimulated to bind to the CRK/CRKL-SH2 domain following tyrosine phosphatase inhibition. Approximately half of the proteins identified were common interactors of both the CRK- and CRKL-SH2 domains. However, both CRK family member SH2 domains exhibited unique binding partners across experimental replicates. These findings reveal an abundance of novel neuronal CRK/CRKL-SH2 domain binding partners and suggest that CRK family SH2 domains possess undescribed docking preferences beyond the canonical pYXXP motif.

KEYWORDS
CRK; CRKL; SH2; LC-MS/MS; Proteomics; Neurodevelopment; Signal Transduction

INTRODUCTION

CT10 regulator of kinase (CRK) and CRK-Like (CRKL) form a family of signaling adaptor proteins that are conserved throughout metazoans and pre-metazoans, playing essential roles in fundamental cellular processes such as proliferation and motility. CRK and CRKL come from two distinct genes that have been shown to play overlapping roles in some signaling pathways.[1] Interestingly, CRK and CRKL have also been shown to have unique roles *in vivo*, as knockout mice for CRK show differing phenotypes than knockout mice for CRKL.[2,3] Both CRK family members possess a Src homology 2 (SH2) domain, which binds phosphorylated tyrosine residues in YXXP motifs (Y=tyrosine, P=proline, X=any amino acid), and two SH3 domains, which bind proline-rich sequences. CRK is known to express as two different isoforms, both have an SH2 domain, but one with two SH3 domains and the other with only one SH3 domain. CRKL only has one isoform containing one SH2 domain and two SH3 domains.[5] Through these modular binding domains, CRK and CRKL facilitate the formation of cellular signaling complexes required for numerous signal transduction pathways.[1,4-7]

CRK and CRKL are particularly important for the proper positioning of cortical neurons during embryonic brain development via Reelin signaling.[4] Src family kinases are activated within a migrating neuron following the binding of Reelin to the Reelin receptors VLDLR and ApoER2. Activated src family kinases then phosphorylate tyrosine residues of receptor-bound Disabled-1 (DAB1) in YXXP motifs, which leads to the translocation of CRK/CRKL and their SH3-bound guanine nucleotide exchange factors complexes to membrane-localized small GTPases, which regulate cytoskeletal dynamics.[1,4,7]

In addition to DAB1, several known CRK/CRKL-SH2 interactors have been characterized, including members of the CRK-associated substrate (CAS) and discoidin, CUB, and LCCL domain-containing protein (DCBLD) families (**Supplementary Table 1**).[8-10] Further, in a recent *in silico* proteomics screen we predicted several CRK- and CRKL-SH2 binding partners, although many of these have yet to be shown to interact experimentally.[9] To identify new neuronal CRK/CRKL-SH2 interactors we made use of a murine neuroblastoma cell line, Neuro2A (or N2A). N2A cells were treated with or without tyrosine phosphatase inhibitors and extracts were incubated with either the CRK- or the CRKL SH2 domain. Bound proteins were ultimately identified using liquid chromatography–tandem mass spectrometry (LC-MS/MS).

METHODS.

Materials.

Penicillin/Streptomycin (100X) and Dulbecco's Modified Eagle Medium (DMEM) were acquired from Mediatech (Manassas, VA, USA). Fetal bovine serum (FBS) and cosmic calf serum (CCS) were obtained from Hyclone (Logan, Utah, USA). The bovine serum albumin (BSA) standard for Bradford assays and the Bradford Reagent were purchased from Amresco Life Sciences, LLC (Cleveland, OH, USA). Enhanced chemiluminescence (ECL) reagents used for Western blot development were purchased from Pierce (Rockford, IL, USA), and x-ray film was from Denville scientific (Metuchen, NJ, USA). Nitrocellulose membranes were from GVS Life Sciences (Sanford, ME, USA). All additional reagents were purchased from Sigma (St. Louis, MO, USA) unless otherwise noted.

Plasmids and antibodies.

The bacterial expression plasmid encoding the fusion of glutathione-S-transferase (GST) with the CRKL-SH2 domain (GST-CRKL-SH2) was a gift from A. Imamoto (U. of Chicago). The bacterial expression plasmid encoding GST-CRK-SH2 was acquired from Addgene (Cambridge, MA; plasmid #46418), and was originally constructed by Bruce Mayer (U. of Connecticut).

The α-tubulin (mouse) antibody (Cell Signaling Technology; Danvers, MA, USA) and α-phosphotyrosine (pY; mouse) antibody (EMD Millipore; Burlington, MA, USA) were diluted 1:1,000 in Tris-Buffered Saline (150 mM NaCl, 40 mM Tris, pH 7.5) containing 0.05% Tween 20 (TBS-T) with 1.5% BSA and 0.005% sodium azide for Western blotting. The secondary α-mouse-HRP antibody, diluted 1:10,000 in TBS-T for Western blotting, was obtained from Jackson Immunolabs (Westgrove, PA, USA).

Cell culture, stimulation, and cell lysis.

N2A cells were grown at 37 ˚C in 5% atmospheric CO_2 in DMEM with 5% FBS, 5% CCS, 50 U/mL penicillin, and 50 μg/mL streptomycin. Cells were either treated or not treated with 8.8 mM H_2O_2 for 15 minutes prior to lysis. Cells were placed on ice, washed with cold phosphate-buffered saline (PBS), then lysed in lysis buffer (25 mM Tris pH 7.2, 137 mM NaCl, 10% glycerol, 1% Igepal, 25 mM NaF, 10 mM $Na_2H_2P_2O_7$, 1mM Na_3VO_4, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 10 mg/mL each leupeptin and pepstatin-A). Lysates were centrifuged at 4 ˚C, and the supernatant was reserved at -20˚ C for pulldown assays. A Bradford assay was used to normalize protein levels in the cell extract.

GST-CRKL-SH2 pull-down assay, Western blotting, and SDS-PAGE.

Purification of GST, and the GST-CRK-SH2 and GST-CRKL-SH2 fusion proteins on glutathione resin was previously described.[10] N2A whole cell extracts (WCEs) used in pulldown assays for LC-MS/MS analysis were pre-cleared on a column packed with GST-bound glutathione resin. N2A extracts (1 mg total protein for Western blotting, 2-3 mg for LC-MS/MS analysis) were incubated with GST-CRK-SH2 or GST-CRKL-SH2 resin for 1-2 days at 4 ˚C, rocking. The resin was then washed three times with lysis buffer. Bound proteins were eluted and denatured in sample buffer (150 mm Tris pH 6.8, 2% SDS, 5% β-mercaptoethanol, 7.8% glycerol, 0.01% bromophenol blue) at 95 ˚C for 5 min. Denatured pulldowns and WCEs (15 μg per lane) were separated on a 10% acrylamide gel (30% w/v and 37.5:1 acrylamide:bis-acrylamide) with 4.2% acrylamide stacking gels. The current was kept at 20 mA per gel through the stacking layers and 30 mA through the separating gels. Gels were either transferred to nitrocellulose membranes for Western blot analysis or stained with Coomassie Brilliant Blue in preparation for mass spectrometry.

For Western blotting, total protein levels on transfers were determined by staining membranes with a reversible Ponceau stain (0.5% Ponceau and 1% acetic acid in H_2O). Membranes were then blocked with TBS-T containing 5% w/v non-fat dry milk and incubated in primary antibodies overnight at 4 ˚C. Membranes were then washed with TBS-T three times (ten minutes each wash) and incubated overnight with the HRP-conjugated secondary antibody solution at 4 ˚C. After three final washes in TBS-T, blots were developed with ECL reagents and x-ray film.

Peptide preparation and analysis by LC-MS/MS.

Pulldowns for LC-MS/MS analysis were conducted in triplicate (complete biological replicates using separate cell cultures and conducted by separate personnel) each with GST-CRK-SH2 and GST-CRKL-SH2. Each lane of a given Coomassie-stained gel

was divided into four regions by molecular weight (**Figure 2A**), and then cut into 1-mm cubes. Gel pieces were de-stained in 50 mM ammonium bicarbonate and 50% acetonitrile at 37 ˚C, and dehydrated in 100% acetonitrile. Proteins were digested with trypsin in-gel at 37 ˚C for 18 hours. Tryptic peptides were extracted with 2.5% formic acid in 50% acetonitrile and further with 100% acetonitrile. Peptides were dried under vacuum centrifugation.

Peptides were re-suspended in Solvent A (2.5% acetonitrile, 2.5% formic acid) and separated via HPLC (300 nL/min) using the Easy n-LC 1200 prior to MS/MS analysis on the Q Exactive Plus mass spectrometer fitted with a Nanospray Flex ion source and supplied with Thermo Xcalibur 4.0 software (Thermo Fisher Scientific). Chromatography columns (15 cm x 100 μm) were packed in-house with 2.7 μm C18 packing material (Bruker, Halo, pore size = 90 Å). Peptides were eluted using a 0-50% gradient of Solvent B (80% acetonitrile, 0.15% formic acid) over 60 min and into the mass spectrometer by electrospray ionization. This gradient was followed by 10 min at 100% Solvent B before a 15-min equilibration in 100% Solvent A. The precursor scan (scan range = 360-1700 m/z, resolution = $7.0x10^4$, AGC = 10^6, maximum IT = 100 ms, lock mass=371.1012 m/z) was followed by ten collision-induced dissociation (CID) tandem mass spectra of the top ten ions in the precursor scan (resolution = 3.5×10^4, AGC = 5.0×10^4, maximum IT = 50 ms, isolation window = ±1.6 m/z, normalize collision energy = 26%, dynamic exclusion=30 s). Raw spectra were searched for matches within a forward and reverse mouse proteome (Uniprot, 2011) using SEQUEST,[11,12] requiring tryptic peptides and permitting the following differential modifications: phosphorylation of serine, threonine and tyrosine (+79.9663 Da), oxidation of methionine (+15.9949 Da), and acrylamidation of cysteine (+71.0371 Da). Peptides were filtered by mass accuracy (tolerance = ±4 ppm), cross correlation (XCorr) score (for z=+1, XCorr ≥ 1.8; z=+2, XCorr ≥ 2.0; z=+3, XCorr ≥ 2.2; z=+4, XCorr ≥ 2.4; z=+5, XCorr ≥ 2.6), and unique ΔCorr (≥ 0.15). Proteins were considered identified if 8 or more total spectra were mapped to peptides within that protein across the three replicates.

Bioinformatic analysis.
The number of YXXP motifs per protein in human and mouse sequences were obtained from Scansite (scansite4.mit.edu) as previously described.[9,13] The number of phosphorylated YXXP motifs in select proteins were obtained using a motif search in the Protein Search section of PhosphoSitePlus (phosphosite.org).[12] Clustal Omega was used to create multiple sequence alignments of select proteins across representative vertebrates.[15,16]

RESULTS AND DISCUSSION

CRK and CRKL are known to play important roles in several fundamental developmental processes; however, it is hypothesized that they serve additional roles in signaling pathways that have not yet been described. CRK and CRKL functional domains are highly conserved (**Figure 1A**, **Supplementary Figure 1**), although it remains possible that CRK and CRKL could possess distinct interacting partners, thereby carrying out divergent functions. SH2-mediated interactions between CRK/CRKL and phosphotyrosine residues are typically induced by specific signaling events involving the activation of receptor or non-receptor tyrosine kinases. One way to stimulate general tyrosine phosphorylation in cultured cells is the addition of hydrogen peroxide (H_2O_2) to cell culture medium, simultaneously inhibiting tyrosine phosphatases and activating tyrosine kinases through reactive oxygen species.[17] To identify neuronal SH2-specific binding partners of CRK and CRKL which are regulated by endogenous kinases and phosphatases we either stimulated with H_2O_2 or left untreated N2A cells. Extracts from these cultures were used to conduct a proteomics screen by incubating them with glutathione S-transferase (GST) fusions of the CRK- or CRKL-SH2 domains. Proteins that bound to GST-CRK-SH2 or GST-CRKL-SH2 were separated by SDS-PAGE and subjected to in-gel tryptic digestion. Peptides were subjected to LC-MS/MS leading to protein identification.

Figure 1B shows the successful generation of the fusion proteins via SDS-PAGE in IPTG-induced bacterial extracts and concentrated on glutathione resin. To verify that tyrosine phosphorylation could be induced in N2A cells by H_2O_2, N2A cells were either left untreated or were treated with 8.8 mM H_2O_2 for 15 minutes prior to lysis. N2A extracts were then subjected to SDS-PAGE and Western blotting with a phosphotyrosine antibody (α-pY). Indeed, cells that were treated with H_2O_2 showed a stronger signal in the α-pY blot compared to the untreated controls (**Figure 1C**). Tubulin levels (α-tubulin) served as a loading control. We then tested whether the GST-SH2 fusion constructs could extract proteins containing phosphotyrosine residues from N2A cells. Pulldown assays were conducted by incubating stimulated and unstimulated N2A extracts with GST-CRK-SH2 or GST-CRKL-SH2 resin prior to SDS-PAGE and immunoblotting (α-pY). Resin coated with GST alone was incubated with the same extracts as a control. H_2O_2-treated lanes demonstrate inducible binding of tyrosine phosphorylated proteins to CRK- and CRKL-SH2 fusion proteins, but not to GST (**Figure 1D**). Notably, several phosphotyrosine-containing proteins bound the CRK- and CRKL-SH2 domains in untreated N2A cells, suggesting that several YXXP-containing proteins were phosphorylated in N2A cells prior to H_2O_2 treatment.

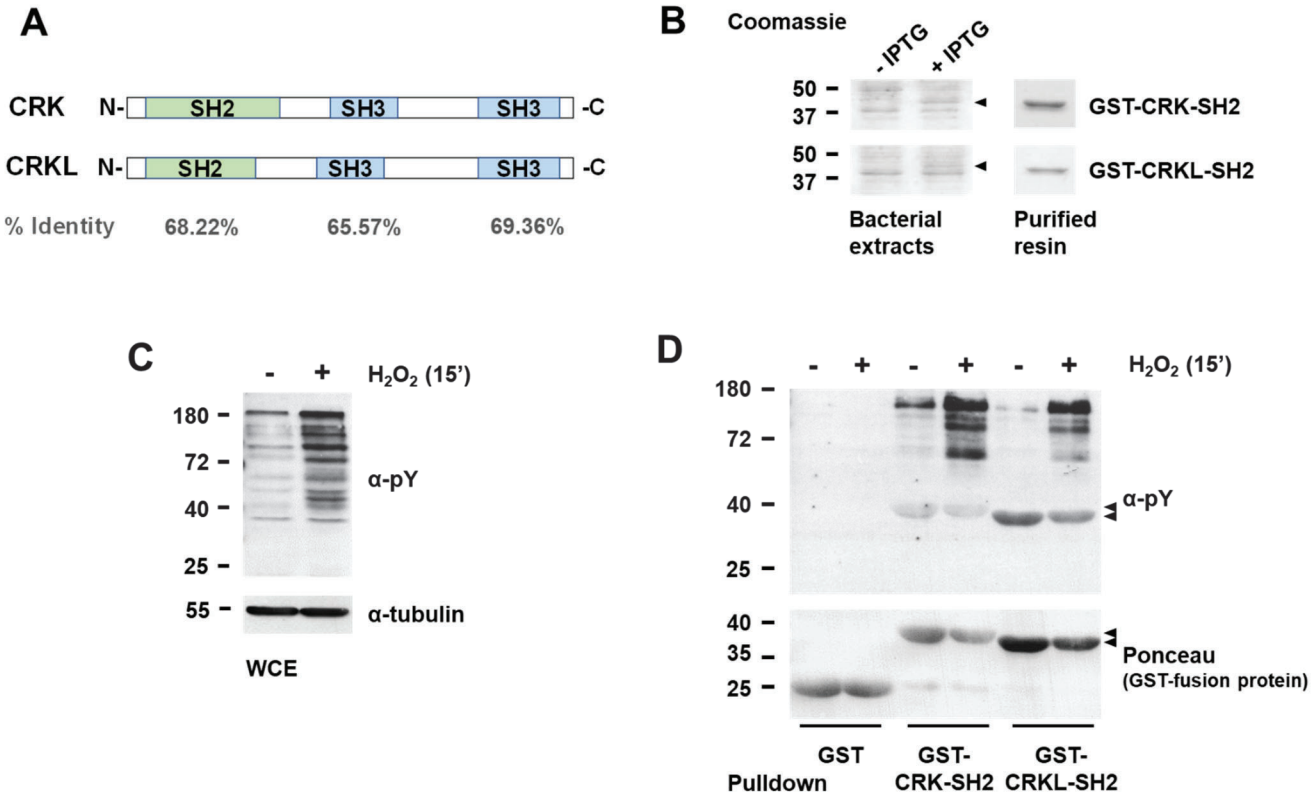

Figure 1. Generation and validation of GST-CRK-SH2 and GST-CRKL-SH2 for pulldown assays in N2A cells. A) Domains structure and conservation of CRK family members. A full alignment of amino acid sequences (ClustalOmega) can be found in **Supplementary Figure 1**. B) Bacterial lysate prior to and following IPTG induction of GST-fusion protein expression. Proteins were separated via SDS-PAGE and stained with Coomassie Brilliant Blue. GST-CRK/CRKL-SH2 expression is visible in the induced lysate (+ IPTG) at 40-45 kDa (indicated with arrows). After incubation with bacterial extracts, glutathione resin was subjected to SDS-PAGE and Coomassie staining, showing successful concentration of the purified GST-fusion protein on the resin. C) N2A cells were either left untreated or were treated with H_2O_2 (8.8 mM) for 15 min prior to lysis. Protein extracts were separated by SDS-PAGE and transferred to a nitrocellulose membrane for immunoblotting. The phosphotyrosine immunoblot (α-pY) shows an induction of tyrosine phosphorylation following H_2O_2 treatment. The α-tubulin blot serves as a loading control. D) Pulldown assays demonstrate functional SH2 domains. N2A extracts from (C) were incubated with GST-alone, GST-CRK-SH2 or GST-CRKL-SH2 prior to SDS-PAGE and Western blotting. The H_2O_2-treated extracts showed an increased phosphotyrosine signal relative to untreated extracts in GST-CRK/CRKL-SH2 pulldowns, while no phosphotyrosine signal was observed in the pulldowns with GST alone. The Ponceau stain shows GST-fusion protein levels around 40 kDa (SH2 fusions are indicated with arrows and background binding of α-pY is observed). GST alone runs at approximately 25 kDa.

To identify the CRK-/CRKL-SH2-interacting phosphoproteins observed in **Figure 1D**, GST-CRK-SH2 and GST-CRKL-SH2 pulldown assays were conducted to analyze via LC-MS/MS. For mass spectrometry assays, pulldowns were conducted as described above in triplicate. An example Coomassie-stained SDS-PAGE gel from one of the GST-CRK-SH2 replicates is shown in **Figure 2A**. Coomassie-stained gels were then cut into four regions by molecular weight (dashed lines in **Figure 2A**), subjected to an in-gel tryptic digest, and extracted peptides were analyzed by LC-MS/MS.

LC-MS/MS analysis of the GST-CRK-SH2 and GST-CRKL-SH2 pulldowns from unstimulated and stimulated N2A cell extracts yielded a total of 135 and 234 binding partners of CRKL-SH2 and CRK-SH2, respectively (**Supplementary Table 2**). Of these, 121 proteins were common interactors of CRK and CRKL, while each SH2 domain exhibited certain unique protein interactors in our study (**Figure 2B**). Spectral counts of proteins induced to bind the CRK- or CRKL-SH2 domain following H_2O_2 treatment (123 in total) are included in Supplementary **Table 3**. To date, no distinct interactor preferences between CRK family members have been described. However, further analysis will be required to determine whether the unique binding partners observed here truly are due to preferential differences between CRK family SH2 domains.

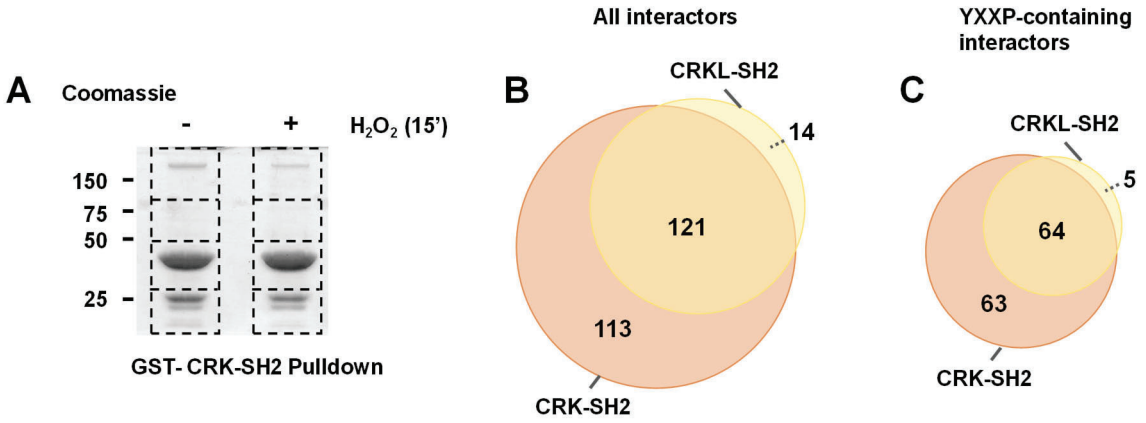

Figure 2. GST-CRK-/CRKL-SH2 pulldown assay for the identification of interactors via LC-MS/MS. A) The Coomassie-stained gel following a GST-CRK-SH2 pulldown from N2A extracts was divided into 4 regions of molecular weight. Each individual sample was subjected to an in-gel tryptic digest and analyzed via LC-MS/MS to identify bound proteins. N2A extracts were precleared with GST alone. RAW files were searched for tryptic matches in a 2011 Uniprot mouse proteome via SEQUEST. Proteins identified by eight or more peptides across three trials were considered SH2-specific interactors. Venn diagrams of all (B) and only YXXP-containing (C) CRK- and/or CRKL-SH2 interactors display the extent of overlapping binding partners identified in N2A cells. Full lists of interactors can be found in **Supplementary Tables 2–4**.

Of the 248 unique proteins identified to interact with the CRK- and/or CRKL-SH2 domain, 132 contain at least one YXXP motif (**Figure 2C**, **Supplementary Table 4**), 77 of which were induced to bind CRK-SH2, CRKL-SH2, or both CRK family members following H_2O_2 treatment (**Table 1**). Ten of the inducible YXXP-containing interactors found in **Table 1** are known to bind to CRK or CRKL (in bold, **Table 1**),[9] although only five of these have previously been shown to bind specifically to the CRK/CRKL-SH2 domain: DOK1, CBL, LASP1, PXN, and DCBLD2 (**Supplementary Table 1**). Importantly, additional known and novel YXXP-containing proteins were found to bind CRK- and/or CRKL-SH2, although they were not specific to the cells treated with H_2O_2 (**Supplementary Table 4**). For example, BCAR1 bound to the SH2 domains of both family members regardless of stimulation, suggesting that BCAR1 YXXP motifs may be highly phosphorylated in N2A cells.

Accession		Gene Symbol	CRK-SH2	CRKL-SH2	# YXXP motifs		# pYXXP IDs	Priority Score
H. sapiens	*M. musculus*				*H. sapiens*	*M. musculus*		*Schmoker, et al. (2018)*
Q99704	**P97465**	**Dok1****	**x**		**6**	**6**	**2886**	**11.5**
P49023	**Q8VI36**	**Pxn****	**x**		**3**	**3**	**2656**	**10.5**
P22681	**P22682**	**Cbl****	**x**		**3**	**3**	**761**	**10**
Q14847	**Q61792**	**Lasp1****	**x**	**x**	**3**	**3**	**517**	**9.5**
Q05397	**P34152**	**Ptk2***	**x**		**5**	**5**	**172**	**8**
Q96PD2	**Q91ZV3**	**Dcbld2****	**x**	**x**	**8**	**8**	**593**	**7.5**
O15357	**Q6P549**	**Inppl1***	**x**		**2**	**2**	**351**	**7**
P19174	Q62077	Plcg1	x		3	3	122	7
Q16555	O08553	Dpysl2	x		2	2	329	6.5
Q9UDY2	Q9Z0U1	Tjp2	x		2	3	943	6.5
Q07157	P39447	Tjp1	x		4	4	355	6.5
Q13813	P16546	Sptan1	x	x	5	5	35	6
O60716	P30999	Ctnnd1	x		3	2	1905	6
Q9UQC2	Q9Z1S8	Gab2	x		5	5	754	6
Q13835	P97350	Pkp1		x	3	3	456	6
Q13671	Q921Q7	Rin1	x		3	3	1074	4.5
P21333	**Q8BTM8**	**Flna***	**x**	**x**	**2**	**3**	**1**	**4**

O75369	**Q80X90**	**Flnb***	**x**	**x**	**4**	**4**	**60**	**4**
Q13191	**Q3TTA7**	**Cblb***	**x**		**3**	**3**	**239**	**4**
Q14203	O08788	Dctn1		x	1	1	0	3.5
P49327	P19096	Fasn	x	x	1	2	95	3.5
Q00610	Q68FD5	Cltc	x	x	5	5	18	3.5
P61160	P61161	Actr2		x	1	1	0	3
P61158	Q99JY9	Actr3		x	1	1	15	3
P52272	Q9D0E1	Hnrnpm		x	1	1	264	3
O43707	P57780	Actn4	x	x	1	1	0	3
P78371	P80314	Cct2	x	x	1	1	25	3
P48643	P80316	Cct5	x	x	1	1	1	3
Q99832	P80313	Cct7	x	x	1	1	0	3
P50990	P42932	Cct8	x	x	1	1	0	3
Q14195	Q62188	Dpysl3	x	x	1	1	3	3
P15924	E9Q557	Dsp	x	x	1	1	0	3
P46940	Q9JKF1	Iqgap1	x	x	1	1	0	3
P55884	Q8JZQ9	Eif3b	x	x	2	2	64	3
P12814	Q7TPR4	Actn1	x	x	3	3	5	3
P13639	P58252	Eef2	x	x	4	4	148	3
P23588	Q8BGD9	Eif4b	x		1	1	63	3
Q8TEW0	Q99NH2	Pard3	x		2	1	9	3
Q9Y490	P26039	Tln1	x		1	1	6	3
P67809	P62960	Ybx1	x		1	1	168	3
Q9Y285	Q8C0C7	Farsa	x		2	2	105	3
P48147	Q9QUR6	Prep	x		2	2	131	3
Q16832	Q62371	Ddr2	x		3	3	92	3
P55196	Q9QZQ1	Mllt4	x		3	4	21	3
P02545	P48678	Lmna		x	1	1	1	3
P61247	P97351	Rps3a		x	1	1	418	3
Q14204	Q9JHU4	Dync1h1		x	7	7	4	3
Q04637	Q6NZJ6	Eif4g1	x	x	2	2	19	0.5
P55072	Q01853	Vcp	x	x	3	3	10	0.5
P26640	Q9Z1Q9	Vars	x	x	5	5	24	0.5
Q14444	Q60865	Caprin1		x	1	1	0	0
P08243	Q61024	Asns	x	x	1	1	0	0
O95817	Q9JLV1	Bag3	x	x	1	1	11	0
Q14008	A2AGT5	Ckap5	x	x	1	1	0	0
P06733	P17182	Eno1	x	x	1	1	4	0
Q92598	Q61699	Hsph1	x	x	1	1	1	0
P11940	P29341	Pabpc1	x	x	1	1	22	0
P62701	P62702	Rps4x	x	x	1	1	0	0

Q8N5H7	Q9QZS8	Sh2d3c	x	x	1	1	1	0
Q13428	O08784	Tcof1	x	x	1	1	1	0
Q8WWM7	Q7TQH0	Atxn2l	x	x	2	2	52	0
P60842	P60843	Eif4a1	x	x	2	2	0	0
Q14240	P10630	Eif4a2	x	x	2	2	0	0
P35637	P56959	Fus	x	x	2	2	0	0
P26599	P17225	Ptbp1	x	x	2	2	1	0
P53396	Q91V92	Acly	x	x	3	3	2	0
A5YKK6	Q6ZQ08	Cnot1	x	x	6	6	8	0
P14868	Q922B2	Dars	x		1	1	1	0
P41250	Q9CZD3	Gars	x		1	1	0	0
P14625	P08113	Hsp90b1	x		1	1	1	0
O95347	Q8CG48	Smc2	x		1	1	0	0
P40939	Q8BMS1	Hadha	x		2	2	2	0
P34932	Q61316	Hspa4	x		2	2	0	0
Q9Y2A7	P28660	Nckap1	x		2	2	0	0
P54577	Q91WQ3	Yars	x		2	2	0	0
Q96F07	Q5SQX6	Cyfip2	x		3	3	0	0
Q9P2J5	Q8BMJ2	Lars	x		5	5	7	0
P62241	P62242	Rps8		x	1	1	18	0

Table 1. YXXP-containing proteins induced to bind CRK-/CRKL-SH2 in H_2O_2-stimulated N2A cells. The number of YXXP motifs (scansite4.mit.edu), the number of experimental identifications of phosphorylated tyrosine residues in YXXP motifs (phosphosite.org) and priority scores from Schmoker, *et al.* (2018) are included for each protein.[9] Known general CRK/CRKL interactors (*) and SH2-specific interactors (**) are indicated. Domain structures and multiple sequence alignments of select potentially novel SH2-specific interactors (underlined gene symbols) are included in **Supplementary Figure 2**. A full list of CRK-/CRKL-SH2 interactors is included in **Supplementary Table 4**.

Interactors in **Table 1** were cross-referenced with predicted CRK/CRKL-SH2 interactors.[7] Schmoker, *et al.* (2018) developed an *in silico* screen to predict and prioritize domain/motif interactions, using the CRK/CRKL-SH2/phospho-YXXP interaction as a model.[9] In that study, priority scores were calculated for each protein using signature characteristics of known interactors, including phospho-motif enrichment parameters, known interacting partners and participation in related signaling pathways.[9] Priority scores calculated by Schmoker, *et al.* (2018) are included for interactors in **Table 1**. Several high-priority novel CRK/CRKL-SH2 interactors were further investigated for their potential to interact with the CRK family of signaling adaptors. Select high-priority interactors involved in developmental processes, particularly those related to neuronal development and cell motility (underlined in **Table 1**), were assayed for conservation of their YXXP motifs across representative vertebrates. Domain structures and multiple sequence alignments of these novel CRK family SH2 interactors are included in (**Supplementary Figure 2**).

Several peptides containing phosphorylated tyrosine residues in YXXP motifs were identified in the pulldown assays. Fragmentation spectra of phosphopeptides were manually validated to confirm localization of the phosphate group on the YXXP tyrosine (**Supplementary Figure 3**). Several of the phosphorylated YXXP-containing peptides were from the known CRK-/CRKL-SH2 interactors BCAR1, DOK1, DCBLD2, and PXN (**Supplementary Table 5**). Phospho-YXXP-containing peptides of eight of the 16 BCAR1 YXXP motifs were identified in both stimulated and unstimulated N2A extracts. It is likely that additional proteins that remain highly phosphorylated in N2A cells, basally, could interact with CRK/CRKL-SH2 in this cell type regardless of H_2O_2 stimulation. Therefore, the list of interactors included in **Table 1** is likely not a complete list of potential phospho-YXXP-dependent CRK-/CRKL-SH2 interactors identified in this study. In addition to these known binding partners, CRMP1 was the only novel phospho-YXXP-containing protein found in complex with CRK- or CRKL-SH2 (**Supplementary Table 5**, **Supplementary Figure 3**). Interestingly, CRMP1 bound to CRK-SH2 regardless of stimulation (**Supplementary Table 4**). Even so, the identification of these phospho-YXXP-containing peptides lends strong evidence toward the interaction of the proteins containing these peptides with CRK/CRKL-SH2 through the identified phospho-YXXP sites (**Supplementary Table 5**).

CONCLUSIONS

In this study, we aimed to identify novel CRK/CRKL-SH2 binding partners that could play important roles in neurodevelopment. A proteomics screen produced 77 inducible CRK/CRKL-SH2 interactors in a neuronal cell line (N2A), with potential to interact specifically via their YXXP motifs (**Table 1**). Given the high conservation of the CRK family SH2 domains (**Figure 1A**, **Supplementary Figure 1**), and the absence of any known unique sequence preferences among these domains, we expected that many of the CRK/CRKL-SH2 interactors identified would overlap. However, we found that although approximately half of the interactors identified were common to CRK-SH2 and CRKL-SH2, each family member possessed a number of unique interactors in N2A cells (**Figure 2B,C**). Although the list of unique CRKL-SH2 interactors was smaller than that of CRK-SH2, this is likely a reflection of the fewer total number of CRKL-SH2 interactors identified (**Supplementary Table 2**). Indeed, a weaker induction of GST-CRKL-SH2 expression prior to purification on glutathione resin (**Figure 1B**) resulted in a lower concentration of fusion protein for use in pulldown assays, relative to GST-CRK-SH2. Despite this, five unique YXXP-containing proteins were identified as unique interactors of CRKL-SH2 (**Figure 2C**), namely, DYNC1H1, PKP1, LMNA, RPS3A, and RPS8. CRK and CRKL are known to play similar roles in developmental processes, however, certain distinctions have been made between CRK family members.[6] Although further studies will be necessary to determine whether the distinct CRK-SH2 and CRKL-SH2 interactors identified here truly are unique to each family member, our findings suggest that certain sequence preferences may exist for CRK family SH2 domains. In addition, several predicted CRK/CRKL-SH2 interactors from our previous *in silico* screen[9] were experimentally validated in this study, providing novel avenues of research regarding CRK family signaling in neuronal development.

ACKNOWLEDGEMENTS

Funding for this project came from the following sources: UVM College of Arts and Sciences and Biology Department; U.S. National Science Foundation IOS grant 1656510; the Vermont Genetics Network through U.S. National Institutes of Health grant 8P20GM103449 from the INBRE program of the NIGMS. The GST-CRK-SH2 and GST-CRKL-SH2 plasmid were kind gifts from Bruce Mayer (U. of Connecticut) and Akira Imamoto (U. of Chicago), respectively.

REFERENCES

1. Park, T.-J. and Curran, T. (2008) Crk and Crk-like play essential overlapping roles downstream of disabled-1 in the Reelin pathway. *Journal of Neuroscience.* 28, 13551-13562. *https://doi.org/10.1523/JNEUROSCI.4323-08.2008*
2. Guris, D. L., Fantes, J., Tara, D., Druker, B. J., & Imamoto, A. (2001). Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. *Nat Genet, 27*(3), 293-298. doi:10.1038/85855
3. Park, T. J., Boyd, K., & Curran, T. (2006). Cardiovascular and craniofacial defects in Crk-null mice. *Mol Cell Biol, 26*(16), 6272-6282. doi:10.1128/mcb.00472-06
4. Ballif, B. A., Arnaud, L., Arthur, W. T., Guris, D., Imamoto, A. and Cooper, J. A. (2004) Activation of a Dab1/CrkL/C3G/Rap1 pathway in Reelin-stimulated neurons. *Current biology.* 14, 606-610. *https://doi.org/10.1016/j.cub.2004.03.038*
5. Feller, S. M. (2001) Crk family adaptors–signaling complex formation and biological roles. *Oncogene.* 20, 6348. *https://doi.org/10.1038/sj.onc.1204779*
6. Kobashigawa, Y. and Inagaki, F. (2012) "Structural biology: CrkL is not Crk-like." *Nature chemical biology.* 8, 504. *https://doi.org/10.1155/2014/372901*
7. Ballif, B. A., Arnaud, L. and Cooper, J. A. (2003) Tyrosine phosphorylation of Disabled-1 is essential for Reelin-stimulated activation of Akt and Src family kinases. *Molecular brain research.* 117, 152-159. *https://doi.org/10.1016/S0169-328X(03)00295-X*
8. Aten, T. M., Redmond, M. M., Weaver, S. O., Love, C. C., Joy, R. M., Lapp, A. S., Rivera, O. D., Hinkle, K. L. and Ballif, B. A. (2013) Tyrosine phosphorylation of the orphan receptor ESDN/DCBLD2 serves as a scaffold for the signaling adaptor CrkL. *FEBS letters.* 587, 2313-2318. *https://doi.org/10.1016/j.febslet.2013.05.064*
9. Schmoker, A. M., Driscoll, H. E., Geiger, S. R., Vincent, J. J., Ebert, A. M. and Ballif, B. A. (2018) An *in silico* proteomics screen to predict and prioritize protein-protein interactions dependent on post-translationally modified motifs. *Bioinformatics.* 1, 9. *https://doi.org/10.1093/bioinformatics/bty434*
10. Schmoker, A. M., Weinert, J. L., Kellett, K. J., Johnson, H. E., Joy, R. M., Weir, M. E., Ebert, A. M. and Ballif, B. A. (2017) Dynamic multi-site phosphorylation by Fyn and Abl drives the interaction between CRKL and the novel scaffolding receptors DCBLD1 and DCBLD2. *Biochemical Journal.* 474, 3963-3984. *https://doi.org/10.1042/BCJ20170615*
11. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R. and Magrane, M. (2004) UniProt: the universal protein knowledgebase. *Nucleic acids research.* 32, D115-D119. *https://doi.org/10.1093/nar/gkh131*
12. Eng, J. K., McCormack, A. L. and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom.* 5, 976-989. *https://doi.org/10.1016/1044-0305(94)80016-2*
13. Obenauer, J. C., Cantley, L. C. and Yaffe, M. B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31, 3635-3641. *https://doi.org/10.1093/nar/gkg584*

14. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V. and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512-520. *https://doi.org/10.1093/nar/gku1267*

15. Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. and Lopez, R. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics.* 23, 2947-2948. *https://doi.org/10.1093/bioinformatics/btm404*

16. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. and Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology.* 7, 539. *https://doi.org/10.1038/msb.2011.75*

17. Zick, Y. and Sagi-Eisenberg, R. (1990) A combination of hydrogen peroxide and vanadate concomitantly stimulates protein tyrosine phosphorylation and polyphosphoinositide breakdown in different cell lines. *Biochemistry.* 29, 10240-10245. *https://doi.org/10.1021/bi00496a013*

ABOUT THE STUDENT AUTHORS

This work was conducted as part of a course-based undergraduate research experience (CURE) laboratory course at the University of Vermont (UVM) in the spring semester of 2019. The students in this course have since graduated from UVM with the following degrees: Caroline Dumas, B.S. in Neuroscience; Shannon Bennett, B.A. in Biology; Amara Chittenden, B.S. in Biological Sciences; Chelsea Darwin, B.S. in Biological Sciences; Helena Gaffney, B.A. in Biology; Hannah Lewis, B.S. in Biological Sciences; Eliana Moskovitz, B.S. in Biochemistry; Jonah Rehak, B.S. in Biological Sciences; Anna Renzi, B.A. in Anthropology and B.A. in Biology; Claire Rothfelder, B.A. in Biology; Adam Slamin, B.S. in Biological Sciences; Megan Tammaro, B.S. in Biological Sciences.

PRESS SUMMARY

The CRK family of adaptor proteins, composed of CRK and CRKL, play important roles in embryonic brain development. These adaptors facilitate protein-protein interactions by transporting signaling effectors to phosphorylated tyrosine residues near the cell membrane, forming protein complexes required for cell proliferation and migration. This work aimed to identify novel protein interactors of the phosphotyrosine binding domain of CRK family members in a neuronal cell line using mass spectrometry-based proteomics. The screen identified 6 known and 126 novel interactors of the CRK/CRKL phosphotyrosine binding domain, 77 of which were induced to bind in conditions with increased tyrosine phosphorylation. Although approximately half of these interactors were common between family members, several unique CRK or CRKL binding partners were confirmed across experimental replicates. These findings reveal an abundance of novel neuronal CRK/CRKL interactions that could be essential to neuronal development.