Supplementary Information for:

Redefining the specificity of phosphoinositide-binding by human PH domain-containing proteins

Nilmani Singh^{1†}, Adriana Reyes-Ordoñez^{1†}, Michael A. Compagnone¹, Jesus F. Moreno¹, Benjamin J. Leslie², Taekjip Ha^{2,3,4,5}, Jie Chen^{1*}

¹Department of Cell & Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ²Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205; ³Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218; ⁴Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205; ⁵Howard Hughes Medical Institute, Baltimore, MD 21205, USA

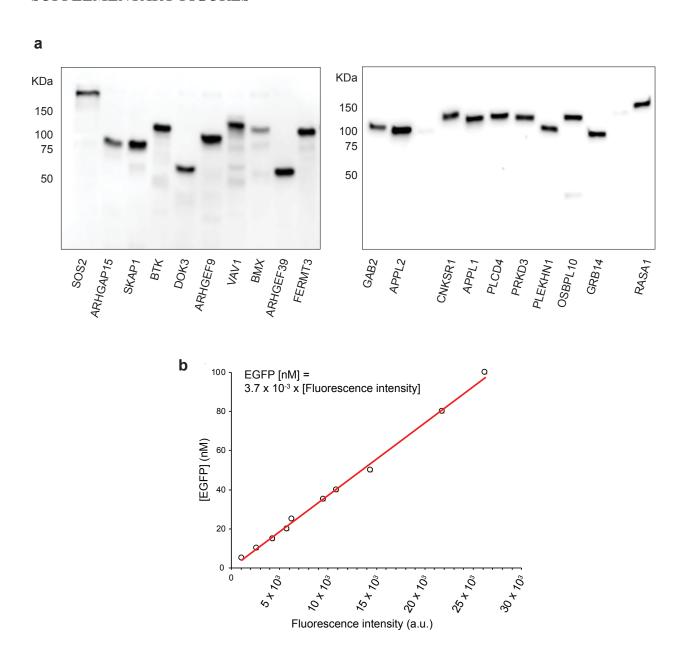
[†]These authors contributed equally to this work.

^{*}Correspondence: <u>jiechen@illinois.edu</u>.

PH Domain	Lipid binding by PH domain	Lipid binding by full-length protein
ACAP1_PH	PI(3,4)P ₂	PI(3,4)P ₂ , PI(4)P
ADAP1_PH1	None	PI(3,4,5)P ₃ , PI(4,5)P ₂ , PI(3,4)P ₂
ADAP1_PH2	None	PI(3,4,5)P ₃ , PI(4,5)P ₂ , PI(3,4)P ₂
ARHGEF3-PH	Various PIPs, see Figure 3	PI(4,5)P ₂ , PI(3,5)P ₂
ARHGEF5_PH	None	PI(4,5)P ₂ , PI(3,4)P ₂
ARHGEF9_PH	None	PI(5)P
DOK2-PH	PI(3,4,5)P3, PI(3)P, PI(4)P	PI(3)P
FERMT3-PH	None	PI(3,4,5)P ₃ , PI(3,4)P ₂
PLEKHA1-PH1	None	PI(3,4)P ₂
PLEKHA1-PH2	PI(3,4)P ₂	PI(3,4)P ₂
SPATA13-PH	None	Bound all lipid vesicles
VAV1-PH	None	PI(3,4,5)P ₃ , PI(4,5)P ₂ , PI(3,4)P ₂ , PI(3,5)P ₂ , PI(5)P

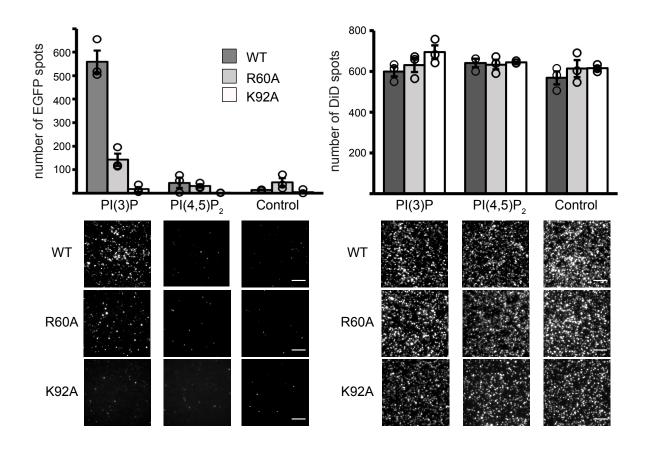
Supplementary Table 1. SiMPull assay results for selected PH domains. All fusion proteins had EGFP at the N-terminus except that ARHGEF3-PH was analyzed with both N-terminus and C-terminus fusions (see Figure 3 and Supplementary Figure S3). Experimental details and data processing were as described for Table 1. At least three independent experiments were performed with similar outcome for each PH domain.

SUPPLEMENTARY FIGURES



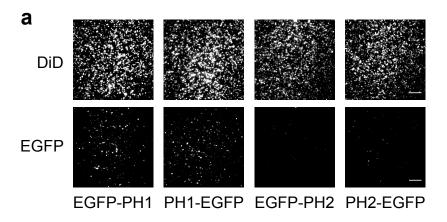
Supplementary Figure 1. Characterization of EGFP-fusion proteins for lipid-SiMPull assay.

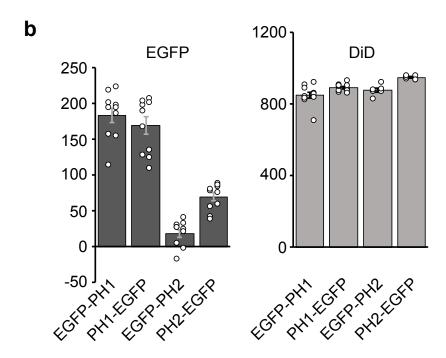
(a) Examples of EGFP-fusion proteins transiently expressed in HEK293 cells. Cell lysates were subjected to western blotting with an anti-GFP antibody. Each fusion protein was analyzed for expression in at least 3 independent experiments. (b) Fluorescence of pure recombinant EGFP (ex 488 / em 520) was measured at 10 different concentrations on a Spectramax GeminiXPS plate reader (Molecular Devices) to yield a linear standard curve. EGFP [nM] = $3.7 \times 10^{-3} \times \text{cm}$ [Fluorescence Intensity(a.u.)]. $R^2 = 0.9969$. Source data are provided as a Source Data file. The experiment was repeated 3 times with similar results.



Supplementary Figure 2. Determining sensitivity of lipid-SiMPull assay.

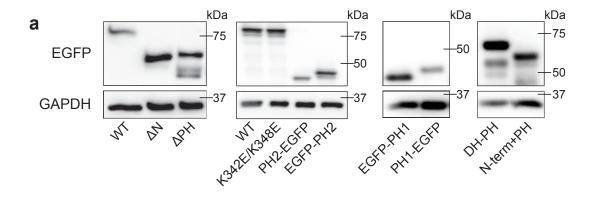
EGFP-p40PhoxPX and mutants were expressed in HEK293 cells, and cell lysates (5 nM EGFP-fusion) were subjected to SiMPull assay with PI(3)P, PI(4,5)P₂ and PC (control) vesicles. The numbers of EGFP and DiD (vesicle) spots were quantified as described in Figure 1 legend and presented in the graphs. Representative images are shown below the graphs. Data are presented as mean \pm SEM (n = 3 independent experiments). Source data are provided as a Source Data file. Scale bars: 5 μ m.

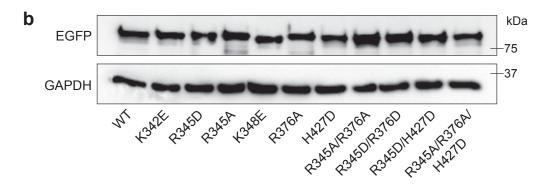




Supplementary Figure 3. Binding of ARHGEF3 PH domain to PI(4,5)P₂.

Various EGFP fusions of ARHGEF3-PH domain were expressed in HEK293 cells, and cell lysates (5 nM EGFP-fusion) were subjected to SiMPull assay with PI(4,5)P₂ vesicles. EGFP was fused at either terminus of the PH as indicated by the names of the constructs. PH1: aa320-455. PH2: aa299-466. (a) Representative images of EGFP and DiD. Scale bars: 5 μ m. (b) Numbers of fluorescent molecules per image area are shown. Data are presented as mean \pm SEM. N \geq 10 images for EGFP; N \geq 6 images for DiD. Source data are provided as a Source Data file, which lists the exact N number for each data point. Three independent experiments were performed with similar results.



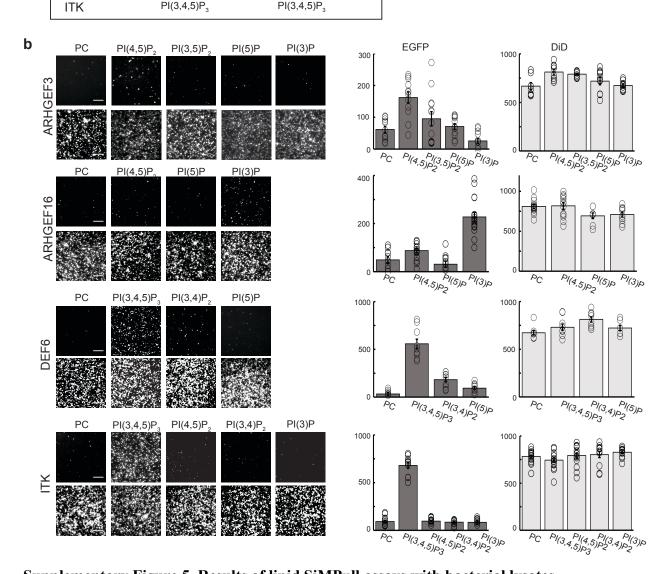


Supplementary Figure 4. Expression of ARHGEF3 truncations and mutants as EGFP fusion proteins.

HEK293 cells were transiently transfected to express various truncations (a) or point mutants (b) of ARHGEF3 fused to EGFP. Cell lysates were analyzed by western blotting with an anti-GFP antibody, and anti-GAPDH blotting served as loading control. EGFP was tagged at either terminus of the PH domain (EGFP-PH or PH-EGFP). All other constructs have the EGFP at their N-terminus. PH1: aa320-455. PH2: aa299-466. DH-PH: aa125-455. N-term+PH: aa1-125+aa320-455. ΔN: aa118-526. ΔPH: aa304-466 deleted. Uncropped and unprocessed western blot images are provided as a Source Data file. Each recombinant protein was analyzed for expression in at least 2 independent experiments.

а				
u	PROTEIN	LIPIDS BOUND WITH MAMMALIAN LYSATES	LIPIDS BOUND WITH BACTERIAL LYSATES	
	ARHGEF3	PI(4,5)P ₂ , PI(3,5)P ₂	PI(4,5)P ₂	
	ARHGEF16	PI(3)P	PI(3)P	
	DEF6	PI(3,4)P ₂	$PI(3,4,5)P_3, PI(3,4)P_2$	

PI(3,4,5)P₃



PI(3,4,5)P₃

Supplementary Figure 5. Results of lipid SiMPull assays with bacterial lysates.

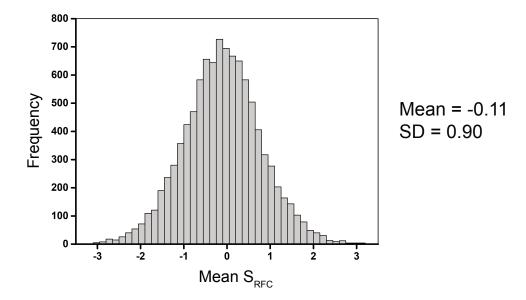
Bacterial lysates expressing SUMO-EGFP-fusion proteins (5 nM EGFP-fusion) were subjected to SiMPull assays. (a) The PH domain-containing proteins assayed and their PIP binding profiles in mammalian versus bacterial cell lysates. (b) Representative images of assays are shown. Scale bars: 5 µm. Numbers of fluorescent molecules per image area are shown at the right. Data are presented as mean \pm SEM. N \geq 10 images for EGFP; N \geq 6 images for DiD. Source data are provided as a Source Data file, which lists the exact N number for each data point. Three independent experiments were performed with similar results.

	β1	β1-β2 Variable loop	β2		β1	β1-β2 Variable loop	β2
ACAP1	HLFK	RASNAFKT	WSRRW	AFAP1-PH1	FLLR	K K R F G O W	TKLLC
ADAP1 PH1	LWKR	G R D N G Q	FLSRK	APPL1	YLNA	RNKTGLVSSTW	DROFY
ADAP1 PH2	YMEK	T G P K Q T E G F	RKRWF	APPL2	YLNL	RNKTGLVTTTW	ERLYF
AFAPIL1-PH1	FLLR	K K R F G Q W	AKQLT	ARHGAP25	LKKO	R S I V K N	WOORY
AFAP1L1-PH2	YLNV	L V N Q G W	KERWC	BMX	LLLK	RSOOKKKMSPNNY	KĒRLF
AKT1	LHKR	G E Y I K T	WRPRY	DOK1	PLFL	O S O R F G T K R W	RKTWA
ARHGAP12	LLNV	TKIAENGKKVRKN	WLSSW	DOK3	ILYO	O H V K F G K K C W	RKVWA
ARHGAP15	YLQK	AKIADGGKKLRKN	WSTSW	DOK4	YVKM	K S R K L G I Y	RRCWL
ARHGAP26	YLYV	Q E K R H F G T S	WVKHY	FERMT2	IKVF	KPKKLTLKGY	KQYWC
ARHGEF3	ELKN	N R G V	KLHVF	FGD5-PH2	YLSR	C K R G K R H W	KKLWF
ARHGEF5	ELTA	LEFSASPGLRRKLN	TRPVH	GRB14	LHAK	E O G K K S	WKKIY
ARHGEF7	QVLI	Q C A G S E E K	NERYL	GRB7	FLQL	RGSGRKLW	KRFFC
ARHGEF9	EMAW	I Y Q P Y G R	NQQRV	GRK2	SKMG	N P F L T	QWQRR
ARHGEF16	ELFL	VEETGLFRKIAS	RPTCY	KIF1B	YLHF	K E P L Y S N W	AKHFV
ARHGEF39	WLLV	V P P H G E	PRPRM	OSBPL8	WLKI	R G T L K S W	TKLWC
BTK	FLKR	SQQKKKTSPLN	FKKRL	PHETA2	MGFL		RTWGG
CNKSR1	WLLL	RKAPGGFMGPR	WRRRW	PLCG2	TVMT	V F S F R K S T P	ERRTV
DEF6	YLWK	R G H L R R N	WAERW	PLEK-PH1	LVKK	G S V F N T W	KPMWV
DOK2	FLYL	Q Q Q Q T F G K K W R	RFGAS	PLEK-PH2	LLKQ	G H R R R K N	WKVRK
EXOC8	DLVE	Y D A D H M A Q	LQRVH	PLEKHF2	VLTK	L C R K K P	KARQF
FERMT3	LRIF	RIPRRPRKLTLKGY	RQHWV	PLEKHN1-PH1	KVQL	R F Q H S Q D V	SDCYL
GAB1	WLRK	SPPEKKLKRYA	WKRRW	PLEKHN1-PH2	RVKL	Q H L P A Q E Q	WDRLL
GAB2	WLRK	SPPEKKLRRYA	WKKRW	PLEKHO2	WIKK	S S G G L L G F W	KDRYL
ITK	QLIK	KSQQKRRTSPSN	FKVRF	PRKD3	VHYT	S R D N L R	KRHYW
MCF2L	SFSV	WTDHKRGHTKVKELARFK	PMQRH	PSD4	ILAR	KMHQDADGKKKRGW	KMFHT
NGEF	ELQQ	MSGPKTSRTLRTKKL	FHEIY	RASA1	YLLK	K G K G K R W	KNLYF
OSBPL10	VLSK	Y T N L L Q G	WQNRY	RASGRF1-PH1	YLSK	R S S D N T K W	QTKWF
OSBPL5	SLKI	RGTLKS	WTKLW	SOS2	PLTR	I G A K	HERHI
PLCD4	PMRK	V R S K S W	KKLRY	STAP1	FLLI	K R S G Y R E Y	EHYWT
PLEKHA1	CVKQ	G A V M K N	WKRRY	(D:d ===	المصناطا	Dhaanhainaaitidaa in Linid CiM	DII)
PLEKHB2	LRQS	T I L K	RWKKN	(Dia na	ot bind i	Phosphoinositides in Lipid-SiM	ruii)
PLEKHF1	VLTK	E C R K K	AKPRI				
PLEKHJ1	ELGM	R G P K K G S V	LKRRL				
PSD2	VLTR	KTHADMDGKR-TPRGRRG	WKKFY				
SBF2	TLYK	R G A L L K G	WKPRW				
SKAP1	EKKS	KDHSFFGSEW	QKRWC				
/D		and a line of the line of CIMD.					

(Bound Phosphoinositides in Lipid-SiMPull)

Supplementary Figure 6. A reported sequence motif is present in the PH domains of both PIP-binding and non-binding PH domain-containing proteins.

Sequence alignment of PIP-binding (left) and non-binding (right) PH domains in the β 1- β 2 region. Bold: PH domains with reported crystal structures. Red: residues conforming to the $KX_n(K/R)XR$ motif. Three PH domains (AFAP1-PH2, FGD5-PH1 and RASGRF1-PH2) were not shown in the alignment due to an unusually short β 1 sheet.



Supplementary Figure 7. Determining a cut-off S_{RFC} score for PIP-binding prediction.

The sequences of 242 human PH domains were scrambled 10,000 times, and the S_{RFC} for each scrambled sequence was calculated. The distribution of mean S_{RFC} is shown, with the overall mean and standard deviation indicated. Source data are provided as a Source Data file.