Title
Functional characterization of the CNPY4 and PEAK3 proteins as regulators of cell signaling

Author
Lo, Megan

Publication Date
2020

Peer reviewed|Thesis/dissertation

Functional characterization of the CNPY4 and PEAK3 proteins as regulators of cell signaling

by
Megan Lo

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Natalia Jura
E438674A382B42F...

Natalia Jura
Chair

DocuSigned by:
Ophir Klein, MD, PhD.
[illegible]

Ophir Klein, MD, PhD.

DocuSigned by:
Frank Mccormick
[illegible]

Frank Mccormick

DocuSigned by:
Jason E. Gestwicki
[illegible]

Jason E. Gestwicki

DocuSigned by:
David Raleigh
7A381335D7C9402...

David Raleigh
Committee Members

Dedicated to my parents,

Leonard Yuan-Chang Lo and Yuri Fujimoto Lo

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors Natalia Jura and Ophir Klein for their guidance and support throughout my doctoral training. Natalia, thank you for your patience and for your trust; you encouraged me to explore whatever area of science my projects took me, no matter the investment of time or resources they required, and I am forever grateful for the scientific curiosity you imparted upon me. Ophir, thank you for your wisdom and for your steadiness as a mentor; you never gave up on these project or on my ability as a scientist, and I thank you for the bigger-picture perspective you always helped steer me towards. I would also like to thank all members of the Jura and Klein labs, past and present. I have had the privilege of overlapping with so many talented, kind, and amazing people during my time in these labs. Your kind words of advice helped me navigate challenging scientific problems, and your unwavering camaraderie encouraged me to never give up.

I am grateful to the UCSF community, in particular the members of my thesis committee: Frank McCormick, Jason Gestwicki, and David Raleigh. Thank you for taking the time to help guide me through graduate school and for imparting words of wisdom--both scientific and personal--upon me year after year. The mentorship I received from each and every one of you was invaluable to my growth as a member of the scientific community. I would also like to recognize the PSPG program for providing me with the opportunity to study at UCSF and for fostering an intellectually driven environment that promotes collaboration and cutting-edge research. I would especially like to thank the PSPG directors, Deanna Kroetz and Nadav Ahituv, and the PSPG administrators, Rebecca Dawson and Nicole Flowers. Your early faith in me allowed me to venture to labs unexplored by the PSPG program, and your support and help throughout my graduate career made for an incomparable learning experience at UCSF.

Thank you to all of my friends, from those I made at Caltech to those I made at UCSF and to everyone else in between. The late night dim sum/In-N-Out/McDonalds runs helped me make it through the long days (and nights), and the world of questionable reality television you introduced to me kept me from constantly fretting over my experiments. Even when I was too busy to think about anything other than my science, you took me out to Kpop concerts, hot pot parties, and bad karaoke sessions. You dragged me out to get some fresh air and much needed exercise on hikes that let me discover just how beautiful California is. I feel incredibly fortunate to have been surrounded by so many funny, kind, and thoughtful people with whom I could share my highest of highs and my lowest of lows. Thank you for all you have done to help mold me into the person I am today.

Thank you to my entire family, especially to my parents. You have always supported me in all my endeavors, no matter how crazy or strange they seemed, and for that, I am eternally grateful. You fostered my boundless curiosity and my love for science and experimenting from an early age, and you opened a countless number of doors to allow me to pursue my dreams. It is no exaggeration to say that I would not be here today if it was not for your constant love and support. Thank you for always believing in me.

Finally, I would like to thank Harris Chu, my boyfriend/fiancé/husband and, most importantly, my forever partner through the adventure that was graduate school and that is life. Thank you for spending your weekends, midnights, Christmases, and a literally countless number of hours of your life in lab with me. Without your relentless support and unwavering dedication, I would not have been able to achieve even a fraction of what I have accomplished. Thank you, and I love you.

Portions of this dissertation are reproductions of material previously published in:

Lo, M., Sharir, A., Agnew, C., Torosyan, H., Swaney, D.L., Krogan, N.J., Raleigh, D.R., Jura, N., Klein, O.D. (2020). "Fine-Tuning of the SHH Pathway by the Novel Regulator CNPY4." (in preparation)

Lopez, M.L.*, **Lo, M.***, Kung, J.E., Dudkiewicz, M., Jang, G.M., Vol Dollen, J., Johnson, J.R., Krogan, N.J., Pawlowski, K., Jura N. (2019). " PEAK3/C19orf35 pseudokinase, a new NFK3 kinase family member, inhibits CrkII through dimerization." *Proc Nat Acad Sci USA*. 116(31): 15495 - 15504. (*) denotes equal contribution.

Functional characterization of the CNPY4 and PEAK3

proteins as regulators of cell signaling

by

Megan Lo

ABSTRACT

Cells communicate with their environment through the transmission of external stimuli internally by a complex, interconnected network of signaling pathways. The magnitude, duration, and location of these pathways vary throughout the stages of development and adulthood to give rise to discrete features and to maintain homeostasis at the cellular, tissue, and organismal level. Through a countless number of studies, our knowledge of cellular signaling pathways has grown immensely in the past few decades. However, an area of study that remains largely understudied is the roles cellular regulators play in fine-tuning the output and dynamics of these pathways. Mounting evidence from studies on malignancies such as cancer, which take advantage of and hijack cellular signaling pathways, have illustrated the need to characterize such regulatory factors in addition to the pathway driving factors that have been the target of much research to date. Increasing difficulties in pharmaceutical development highlight the need to expand the range of druggable targets, and recent studies suggest that these proteins are well posed to become the next generation of drug targets. Here, we present the functional characterization of two understudied regulatory proteins: Canopy 4 (CNPY4) and PEAK3. We show that CNPY4 is a member of the saposin-like protein (SAPLIP) family that intersects the Sonic hedgehog (SHH) signaling pathway. Silencing of *Cnpy4* leads to severe hyperactivation of

the SHH pathway and alteration of accessible cholesterol levels in the cell membrane. These changes likely contribute to the severe congenital malformations observed in Cnpy4 *knockout* mice. We further demonstrate that PEAK3, a novel pseudokinase with no previously reported function, modulates cellular motility and migration through antagonization of the adaptor protein CrkII. PEAK3 regulation of CrkII is dependent on its ability to dimerize through a structural motif termed the split helical dimerization domain and the presence of an intact DFG motif. Furthermore, our initial studies suggest that PEAK3 antagonization of CrkII may itself be controlled by its interaction with the 14-3-3 family of regulatory proteins. Together, our studies provide the first functional evidence that CNPY4 is a regulator of the SHH pathway and that PEAK3 is a regulator of the CrkII pathway.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 3

LIST OF ABBREVIATIONS

CNPY	Canopy
CrkII	v-crk sarcoma virus CT10 oncogene homolog adaptor protein
E	Embryonic day
ER	Endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
GLI	Glioma-associated oncogene
HH	Hedgehog
MEF	Mouse embryonic fibroblast
NKF3	New kinase family 3
PEAK	Pseudopodium-enriched atypical kinase
PFO	Perfringolysin O
Ptc1	Patched 1 (gene)
PTCH1	Patched 1 (protein)
qRT-PCR	Quantitative real-time polymerase chain reaction
RTK	Receptor tyrosine kinase
SAG	Smoothened agonist
SANT1	Smoothened antagonist
SAPLIP	Saposin-like protein
SEC	Size exclusion chromatography

SgK	Sugen kinase
SHH	Sonic hedgehog
siRNA	Small interfering ribonucleic acid
SMO	Smoothened
SUFU	Suppressor of Fused

Chapter 1: Fine-tuning the sonic hedgehog pathway by Canopy 4

ABSTRACT

The Sonic hedgehog (SHH) pathway is critical for metazoan development and adult tissue homeostasis. Aberrant signaling through this pathway can lead to cancer and is a major cause of congenital malformations, such as digit anomalies and holoprosencephaly. Signal transduction is initiated by binding of the SHH ligand to the Patched 1 (PTCH1) receptor on primary cilia, which releases a series of inhibitory interactions, beginning with the receptor Smoothened (SMO). Multiple levels of control over the pathway have been identified, including ciliary lipid composition and downstream regulatory proteins. However, many of the cellular mechanisms underlying regulation of SHH signaling remain undefined. Here, we identify Canopy 4 (CNPY4), a putative Saposin-like protein, as a novel regulator of the SHH pathway. *Cnpy4*$^{-/-}$ mice exhibit multiple embryonic defects in line with perturbations of SHH signaling, most notably changes in digit number. Consistent with these observations, SHH signaling was upregulated in these embryos and in cells treated with *Cnpy4* siRNA. Epistatic analysis revealed that CNPY4 likely functions at the level of PTCH1 and SMO, and absence of CNPY4 alters membrane levels of accessible cholesterol, the putative ligand for SMO activation. Our data demonstrate that CNPY4 is a negative regulator that fine-tunes the initial steps of the SHH pathway, thus uncovering a previously unknown function for the poorly characterized CNPY family of proteins.

INTRODUCTION

The hedgehog (*Hh*) gene was first identified in *Drosophila* as a regulator of larval segmentation in 1980[1], after which three mammalian homologs were discovered: desert hedgehog (*Dhh*), Indian hedgehog (*Ihh*), and sonic hedgehog (*Shh*)[2-5]. Of the three, *Shh* is the most widely expressed and is found in the epithelium and epithelial-mesenchymal boundaries of various tissues, including the tooth, gut, lung, and limb, where it plays a central role in their morphogenesis[6]. Precise regulation of *Shh* signaling is therefore critical for proper development and patterning of these tissues. Even slight alterations to the pathway have been linked to a multitude of severe congenital abnormalities, including polydactyly and holoprosencephaly[7]. Perturbations and silencing of *Shh* pathway genes can also lead to the development of diseases, most notably cancers such as basal cell carcinoma, the most prevalent form of skin cancer, and medulloblastoma, the most common pediatric brain cancer[7,8].

HH activity is tightly controlled through multiple regulatory mechanisms at the cellular level. These include sequential inhibitory interactions that hold the pathway in an inactive state, as well as an indirect two-receptor activation mechanism[9]. Further regulation of the pathway occurs through the precise spatiotemporal localization of pathway components, which in vertebrates occurs at the primary cilium, an antenna-like organelle that protrudes off the surface of most cells[10,11]. Signaling is initiated by binding of the secreted HH ligand to the transmembrane receptor Patched 1 (PTCH1), which resides in and at the base of primary cilia[12-17]. HH binding to PTCH1 releases its inhibition on the G-protein coupled receptor Smoothened (SMO), allowing SMO to concentrate in cilia[17]. There, SMO is likely activated by a ligand, whose presence and identity remain unresolved[18]. Activation of SMO inhibits the partial proteolysis of the glioma-associated oncogene (GLI) family of transcription factors (GLI1, 2, and

3) into their repressor forms and promotes their dissociation from the protein Suppressor of Fused (SUFU)[19]. The full-length, active GLI proteins can then translocate into the nucleus and initiate transcription of key developmental genes[20]. This activation also upregulates transcription of pathway genes including *Ptc1* and *Gli1*, *2*, and *3*, leading to a complex signaling feedback loop[21,22].

An additional aspect of HH pathway regulation, which has only just begun to be fully appreciated, is the interaction between pathway components and lipids. Depletion of cellular lipids and inhibition of sterol biosynthesis either genetically or pharmacologically hinders HH signal transduction and can lead to severe developmental defects[23-29]. The precise molecular mechanisms underlying HH regulation, however, remain poorly linked to current evidence of lipid involvement in pathway modulation. The proteolyzed, N-terminal fragment of the HH ligand (HH-N), which constitutes the active segment, is covalently modified by two lipid moieties, cholesterol and palmitate[30,31]. These modifications serve to restrict the diffusion and solubility of the ligand to facilitate its secretion and membrane incorporation[32-39]. It was not until recently that the palmitate group was further shown to facilitate HH-N interaction with PTCH1 by docking into the groove between the two extracellular domains of PTCH1, suggesting an active role for this lipid moiety in receptor modulation[40]. While not covalently modified, SMO can be activated by binding of a number of lipid molecules, including cholesterol and a subset of its metabolites termed oxysterols[41-44]. Lipids further contribute to HH signaling through specialization of the ciliary membrane lipid composition[44,45]. This specialization likely facilitates proper trafficking of proteins into and out of the cilia, and it may play an important role in the ability of PTCH1 to efflux sterols to regulate SMO activation[45]. However, further research is required to determine if this is the case. Taken together, it becomes evident that there likely

remains undiscovered regulators of the pathway that could link these lipid-based observations to the cellular modulation of HH activity.

Saposin and saposin-like (SAPLIP) proteins comprise a large family that frequently functions via interactions with lipids and/or membranes to regulate (i) membrane binding, (ii) membrane lipid extraction, or (iii) membrane permeabilization[46]. These proteins are characterized by a structural motif termed the saposin-like fold, which consists of a compact, alpha-helical bundle held together by three, precisely-spaced disulfide bonds[46-48]. Despite this shared feature, SAPLIP proteins otherwise have low sequence identity (<10%) and possess a wide range of unique cellular functions[46-48]. Recently, an understudied SAPLIP family composed of four ER-resident proteins named Canopy (CNPY1-4) (**Fig. 1.1**) have been linked to a variety of cellular processes connected to HH signaling, including the regulation of fibroblast growth factor (FGF) signaling[49,50], the unfolded protein response[51], and lipoprotein receptors, of which PTCH1 is hypothesized to be a member[52,53]. Furthermore, cnpy1 regulates the development of the Kupffer's vesicle, an organ in zebrafish analogous to the mouse node, which controls left-right asymmetry through SHH and IHH signaling[50,54,55], suggesting that CNPY proteins may participate in the regulation of HH signaling.

RESULTS

We explored the role of CNPY proteins by breeding *Cnpy4* knockout mouse lines using embryos from Lexicon (http://www.lexicongenetics.com) (**Fig 1.2 A, B**), as previous characterizations on *Cnpy2* and *Cnpy3* knockout mice and *cnpy1* knockdown zebrafish have been previously reported[49-51,56]. Knockdown of *Cnpy4* resulted in a multitude of developmental

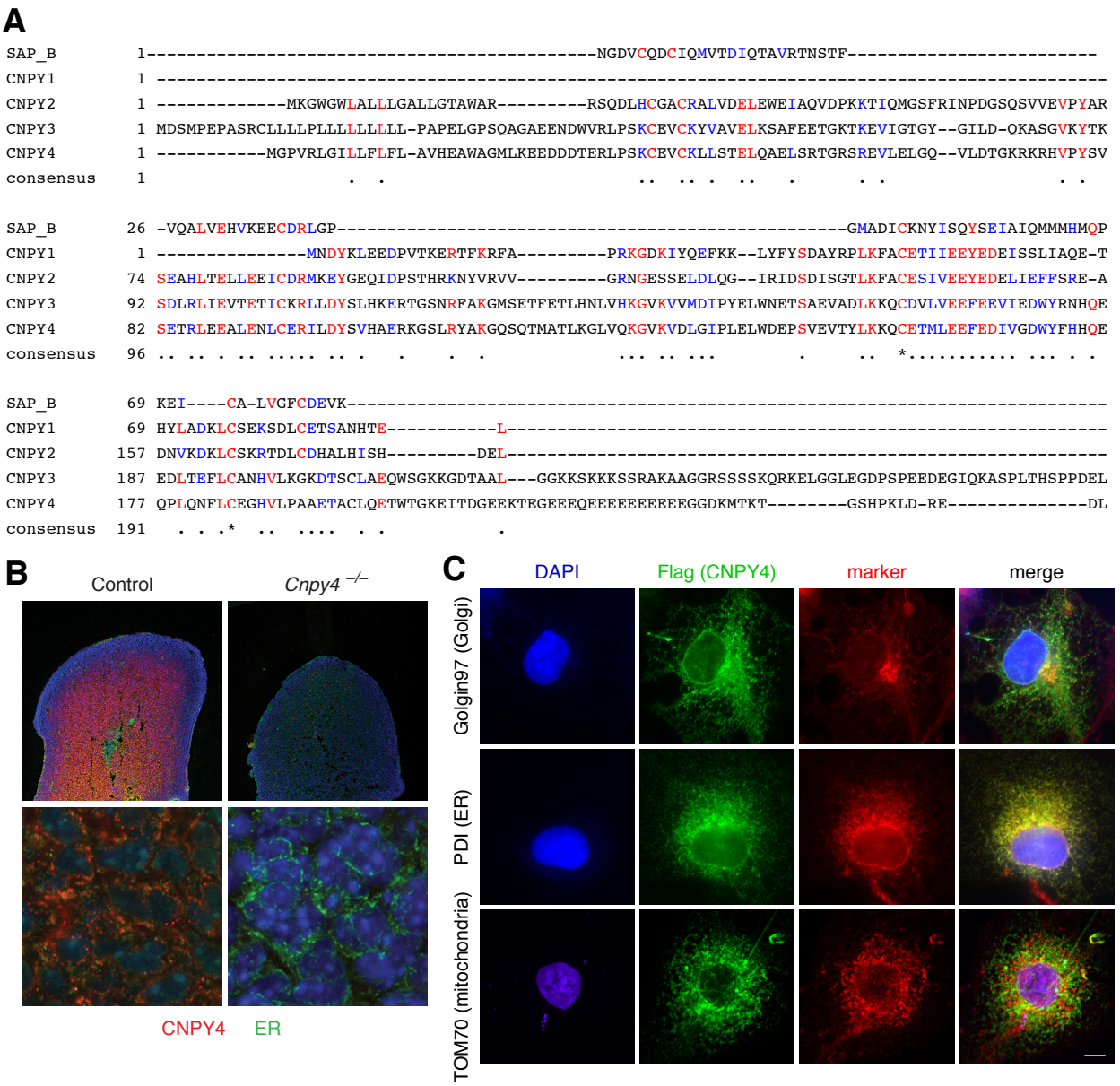


Figure 1.1. CNPY4 is an ER-resident member of the SAPLIP family. (A) Protein sequence alignment for human Saposin B, CNPY1, CNPY2, CNPY3, and CNPY4. Alignment and conservation was determined using SIM-Alignment tool (ExPASy). **(B)** Immunostaining of CNPY4 in control and *Cnpy4* null digits at E11.5. **(C)** Immunofluorescence of COS-7 cells transiently co-transfected with Flag-tagged CNPY4. CNPY4 was detected with an anti-Flag antibody (green), the nucleus with DAPI (blue), and organelle markers with the indicated antibody (red). Scale bar corresponds to 10 μm.

malformations (**Fig. 1.2 C - E**). Of the *Cnpy4*$^{-/-}$ embryos examined, 85% exhibited abnormalities in hindlimb digit number, ranging from formation of one or two supernumerary digits on the

anterior side of the limb (termed preaxial polydactyly) to a loss of up to three posterior digits (**Fig. 1.2 D, E**). Approximately 20% of the mutants exhibited other anomalies, including rostral and/or caudal neural tube closure defects, splayed vertebrae, and rib bifurcations (**Fig. 1.2 C**). Due to the high penetrance of the limb phenotype and the central role of *Shh* in digit number, we focused on the limb abnormalities.

Limb bud development is tightly regulated by the intricate interplay between a number of signaling networks, including SHH, FGF, bone morphogenetic protein, and WNT[57]. In particular, the FGF signaling pathway, which controls patterning of the proximal-distal axis, and the SHH signaling pathway, which regulates patterning of the anterior-posterior axis from the zone of polarizing activity, are essential for establishing and maintaining the morphogen signal gradient that gives rise to patterning in the limb bud[58]. Given the importance of SHH in anterior-posterior limb patterning and its known role in the development of other tissues that showed defects in *Cnpy4*$^{-/-}$ mice, such as the neural tube and rib[59,60], we first explored whether *Cnpy4* modulates the SHH pathway by examining the expression of *Shh* and its downstream effector *Gli1* during limb development. *Shh* and *Gli1* expression expanded anteriorly in early forelimb buds of *Cnpy4* mutants (embryonic day (E) 10.5 - E11.5), and ectopic *Shh* and *Gli1* expression was observed in anterior domains at later developmental stages (E12.5) (**Fig. 1.3 A**); interestingly, such changes are also observed in other mutants with preaxial polydactyly[61-63].

In order to assess the changes we observed in SHH signaling at the cellular level, we utilized a luciferase reporter assay to measure the levels of *Gli1* mRNA transcription in NIH3T3 cells in which *Cnpy4* was transiently knocked down by siRNA (**Fig. 1.3 B**). Knockdown of *Cnpy4* resulted in elevated levels of basal HH activity (**Fig. 1.3 C, D**) and potentiated signaling to levels nearly four times greater than that achieved using saturating amounts of ligand

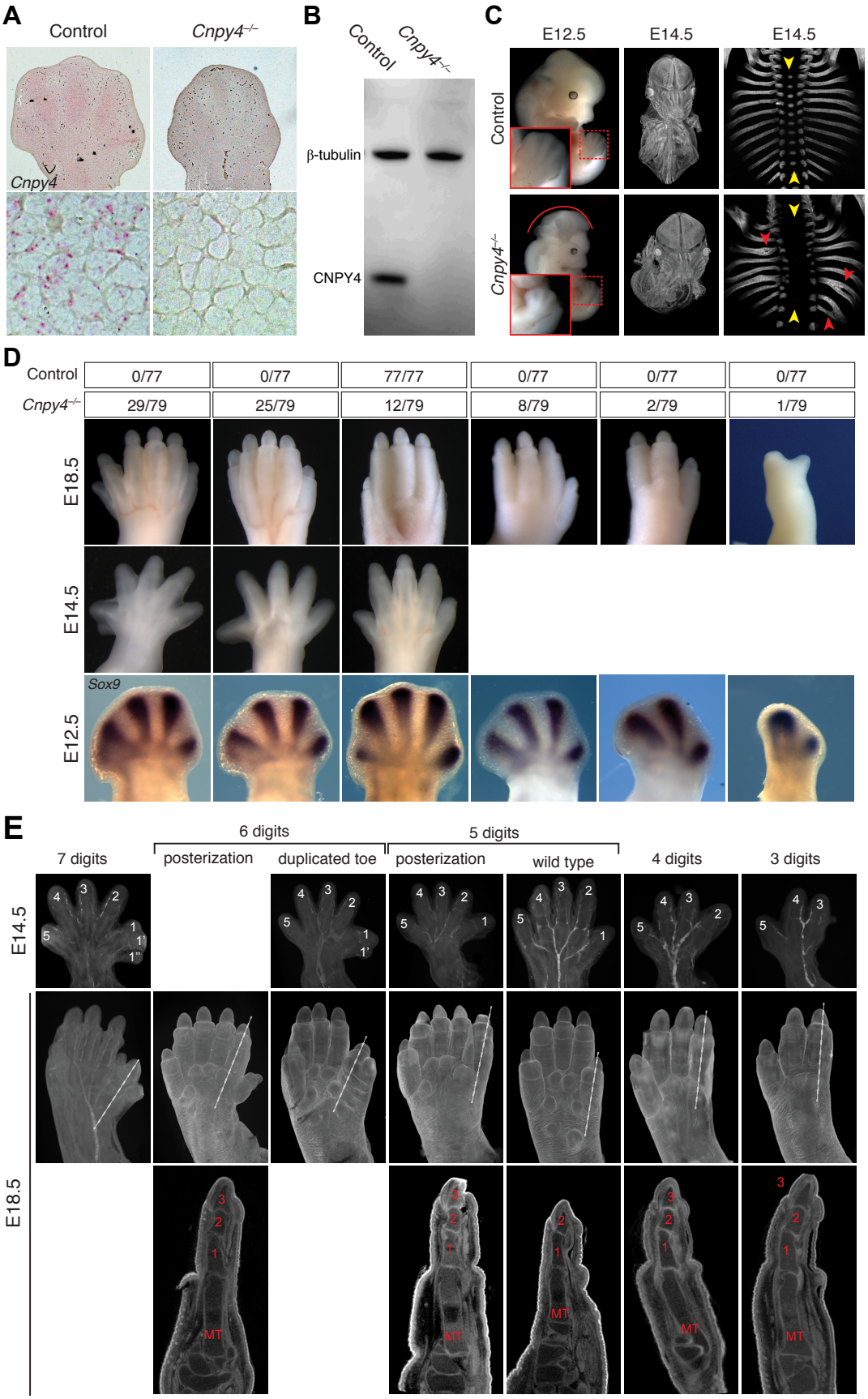

A
Control
Cnpy4−/−
Cnpy4
B
Control
Cnpy4−/−
β-tubulin
CNPY4
C
E12.5
E14.5
E14.5
Control
Cnpy4−/−
D
Control
Cnpy4−/−
0/77
0/77
77/77
0/77
0/77
0/77
29/79
25/79
12/79
8/79
2/79
1/79
E18.5
E14.5
E12.5
Sox9
E
7 digits
6 digits
posterization
duplicated toe
5 digits
posterization
wild type
4 digits
3 digits
E14.5
E18.5
MT

Figure 1.2. Developmental defects in a *Cnpy4* knockout mouse model. (A) *In situ* hybridization of *Cnpy4* in hindlimbs of wild-type embryos at E12.5. **(B)** Cells from control and mutant embryos were lysed and protein extracts run on a gel, which was blotted with anti-CNPY4 and β-tubulin loading control antibodies. **(C)** Whole mount pictures, micro computed tomography and skeleton preparations. (Left) cranial neural tube defect (exencephaly); (middle) shortening and kinking of the body axis; (right) abnormal rib morphology with fusions and bifurcations and truncation of the sternum. **(D, E)** Dorsal whole mount, whole mount *in situ* hybridization (D) and ventral mCT (E) of wild-type and *Cnpy4* mutant limbs at various embryonic stages.

alone (**Fig. 1.3 C, D**). This effect was independent of the ligand used to activate the pathway, including recombinant SHH that binds PTCH1 to activate HH signaling, the chemical SMO agonist (SAG), and several oxysterol compounds (20(S)-hydroxycholesterol and 24(S), 25-epoxycholesterol), which bind and activate SMO[44]. Direct analysis of *Gli1* transcript levels by qRT-PCR corroborated these findings (**Fig. 1.3 E - H**). The hyperactivity of the HH pathway in cells lacking CNPY4 is consistent with the developmental abnormalities observed in *Cnpy4* knockout animals and suggests that CNPY4 functions as a negative regulator of the HH pathway.

The hyperactivation we observed in HH signaling in cells with knockdown of *Cnpy4* was reminiscent of the effects seen in cells with alterations to their cilia[64-66]. Gross morphological differences in the cilia, changes in overall levels of cell ciliation, and improper trafficking of proteins into and out of the cilia are all linked to aberrant HH activity during development[10,11,45,64-66]. We therefore asked if defects in cilia number or morphology could explain the hyperactivation of the HH pathway observed by staining for acetylated tubulin, a marker of the cilium, in *Cnpy4* deficient NIH3T3 cells and in mouse embryonic fibroblasts (MEFs) derived from *Cnpy4*$^{-/-}$ embryonic limb buds (**Fig. 1.4 A - D**). However, both *Cnpy4* null NIH3T3 cells and *Cnpy4*$^{-/-}$ MEFs did not show significant differences in their ability to ciliate or in the number of cells ciliated compared to control cells (**Fig. 1.4 E, F**). Furthermore, the length and morphology of their cilia were not drastically altered (**Fig. 1.4 A, B, G, H**). Trafficking of

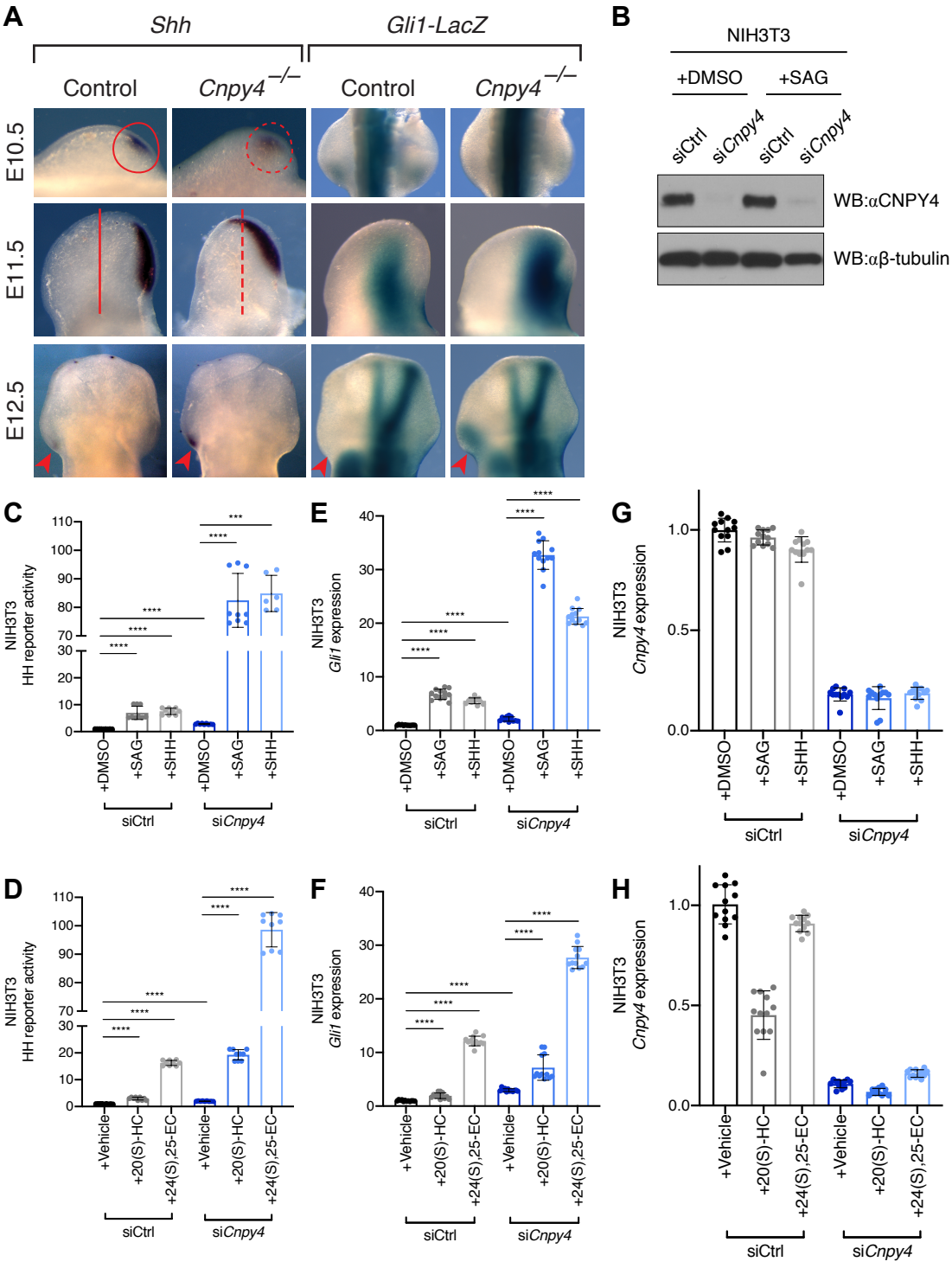


Figure 1.3. CNPY4 alters SHH pathway expression and signaling. (A) *In situ* hybridization of *Shh* and *Gli1-lacZ* expression in hindlimb buds showed that the *Shh* domain is enlarged (circles and lines) and ectopic expression of both *Shh* and *Gli1* (arrowheads) in *Cnpy4* mutants. **(B)** Western blot analysis of CNPY4 and β-tubulin loading control protein levels in NIH3T3 cells treated with *Cnpy4* or control siRNA. **(C, D)** Luciferase reporter assay in ciliated NIH3T3 cells treated with *Cnpy4* or control siRNA and stimulated with SAG or recombinant SHH (C) and

20(S)-hydroxycholesterol (HC) or 24(S), 25-epoxycholesterol (EC) (D). Quantifications were normalized to the average value of control siRNA treated cells stimulated with DMSO or vehicle. **(E - H)** qRT-PCR assessment of *Gli1* (E, F) or *Cnpy4* (G, H) expression in ciliated NIH3T3 cells treated with *Cnpy4* or control siRNA and stimulated with SAG or recombinant SHH (E, G) and 20(S)-hydroxycholesterol or 24(S), 25-epoxycholesterol (F, H). Significance calculations were performed as described in Methods and Materials with *** = $p < 0.0005$ and **** = $p < 0.0001$

SMO into the cilia upon stimulation with SAG also remained unimpaired in *Cnpy4* silenced cells (**Fig. 1.4 I**). Thus, the effect CNPY4 exerts on the HH pathway was likely due to mechanisms other than direct effects on cilia.

We therefore used our *in vitro* system to probe HH pathway genes for epistasis with *Cnpy4* to gain insight into how it may be modulating the pathway. Knockout of *Ptc1*, which at a basal state inhibits HH signal transduction, results in constitutive activation of HH signaling[17]. Remarkably, $Ptc1^{-/-}$ MEFs in which *Cnpy4* was additionally knocked-down displayed nearly four-fold elevated levels of HH signaling activity compared to control treated cells (**Fig. 1.5 A - D**). Similar potentiation was observed in NIH3T3 cells treated with *Cnpy4* siRNA compared to control cells upon ligand stimulation (**Fig. 1.3 C - F**), suggesting that CNPY4 likely intersects the HH pathway either downstream or at the level of PTCH1. Knockout of *Sufu*, a negative regulator of the pathway downstream of PTCH1, also leads to elevated basal activity of the HH pathway (**Fig. 1.5 E - H**). However, in contrast to the effect of *Cnpy4* knockdown in $Ptc1^{-/-}$ MEFs, silencing of *Cnpy4* in $Sufu^{-/-}$ MEFs resulted in a comparatively modest increase of *Gli1* mRNA transcription, indicating that CNPY4 was likely operating upstream of SUFU.

SMO is situated in the HH pathway downstream from PTCH1 and upstream from SUFU[9]. As SMO is required for HH signal transduction, $Smo^{-/-}$ MEFs are unable to transduce HH activity even in the presence of stimuli. We therefore examined whether CNPY4-mediated

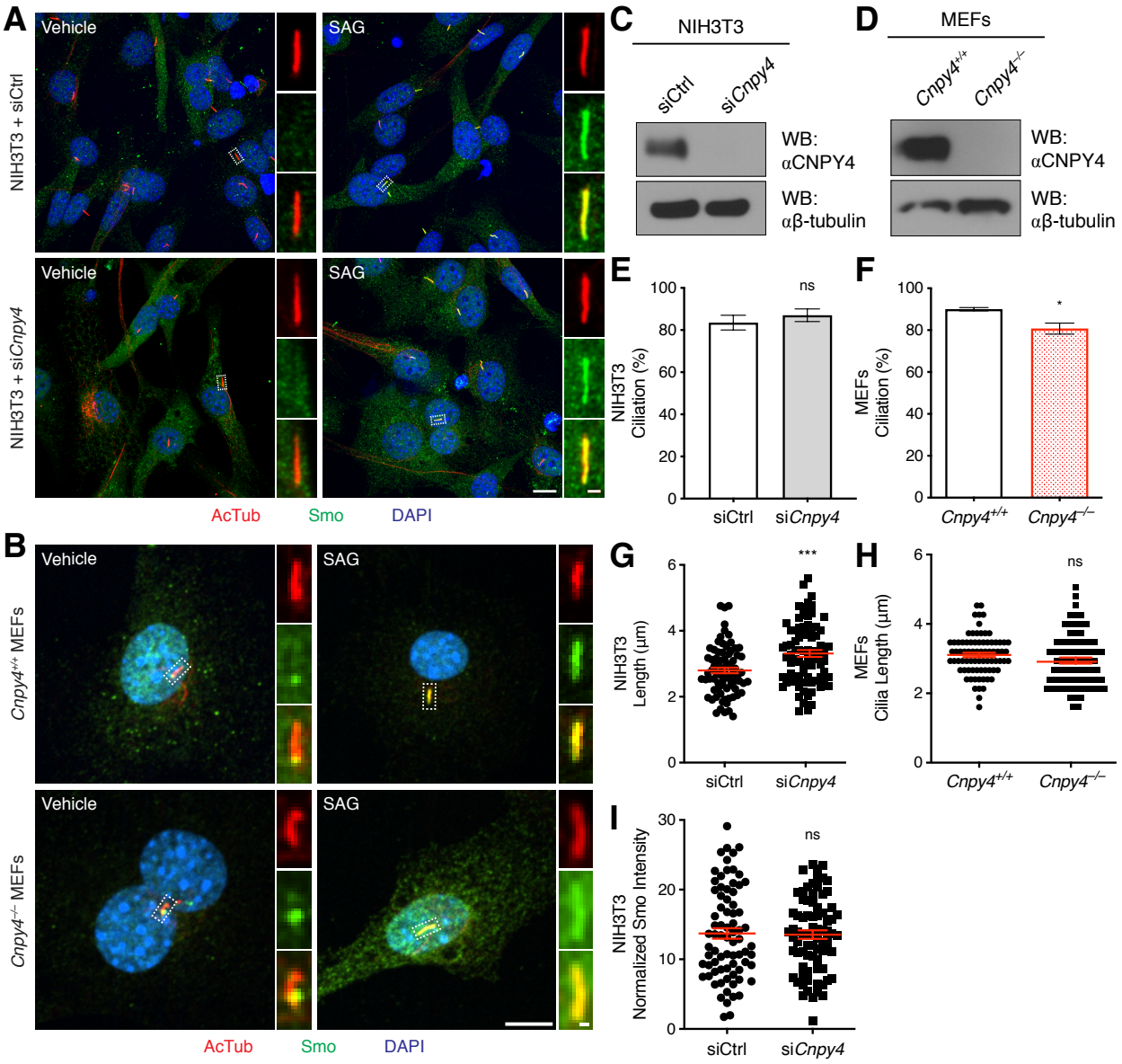


Figure 1.4. Loss of CNPY4 has little effect on the primary cilia. (A, B) Immunofluorescence of primary cilia (acetylated tubulin, red), SMO (SMO, green), and the nuclei (DAPI, blue) in ciliated NIH3T3 cells treated with *Cnpy4* or control siRNA (A) or ciliated control and *Cnpy4* null MEFs (B) treated with DMSO (vehicle) or SAG. The scale bar represents 10 µm. Inset scale bar represents 1 µm. **(C, D)** Western blot analysis of CNPY4 and β-tubulin loading control protein levels in NIH3T3 cells treated with *Cnpy4* or control siRNA (C) or in control and *Cnpy4* null MEFs (D) used for immunofluorescence. **(E, F)** Quantification of the number of NIH3T3 (E) and MEF (F) cells ciliated as assessed by acetylated tubulin immunofluorescence. **(G, H)** Quantification of the length of cilia in ciliated NIH3T3 (G) and MEF (H) cells, as measured in FIJI using the acetylated tubulin channel. **(I)** Quantification of SMO trafficking into cilia in NIH3T3 cells. Analyses were done in FIJI by measuring fluoresence intensity of SMO over the length of the cilia in the appropriate channel and subtracting background fluorescence measured over the same length. All significance calculations were performed as described in Methods and Materials with ns = $p > 0.05$, * = $p < 0.05$, and *** = $p < 0.0005$.

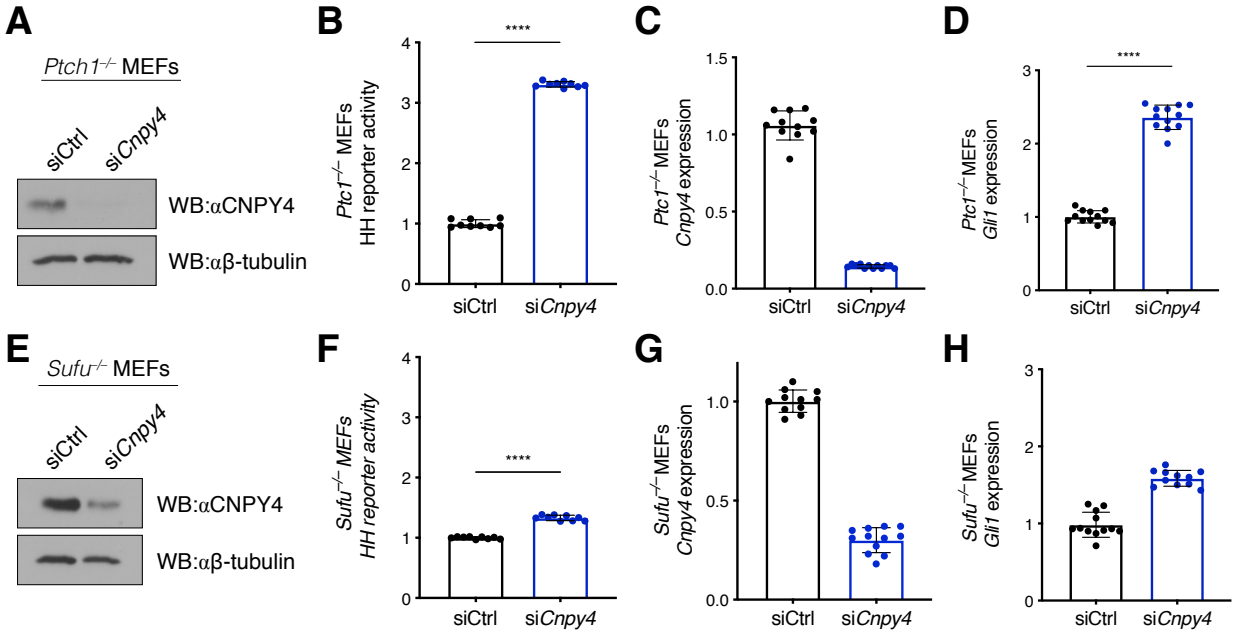


Figure 1.5. *Cnpy4* epistatically interacts with the *Shh*-dependent gene *Ptc1*. (A) Western blot analysis of CNPY4 and β-tubulin loading control proteins levels in *Ptc1*$^{-/-}$ MEF cells with *Cnpy4* or control siRNA treatment. **(B)** Luciferase reporter assay in ciliated *Ptc1*$^{-/-}$ MEFs treated with *Cnpy4* or control siRNA. Quantifications were normalized to the average value of control siRNA treated cells. **(C, D)** qRT-PCR assessment of *Cnpy4* (C) and *Gli1* (D) expression in *Ptc1*$^{-/-}$ MEFs. **(E)** Western blot analysis of CNPY4 and β-tubulin loading control proteins levels in *Sufu*$^{-/-}$ MEFs with *Cnpy4* or control siRNA treatment. **(F)** Luciferase reporter assay in ciliated *Sufu*$^{-/-}$ MEFs treated with *Cnpy4* or control siRNA. Quantifications were normalized to the average value of control siRNA treated cells. **(G, H)** qRT-PCR assessment of *Cnpy4* (G) and *Gli1* (H) expression in *Sufu*$^{-/-}$ MEFs. All significance calculations were performed as described in Methods and Materials with **** = $p < 0.0001$.

potentiation of HH signaling is dependent upon SMO through genetic (**Fig. 1.6 A - C**) and pharmacological (**Fig. 1.6 D - G**) manipulation. While knockdown of *Cnpy4* in *Smo*$^{-/-}$ MEFs elevated basal levels of HH signaling to a modest extent (**Fig. 1.6 A, C**), similar to levels we observed in unstimulated, wild type cells (**Fig. 1.3 C - F**), we did not observe further potentiation of HH signaling upon SAG stimulation. The same effects were seen when SMO was pharmacologically inhibited by its antagonist SANT-1, which directly competes with SAG for binding to SMO, in NIH3T3 cells (**Fig. 1.6 D, F**). Together these findings point to an essential

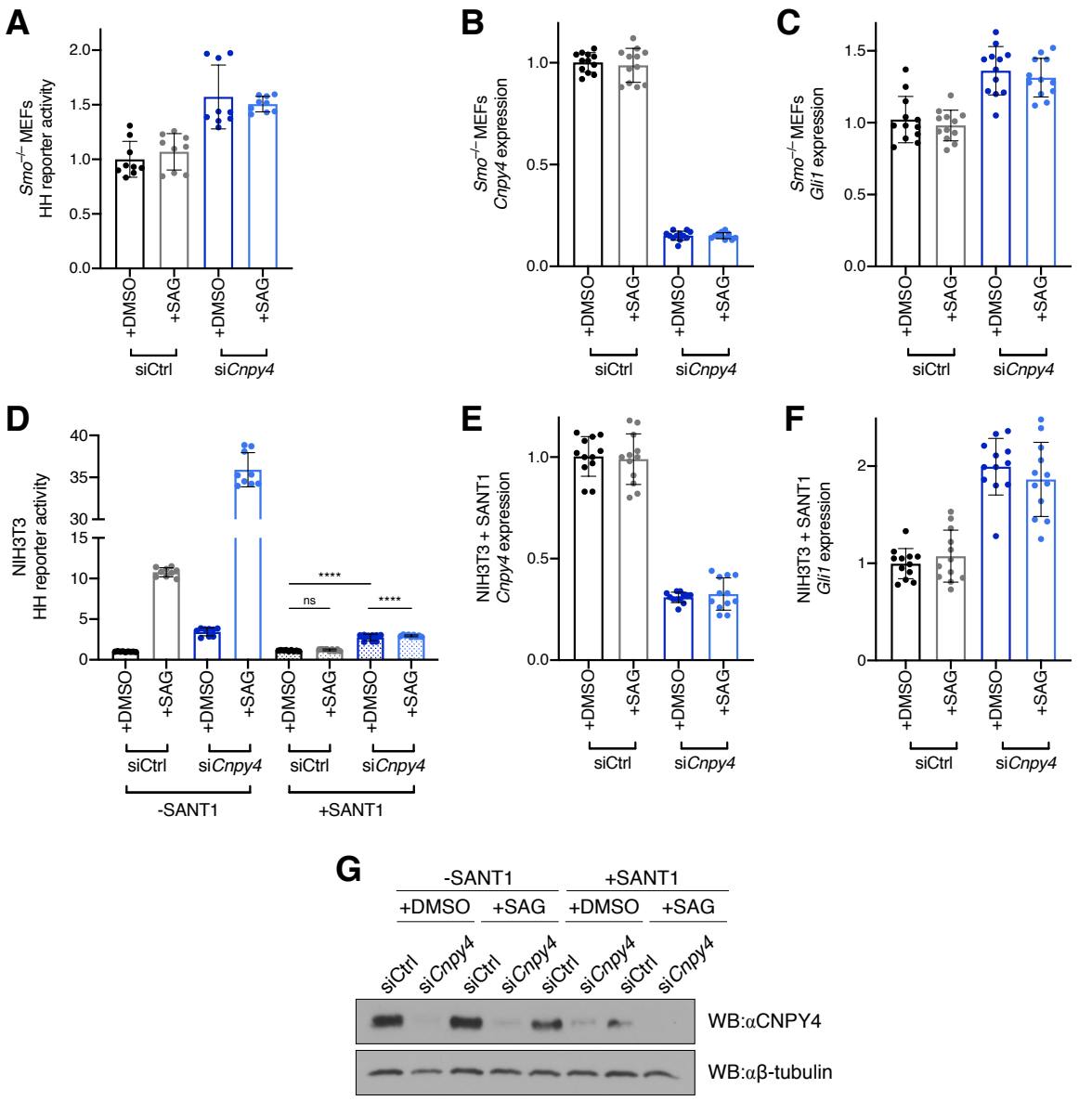
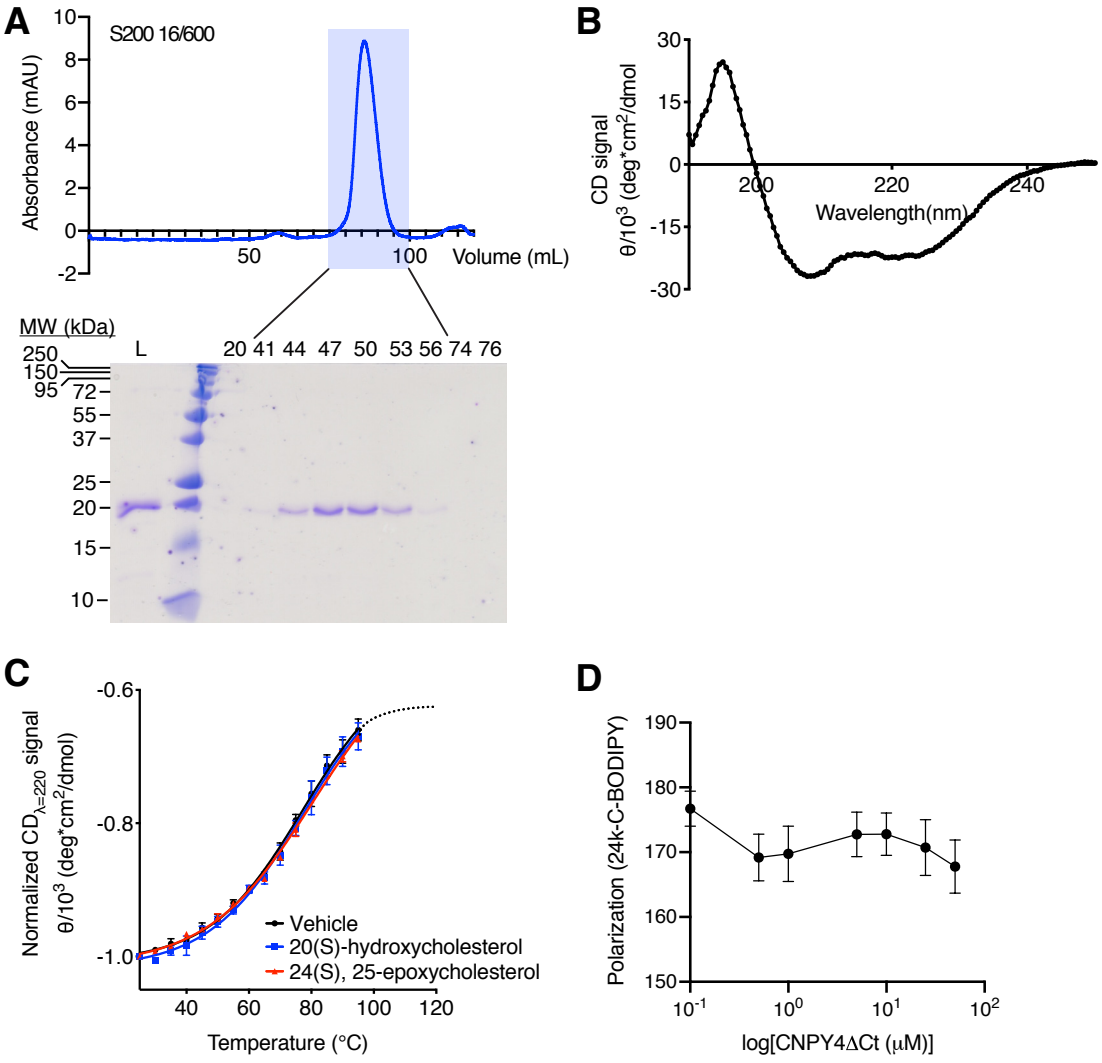
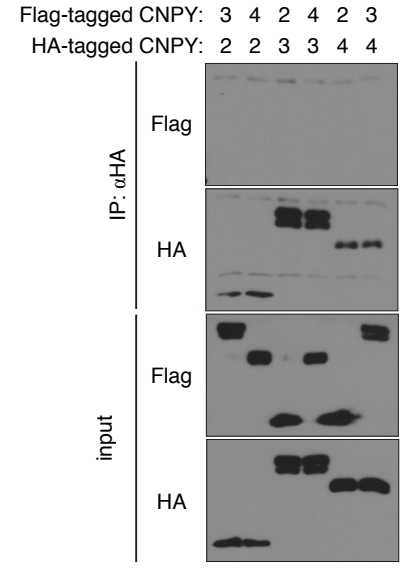


Figure 1.6. SHH signal regulation by CNPY4 requires SMO. (A) Luciferase reporter assay in ciliated *Smo*$^{-/-}$ MEF cells treated with *Cnpy4* or control siRNA and stimulated with SAG. Quantifications were normalized to the average value of control siRNA treated cells stimulated with DMSO. **(B, C)** qRT-PCR assessment of *Cnpy4* (B) or *Gli1* (C) expression in ciliated *Smo*$^{-/-}$ MEFs treated with *Cnpy4* or control siRNA and stimulated with SAG. **(D)** Luciferase reporter assay in ciliated NIH3T3 cells treated with *Cnpy4* or control siRNA and SANT-1 and stimulated with SAG. Quantifications were normalized to the average value of control siRNA treated cells stimulated with DMSO. **(E, F)** qRT-PCR assessment of *Cnpy4* (E) or *Gli1* (F) expression in ciliated NIH3T3 cells treated with *Cnpy4* or control siRNA and SANT-1 and stimulated with SAG. All significance calculations were performed as described in Methods and Materials with **** = $p < 0.0001$. **(G)** Western blot analysis of CNPY4 and β-tubulin loading control proteins levels in NIH3T3 cells treated *Cnpy4* or control siRNA with SANT-1 and SAG treatment.

role of SMO in the SAG-dependent potentiating effect of CNPY4 loss on HH signaling. However, the reproducible increase in basal HH signaling we consistently observed upon silencing of *Cnpy4*, independent of the genetic background, suggested that CNPY4 may act upon the pathway in a pleiotropic manner. We hypothesized that rather than regulating a specific node, CNPY4 could more broadly regulate upstream components of the pathway. As SMO and PTCH1 are both transmembrane receptors, we postulated that CNPY4, as a SAPLIP protein, may modulate the lipid environment in which both of these receptors are located.

The ciliary membrane in which PTCH1 and SMO reside is highly enriched in cholesterol precursors and oxysterols in comparison to other membrane compartments[44]. We therefore probed the ability of CNPY4 to interact with several of these oxysterol compounds *in vitro.* We purified a recombinant construct of human CNPY4 lacking its signal sequence and C-terminal tail (CNPY4ΔCt), which is predicted to be largely unstructured, from T7 SHuffle *E. coli* cells (**Fig. 1.7 A**). These cells are engineered to facilitate the formation of disulfide bonds[67], and we confirmed proper folding of the protein by circular dichroism. Purified CNPY4ΔCt is well-folded and predominantly alpha helical, as expected for a SAPLIP protein (**Fig. 1.7 B**). However, recombinant CNPY4ΔCt did not display measurable binding to a number of oxysterols (24-ketocholesterol, 24(S), 25-epoxycholesterol, and 20(S)-hydroxycholesterol) known to be specifically enriched in the ciliary membrane and directly involved in HH pathway activation[44] (**Fig. 1.7 C, D**). As the ability of most SAPLIP proteins to interact with lipids is directly tied to their oligomeric state[46,47,68-70], we tested if recombinant CNPY4ΔCt is a dimer. Size exclusion chromatography was consistent with CNPY4ΔCt being monomeric (**Fig. 1.7 A**), in contrast to the majority of other structurally characterized SAPLIP proteins[46-48,68-72]. As SAPLIP homodimerization is driven through the saposin-fold, we did not expect that the inability of


A
S200 16/600
Absorbance (mAU)
Volume (mL)
MW (kDa)
L
20 41 44 47 50 53 56 74 76
B
CD signal
θ/10³ (deg*cm²/dmol
Wavelength(nm)
C
Normalized CD λ=220 signal
θ/10³ (deg*cm²/dmol)
Vehicle
20(S)-hydroxycholesterol
24(S), 25-epoxycholesterol
Temperature (°C)
D
Polarization (24k-C-BODIPY)
log[CNPY4ΔCt (μM)]


E
Flag-tagged CNPY: 3 4 2 4 2 3
HA-tagged CNPY: 2 2 3 3 4 4
IP: αHA
Flag
HA
input
Flag
HA

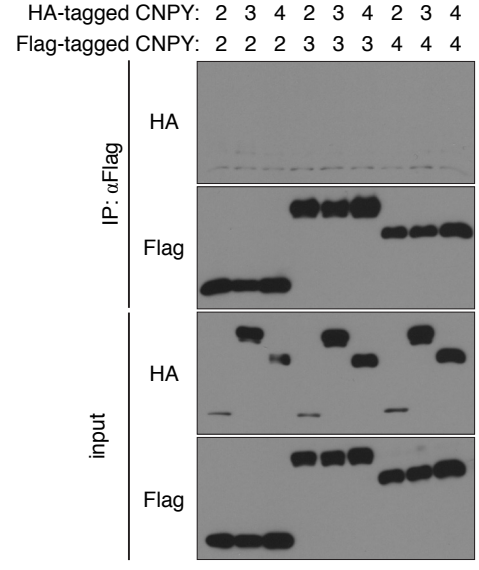

F
HA-tagged CNPY: 2 3 4 2 3 4 2 3 4
Flag-tagged CNPY: 2 2 2 3 3 3 4 4 4
IP: αFlag
HA
Flag
input
HA
Flag

Figure 1.7. Recombinant CNPY4 does not bind oxysterols involved in SHH activation. (A) Size exclusion chromatography profile for hCNPY4ΔCt purification with accompanying protein gel stained with Coomassie. **(B)** Circular dichroism of hCNPY4ΔCt at room temperature. **(C)** Thermal melt of hCNPY4ΔCt incubated with vehicle, 20(S)-hydroxycholesterol, or 24(S), 25-epoxycholesterol analyzed by circular dichroism at a wavelength of 222 nm. **(D)** Fluorescence polarization of a BODIPY-labeled 24-ketocholesterol probe incubated with increasing concentration of purified hCNPY4ΔCt. **(E, F)** Co-immunoprecipitation of Flag-tagged and HA-tagged variants of wild type CNPY2, CNPY3, and CNPY4 to probe for homo- and heterotypic association. Proteins were transiently expressed in HEK293 cells and pulled-down using either an anti-HA (E) or an anti-Flag (F) antibody. Protein levels were detected with the indicated antibodies by Western blot analysis.

recombinant CNPY4ΔCt to dimerize was due to truncation of its C-terminal tail or signal sequence. Co-immunoprecipitation of two differentially tagged constructs of full-length CNPY4 in HEK293 cells confirmed that CNPY4 was indeed unable to form homotypic associations as well as heterotypic associations with other CNPY family members (**Fig. 1.7 E, F**).

We therefore inspected the effect *Cnpy4* silencing had on overall membrane composition in cells. At the cell membrane, cholesterol exists in three major pools: one for essential membrane integrity, a second termed "accessible" cholesterol that has high chemical activity, and a third sphingomyelin-bound pool of cholesterol with low chemical activity[73]. Recently, PTCH1 was proposed to control the levels of accessible cholesterol and/or sphingomyelin via a molecular pump mechanism and thus modulate SMO activation[74]. We hypothesized that increasing the levels of accessible cholesterol in the membrane could be a mechanism behind the hyperactivation of SHH signaling observed in the absence of CNPY4. To directly measure the levels of accessible cholesterol in the plasma membrane of intact cells, we used a modified protein probe derived from the bacterial toxin Perfringolysin O (PFO*) coupled to a fluorescent tag[73]. Remarkably, NIH3T3 cells in which *Cnpy4* was knocked down displayed significantly elevated levels of accessible cholesterol compared to control treated cells (**Fig. 1.8 A, B**). Additionally, MEFs derived from embryonic limb buds of *Cnpy4* null animals displayed greatly

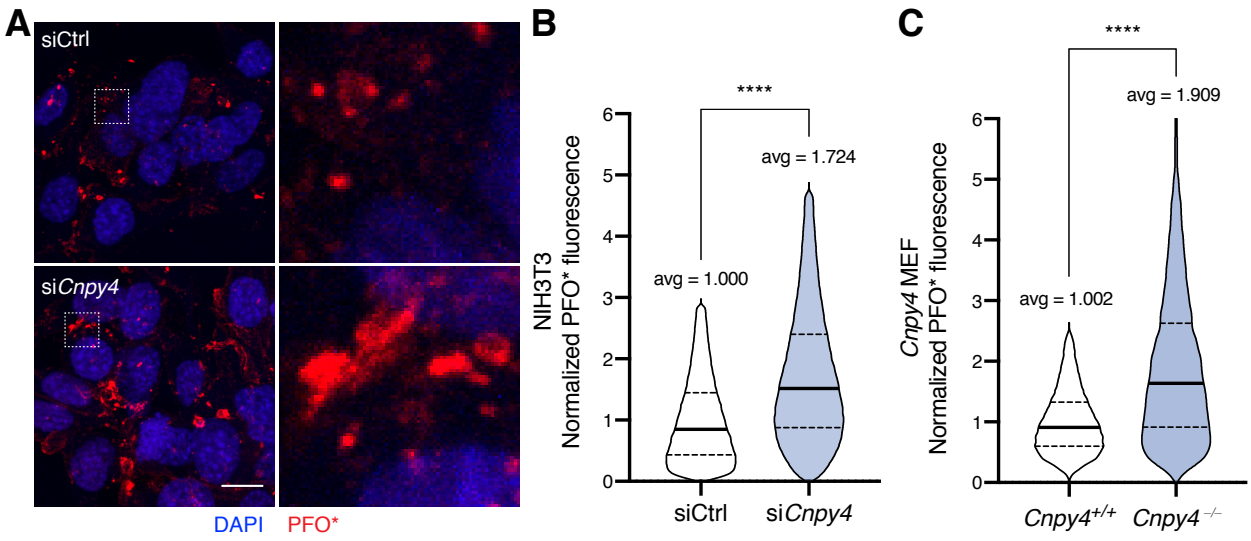


Figure 1.8. CNPY4 modulates levels of accessible cholesterol. (A) Immunofluorescence of accessible cholesterol (PFO*-AF647, red) and the nucleus (DAPI, blue) of NIH3T3 cells treated with *Cnpy4* or control siRNA. **(B, C)** FACS analysis of NIH3T3 cells treated with *Cnpy4* or control siRNA (B) or *Cnpy4*$^{+/+}$ and *Cnpy4*$^{-/-}$ MEFs (C) stained with PFO*-AF647 for accessible cholesterol. Outliers were removed and quantifications were normalized to the average value of control siRNA treated cells. Significance calculations were performed as described in Methods and Materials with **** = $p < 0.0001$

elevated levels of accessible cholesterol in a basal state (**Fig. 1.8 C**). Taken together, these data suggest that CNPY4 may directly influence signal transduction of the HH pathway through modulation of accessible cholesterol levels in the membrane.

The hyperactivation of HH signaling *Cnpy4* knockdown induced led us to question whether signaling networks that crosstalk with the HH pathway would also be affected by silencing of *Cnpy4*. In particular, the FGF pathway intersects HH signaling in multiple ways such as its contribution to ciliogenesis[75], which directly influences HH signaling capacity[76]. Furthermore, as FGF works in concert with HH signaling during development to give rise to the limb bud[57, 58], the feature in which we observed the most severe phenotypes in *Cnpy4*$^{-/-}$ embryos, we chose to investigate the effect CNPY4 has on FGF signaling. *In situ* hybridization experiments on *Fgf8* and *Etv5*, a downstream effector of the FGF pathway, showed varied levels

and localization of mRNA expression in *Cnpy4*$^{-/-}$ limb buds compared to wild-type buds (**Fig. 1.9 A**). Futhermore, *Cnpy4*$^{-/-}$ MEFs displayed altered signaling kinetics when stimulated with FGF1 ligand compared to wild-type MEFs (**Fig. 1.9 B - D**). Notably, Ser473 phosphorylation of Akt, which lies downstream of FGF receptor (FGFR) activation by an FGF ligand[77], diminishes at a significantly quicker rate post-stimulation (**Fig. 1.9 B, C**). As the kinetics and magnitude of Akt phoshorylation appear unaltered at early timepoints (<60 minutes post-stimulation), it is unclear whether the difference observed at later timepoints is due to quicker internalization of FGFR or elevated rates of dephosphorylation and/or degradation of phosphorylated Akt. As the phosphorylation kinetics of ERK, which is also upregulated by FGFR activity[77], was not significantly affected (**Fig. 1.9 B, D**), it is possible that the effect observed on Akt phosphorylation is due to an Akt-specific event rather than a consequence of FGFR alteration. However, further experiments are required to determine if this is the case.

As the main downstream output of Akt phosphorylation is cell viability[77], we examined the effect CNPY4 has on FGF-mediated survival using Ba/F3 cells, which do not express quantifiable protein levels of receptor tyrosine kinases (RTK) such as FGFR[78]. Instead, these cells are dependent on the interleukin-3 (IL-3) pathway to proliferate and survive[78]. However, stable expression of an RTK and supplementation of the growth media with the appropriate growth factor ligand can wean Ba/F3 cells off of their IL-3 dependency[79]. Ba/F3 cells in which FGFR1 was stably expressed showed elevated levels of cell survival upon removal of IL-3 and supplementation with FGF1 and its co-ligand heparin compared to parental cells (**Fig. 1.9 E**). Remarkably, stable co-expression of FGFR1 and CNPY4 increased the survival rate to an even higher level, supporting the conclusions from our earlier experiments that CNPY4 intersects the Akt arm of the FGFR pathway (**Fig. 1.9 E**). While co-immunoprecipitation studies in HEK293

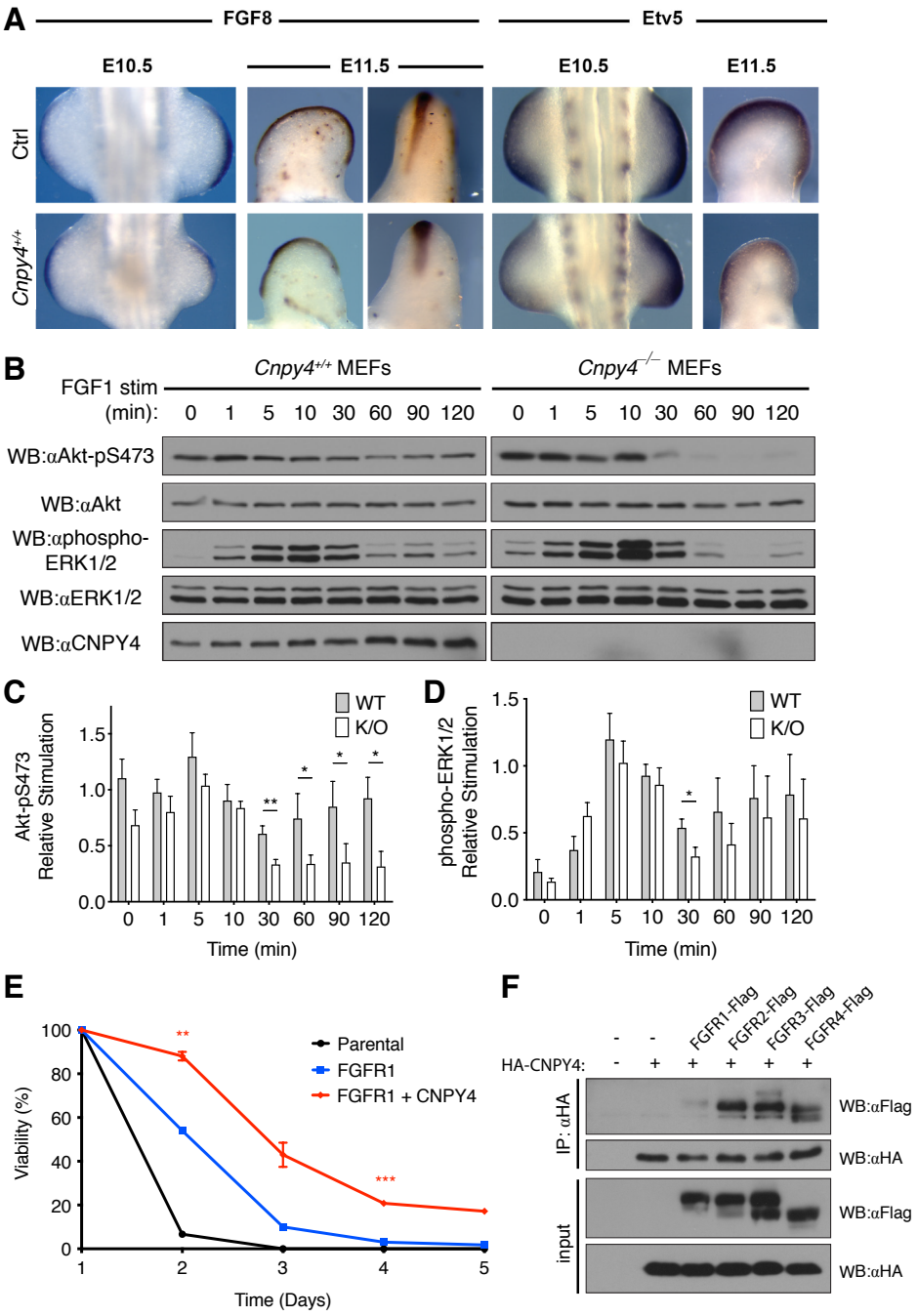


Figure 1.9. CNPY4 modulates FGFR-mediated signaling. (A) *In situ* hybridization of *Fgf8* and *Etv5* in hindlimb buds of *Cnpy4*$^{+/+}$ and *Cnpy4*$^{-/-}$ embryos at E10.5 and E11.5. **(B)** FGF1 stimulation of *Cnpy4*$^{+/+}$ and *Cnpy4*$^{-/-}$ MEF cells. Protein levels in lysates were normalized using a BCA assay and were detected with the indicated antibodies by Western blot analysis. **(C, D)** Quantification of Akt-pS473 (C) and phospho-ERK (D) upon FGF1 stimulation of MEF cells. Data were doubly normalized against the corresponding non-phosphorylated species as a loading control and to the zero-time point. Significance calculations were performed as described in Methods and Materials with * = $p < 0.05$ and ** = $p < 0.01$. **(E)** FACS analysis of parental, FGFR1, and FGFR1 + CNPY4 stably expressing Ba/F3 cell viability. IL-3 was removed from the

growth media and 10 ng/mL FGF-1 and 50 μg/mL heparin were supplemented in low serum (0.1% FBS) media instead. Significance calculations were performed as described in Methods and Materials with ** = $p < 0.01$ and *** = $p < 0.0005$. **(F)** Co-immunoprecipitation of an HA-tagged CNPY4 construct with Flag-tagged FGFR1, 2, 3, or 4. Proteins were transiently expressed in HEK293 cells and pulled-down using an anti-HA antibody. Protein levels were detected with the indicated antibodies by Western blot analysis.

cells demonstrated that exogenous CNPY4 is able to interact with all members of the FGFR family to varying degrees (**Fig. 1.9 F**), further studies are required to fully understand whether CNPY4 intersects the FGF pathway directly or can merely alter FGFR-mediated signaling via crosstalk with the SHH pathway or modulation of the cell membrane lipid composition.

DISCUSSION

The role cholesterol and lipids play in the cellular control of SHH activity is only beginning to be fully appreciated. Although there has been a surge of new data suggesting that lipids are able to modulate HH pathway activation and signal transduction[23-29, 41-45], it remains unclear what the molecular mechanisms are that enable them to do so. One specific question that is critical to understand is whether PTCH1 can modulate the lipid membrane composition to control activation of SMO, and, if so, how it is able to do so. Recent studies have revealed that the PTCH1 structure closely resembles that of the resistance-nodule-division family of transporters[40,80]. Furthermore, they uncovered multiple steroid-shaped densities residing in and around PTCH1, with one such density found in its sterol-sensing domain (SSD)[40,80]. This domain is similar to those found in proteins involved in cholesterol metabolism and signaing such as Niemann-Pick C1 and HMG-CoA reductase, as well as another HH receptor, Dispatched[81,82]. However, the exact role the SSD plays in PTCH1 function, as well as other proteins', remains

poorly understood[40,80]. It is possible that CNPY4 may serve an important role in the ability of PTCH1 to efflux sterols across the membrane through interaction with this domain, however additional studies would be needed to test this hypothesis.

More likely, perhaps, is the possibility that CNPY4 function is tied directly to membrane regulation rather than to the HH pathway specifically. Immunoprecipitation followed by mass spectrometry analysis of the CNPY4 interactome revealed that CNPY4 can bind the cholesterol biosynthesis enzyme Emopamil-Binding Protein (EBP) (also known as 3-Beta-Hydryoxysteriod-Delta(8), Delta(7)-Isomerase) (**Fig. 1.10 A**). Co-immunoprecipitation of EBP confirmed its ability to engage FLAG-tagged CNPY4 in cells (**Fig. 1.10 B**), suggesting that CNPY4 may indirectly mediate HH signaling through altering cholesterol biosynthesis in the ER via its interaction with EBP. It is then possible that the effect CNPY4 has on membrane composition and dynamics could have a wider-reaching outcome than just HH pathway perturbation, as the lipid membrane is an inalienable part of all transmembrane receptor signaling[83-86]. Our preliminary analyses of the effect CNPY4 exerts on FGFR signaling (**Fig. 1.9**) and previous reports on the role of CNPY4 in toll-like receptor signaling indicate that this could be the case[87].

Taken together, our study sheds further light on the multiple layers of HH signal regulation and provides a crucial link between lipid biology and SHH activation. Furthermore, it may reveal another layer of complexity of network crosstalk during development that was previously unknown. Additional studies on CNPY4 function may provide important clues on how dynamics within the ER, where CNPY4 resides, are able to modulate other cellular compartments, such as the cell membrane. Such experiments are critical for understanding the full range of cellular regulation that occurs to modulate signal transuction. Lastly, our identification of CNPY4 as a previously undescribed regulator of lipid membrane dynamics

A

Hit #	Name	Uniprot ID	Fold change
1	CNPY4	Q8N129	464.709
2	Kalirin	O60229	84.325
3	SPAG5	Q96R06	20.666
4	ACAD9	Q9H845	19.694
5	SLIT2	O94813	7.872
6	EBP	Q15125	7.671
7	K2026	Q5HYC2	6.647
8	ARHG3	Q9NR81	6.506
9	PML	P29590	5.593
10	LRC58	Q96CX6	5.176
11	ARL8b	Q9NVJ2	4.788
12	ABCD3	P28288	4.77
13	LRRC1	Q9BTT6	4.746
14	COQ9	O75208	4.524
15	PGES2	Q9H7Z7	4.425
16	GNPAT	O15228	4.38
17	COX41	P13073	4.328
18	CNPY3	Q9BT09	4.267
19	ADT3	P12236	4.205
20	PP1G	P36873	4.062
21	Zn536	O15090	4.06
22	Prohibitin	P35232	4.02
23	HEMH	P22830	3.919
24	VDAC1	P21796	3.75
25	AAAT	Q15758	3.671

B

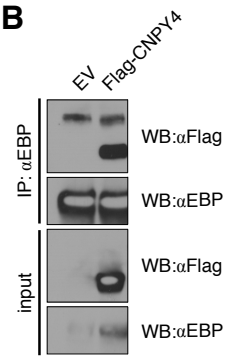


Figure 1.10. CNPY4 interactome as analyzed by IP/MS. (A) Top interactors of CNPY4 identified by the IP/MS analysis, including their Uniprot ID and abundance score. **(B)** Co-immunoprecipitation of an empty vector control or Flag-tagged CNPY4 with endogenous EBP. Proteins were transiently expressed in HEK293 cells and pulled-down using an anti-EBP antibody. Protein levels were detected with the indicated antibodies by Western blot analysis.

provides the first link for this protein family to other SAPLIP proteins, whose primary functions are directly tied to lipid and membrane modulation. Moreover, it may assist with unraveling the cellular mechanism behind previously reported functions of other CNPY proteins, such as the role cnpy1 plays as a putative positive feedback regulator of FGF signaling in zebrafish[49, 50] and the ability of CNPY2 to iniatiate the PERK-CHOP pathway[51].

ACKNOWLEDGMENTS

We would like to thank the labs of David Raleigh, Pao-Tien Chuang, and Jeremey Reiter for the NIH3T3, *Sufu*$^{-/-}$, *Ptch*$^{-/--}$, and *Smo*$^{-/-}$ MEF cell lines, Maia Kinnebrew and Rajat Rohatgi for the PFO* probe, Shaoqing Zhang and Haifan Wu for their guidance on CD experiments, Sarah Findakly-Oshima for help with confocal imaging, Hayarpi Torosyan for assistance with purification and biophysical assays, Danielle Swaney and Nevan Krogan for their work on the IP/MS data, Amnon Sharir for work on the mice data, and Christopher Agnew for mentorship on constructs, purifications, and crystallography. We would also like to thank Jennifer Kung for critical reading of the text and all members of the Jura and Klein lab for their helpful discussions.

MATERIALS AND METHODS

Mice maintenance, *in situ* hybridization, *lacZ* staining and mCT scans. Mice were kept in a pathogen-free University of California (UCSF) facility in accordance with the guidelines specified by Institutional Animal Care and Use Committee and Laboratory Animal Resource Center (LARC). All experimental procedures were done with approval from LARC at UCSF. Mice were perfused using 1x PBS and then 4% paraformaldehyde (PFA) diluted in 1x PBS after euthenization. Tissues were sectioned and paraffin stained as previously described[88]. In situ hybridization, lacZ staining, and mCT scans were additionaly done as previously described[88].

Cell culture and drug treatments. Cells were cultured in Dulbecco's modified Eagle media (Gibco) supplemented with 10% FBS (Hyclone) and penicillin streptomycin (Gibco) and incubated at 3°7C with 5% CO_2. All cells lines were regularly tested for mycoplasma contamination using the MycoAlert mycoplasma detection kit (Lonza). Stimulations were

performed in low-serum OptiMEM (Life Technologies) with 100 nM SAG (EMD Millipore), 1 μg/mL recombinant SHH (R&D Systems), 30 μM 20(S)-hydroxycholesterol (Cayman Chemicals), 30 μM 24(S), 25-epoxycholesterol (Avanti Polar Lipids), or 25 μM SANT-1 (Selleckchem). Incubations with SAG, SHH, and SANT-1 were done for 24 hours and oxysterols were done for 30-36 hours.

Immunofluorescence and data analysis. Staining and imaging of tissues were performed as previously described[88]. COS-7 cells were plated onto glass coverslips and transfected the following day. Cells were fixed in 3.7% PFA solution diluted in 1x PBS at room temperature with rocking and then incubated with a 0.1% Triton X-100 and 2.5% BSA solution in 1x PBS to permeabilize cells and to block for non-specific antibody interaction. Primary antibodies were diluted in blocking buffer and incubated overnight at 4°C then washed out three times with 0.1% Triton X-100 in 1x PBS. Secondary antibodies were diluted in blocking buffer and incubated for 2 hours at root temperatures before subsequent washes. DAPI staining was conducted for 10 minutes following the last wash before cells were mounted onto glass coverslips with Prolong Gold AntiFade Mountant (Life Technologies). Images were acquired on either a Nikon Elipse Ti with a CSU-X1 spinning disc confocal and Andor Clara interline CCD camera with a Nikon Plan Apo 60x oil objective or a Zeiss LSM 800 confocal laser scanning microscope with a 63x oil objective. Cell length calculations and SMO intensity analysis was done on Fiji.

MEF generation. Embryos were isolated and washed in 1x PBS twice. Limb buds were separated using sterile tweezers from each embryo and washed with DMEM before incubation with 0.25% Trypsin/EDTA (Gibco) at 37°C for 10 minutes. Trypsin was quenched by addition of DMEM supplemented with FBS and penicillin streptomycin. Cells were pipetted up and down at least 10 times to further dissociate cells before being transferred into fresh 15 mL tubes. Cells

were gently pelleted at 200xg for 5 minutes at room temperature. Supernatant was carefully aspirated and cells were resuspended in fresh media and plated in 6-cm plates (Gibco). Additional cell debris was aspirated off and fresh media added daily until cells reached confluency, upon which they were split and expanded once before being pooled and flash frozen.

siRNA transfection. 22.5 pmol of siRNA SMARTpool (Dharmacon) were transiently transfected into indicated cells using lipofectamine RNAiMax (Invitrogen) according to the manufacturer's protocols. Cells were transfected for 72 hours before cell analysis. Confirmation of mRNA silencing was done by qRT-PCR analysis and confirmation of protein knockdown was performed via Western blotting.

qRT-PCR analysis. Cells were grown in either 6- or 12-well plates and treated with indicated expression conditions. RNA was extracted from cells using the RNEasy Mini kit (Qiagen) and reverse-transcribed to produce cDNA using the iScript cDNA synthesis kit (Bio-Rad). qRT-PCR was performed using Power-Up SYBR Green Master Mix (Applied Biosystems) on an Invitrogen real-time PCR machine. mRNA transcript relative abundances were calculated using the ΔΔCt method against *Gapdh*.

Table 1.1 qRT-PCR primers.

Gene	**Forward**	**Reverse**
Gapdh (mouse)	tgcccccatgtttgtgatg	tgtggtcatgagcccttcc
Cnpy4 (mouse)	gacaaaagaggaggaagatgacacag	ccaggatccgctcgcacaaattctcc
Gli1 (mouse)	ggtgctgcctatagccagtgtcctc	gtgccaatccggtggagtcagaccc

Luciferase-based reporter assays. Cells were plated in 6-well plates and transfected with siRNA as described above at least 16 hours post-plating. 396 ng of Gli1-responsive Firefly luciferase reporter plasmid, 4 ng of a control Renilla luciferase reporter plasmid under the control of a constitutively active TK promoter, and 1 μg of pcDNA3.1+ empty vector were transfected into cells at least 6 hours post-siRNA transfection using lipofectamine LTX with Plus reagent (Invitrogen) according to the manufacturer's protocol. 16 hours-post transfection, cells were recovered with fresh media for 24 hours. Stimulation with indicated ligand was performed in low-serum OptiMEM media (Gibco) for 24-36 hours. Luciferase assays were conducted using the Dual Luciferase Reporter Assay System (Promega) and measured on a GloMax 96 Microplate Luminometer with Dual Injectors (Promega).

PFO* staining and FACS analysis. Cells were grown in 6-wells and treated with indicated conditions. Cells were lifted with 0.5% Triton-EDTA and gently pelleted by centrifugation at 200xg for 5 minutes. Pellets were washed gently two times with 1x PBS before incubation in blocking buffer (10 mg/mL BSA in 1x PBS) for 10 minutes on ice. Cells were pelleted once more before incubation with 5 μg/mL PFO* probe diluted in blocking buffer for 30 minutes on ice. Cells were gently washed one time with 1x PBS before analysis by FACS. Fluorescent intensity measurements by flow cytometry were performed on a Sony Cell Sorter SH800 using a 638 nm laser for excitation. Live and singlet populations were selected based on forward and side scatter. No further gating was used to select cell populations.

Statistical analysis. All statistical analysis were performed using Prism 8 (GraphPad). Outliers were identified using the identify outliers function. Significance analysis for luciferase assay and qRT-PCR analyses were done using the Mann-Whitney non-parametric test and for ciliation and FACS analyses were performed using the Welch's t-test.

Recombinant protein expression and purification. CNPY4 constructs were synthesized by Genscript and subcloned into a pET28b plasmids with a 10xHis tag sequence. Cloning verification was done by DNA sequencing (Elim biotechnology). Constructs were transformed into SHuffle T7 competent *E. coli* cells and underwent antibiotic selected on Kanamycin plates for 16 hours at 37°C. A single colony was used to inoculate a Luria broth starter culture supplemented with appropriate antibiotic for 16 hours at 37°C, 220 rpm shaking. 10 mL of starter culture was used to inoculate 900 mL of Terrific broth supplemented with appropriate antibiotic and 100 mL of 10x phosphate buffer (0.17 M KH_2PO_4, 0.72 M K_2HPO_4). Cells were grown at 37°C, 220 rpm shaking to an OD_{600} of 0.6 - 0.8 before being induced with isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.5 mM. Cultures were allowed to grow for an additional 20 hours at 18°C, 220 rpm shaking, and then were spun down in an Avanti centrifuge equipped with an JA 8.5 rotor at 4000xg, 40 minutes, 4°C. Pellets were flash frozen for later purification or resuspended in binding buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 20 mM Imidazole, pH 8.0, 5% glycerol) supplemented with DNaseI and cOmplete mini EDTA-free protease inhibitor cocktail (Roche) and lysed via sonication at 30% amplitude, 4 seconds on, 2 seconds off, for a total of 5 minutes. Lysates were clarified in an Avanti centrifuge equipped with an JLA 25.50 rotor at 20,000xg, 40 minutes, 4°C. Clarified lysates were incubated with Ni NTA 6 Fast Flow beads (GE Life Sciences) for 16 hours at 4°C with rotating before being applied to a gravity flow Econo-column (Bio-Rad). Beads were washed thoroughly with 20 column volumes of binding buffer followed by 10 column volumes of binding buffer supplmeneted with an additional 12.5 mM Imidazole. Protien was eluted in 5 column volumes of elution buffer (binding buffer with 250 mM Imidazole). The elution was buffer exchanged back into low Imidazole binding buffer and incubated with 1 mg of recombinant 3C protease for 16

hours at 4°C. The protein was applied over new NiNTA 6 Fast Flow beads and the flow through was collected. The protein was then diluted 10x into mono Q binding buffer (50 mM HEPES, pH 8.0) and applied to a monoQ 5/50 GL column (GE Life Sciences) connected to an Akta Pur system (GE Life Sciences) using a superloop (GE Life Sciences). Protein was eluted with a linear gradient of elution buffer (50 mM HEPES, pH 8.0, 500 mM NaCl). Elutions were concentrated using an Amicon Ultra-15 10k MWCO centrifugal filter (Millipore) before being loaded onto a Superdex 200 10/300 GL column (GE Life Sciences) or a Superdex 200 16/600 column (GE Life Sciences) with an isocratic gradient of size exclusion chromatography (SEC) buffer (50 mM Bicine, pH9.0, 150 mM NaCl).

Circular dichroism. Purified CNPY proteins were analyzed on a Jasco J-810 spectropolarimeter at 1 nm steps. Proteins were analyzed at an approximate concentration of 2 μM in a 50 mM sodium phosphate buffer, pH7.0 at 25°C. Thermal melt data was collected at 222 nm with a temperatures range of 25°C to 95°C in increments of 5°C. CNPY4ΔCt was additionally incubated with 30 μM of 20(S)-hydroxycholesterol and 24(S), 25-epoxycholesterol prior to thermal melt analysis for assessment of binding capacity. Data was fitted using the log(agonist) vs. response -- Variable slope non-linear analysis on Prism (GraphPad), and the LogEC50 from the analysis was reported as the melting temperature.

Fluorescence polarization. Purified human CNPY4ΔCt in SEC buffer (50 mM HEPES, pH 8.0, 150 mM NaCl) were analyzed for binding to a BODIPY-labeled 24 keto-cholesterol synthesized by the Renslo lab at UCSF. Protein was at a 100-fold molar excess at the indicated concentrations. 1% Tween-20 was added to the reaction mixture. Experiments were performed with a reaction volume of 20 μL in triplicate using a black-bottom 384-well plates (Corning) on an Analyst AD plate reader (Molecular Devices). Excitation and emission wavelengths used for

the kinetic experiments were 485 nm and 528 nm, respectively. As no significant difference was observed, signal was averaged across all time points for each triplicate with standard error calculated for each data point.

Immunoprecipitation. HEK293 cells were seeded onto 6-cm plates and transfected the following day using lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol for 24 hours. The cells were then washed two times on ice with 1xPBS before lysis buffer (0.5% Triton X-100, 0.5% NP-40, 150 mM NaCl, 50 mM Tris pH 8.0, 1 mM NaF, 1 mM $Na(VO_4)_3$, 1 mM EDTA, cOmplete mini EDTA-free protease inhibitor cocktail (Roche)) application. Cells were immediately scraped and transferred into 1.5 mL tubes to lyse for 45 minutes at 4°C with rotating. Lysates were clarified by centrifugation for 10 minutes at 15,000 rpm. The clarified lysates were pre-cleared with washed Protein A beads (Novex) for 30 minutes at 4°C with rotating before overnight incubation with pre-complexed antibody/protein A beads at 4°C with rotating. The protein-bound beads were washed three times with lysis buffer. To elute the bound protein, beads were incubated with SDS-loading buffer and were boiled at 95°C for 10 minutes.

FGF stimulation and analysis. MEFs generated from *Cnpy4*$^{+/+}$ and *Cnpy4*$^{-/-}$ embryos were plated into 6-well plates. The following day, media was replaced with serum-starvation media (DMEM supplemented with 0.1% FBS) for 24 hours. Cells were stimulated with 5 ng/mL FGF-1 (Peprotech) for indicated length of time. Post-stimulation, cells were immediately put on ice and lysed with RIPA buffer (1% NP-40, 0.1% SDS, 1% sodium deoxycholate, 150 mM NaCl, 50 mM Tris pH 8.0, 1 mM NaF, 1 mM $Na(VO_4)_3$, 1 mM EDTA, cOmplete mini EDTA-free protease inhibitor cocktail (Roche)). Lysates were normalized using a BCA assay (Fisher Scientific) before Western blot analysis. Western blot band intensity for phosphorylated proteins was calculated using Fiji and normalized first to the corresponding unphosphorylated protein

levels (i.e. for pS473 Akt to total Akt and for phospho-Erk to total Erk) and subsequently to the 0 minute time point phosphorylated protein level.

Stable Ba/F3 cell lines generation and viability assays. FGFR-mCherry fusion constructs were subcloned into pMSCV backbones with a puromycin selection marker. CNPY4 constructs were subcloned into pMIGII backbones. Cloning verification was done by DNA sequencing (Elim biotechnology). Plat-E cells were seeded onto 6-cm plates in 10% FBS supplemented DMEM. Cells were transfected the following day using Fugene 6 (Promega) according to the manufacturer's specifications. Viral supernatant was collected 48 hours post-transfection and filtered through a 0.22 μm filter. Ba/F3 cells were then spinoculated with filtered viral supernatants, 1 ng/mL IL-3 (Peprotech), and 8 μg/mL Polybrene (Sigma Aldrich) for 1 hour at room temperature. Cells were allowed to recover for 24 hours before selection with 1 μg/mL of puromycin for 3 days. Cells were further selected via FACS to ensure for proper integration of the plasmids. 4 x 10^5 stably expressing Ba/F3 cells were pelleted and resuspended in 4 mL of DMEM supplemented with 0.1% FBS (Hyclone), 50 ng/mL FGF-1 (Peprotech), and 10 μg/mL heparin sulfate (Sigma). 1 mL of media was removed every 24 hours and incubated with SYTOX Blue dead cell stain (ThermoFischer) before being measured using a BD FACSAria machine at 450/60 nm. Live cell and singlet populations were selected based on forward and side scatter. No further gating was used to select cell populations. 1 mL of supplemented media was added to remaining cells after removal of sample.

REFERENCES

1. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801 (1980).
2. Echelard, Y. *et al.* Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
3. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
4. Krauss, S., Concordet, J. P. & Ingham, P. W. A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* **75**, 1431–1444 (1993).
5. Roelink, H. *et al.* Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord. *Cell* **76**, 761–775 (1994).
6. Fietz, M. J. *et al.* The hedgehog gene family in Drosophila and vertebrate development. *Dev. Suppl.* 43–51 (1994).
7. Nieuwenhuis, E. & Hui, C. C. Hedgehog signaling and congenital malformations. *Clin. Genet.* **67**, 193–208 (2004).
8. Raleigh, D. R. & Reiter, J. F. Misactivation of Hedgehog signaling causes inherited and sporadic cancers. *J. Clin. Invest.* **129**, 465–475 (2019).
9. Kong, J. H., Siebold, C. & Rohatgi, R. Biochemical mechanisms of vertebrate hedgehog signaling. *Development* **146**, (2019).
10. Huangfu, D. *et al.* Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature* **426**, 83–87 (2003).
11. Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate

development. *Nat. Rev. Genet.* **11**, 331–344 (2010).

12. Nakano, Y. *et al.* A protein with several possible membrane-spanning domains encoded by the Drosophila segment polarity gene patched. *Nature* **341**, 508–513 (1989).
13. Hooper, J. E. & Scott, M. P. The Drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989).
14. Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
15. Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
16. Fuse, N. *et al.* Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10992–10999 (1999).
17. Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317**, 372–376 (2007).
18. Corbit, K. C. *et al.* Vertebrate Smoothened functions at the primary cilium. *Nature* **437**, 1018–1021 (2005).
19. Préat, T. Characterization of Suppressor of fused, a complete suppressor of the fused segment polarity gene of Drosophila melanogaster. *Genetics* **132**, 725–736 (1992).
20. Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 416–429 (2013).
21. Dai, P. *et al.* Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J. Biol. Chem.* **274**, 8143–8152 (1999).
22. Ågren, M., Kogerman, P., Kleman, M. I., Wessling, M. & Toftgård, R. Expression of the PTCH1 tumor suppressor gene is regulated by alternative promoters and a single

functional Gli-binding site. *Gene* **330**, 101–114 (2004).

23. Guy, R. K. Inhibition of sonic hedgehog autoprocessing in cultured mammalian cells by sterol deprivation. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7307–7312 (2000).

24. Maity, T., Fuse, N. & Beachy, P. A. Molecular mechanisms of Sonic hedgehog mutant effects in holoprosencephaly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17026–17031 (2005).

25. Cooper, M. K. *et al.* A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nat. Genet.* **33**, 508–513 (2003).

26. Chevy, F., Illien, F., Wolf, C. & Roux, C. Limb malformations of rat fetuses exposed to a distal inhibitor of cholesterol biosynthesis. *J. Lipid Res.* **43**, 1192–1200 (2002).

27. Cooper, M. K., Porter, J. A., Young, K. E. & Beachy, P. A. Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science* **280**, 1603–1607 (1998).

28. Gofflot, F. Molecular mechanisms underlying limb anomalies associated with cholesterol deficiency during gestation: implications of Hedgehog signaling. *Hum. Mol. Genet.* **12**, 1187–1198 (2003).

29. Stottmann, R. W. *et al.* Cholesterol metabolism is required for intracellular hedgehog signal transduction in vivo. *PLoS Genet.* **7**, e1002224 (2011).

30. Lee, J. J. *et al.* Autoproteolysis in hedgehog protein biogenesis. *Science* **266**, 1528–1537 (1994).

31. Chen, X. *et al.* Processing and turnover of the Hedgehog protein in the endoplasmic reticulum. *J Cell Biol* **192**, 825–838 (2011).

32. Porter, J. A., Young, K. E. & Beachy, P. A. Cholesterol modification of hedgehog signaling proteins in animal development. *Science* **274**, 255–259 (1996).

33. Porter, J. A. *et al.* Hedgehog patterning activity: role of a lipophilic modification mediated

by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21–34 (1996).

34. Pepinsky, R. B. *et al.* Identification of a palmitic acid-modified form of human Sonic hedgehog. *J. Biol. Chem.* **273**, 14037–14045 (1998).

35. Mann, R. K. & Beachy, P. A. Novel Lipid Modifications of Secreted Protein Signals. *Annu. Rev. Biochem.* **73**, 891–923 (2004).

36. Burke, R. *et al.* Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell* **99**, 803–815 (1999).

37. Creanga, A. *et al.* Scube/You activity mediates release of dually lipid-modified Hedgehog signal in soluble form. *Genes Dev.* **26**, 1312–1325 (2012).

38. Tukachinsky, H., Kuzmickas, R. P., Jao, C. Y., Liu, J. & Salic, A. Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep* **2**, 308–320 (2012).

39. Williams, K. P. *et al.* Functional antagonists of sonic hedgehog reveal the importance of the N terminus for activity. *J. Cell. Sci.* **112 (Pt 23)**, 4405–4414 (1999).

40. Qi, X., Schmiege, P., Coutavas, E., Wang, J. & Li, X. Structures of human Patched and its complex with native palmitoylated sonic hedgehog. *Nature* **560**, 128–132 (2018).

41. Byrne, E. F. X. *et al.* Structural basis of Smoothened regulation by its extracellular domains. *Nature* **535**, 517–522 (2016).

42. Luchetti, G. *et al.* Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling. *eLIFE* **5**, e20304 (2016).

43. Huang, P. *et al.* Structural Basis of Smoothened Activation in Hedgehog Signaling. *Cell* **174**, 312–324 (2018).

44. Raleigh, D. R. *et al.* Cilia-Associated Oxysterols Activate Smoothened. *Mol. Cell* **72**,

316–327 (2018).

45. Garcia, G., Raleigh, D. R. & Reiter, J. F. How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In. *Curr. Biol.* **28**, R421–R434 (2018).

46. Bruhn, H. A short guided tour through functional and structural features of saposin-like proteins. *Biochem. J.* **389**, 249–257 (2005).

47. Munford, R. S., Sheppard, P. O. & O'Hara, P. J. Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. *J. Lipid Res.* **36**, 1653–1663 (2002).

48. Kishimoto, Y., Hiraiwa, M. & O'Brien, J. S. Saposins: structure, function, distribution, and molecular genetics. *J. Lipid Res.* **33**, 1255–1267 (2002).

49. Hirate, Y. & Okamoto, H. Canopy1, a Novel Regulator of FGF Signaling around the Midbrain-Hindbrain Boundary in Zebrafish. *Curr. Biol.* **16**, 421–427 (2006).

50. Matsui, T. *et al.* Canopy1, a positive feedback regulator of FGF signaling, controls progenitor cell clustering during Kupffer's vesicle organogenesis. *Proc. Nalt. Acad. Sci. USA* **108**, 9881–9886 (2011).

51. Hong, F. *et al.* CNPY2 is a key initiator of the PERK-CHOP pathway of the unfolded protein response. *Nat. Struct. Mol. Biol.* (2017).

52. Callejo, A., Culi, J. & Guerrero, I. Patched, the receptor of Hedgehog, is a lipoprotein receptor. *Proc. Natl. Acad. Sci. USA* **105**, 912–917 (2008).

53. Do, H. T. *et al.* Fibroblast Growth Factor-21 (FGF21) Regulates Low-density Lipoprotein Receptor (LDLR) Levels in Cells via the E3-ubiquitin Ligase Mylip/Idol and the Canopy2 (Cnpy2)/Mylip-interacting Saposin-like Protein (Msap). *J. Biol. Chem.* **287**, 12602–12611 (2012).

54. Okabe, N., Xu, B. & Burdine, R. D. Fluid dynamics in zebrafish Kupffer's vesicle. *Dev.*

Dyn. **237**, 3602–3612 (2008).

55. Zhang, X. M., Ramalho-Santos, M. & McMahon, A. P. Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. *Cell* **106**, 781–792 (2001).

56. Mutoh, H. *et al.* Biallelic Variants in CNPY3, Encoding an Endoplasmic Reticulum Chaperone, Cause Early-Onset Epileptic Encephalopathy. *Am. J. Hum. Genet.* **102**, 321–329 (2018).

57. Towers, M., Wolpert, L. & Tickle, C. Gradients of signalling in the developing limb. *Curr. Opin. Cell Biol.* **24**, 181–187 (2012).

58. Cooper, K. L. *et al.* Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511**, 41–45 (2014).

59. Fogel, J. L., Lakeland, D. L., Mah, I. K. & Mariani, F. V. A minimally sufficient model for rib proximal-distal patterning based on genetic analysis and agent-based simulations. *eLIFE* **6**, e29144 (2017).

60. Dessaud, E., McMahon, A. P. & Briscoe, J. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**, 2489–2503 (2008).

61. Lettice, L. A. & Hill, R. E. Preaxial polydactyly: a model for defective long-range regulation in congenital abnormalities. *Curr. Opin. Genet. Dev.* **15**, 294–300 (2005).

62. VanderMeer, J. E. & Ahituv, N. cis-regulatory mutations are a genetic cause of human limb malformations. *Dev. Dyn.* **240**, 920–930 (2011).

63. Al-Qattan, M. M. Zone of Polarizing Activity Regulatory Sequence Mutations/Duplications with Preaxial Polydactyly and Longitudinal Preaxial Ray

Deficiency in the Phenotype: A Review of Human Cases, Animal Models, and Insights Regarding the Pathogenesis. *Biomed. Res. Int.* **2018**, 1573871 (2018).

64. Wheway, G., Nazlamova, L. & Hancock, J. T. Signaling through the Primary Cilium. *Front. Cell Dev. Biol.* **6**, 1358 (2018).

65. Reiter, J. F. & Leroux, M. R. Genes and molecular pathways underpinning ciliopathies. *Nat. Rev. Mol. Cell Biol.* **18**, 533–547 (2017).

66. Kim, J. *et al.* The role of ciliary trafficking in Hedgehog receptor signaling. *Sci Signal* **8**, ra55 (2015).

67. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 1–1 (2012).

68. Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. & Privé, G. G. Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proc. Natl. Acad. Sci. USA* **100**, 38–43 (2002).

69. Ahn, V. E., Leyko, P., Alattia, J.-R., Chen, L. & Privé, G. G. Crystal structures of saposins A and C. *Protein Sci.* **15**, 1849–1857 (2006).

70. Rossmann, M. *et al.* Crystal Structures of Human Saposins C and D: Implications for Lipid Recognition and Membrane Interactions. *Structure* **16**, 809–817 (2008).

71. Parkash, V. *et al.* The structure of the conserved neurotrophic factors MANF and CDNF explains why they are bifunctional. *Protein Eng. Des. Sel.* **22**, 233–241 (2009).

72. Willis, C. *et al.* Insights into the Membrane Interactions of the Saposin-Like Proteins Na-SLP-1 and Ac-SLP-1 from Human and Dog Hookworm. *PLoS ONE* **6**, e25369 (2011).

73. Das, A., Brown, M. S., Anderson, D. D., Goldstein, J. L. & Radhakrishnan, A. Three

pools of plasma membrane cholesterol and their relation to cholesterol homeostasis. *eLIFE* **3**, e19316 (2014).

74. Kinnebrew, M. *et al.* Cholesterol accessibility at the ciliary membrane controls hedgehog signaling. *eLIFE* **8**, e50051 (2019).

75. Thomas, J. *et al.* Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. *Biol. Cell* **102**, 499–513 (2012).

76. Bangs, F. & Anderson, K. V. Primary Cilia and Mammalian Hedgehog Signaling. *CSH Perspect. Biol.* **9**, a028175 (2017).

77. Ornitz, D. M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *WIREs Dev. Biol.* **4**, 215–266 (2015).

78. Warmuth, M., Kim, S., Gu, X.-J., Xia, G. & Adrián, F. Ba/F3 cells and their use in kinase drug discovery. *Curr. Opin. Oncol.* **19**, 55–60 (2007).

79. Jiang, J. *et al.* Epidermal Growth Factor–Independent Transformation of Ba/F3 Cells with Cancer-Derived Epidermal Growth Factor Receptor Mutants Induces Gefitinib-Sensitive Cell Cycle Progression. *Cancer Res.* **65**, 8968–8974 (2005).

80. Gong, X. *et al.* Structural basis for the recognition of Sonic Hedgehog by human Patched1. *Science* (2018).doi:10.1126/science.aas8935

81. Goldstein, J. L., DeBose-Boyd, R. A. & Brown, M. S. Protein Sensors for Membrane Sterols. *Cell* **124**, 35–46 (2006).

82. Kuwabara, P. E. & Labouesse, M. The sterol-sensing domain: multiple families, a unique role? *Trends Genet.* **18**, 193–201 (2002).

83. Casaletto, J. B. & McClatchey, A. I. Spatial regulation of receptor tyrosine kinases in development and cancer. *Nat. Rev. Cancer* **12**, 387–400 (2012).

84. Bocharov, E. V., Sharonov, G. V., Bocharova, O. V. & Pavlov, K. V. Conformational transitions and interactions underlying the function of membrane embedded receptor protein kinases. *Biochim. Biophys. Acta Biomembr.* **1859**, 1417–1429 (2017).

85. Sunshine, H. & Iruela-Arispe, M. L. Membrane lipids and cell signaling. *Curr. Opin. Lipidol.* **28**, 408–413 (2017).

86. Bethani, I., nland, S. S. S. A., Dikic, I. & Acker-Palmer, A. Spatial organization of transmembrane receptor signalling. *EMBO J.* **29**, 2677–2688 (2010).

87. Konno, K. *et al.* A molecule that is associated with Toll-like receptor 4 and regulates its cell surface expression. *Biochem. Biophys. Res. Co.* **339**, 1076–1082 (2006).

88. Sharir, A. *et al.* A large pool of actively cycling progenitors orchestrates self-renewal and injury repair of an ectodermal appendage. *Nat. Cell Biol.* **21**, 1102–1112 (2019).

Chapter 2: Structural determination of Canopy proteins

ABSTRACT

Saposin and saposin-like (SAPLIP) proteins are characterized by a unique structural element comprised of an alpha helical bundle held together by three specifically spaced disulfide bonds. Although much remains unknown about SAPLIP proteins, they have been shown to play essential cellular roles through their interactions with lipids and lipid membranes. The Canopy (CNPY) proteins are a family of signaling regulators, which have been proposed to be members of the SAPLIP superfamily. However, there is little biophysical evidence and no structural evidence that this is the case. Here, we describe efforts to determine the structure of a CNPY protein by X-ray crystallography. Analysis by circular dichroism indicates that the secondary structure of purified CNPY2 is consistent with that of a SAPLIP protein. Interestingly, CNPY2 appears to be monomeric, consistent with our previous analyses on other members of the family but in contrast with saposins and the majority of SAPLIP proteins characterized thus far, which form dimers. Although dimerization has been shown to be critical for the function of these SAPLIP proteins, our previous analyses on CNPY4 suggest that it retains the ability to modulate lipid membranes even as a monomeric protein. While further optimization is required for structural determination, our studies indicate that the CNPY proteins are indeed part of the SAPLIP family. Furthermore, they reveal that the CNPY proteins may have evolved a novel, dimerization-independent mechanism for regulating lipids.

INTRODUCTION

Saposin and saposin-like (SAPLIP) proteins form a diverse superfamily uniquely characterized by the presence of a structural feature that goes by many names, including the saposin fold, sap domain, saposin-like domain, and SAPLIP domain[1-3]. This unifying structure is comprised of four to five alpha helices held together by six, specifically spaced cysteine residues that form three disulfide bonds. Two bonds hold the N- and C-termini together, and a third is formed in the center of the protein, resulting in a clamshell-like structure[3]. As little else defines the classification of proteins into the SAPLIP superfamily, several hundreds of proteins have been categorized as such. Some of these proteins are comprised solely of a saposin fold, though many contain additional domains. These proteins, which together share sequence similarity as low as 10%, are found in organisms as far back as unicellular amaebozoa[1]. There appear to be no SAPLIP proteins in bacteria, as the few that are classified as such in databases like InterPro lack the characteristic spacing of their cysteine residues and are thus likely misannotated[1]. Plants possess proteins with a modified, circular permutation of the classic saposin fold, which are referred to as swaposins[4]. As swaposins retain the same tertiary structure as other SAPLIP proteins and possess similar functions, they have remained grouped with the SAPLIP superfamily[1]. Though there are several classified SAPLIP proteins in fungi and a few families found in protists as well, the vast majority of SAPLIP proteins are found in metazoa[1].

Despite their low sequence similarity, SAPLIP proteins appear to possess a shared ability to regulate lipids and lipid membranes, typically through one of three common mechanisms: (1) membrane binding that results in local disorder, (2) membrane perturbation, such as lipid extraction for enzymatic activity, and (3) membrane permeabilization as a means for cell killing[1]. Much of our understanding of these functions is derived from molecular analysis of diseases

associated with saposins. Saposins, the founding members of the SAPLIP family, are lysosomal-resident cofactors for sphingolipid-degrading enzymes[1-3]. Deficiency of any one of the four saposin proteins (A-D) or the precursor from which all four are derived, presaposin, results in lysosomal storage diseases, which occur due to impairment of enzymatic activity within the lysosome[5-6]. Complementary *in vitro* studies have helped identify the diverse set of sphingolipids these proteins act upon and have demonstrated how they facilitate certain vesicle dynamics, such as vesicular fusion and clearance, from within the lysosome[7-11]. Several other identified functions include those of the SAPLIP surfactant protein SP-B, which clusters upon membranes as a result of increasing protein/phospholipid ratios to induce membrane destabilization and fusion[8-11], and the amoebapore and granulysin proteins, which both induce membrane permeabilization[12-14]. Interestingly, despite sharing similar cellular functions and possessing a saposin fold, the amoebapore and granulysin proteins share less than 20% sequence identity[1]. Indeed, little information regarding function can be gleaned from the primary sequence of SAPLIP proteins and is likely one of many reasons for our lack of knowledge on this diverse protein family.

Biophysical and structural studies over the past 18 years have begun to shed light on the unique ways by which the saposin fold facilitates the interactions between SAPLIP proteins and lipids. Many SAPLIP proteins require dimerization for full activity, including the saposin proteins. Low pH levels, such as those found within the lysosome, induce a closed dimerization state of the saposin proteins, which enables them to form a shell-like structure around extracted lipids for enzymatic presentation[15-17]. Alteration or inhibition of dimerization through mutation inhibits this function[16]. Similarly, NMR structures revealed that amoebapore A undergoes pH-dependent dimerization catalyzed by an exposed surface histidine residue, which is critical for its proper membrane insertion and pore formation[18,19]. SP-B too is able to dimerize, however it does

so in a covalent manner driven by a seventh cysteine residue[1]. Despite this common regulatory mechanism, not all SAPLIP proteins appear to require dimerization for activity. Granulysin and its porcine homolog, NK-lysin, while performing a similar function as the amoebapore proteins, bind and lyse membranes as monomers[14,20,21]. Additionally, structural studies on the human hookworm SAPLIP *Na*-SLP-1 proposed that it too binds membranes in a monomeric state[22].

Another common method of membrane interaction utilized by SAPLIP proteins that has been uncovered by biophysical and structural studies is the use of charged residues to drive electrostatic interactions[1]. Interestingly, both monomeric and dimeric SAPLIP proteins exploit such residues for membrane interaction, albeit through very different mechanisms[1]. NK-lysin, for example, which is monomeric and acts in a pH-independent manner, contains a significant number of positively charged surface residues. These residues cluster on one side of protein, which allows it to "coat" the membrane as monomers and permeabilize it through an action termed molecular electroporation[20, 21]. In contrast, amoebapore A, which like NK-lysin functions in cell permeabilization, has been shown to assemble in a pH-dependent manner into a trimer of dimers upon insertion into the membrane[18]. Modeling based on its solved NMR structure suggests a barrel-like formation with a highly basic core that is able to form specifically sized pores within membranes[18,23]. However, like dimerization, many SAPLIP proteins do not appear to utilize electrostatic residues to drive interaction with the membrane, such as the canine hookworm SAPLIP *Ac*-SLP-1, which structural studies hypothesize uses a surface tryptophan residue to drive membrane association[22]. Even from these limited studies, it is clear that the saposin fold, while virtually identical at the tertiary level in all solved structures of SAPLIP proteins except for saposin B[1], has evolved unique mechanisms of action based on specific primary sequence. It is therefore clear that structural and biophysical analyses, in addition to

cellular and biochemical experiments on cellular function, are critical for full understanding of SAPLIP mechanism.

Our work on CNPY4 indicated that this understudied SAPLIP protein regulates sonic hedgehog (SHH) pathway signaling by altering the levels of accessible cholesterol in the cell membrane. However, it remained unclear if CNPY4 can directly interact with this lipid or with the membrane and, if so, how it could do so as an ER-resident protein. Our preliminary studies using protein lipid overlay assays to assess lipid binding indicated that CNPY4 lacking its C-terminal tail could not efficiently bind to cholesterol or other sphingolipids, including sphingomyelin, which is hypothesized to control the levels of accessible cholesterol in the membrane. Additional binding studies with oxysterols yielded similar results (**Fig. 1.7 C, D**), suggesting CNPY4 was not directly interacting with lipids or the membrane. We therefore sought to determine the molecular structure of a CNPY protein in order to aid our analysis of their function and to elucidate the mechanism behind the modulation CNPY4 exerts on accessible cholesterol levels in the cell membrane.

RESULTS

As expression of full-length human CNPY4 in both *E. coli* and Sf9 insect cells proved to be unsuccessful, we expressed a construct of CNPY4 with a truncated C-terminal tail (CNPY4ΔCt), which is predicted to be largely unstructured (**Fig. 1.7 A**). However, crystallographic trials with purified protein proved to be unsuccessful. We therefore used murine CNPY2, which is >97% identical to human CNPY2 and has a similar domain architecture as CNPY4ΔCt, for structural studies. Recombinant full-length mCNPY2 lacking its signal sequence

(residues 1 - 20) with an N-terminal 6xHis-tag was expressed in T7 SHuffle *E. coli* cells, which constitutively express the bacterial disulfide bond isomerase DsbC. Cells were lysed using a sonicator, and the His-tag was cleaved from isolated proteins, which was then purified to a purity of >90% as analyzed by size exclusion chromatography (SEC) and gel electrophoresis with Coomassie staining (**Fig. 2.1 A**). The final SEC step was performed in a high pH buffer (50 mM Bicine, pH9.0) for two reasons. Firstly, thermal stability screens in a range of pH and salt concentrations indicated that the protein was most stable at high pH (>7.5) values (**Fig. 2.1 B**). Secondly, lower pH (<7.5) buffers caused the appearance of a doublet to occur when visualized by gel electrophoresis under reducing conditions over time (**Fig. 2.1 C, D**). Tryptic mass spectrometry was unable to identify a difference in the sequence of this band, suggesting that it was likely not the result of degradation of the protein. While further experiments are necessary to determine the difference in the two visualized bands, we decided to proceed with a high pH buffer to maximize homogeneity of the purified protein sample.

Purified full-length mCNPY2 was analyzed via circular dichroism (CD) to assess folding and thermal stability. Recombinant mCNPY2 is predominantly alpha helical and has a high melting temperature (Tm ~ 66.9°C), consistent with other SAPLIP proteins[1, 2] (**Fig. 2.2 A, B**). Analysis by SEC and analytical ultracentrifugation indicated that the purified protein was a monomer under the conditions tested (**Fig. 2.2 C, D**). As saposin and SAPLIP dimerization can be induced by changes in pH, we analyzed recombinant mCNPY2 at a range of pH values (pH 4.0 - 9.0) (**Fig. 2.2 C**). Decreasing the pH did not induce dimerization of the purified mCNPY2 protein, suggesting that either CNPY2 does not dimerize or that it can only do so under conditions different than those tested. These findings were consistent with our previous co-immunoprecipitation experiments with human CNPY2 from cells with differentially tagged

CNPY2 or CNPY3 and CNPY4 (**Fig. 1.7 E, F**), which indicated that CNPY2 does not engage in homotypic or heterotypic oligomers.

To gain insights into the structures of CNPY proteins, we sought to obtain a crystal structure of CNPY2. Crystallization screens were performed using the purified full-length

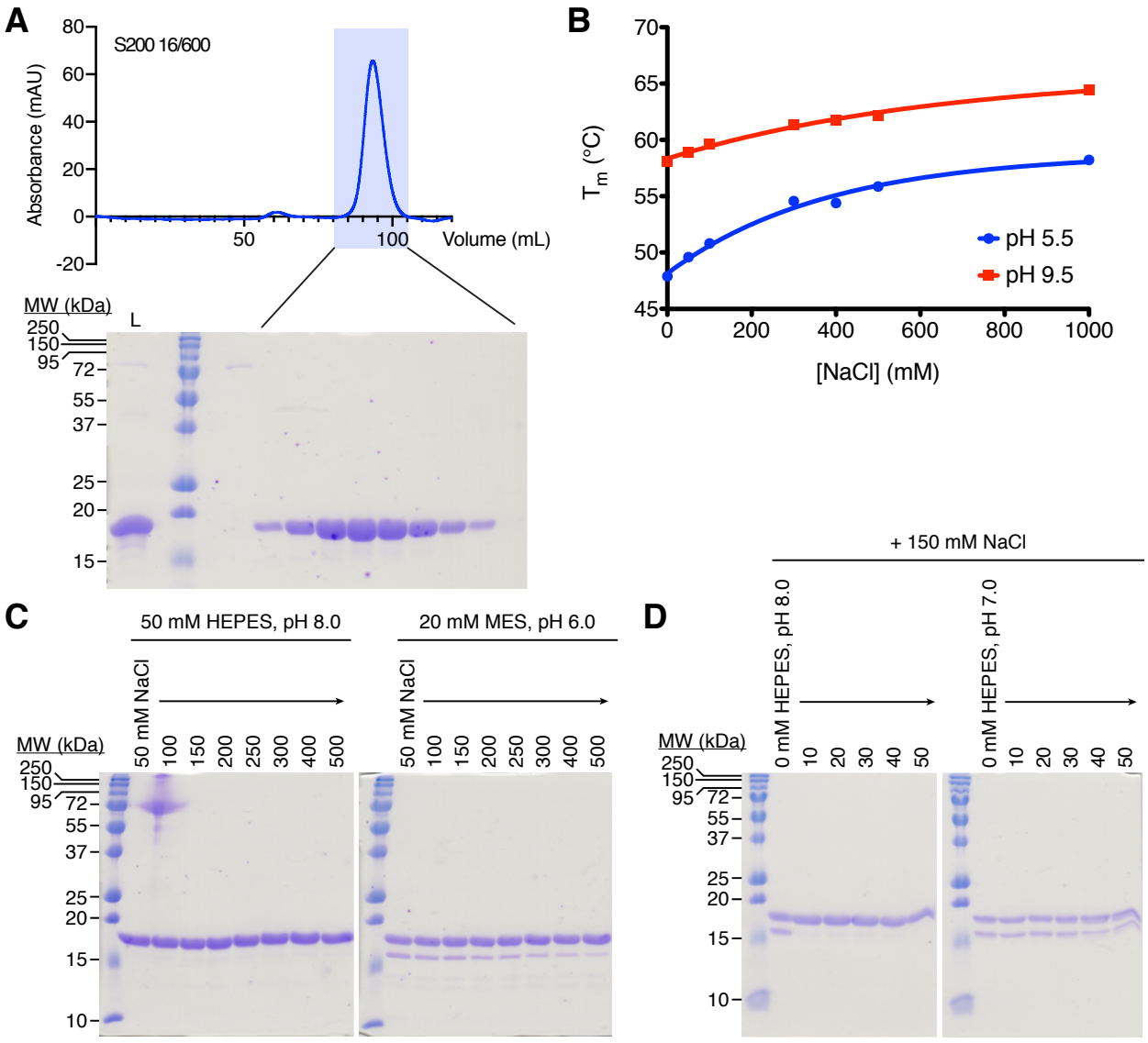


Figure 2.1. pH is critical for mCNPY2 purification. (A) Size exclusion chromatography profile for mCNPY2 purification with accompanying gel stained with Coomassie. **(B)** Thermal melts of purified mCNPY2 protein in increasing pH and salt buffers. **(C, D)** Gel electrophoresis analysis of purified mCNPY2 stability in buffers with varying pH and salt concentrations.

mCNPY protein. mCNPY2 crystallized in one condition (pHClear II Suite: 1 M LiCl, 0.1 M HEPES, pH 7.0, 30% PEG 6000, room temperature, 1:1 ratio, sitting drop) (**Fig. 2.2 E**). The crystals were confirmed by UV analysis to contain protein (**Fig. 2.2 E**). Crystallization was optimized around this condition in 24-well plate format (**Fig. 2.2 F**). Several crystals were picked for structural determination, with the best crystals diffracting to a limit of 1.9Å (**Fig. 2.2 G**). Preliminary analysis indicated that the space group of the crystal lattice was C2, which necessitated 180° of data collection. However, CNPY2 crystals exhibited a high level of anisotropy, and data collected at angles greater than 60° showed diffraction spot smearing and diffracted to a limit less than 4Å (**Fig. 2.2 G**). This was likely also caused by a high degree of mosaicity of the crystals used for X-ray diffraction, as the crystal form was thin, overlapping hexagonal plates. Optimization by seeding and additive screens proved to be unsuccessful in growing crystals of a different form. We further encountered issues with reproducing crystals in custom 24-well plates when switching our source of PEG 6000. Together, these issues led us to pursue alternative avenues of structural determination for CNPY2.

We reasoned that the long inserts between the alpha helices of CNPY2, which are predicted to be largely unstructured, might be hampering efforts to crystallize the full-length CNPY2 in a different crystal form. We therefore screened multiple constructs in which the insert regions were truncated or entirely deleted. mCNPY2 with a truncation of the three residues C-terminal of the signal sequence (residues 21 - 23) (CNPY2Δ4) had the highest levels of expression of the variants screened. mCNPY2Δ4 was expressed and purified under the same conditions as WT (**Fig. 2.3 A**). CD analysis of purified mCNPY2Δ4 indicated that there were very few biophysical differences between the two constructs (**Fig. 2.3 B, C**). However, mCNPY2Δ4 crystallized in a variety of commercial conditions chemically disparate to those that

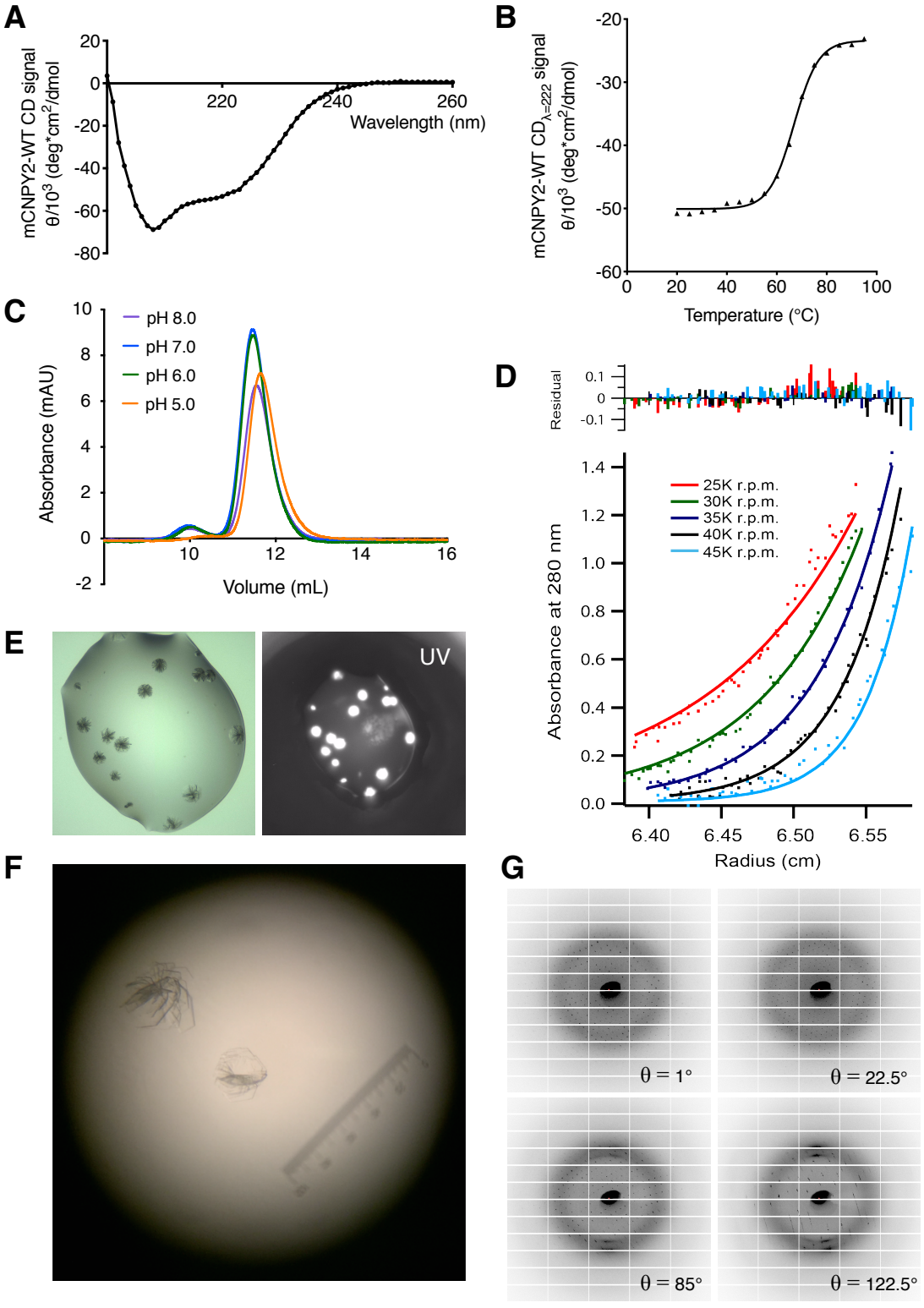


Figure 2.2. mCNPY2 biophysical analysis and structural efforts. (A) Circular dichroism of mCNPY2 at room temperature. **(B)** Thermal melt of mCNPY2 analyzed by circular dichroism at a wavelength of 222 nm. **(C)** Size exclusion chromatography profile for mCNPY2 in various buffers: 50 mM MES, pH 5.0; 50 mM MES, pH 6.0; 50 mM Tris, pH 7.0; 50 mM HEPES, pH 8.0.

Data collected at pH 4.0 is not shown as the protein likely aggregates at this pH as it is lower than the predicted pI (4.88) of the protein. **(D)** Sedimentation velocity analytical ultracentrifugation of mCNPY2. **(E)** Crystal hit condition (pH Clear II Suite, B10: 1 M LiCl, 0.1 M HEPES, pH 7.0, 30% PEG 6000) imaged under visible and UV light. **(F)** Crystal hit in 24-well plate optimized condition (1 M LiCl, 0.1 M HEPES, pH 7.2, 32% PEG 6000). **(G)** Example diffraction data of mCNPY2 crystal at different angles incident to the beam line.

the full- length protein crystallized under (e.g. JCSG Core I Suite: 0.2 M Ca Acetate, 20% PEG 3350) (**Fig. 2.3 D**). Crystals reproduced very readily in both 96- and 24-well format using custom screens around the initial crystal condition. Several crystals from different chemical conditions were picked for structural determination by X-ray crystallography. Unfortunately, these crystals diffracted at best to a limit of ~2.8Å and also exhibited considerable anisotropy and mosaicity, although to levels lower than that of WT. Further optimization, specifically additional truncations of the longer inserts between the alpha helices, is likely necessary for structural determination of CNPY2 by X-ray crystallography.

DISCUSSION

Over the past 20 years, our understanding of saposin and SAPLIP proteins has grown tremendously, largely due to a number of molecular structures. Despite these advancements, functions of novel SAPLIP proteins cannot be easily extrapolated from the existing structures due to the high variation in primary sequence between different SAPLIP proteins. Specifically, it is difficult to deduce if and how a protein interacts with lipids and/or the plasma membrane just by its classification as a SAPLIP protein due to the lack of known consensus binding motifs that mediate these interactions. Indeed, with each new SAPLIP structure published, it becomes clearer that an obvious connection between primary and tertiary structure likely does not exist or

is not easily discernible. These challenges, in part, necessitate structural analysis of each new SAPLIP for complete understanding of its cellular mechanism.

While further work is needed to determine the atomic structure of a CNPY protein, biophysical experiments on recombinant CNPY2 and CNPY4 protein have provided new insights into their function. CD data confirmed that CNPY proteins indeed are SAPLIP proteins that are predominantly alpha helical and thermally stable. Interestingly, both proteins are monomeric, consistent with our earlier cell-based immunoprecipitation studies. While many SAPLIP proteins have been reported to function as obligate dimers, a few known examples function as monomers. However, the mechanisms by which these monomeric SAPLIP proteins typically function primarily relies on direct interaction with lipid membranes, such as the use of electrostatic residues in the case of NK-lysin and *Na*-SLC-1[13,20-22, 24]. However, it is highly likely that CNPY proteins do not directly modulate the membrane and do so indirectly from the ER, within which they reside (**Fig. 1.1**). As lipids are primarily synthesized, matured, and eventually trafficked from the ER[25,26], it is possible that ER-resident SAPLIP proteins could bind to and/or modulate immature lipids in this organelle. Previous studies on another ER-resident SAPLIP protein, mesencephalic astrocyte-derived neurotrophic factor (MANF), uncovered that it is able to bind the sphingolipid sulfatide[27]. The crystal structure of MANF illustrated a relatively even surface charge distribution, suggesting that it may interact with sulfatide through a distinct, non-electrostatic driven mechanism[28]. However, studies have yet to elucidate what the precise mechanism may be. Furthermore, as MANF is also secreted through the classical ER-Golgi pathway[29], it is unclear whether it is able to bind sulfatide within the ER or only in the extracellular space. Taken together, there remains little precedent to anticipate the mechanism that underlies signaling by CNPY proteins from within the ER. While some studies report that

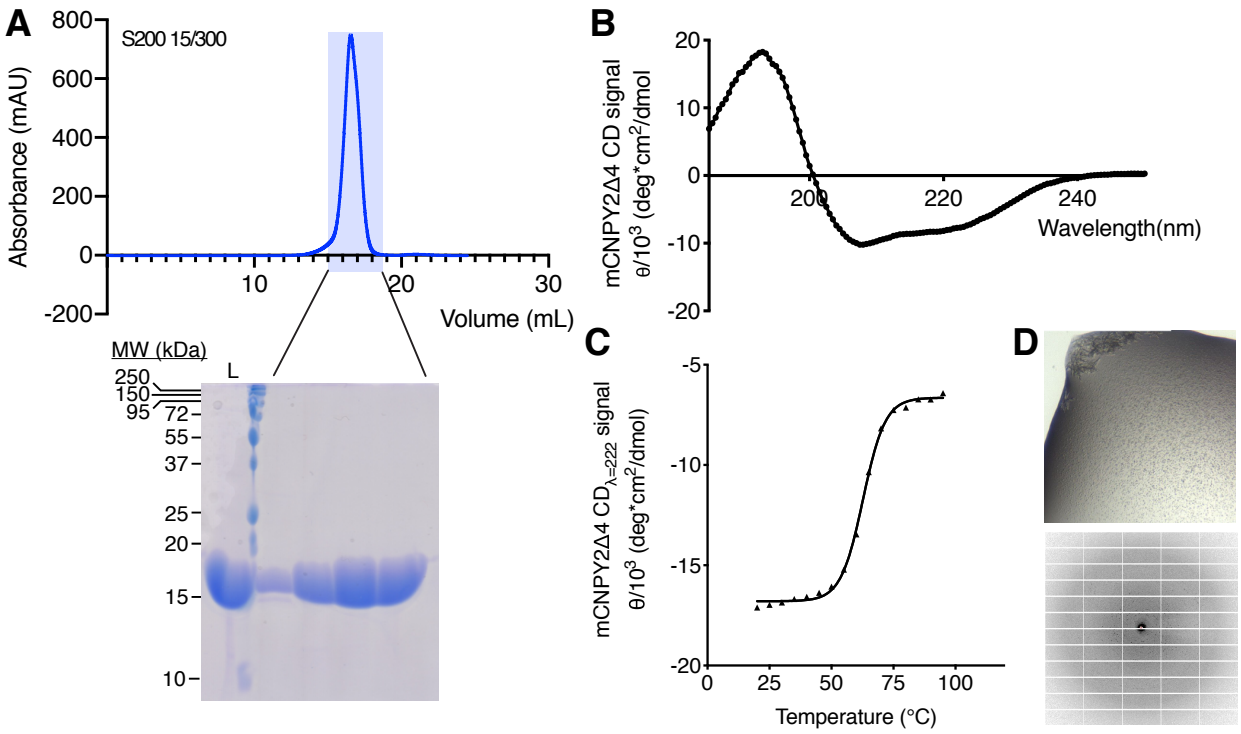


Figure 2.3. mCNPY2Δ4 is biophysically similar to WT mCNPY2. (A) Size exclusion chromatography profile for mCNPY2Δ4 purification with accompanying gel stained with Coomassie. **(B)** Circular dichroism of mCNPY2Δ4 at room temperature. **(C)** Thermal melt of mCNPY2Δ4 analyzed by circular dichroism at a wavelength of 222 nm. **(D)** Crystal hit condition (JCSG Core I Suite, B8: 0.2 M Ca Acetate, 20% PEG3350) and example diffraction.

CNPY2 is also able to be secreted[30], this does not appear to be the case for other members of the CNPY family (**Fig 1.1**), indicating that the effect CNPY4 has on the lipid membrane is unlikely due to a previously described mechanism. While our initial binding studies demonstrated that CNPY4 lacking its C-terminal tail was unable to interact with sphingolipids, cholesterol, and oxysterols that are highly enriched within the ciliary membrane, IP/MS analysis suggested that CNPY4 can interact with the cholesterol biosynthesis protein Emopamil-binding protein (EBP) (**Fig. 1.10 A**). Although further studies are required to determine the effect CNPY4 interaction with EBP has on lipid dynamics within the ER, it is possible that CNPY4 has evolved a unique mechanism for lipid modulation that is disparate from other SAPLIP proteins.

The unique functions of CNPY4 are likely further diversified by a long unstructured C-terminal tail. This tail, which is roughly 40 residues long, contains a large number of acidic residues (17 out of the predicted 42 residues). Interestingly, only CNPY3 possesses a similar tail, but one that contains predominantly basic residues (15 out of the predicted 68 residues). Neither CNPY2 nor CNPY1, which in many mammals appears to be mis-annotated at half the expected length, contain such a feature. It is possible that the long C-terminal tails in CNPY4 and CNPY3 are responsible for their reportedly unique function, which diverges from that of CNPY2, CNPY1, and many other known SAPLIP proteins. CNPY4 has been reported by us to be a regulator of HH signaling and by others to regulate the surface levels of toll-like receptor with CNPY3[31-36]. In contrast, CNPY2 has been reported to perform a variety of dissimilar functions, including neurite outgrowth and initiation of the PERK-CHOP pathway[37-39]. Meanwhile, CNPY1 is annotated as a positive feedback regulator of FGF signaling[40,41]. Our preliminary studies have demonstrated that CNPY4 too appears to be able to regulate FGF signaling despite being the protein least related to CNPY1 in the CNPY family, muddying the relationship between structure and function even further. Taken together, it becomes clear that structural determination may hold the key to uncoupling the complex and diverse roles of the CNPY family of proteins.

ACKNOWLEDGMENTS

We would like to thank Christopher Agnew for his advice on construct design, purification, and crystallography, Chari Noddings for assistance with CNPY4Δ4 purification and crystallography, the scientists at the Advanced Light Source Beamline 8.3.1 at Lawrence Berkeley National Laboratory for their help on x-ray crystallography data collection, Jennifer Kung for critical reading of the manuscript, and all members of the Jura and Klein labs for helpful discussions.

MATERIALS AND METHODS

Recombinant protein expression and purification. CNPY2 constructs were synthesized by Genscript and subcloned into pOPINF plasmids with a 6xHis tag sequence. Cloning verification was done by DNA sequencing (Elim biotechnology). Constructs were transformed into SHuffle T7 competent *E. coli* cells and underwent antibiotic selected on Ampicillin plates for 16 hours at 37°C. A single colony was used to inoculate a Luria broth starter culture supplemented with appropriate antibiotic for 16 hours at 37°C, 220 rpm shaking. 10 mL of starter culture was used to inoculate 900 mL of Terrific broth supplemented with appropriate antibiotic and 100 mL of 10x phosphate buffer (0.17 M KH_2PO_4, 0.72 M K_2HPO_4). Cells were grown at 37°C, 220 rpm shaking to an OD_{600} of 0.6 - 0.8 before being induced with 0.5 M (1000x stock) of isopropyl β-D-1-thiogalactopyranoside. Cultures were allowed to grow for an additional 20 hours at 18°C, 220 rpm shaking, and then were spun down in an Avanti centrifuge equipped with an JA 8.5 rotor at 4000xg, 40 minutes, 4°C. Pellets were flash frozen for later purification or resuspended in binding buffer (50 mM HEPES, pH 8.0, 500 mM NaCl, 20 mM Imidazole, pH 8.0, 5% glycerol) with DNaseI and cOmplete mini EDTA-free protease inhibitor cocktail (Roche) and lysed via sonication at 30% amplitude, 4 seconds on, 2 seconds off, for a total of 5 minutes. Lysates were clarified in an Avanti centrifuge equipped with an JLA 25.50 rotor at 20,000xg, 40 minutes, 4°C. Clarified lysates were incubated with Ni NTA 6Fast Flow beads (GE Life Sciences) for 2 - 16 hours at 4°C with rotating before being applied to a gravity flow Econo-column (Bio-Rad). Beads were washed thoroughly with 20 column volumes of binding buffer followed by 10 column volumes of binding buffer with 12.5 mM Imidazole. Protein was eluted in 5 column volumes of elution buffer (binding buffer with 250 mM Imidazole). The elution was buffer exchanged back into low Imidazole binding buffer and incubated with 1 mg of

recombinant 3C protease for 16 hours at 4°C. The protein was applied over new NiNTA 6Fast Flow beads, and the flow through was collected. The protein was then diluted 10x into mono Q binding buffer (50 mM HEPES, pH 8.0) and applied to a monoQ 5/50 GL column (GE Life Sciences) connected to an Akta Pur system (GE Life Sciences) using a superloop (GE Life Sciences). Protein was eluted with a linear gradient of elution buffer (50 mM HEPES, pH 8.0, 500 mM NaCl). Elutions were concentrated using an Amicon Ultra-15 10k centrifugal filter (Millipore) before being loaded onto a Superdex 200 10/300 GL column (GE Life Sciences) or a Superdex 200 16/600 column (GE Life Sciences) with an isocratic gradient of size exclusion chromatography (SEC) buffer (20 mM Bicine, pH 9.0, 150 mM NaCl).

Differential scanning fluorimetry. CNPY2 was assessed for thermal stability in different buffer pH and NaCl concentrations. Purified protein was diluted to 2 μM in various buffers (50 mM Bis-tris Propane pH 9.5 or 50 mM MES, pH 5.5 with 0, 50, 100, 300, 400, 500, or 1000 mM NaCl), and Sypro Orange (Molecular Probes) was added at a final concentration of 2x. The total reaction volume was 25 μL. Thermal melt was assessed on a CFX1000 real-time thermal cycler (Bio-Rad) with the temperature increased at 1°C/minute from 25°C to 95°C at a fluorescence of 530 nm for Sypro Orange. Data was plotted in Prism (GraphPad) and fitted using the log(agonist) vs. response -- Variable slope non-linear analysis on Prism (GraphPad), and the LogEC50 from the analysis was reported as the melting temperature.

Analytical ultracentrifugation. Purified murine CNPY2 protein was analyzed by equilibrium sedimentation using a Beckman XL-I analytical ultracentrifuge. Analysis was performed in SEC buffer (20 mM HEPES, pH 8.0, 150 mM NaCl) at a protein concentration of approximately 100 μM. Equilibrium radial concentration gradients were acquired for an absorbance of 280 nm at rotor speeds of 25, 30, 35 and 40 K r.p.m. at 25 °C. Data were fit to a single-species or two-

species model of equilibrium sedimentation by a nonlinear least-squares method using IGOR Pro (Wavemetrics) with the best-fitting model accepted.

Circular dichroism. Purified CNPY proteins were analyzed on a Jasco J-810 spectropolarimeter at 1 nm steps. Proteins were analyzed at an approximate concentration of 2 μM in a 50 mM sodium phosphate buffer, pH7.0 at 25°C. Thermal melt data was collected at 222 nm with a temperatures range of 25°C to 95°C in increments of 5°C. CNPY4ΔCt was additionally incubated with 30 μM of 20(S)-hydroxycholesterol and 24(S), 25-epoxycholesterol prior to thermal melt analysis for assessment of binding capacity. Data was fitted using the log(agonist) vs. response -- Variable slope non-linear analysis on Prism (GraphPad), and the LogEC50 from the analysis was reported as the melting temperature.

X-ray crystallography. Purified protein was screened in crystallographic conditions using commercial screens (Qiagen) in both hanging and sitting well drop format, at room temperature and 4°C, and at a ratio of 0.1 μL protein + 0.1 μL mother liquor. Screens were set up using a mosquito nanoliter protein crystallization robot (TTP Labtech). Custom screens around crystal hits were designed using a dragonfly discovery robot (TTP Labtech) for 96-well format and by hand for 24-well format. 24-well format screens were done in the sitting well drop format at room temperature with a ratio of 0.5 μL protein + 0.5 μL mother liquor. Diffraction data were collected at the Advanced Light Source at Lawrence Berkeley National Laboratories on the 8.3.1 beam line.

REFERENCES

1. Bruhn, H. A short guided tour through functional and structural features of saposin-like proteins. *Biochem. J.* **389**, 249–257 (2005).
2. Kishimoto, Y., Hiraiwa, M. & O'Brien, J. S. Saposins: structure, function, distribution, and molecular genetics. *J. Lipid Res.* **33**, 1255–1267 (2002).
3. Munford, R. S., Sheppard, P. O. & O'Hara, P. J. Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. *J. Lipid Res.* **36**, 1653–1663 (2002).
4. Ponting, C. P. & Russell, R. B. Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem. Sci.* **20**, 179–180 (1995).
5. Schnabel, D. *et al.* Simultaneous deficiency of sphingolipid activator proteins 1 and 2 is caused by a mutation in the initiation codon of their common gene. *J. Biol. Chem.* **267**, 3312–3315 (1992).
6. Christomanou, H., Chabás, A., Pámpols, T. & Guardiola, A. Activator protein deficient Gaucher's disease. A second patient with the newly identified lipid storage disorder. *Klin. Wochenschr.* **67**, 999–1003 (1989).
7. Sandhoff, K. & Kolter, T. Biosynthesis and degradation of mammalian glycosphingolipids. *Phil. Trans. R. Soc. Lond. B* **358**, 847–861 (2003).
8. Poulain, F. R., Allen, L., Williams, M. C., Hamilton, R. L. & Hawgood, S. Effects of surfactant apolipoproteins on liposome structure: implications for tubular myelin formation. *Am. J. Physiol.* **262**, L730–9 (1992).
9. Poulain, F. R., Nir, S. & Hawgood, S. Kinetics of phospholipid membrane fusion induced by surfactant apoproteins A and B. *Biochim. Biophys. Acta* **1278**, 169–175 (1996).

10. Chang, R., Nir, S. & Poulain, F. R. Analysis of binding and membrane destabilization of phospholipid membranes by surfactant apoprotein B. *Biochim. Biophys. Acta* **1371**, 254–264 (1998).

11. Hawgood, S. Surfactant Protein B: Structure and Function. *Neonatology* **85**, 285–289 (2004).

12. Leippe, M., Andrä, J., Nickel, R., Tannich, E. & Müller-Eberhard, H. J. Amoebapores, a family of membranolytic peptides from cytoplasmic granules of Entamoeba histolytica: isolation, primary structure, and pore formation in bacterial cytoplasmic membranes. *Mol. Microbiol.* **14**, 895–904 (1994).

13. Bruhn, H., Riekens, B., Berninghausen, O. & Leippe, M. Amoebapores and NK-lysin, members of a class of structurally distinct antimicrobial and cytolytic peptides from protozoa and mammals: a comparative functional analysis. *Biochem. J.* **375**, 737–744 (2003).

14. Anderson, D. H. *et al.* Granulysin Crystal Structure and a Structure-derived Lytic Mechanism. *J. Mol. Biol.* **325**, 355–365 (2003).

15. Vaccaro, A. M. *et al.* pH-dependent conformational properties of saposins and their interactions with phospholipid membranes. *J. Biol. Chem.* **270**, 30576–30580 (1995).

16. Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. & Privé, G. G. Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proc. Natl. Acad. Sci. USA* **100**, 38–43 (2002).

17. Ahn, V. E., Leyko, P., Alattia, J.-R., Chen, L. & Privé, G. G. Crystal structures of saposins A and C. *Protein Sci.* **15**, 1849–1857 (2006).

18. Hecht, O. *et al.* Solution Structure of the Pore-forming Protein of Entamoeba histolytica.

J. Biol. Chem. **279**, 17834–17841 (2004).

19. Xun, Y. *et al.* Cell-free synthesis and combinatorial selective 15N-labeling of the cytotoxic protein amoebapore A from Entamoeba histolytica. *Protein Expr. Purif.* **68**, 22–27 (2009).

20. Miteva, M., Andersson, M., Karshikoff, A. & Otting, G. Molecular electroporation: a unifying concept for the description of membrane pore formation by antibacterial peptides, exemplified with NK-lysin. *FEBS Letters* **462**, 155–158 (1999).

21. Linde, C. M. A. *et al.* Conserved Structure and Function in the Granulysin and NK-Lysin Peptide Family. *Infect. Immun.* 6332–6339 (2005).

22. Willis, C. *et al.* Insights into the Membrane Interactions of the Saposin-Like Proteins Na-SLP-1 and Ac-SLP-1 from Human and Dog Hookworm. *PLoS ONE* **6**, e25369 (2011).

23. Keller, F., Hanke, W., Trissl, D. & Bakker-Grunwald, T. Pore-forming protein from Entamoeba htstolyttca forms voltageand pH-controlled multi-state channels with properties similar to those of the barrel-stave aggregates. *Biochim. Biophys. Acta* 89–93 (1989).

24. Ruysschaert, J. M. *et al.* Lipid membrane binding of NK-lysin. *FEBS Letters* **425**, 341–344 (1998).

25. Fagone, P. & Jackowski, S. Membrane phospholipid synthesis and endoplasmic reticulum function. *J. Lipid Res.* **50**, S311–S316 (2009).

26. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Rep* **18**, 1905–1921 (2017).

27. Bai, M. *et al.* Conserved roles of C. elegans and human MANFs in sulfatide binding and cytoprotection. *Nat. Commun.* **9**, 897 (2018).

28. Parkash, V. *et al.* The structure of the conserved neurotrophic factors MANF and CDNF explains why they are bifunctional. *PEDS* **22**, 233–241 (2009).

29. Apostolou, A., Shen, Y., Liang, Y., Luo, J. & Fang, S. Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced cell death. *Exp. Cell Res.* **314**, 2454–2467 (2008).

30. Guo, J. *et al.* A secreted protein (Canopy 2, CNPY2) enhances angiogenesis and promotes smooth muscle cell migration and proliferation. *Cardiovasc. Res.* **105**, 383–393 (2015).

31. Shibata, T. *et al.* PRAT4A-dependent expression of cell surface TLR5 on neutrophils, classical monocytes and dendritic cells. *Int. Immunol.* **24**, 613–623 (2012).

32. Wakabayashi, Y. *et al.* A protein associated with toll-like receptor 4 (PRAT4A) regulates cell surface expression of TLR4. *J. Immunol.* **177**, 1772–1779 (2006).

33. Kiyokawa, T. *et al.* A single base mutation in the PRAT4A gene reveals differential interaction of PRAT4A with Toll-like receptors. *Int. Immunol.* **20**, 1407–1415 (2008).

34. Takahashi, K. *et al.* A protein associated with Toll-like receptor (TLR) 4 (PRAT4A) is required for TLR-dependent immune responses. *J. Exp. Med.* **204**, 2963–2976 (2007).

35. Hart, B. E. & Tapping, R. I. Cell Surface Trafficking of TLR1 Is Differentially Regulated by the Chaperones PRAT4A and PRAT4B. *J. Biol. Chem.* **287**, 16550–16562 (2012).

36. Konno, K. *et al.* A molecule that is associated with Toll-like receptor 4 and regulates its cell surface expression. *Biochem. Biophys. Res. Commun.* **339**, 1076–1082 (2006).

37. Bornhauser, B. C., Olsson, P.A. & Lindholm, D. MSAP Is a Novel MIR-interacting Protein That Enhances Neurite Outgrowth and Increases Myosin Regulatory Light Chain. *J. Biol. Chem.* **278**, 35412–35420 (2003).

38. Hong, F. *et al.* CNPY2 is a key initiator of the PERK-CHOP pathway of the unfolded

protein response. *Nat. Struct. Mol. Biol.* (2017).

39. Ito, S. *et al.* CNPY2 inhibits MYLIP-mediated AR protein degradation in prostate cancer cells. *Oncotarget* **9**, 17645–17655 (2018).

40. Hirate, Y. & Okamoto, H. Canopy1, a Novel Regulator of FGF Signaling around the Midbrain-Hindbrain Boundary in Zebrafish. *Curr. Biol.* **16**, 421–427 (2006).

41. Matsui, T. *et al.* Canopy1, a positive feedback regulator of FGF signaling, controls progenitor cell clustering during Kupffer's vesicle organogenesis. *Proc. Natl. Acad. Sci. USA* **108**, 9881–9886 (2011).

Chapter 3: Characterization of the novel pseudokinase PEAK3

ABSTRACT

Members of the New Kinase Family 3 (NKF3), PEAK1/SgK269 and Pragmin/SgK223 pseudokinases, have emerged as important regulators of cell motility and cancer progression. Here, we demonstrate for the first time that C19orf35 (PEAK3), a newly identified member of the NKF3 family, is a kinase-like protein evolutionarily conserved across mammals and birds and a novel regulator of cell motility. In contrast to its family members, which promote cell elongation when overexpressed in cells, PEAK3 overexpression does not have an elongating effect on cell shape but instead is associated with loss of actin filaments. Through an unbiased search for PEAK3 binding partners, we identified several regulators of cell motility, including the adaptor protein CrkII. We show that by binding to CrkII, PEAK3 prevents the formation of CrkII-dependent membrane ruffling. This function of PEAK3 is reliant upon its dimerization, which is mediated through a split helical dimerization (SHED) domain conserved among all NKF3 family members. Disruption of the conserved DFG motif in the PEAK3 pseudokinase domain also interferes with its ability to dimerize and subsequently bind CrkII, suggesting that the conformation of the pseudokinase domain might play an important role in PEAK3 signaling. Hence, our data identify PEAK3 as an NKF3 family member with a unique role in cell motility driven by dimerization of its pseudokinase domain.

INTRODUCTION

Protein kinases are conserved in nearly all organisms and play fundamental roles in development and cellular homeostasis by catalyzing the phosphorylation of their substrate proteins. As such, precise regulation of kinase activity is critical for the proper execution of cellular programs. Dysregulation of kinase signaling leads to a variety of congenital abnormalities and is linked to the progression of multiple types of cancer and other malignancies[1]. In fact, kinase domain mutations account for the largest number of known oncogenic drivers[2]. Protein kinases, therefore, are exceptionally effective therapeutic targets and account for one of the largest portions of drug discovery research.

To this day, a substantial portion of the human protein kinome, introduced by Manning and colleagues in 2002[3], has remained in the dark with their physiological functions unknown. Pseudokinases, which represent approximately one tenth of all of kinases, embody the darkest matter in the human kinome[3]. These proteins are characterized by the presence of mutations in active site motifs, such as the HRD and DFG motif, which, in most cases, result in loss of catalytic activity. As such, pseudokinases are thought to signal primarily through protein-protein interaction-based mechanisms, such as protein scaffolding and allosteric activation. Over the last decade, there has been exponential progress in our understanding of these functions[4]. The divergence of pseudokinases from their active counterparts both in structure and function raises an interesting possibility that there may still remain unidentified kinases, which were missed during the original assembly of the kinome due to significant departure from the canonical sequence motifs and which may be of therapeutic interest. In this paper, we characterize a kinase not included in the original human kinome, Chromosome 19 open reading frame 35 (C19orf35), as a banner case for discovery of new functions in the human kinome.

C19orf 35, which was annotated as PEAK3, is a new human kinase-like protein in the New Kinase Family 3 (NKF3) that contains two known pseudokinases: Sugen Kinase (SgK) 269 (also called Pseudopodium Enriched Atypical Kinase 1 (PEAK1)) and SgK223 (also called PEAK2), an ortholog of the rat protein Pragmin and the mouse protein Notch activation complex kinase (NACK)[3,5,6]. From here on forth, we will refer to the proteins as PEAK1 and Pragmin. PEAK1 and Pragmin are large (1746 and 1406 residues, respectively) signaling scaffolds comprised of long N-terminal regions with no predicted domain structure followed by protein kinase-like domains at their C-termini. Though the catalytic HxD motif remains conserved in both PEAK1 (HCD) and Pragmin (HRD), their pseudokinase domains are atypical and carry multiple substitutions in other canonical consensus motifs critical for the catalysis of phosphorylation. These substitutions include mutation of the DFG motif (to NFL in PEAK1 and to NFS in Pragmin), lack of the conserved glutamate in the αC helix, and a degenerate glycine-rich loop. Likely as a consequence of these changes, both kinases were demonstrated to lack the ability to bind nucleotides[7]. Consistent with these data, the recent crystal structures of both PEAK1 and Pragmin revealed nucleotide-binding pockets with significantly distorted architectures, which render them incompatible with ATP binding[5,8,9]. Collectively, these observations support the categorization of PEAK1 and Pragmin as pseudokinases.

Functional studies on PEAK1 and Pragmin have underscored their common role in the regulation of cell morphology and migration. Pragmin was originally described as an effector of Rnd2, a Rho family GTPase expressed primarily in neurons, that participates in negative regulation of neurite outgrowth by activating RhoA[10]. Contradictory to this function, overexpression of Pragmin in cultured epithelial cells leads to an elongated cell morphology and promotes cell migration[11-13]. Like observed for Pragmin, exogenous expression of PEAK1 is

associated with cell elongation and increased cell migration, and knockdown of the protein dramatically reduces cell motility[14-16]. The cellular phenotypes induced by both PEAK1 and Pragmin are largely consistent with the hypothesis that they act as late-stage mediators of EGFR signaling, which switch signaling output of EGFR activation from pro-mitogenic to pro-migratory[17]. Upon EGF stimulation, PEAK1 serves as a scaffold for Grb2-independent signaling complexes that enhance cell motility by promoting the phosphorylation of key focal adhesion proteins including p130Cas, Crk, and Paxillin[14]. PEAK1 is also able to recruit Pragmin to these Grb2-independent complexes, where it enhances phosphorylation of JAK1 at Y1023 to induce STAT3-dependent changes in cell morphology[11,17].

The complex roles of PEAK1 and Pragmin in cell migration pathways likely underlie their oncogenic potential and involvement in the development of a number of tumors[11,14,18-23]. In recent years, Pragmin has emerged as an important player in cancer progression that is necessary for Src-mediated invasion of colon carcinoma cells, NOTCH-dependent tumorigenesis, and the development of pancreatic cancer through JAK1/STAT3 signaling[11,19,20]. Furthermore, genetic ablation of Pragmin in human esophageal carcinoma cells was shown to promote tumor growth, emphasizing the critical role it plays in the advancement of cancer[20]. PEAK1 has also been shown to promote cancer progression and is overexpressed in subsets of pancreatic, breast, and colon cancers[14,20,21,23].

PEAK3 diverges from PEAK1 and Pragmin in that it has a significantly shorter N-terminal domain and a less conserved kinase-like domain. Using sequence analysis across evolution, we show that, although the NKF3 family emerges in early metazoans (e.g. Placozoa (*Trichoplax*) and sponges), PEAK3 does not appear in evolution until much later (in some reptilian lineages) and is evolutionarily conserved in birds and mammals. We present the first

bioinformatic and biochemical evidence that PEAK3 shares characteristic topological features with its NKF3 family members, including a SHED domain that supports its homodimerization. Using an unbiased approach to characterize PEAK3 function in cells, we found that one of the direct binding partners of PEAK3 is the focal adhesion protein CrkII. By binding to CrkII in a dimerization-dependent manner, PEAK3 antagonizes CrkII signaling, exerting an opposite effect on cell morphology to PEAK1 and Pragmin.

RESULTS

PEAK3 is a distinct member of the NKF3 family of atypical protein kinases

We searched for distant homologs of protein kinases in the human proteome using the sensitive sequence comparison algorithm Fold and Function Assignment System (FFAS)[24], previously used by us to discover novel kinase-like proteins in humans (SELO and FAM69/DIA1 families[25,26]) and in bacteria and fungi (COTH[27]). The only novel protein with a significant FFAS score found that was not previously identified as kinase-like was a single uncharacterized protein, denoted as chromosome 19 open reading frame 35 (C19orf35). C19orf35 has the closest similarity to the PEAK1/SgK269 and Pragmin/SgK223 pseudokinases that together constitute the NKF3 family. Recently Lecointre and colleagues reported the identity of C19orf35 as a new member of the NKF3 family, and C19orf35 was annotated in the Universal Protein Resource (Uniprot) as PEAK3[5]. For consistency, from this point on, we refer to C19orf35 as PEAK3. Using PEAK1 or Pragmin as Blast queries results in only partial alignments covering fragments of the kinase-like domain. However, the FFAS method is capable of detecting similarity over the entire length of the putative kinase-like domain (**Fig. 3.1 A**). The kinase-like domain of human PEAK3 shares only approximately 26% sequence identity with

human PEAK1 and Pragmin due to low complexity regions (LCRs), which introduce long gaps in the pairwise alignments. This is most likely why the PEAK3 sequence, although present in the Uniprot database since 2004, was not annotated as kinase-like until recently.

LCRs, defined as areas of protein sequences with biased amino acid composition, located within the kinase-like domain are a distinct feature of NKF3 proteins[28]. These regions are typically implicated in mediating protein-protein interactions[29]. In NKF3 pseudokinases, the LCRs correlate with flexible regions, as judged by missing coordinates corresponding to these sequences in structures of PEAK1 and Pragmin. One example is the PAPAPAPA motif in Pragmin that is located between the HRD and NFL motifs, where the NFL motif corresponds to the DFG motif in active kinases (**Fig. 3.2 A**). The majority of the LCRs in PEAK1, Pragmin, and PEAK3 diverge in sequence from one another and are located in different regions. PEAK3 stands out by having the largest portion of its kinase-like domain sequence (more than 20%) denoted as LCRs. Some of these motifs are relatively well conserved in evolution, such as the PPGPPGSPGP motif that is immediately downstream of the DFG motif in PEAK3 (**Fig. 3.2 B**).

Evolutionary conservation of PEAK3 and its kinase-like domain

The NKF3 family likely appeared at the origin of Metazoa, which is indicated by the presence of homologs in sponges (*Amphimedon*) and placozoans (*Trichoplax*). Most invertebrates (e.g. cnidarians, echinoderms) contain a single member of the family, although in some lineages (e.g. insects and nematodes), the NKF3 family has seemingly been lost. The division into the PEAK1 and Pragmin subfamilies likely occurred between the emergence of chordates (the lancelet *Branchiostoma* has a single family member) and jawed vertebrates (*Gnathostomata* have both PEAK1 and Pragmin). The PEAK3 subfamily diverges from the

Pragmin branch likely at the origin of reptiles (e.g. it is found in crocodilians) and is present in birds and mammals (**Fig. 3.1 B**). Interestingly, PEAK3 is missing in some reptile species such as snakes and lizards. The early members of the NKF3 family, i.e. proteins from sponges and placozoans, exhibit high sequence and structure conservation with those found in more complex metazoans (e.g. vertebrates), suggesting evolutionarily conserved functions of these proteins.

The kinase-like domain of PEAK3 is highly conserved in evolution and carries unique sequence alterations in key catalytic motifs found in active kinases when compared with PEAK1 and Pragmin. Interestingly, these alterations within the putative active site of PEAK3 vary markedly between species (**Fig. 3.1 A; Fig. 3.2**). Mammalian PEAK1 and Pragmin do not possess a conserved DFG motif. However, PEAK1 and Pragmin have intact HxD motifs: HRD in PEAK1 and HCD in Pragmin. In contrast, mammalian PEAK3 contains a conserved DFG motif, while the HxD motif is replaced by LxE. These unique features are not conserved in avian PEAK3 homologs, which have an NFF or SFF sequence in place of DFG, more closely resembling sequences present in PEAK1 (NFS) and in Pragmin (NFL). In PEAK1 and Pragmin, these motifs are almost perfectly conserved irrespective of a species. While HxD is also not conserved in avian PEAK3, the catalytic aspartate (contained within the HxD motif) is present within the QGD sequence that replaces the HxD motif.

In all species, PEAK3 has a conserved EN motif, corresponding to the EN sequence located in PKA at positions 170-171 that coordinates divalent cations[30]. This feature is also present in PEAK1 and Pragmin (**Fig. 3.1 A; Fig. 3.2**). The catalytic lysine, corresponding to K72 in PKA, is conserved in PEAK3 (K204 in PEAK3), but the glutamate (E91 in PKA) that forms a salt bridge with the catalytic lysine in active kinases is missing in PEAK3. In PEAK1 and Pragmin, the catalytic lysine is also present but is “hijacked” by interactions with three residues

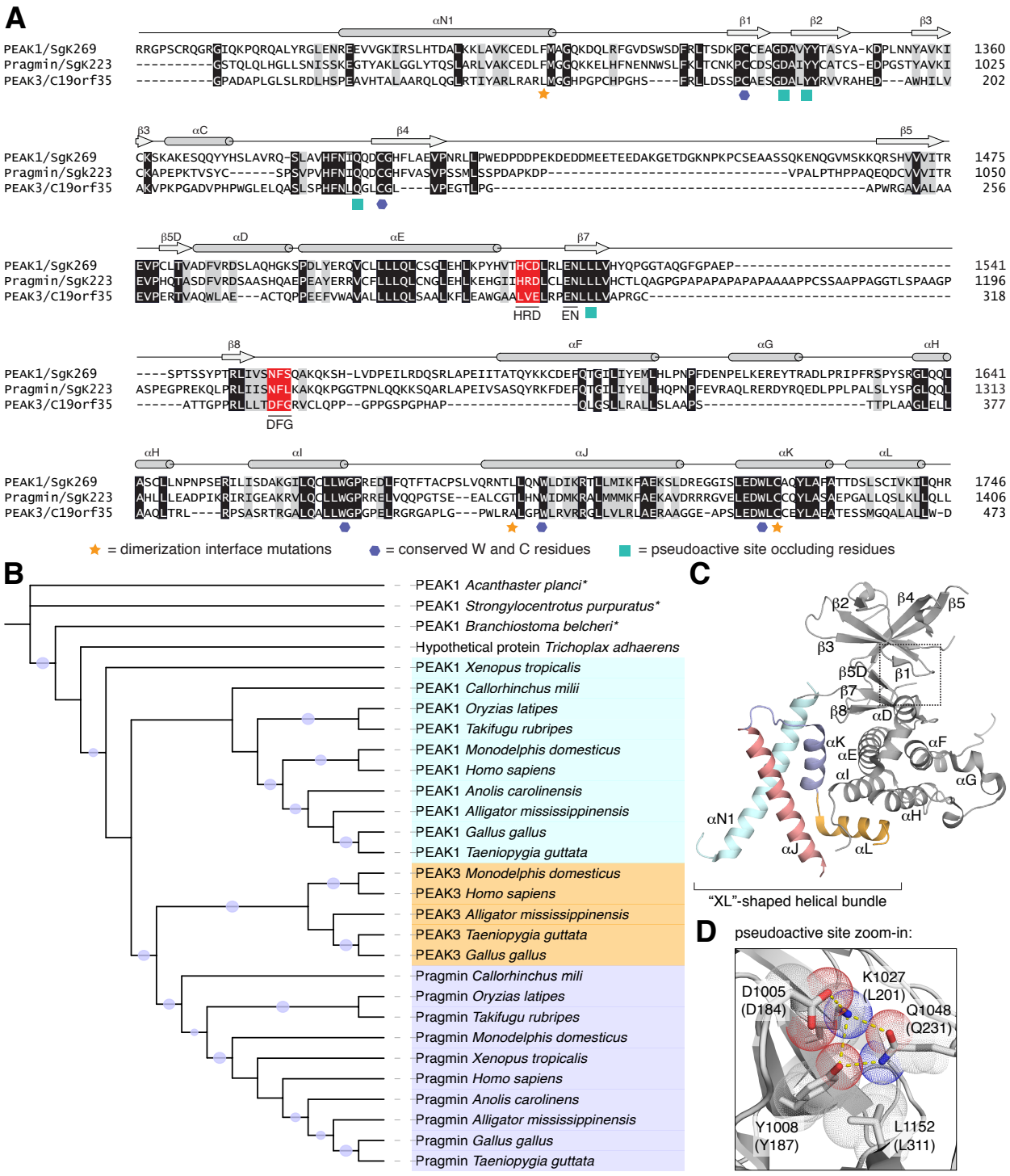


Fig. 3.1. PEAK3 (C19orf35) is a homolog of PEAK1 and Pragmin. (A) Protein sequence alignment for the kinase domains of human PEAK1/SgK269, Pragmin/SgK223 and PEAK3 (C19orf35). Secondary structure elements are denoted based on the Pragmin structure (PDB ID: 5VE6). Shading for conserved residues (black), conservative mutations (grey), and canonical sequence motifs of active kinases (red) are marked. Residues involved in dimerization, conserved W and C residues, and that occlude the pseudo-active site in Pragmin are marked by the indicated

symbols. **(B)** Phylogenetic tree (PhyML) for selected NKF3 kinases. Branches with bootstrap values greater than 70% are marked with circles; * = predicted proteins. **(C)** Cartoon representation of the crystal structure of the Pragmin SHED domain/pseudokinase module (PDB ID: 5VE6). **(D)** Zoomed-in view of the pseudoactive site in Pragmin depicting residues that occlude the canonical nucleotide-binding pocket. Top numbering corresponds to Pragmin residues (PDB ID: 5VE6), bottom numbering in parentheses to predicted corresponding residues in PEAK3. The Pragmin D184 residue was modeled in the active site based on the structure of Pragmin (PDB: 6EWX).

collectively termed as "inhibitory triad" that, in addition to other conserved residues, occlude the nucleotide binding pocket and prevent binding of ATP (**Fig. 3.1 C, D**)[5,8,9]. These residues are well-conserved in PEAK3 and are represented by D184, Y187, L201, Q231, and L311 (**Fig. 3.1 A, D**). Collectively, the extent of mutations in the key catalytic motifs strongly indicates that although PEAK3 diverges from other members of the NKF3 family, it still falls into the category of pseudokinases defined by Manning *et al*.[3] as kinases that lack one or more of the canonical catalytic sequence motifs. Hence, all NKF3 kinases carry pseudokinase characteristics and seem to derive from a common NKF3 pseudokinase ancestor in early Metazoans, as indicated by the sequence variability in the NKF3 proteins and reconstruction of the ancestral NKF3 sequence via the Ancescon method[31].

PEAK3 interactome

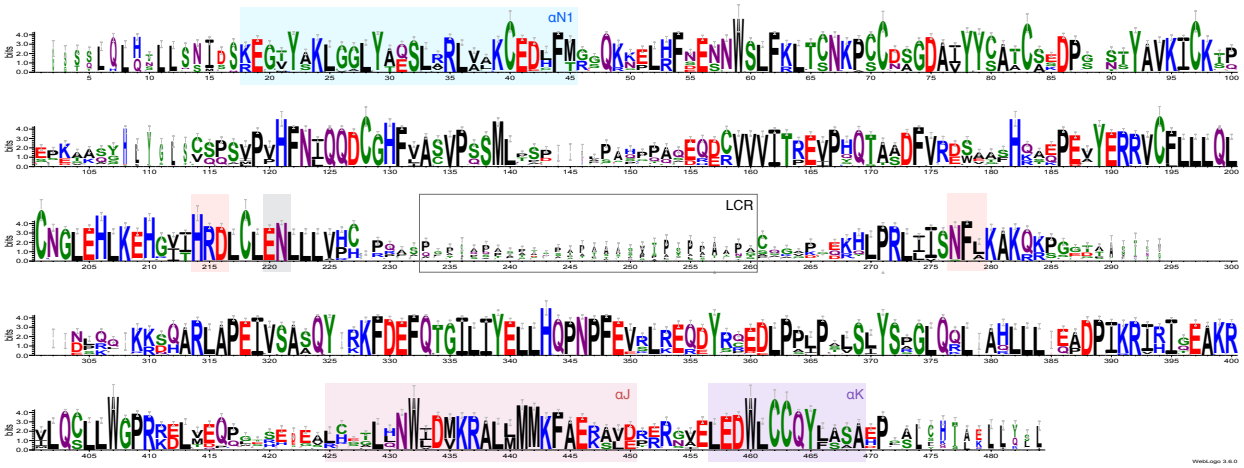
With no prior insights into PEAK3 function, we took an unbiased approach to identify its interacting partners using immunoprecipitation followed by mass spectrometry (IP/MS). Due to the current lack of a suitable antibody for detection of endogenous PEAK3, our analysis was conducted using Flag-tagged human PEAK3 transiently expressed in HEK293T cells. Identified proteins that co-immunoprecipitated with PEAK3-Flag can be categorized into several subgroups: (i) CrkII and CrkL, highly homologous adaptor proteins that regulate cell

proliferation, adhesion and cytoskeletal integrity downstream from receptor tyrosine kinases and integrins[32,33]; (ii) 14-3-3 scaffold proteins (β, γ, η, and τ), which play diverse roles in signaling, including regulation of cell motility, survival and intracellular protein trafficking[34-36]; (iii) guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) for Rho family of small GTPases, including ASAP1 that participates in actin cytoskeletal dynamics and cell movement[37,38]; and (iv) several proteins connected to the regulation of cell death and survival, including the SIAH1 ubiquitin ligase[39], DRAK1[40] and another 14-3-3 scaffold protein, 14-3-3σ, also known as SFN[41] (**Table 3.1)**. Collectively, this analysis suggests that PEAK3 is involved in a number of functions involved in cell proliferation, survival, and motility.

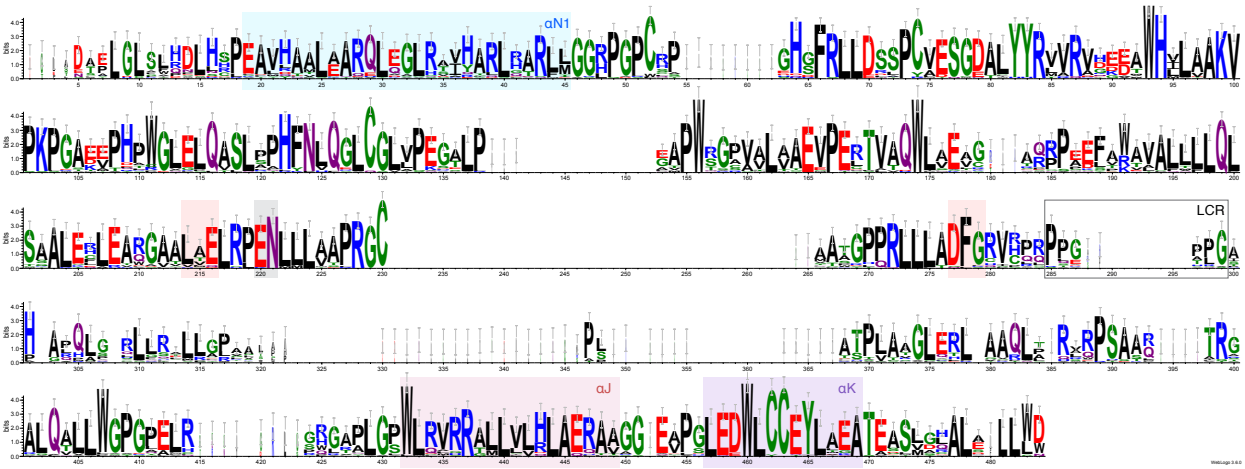
PEAK3 interacts with 14-3-3 and CrkII via its N-terminal domain.

We sought to verify the interactions identified by our IP/MS analysis by identifying the interactor binding sites in PEAK3. Primary sequence analysis of the N-terminal domain of PEAK3 revealed the presence of multiple putative binding sites, including a putative CrkII-binding site, PPPLPK, and a putative 14-3-3-binding site, RTQ(p)SLP, located 59 and 49 residues upstream from the predicted kinase domain, respectively (**Fig. 3.3 A**). These site closely resembles the consensus sequence present in known binding partners, such as BRAF in the case of 14-3-3- and the GEF proteins DOCK180 and C3G in the case of CrkII (**Fig. 3.3 B, C**)[14,42-44]. Furthermore, these putative binding motifs in PEAK3 are highly conserved across evolution, suggesting their potential importance in PEAK3 function (**Fig. 3.3 D**). Interestingly, similar motifs for both 14-3-3 and CrkII binding are also present in PEAK1 and Pragmin, and PEAK1 was previously shown to co-immunoprecipitate with CrkII[14].

A Pragmin/SgK223



B PEAK3/C19orf35



C PEAK1/Sgk269

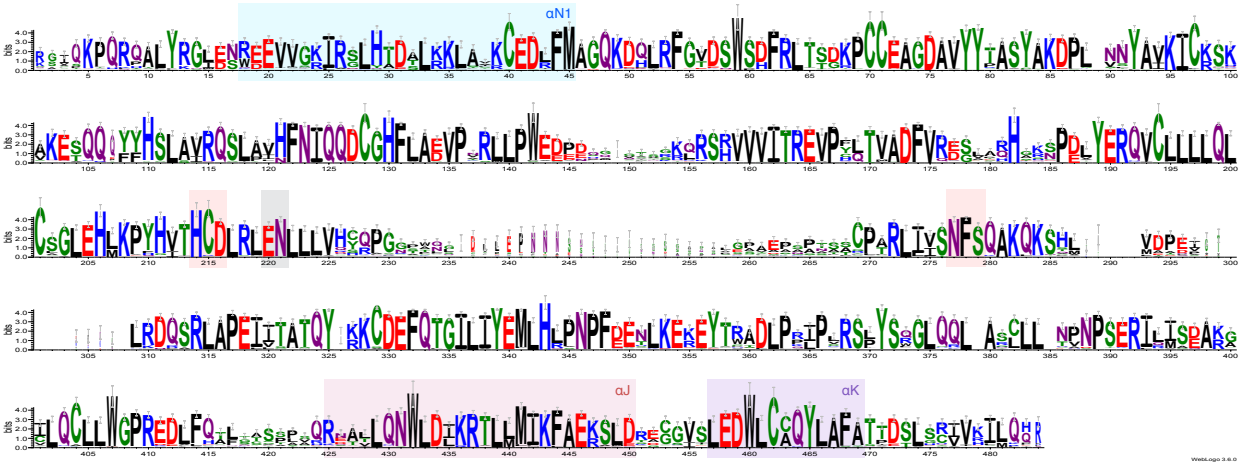


Fig. 3.2. Sequence conservation in the NKF3 family. Aligned sequence logos for **(A)** Pragmin/SgK223 sequences, **(B)** mammalian PEAK3/C19orf35 sequences, and **(C)** PEAK1/SgK269 sequences. The motifs corresponding to the canonical HRD active site (residues 214-216) and to the canonical DFG site (residues 277-279) are highlighted in red. The conserved EN motifs (residues 220-221) are highlighted in grey. The conserved SHED domain helices correspond to residues 18/19-45 (αN1 - highlighted blue), residues 425-250 (Pragmin, PEAK1) or 432-449 (PEAK3) (αJ - highlighted magenta), and residues 457-469 (αK - highlighted purple). Key LCR areas are boxed. Multiple sequence alignment (Promals3D) of the NKF3 family were split into three subfamily alignments with matched column numbering. Sequence logos were created using the WebLogo3 server.

Using co-immunoprecipitation, we verified that transiently expressed PEAK3 in HEK293 cells was able to bind to endogenous 14-3-3 (**Fig. 3.3 E**) and to both endogenously and exogenously expressed CrkII (**Fig. 3.3 F, G**). Mutation of residues in the predicted binding sites (14-3-3 binding site: R66A, S69A, and P71A in RTQSLP; CrkII-binding site: P56A, L59A, and P60A in PPPLPK) completely abolished the ability of mutant PEAK3 (PEAK3Δ14-3-3 and PEAK3-3A) to co-immunoprecipitate with 14-3-3 and CrkII (**Fig. 3.3 E - G**). Interestingly, PEAK3 was able to co-immunoprecipitate with both 14-3-3 and CrkII simultaneously despite the proximity of their binding sites. Furthermore, mutation of either binding site in PEAK3 did not perturb its interaction with the other binding partner (**Fig. 3.3 H**). Only mutation of both sites simultaneously (PEAK3-DK) abolished binding of both 14-3-3 and CrkII (**Fig. 3.3H**).

Given the documented roles of NKF3 family members in the regulation of cellular motility[11-16], we focused further PEAK3 functional studies on its interaction with the adaptor protein CrkII, which was one of the most abundant PEAK3-interacting proteins identified by the IP/MS analysis measured as a high confidence score via the comparative proteomic analysis software suite (ComPASS[45])(**Table 3.1**). CrkII is composed of an SH2 domain followed by two SH3 domains, termed N-terminal SH3 domain ($SH3^N$) and C-terminal SH3 domain ($SH3^C$). The inability of the PEAK3-3A mutant to bind CrkII suggested that PEAK3 interacts specifically

with the $SH3^N$ in CrkII, which has previously been shown to engage similar proline-rich motifs in other CrkII-binding partners[43,44,46]. Indeed, mutation of the $SH3^N$ (CrkII-W170K) but not of the $SH3^C$ domain (CrkII-W276K) rendered CrkII unable to bind PEAK3 (**Fig. 3.3 I**), indicating that PEAK3 likely interacts with CrkII in a canonical manner.

PEAK3 antagonizes CrkII-induced changes in cellular morphology.

One well-characterized role of other NFK3 family members is the regulation of cell morphology and migration. Both PEAK1 and Pragmin localize to actin filaments and focal adhesions, induce cell elongation, and promote cell migration when transiently expressed in cells[11-16,23]. In contrast, we found that PEAK3 distributes diffusely throughout the cytoplasm of COS-7 cells and U2OS cells upon transient transfection and had little to no effect on overall cell shape compared to vector-transfected control cells (**Fig. 3.4 A; Fig. 3.5 A**). However, while the majority of control cells retained prominent actin filaments that traversed the cell[47,48], cells overexpressing PEAK3 exhibited notably fewer stress fibers and possessed shorter, less organized actin filaments, mirroring the phenotype typically observed in cells in which the *CrkII* gene is knocked down (**Fig. 3.4 B**)[49].

Endogenous CrkII participates in actin polymerization and positively regulates cell motility[48-53]. Exogenously expressed CrkII localizes to the cell cortex where it induces a spindle-shaped cell morphology with notable membrane extensions resembling lamellipodia or polarized membrane ruffles[32,53-57]. These morphological changes can be visualized robustly in COS-7 and U2OS cells[49,52,58,59], hence we used these cell lines to measure the functional consequences of PEAK3 overexpression on CrkII-dependent effects on cell morphology. Remarkably, cells co-expressing PEAK3 and CrkII had markedly fewer membrane extensions and largely did not

Table 3.1. Top interactors of PEAK3 as identified by IP/MS. Top interactors of PEAK3 identified by the IP/MS analysis, including their Uniprot ID and abundance score. The details of the analysis are described in Methods.

Hit #	Name	Uniprot ID	ComPASS score
1	C19orf35	Q6ZS72	711.90724
2	c-Crk	P46108	557.91605
3	14-3-3 protein eta	Q04917	161.16351
4	SIAH1	Q8IUQ4	161.05648
5	14-3-3 protein σ	P31947	140.33994
6	14-3-3 protein β/α	P31946	45.04511
7	14-3-3 protein γ	P61981	33.21708
8	14-3-3 protein ε	P62258	32.00987
9	FGD3	Q5JSP0	31.58578
10	14-3-3 protein ζ/δ	P63104	26.21800
11	14-3-3 protein θ	P27348	23.02257
12	STK17A	Q9UEE5	12.89484
13	CrkL	P46109	11.57874
14	ZFP791	Q3KP31	11.16726
15	HSPA1A/HSPA1B	P08107	9.53939
16	ASAP1	Q9ULH1	9.11803
17	HSP7C	P11142	9.01850
18	Tubulin β-5 chain	P07437	8.02081
19	HSPA2	P54652	7.74915
20	Tubulin β-4B chain	P68371	7.59386
21	Tubulin β-2B chain	Q9BVA1	6.97615
22	Tubulin β-4A chain	P04350	6.92820
23	Tubulin β-2A chain	Q13885	6.87992
24	Tubulin α-1B chain	P68363	6.75771
25	HSPA1L	P34931	6.73300

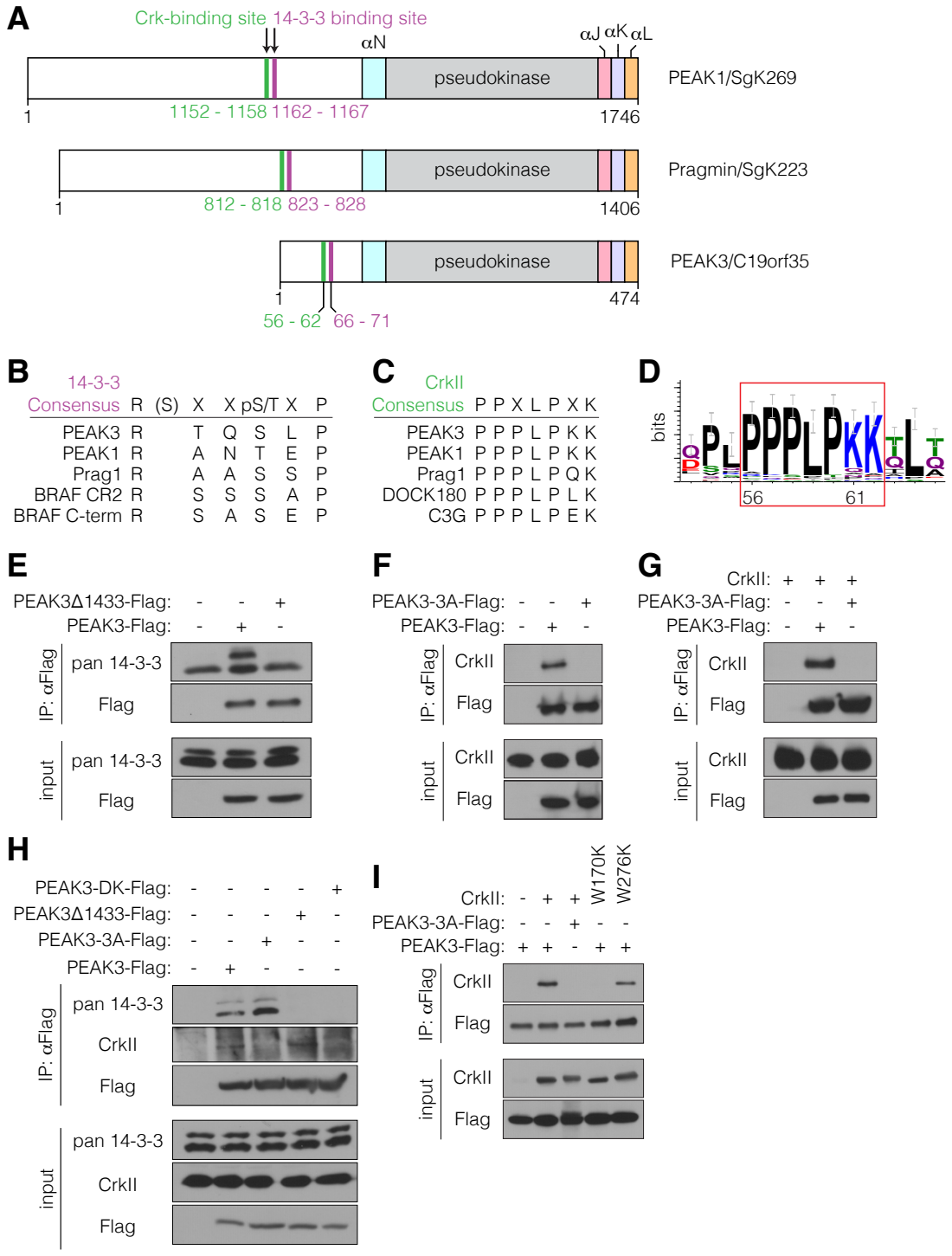


Fig. 3.3. PEAK3 binds 14-3-3 and CrkII. (A) Schematic representation of PEAK1, Pragmin, and PEAK3 domain structure. The locations of the 14-3-3-binding sites, CrkII-binding sites, and helical regions within the SHED domain are highlighted. **(B, C)** Consensus sequence of 14-3-3-binding sites (B) and CrkII-binding sites (C) in selected proteins. **(D)** Sequence logo depicting conservation of the CrkII-binding site in PEAK3 homologs. **(E-I)** Co-immunoprecipitation of Flag-tagged wild-type and mutant PEAK3 transiently expressed in

HEK293 cells. Interaction with endogenous 14-3-3 (E, G), endogenous CrkII (F, G), transiently expressed untagged wild-type CrkII (H), or transiently expressed untagged CrkII variants carrying mutations in the SH3 domains (W170K in $SH3^N$ and W276K in $SH3^C$) (I) were detected with the indicated antibodies by Western blot analysis.

adopt a CrkII-dependent morphology (**Fig. 3.4 A; Fig. 3.5 B**). In contrast, the PEAK3-3A mutant, which does not interact with CrkII, was unable to interfere with the CrkII-dependent phenotype (**Fig. 3.4 A; Fig. 3.5 B**). To quantitatively compare these differences in cellular phenotypes, we developed a metric in which the effect of CrkII on cell morphology is measured as an increase in cell perimeter (**Fig. 3.4 C**). While CrkII overexpression alone significantly increases cell perimeter, there is no change in cell perimeter when CrkII is co-expressed with wild type PEAK3 (**Fig. 3.4 D; Fig. 3.5 C**). PEAK3-CA mutant has no effect on CrkII-dependent increase in cell perimeter, supporting a conclusion that PEAK3 negatively regulates CrkII as a result of their direct interaction.

Negative regulation of CrkII by PEAK3 requires the C-terminal domain.

The PEAK3 protein can be arbitrarily subdivided into two distinct domains: the N-terminal domain that contains the CrkII-binding site and the C-terminal domain that contains the pseudokinase domain. We exogenously expressed the N-terminal domain of PEAK3 to determine if it alone is sufficient for the negative regulation of CrkII. Unexpectedly, a construct containing only the N-terminal domain (PEAK3 ΔPK) did not antagonize CrkII function (**Fig. 3.6 A-C**). Even more surprisingly, the PEAK3 ΔPK mutant was also unable to interact with CrkII, despite the presence of the intact CrkII-binding motif (**Fig. 3.6 D**). Hence, while the CrkII-binding motif is necessary to facilitate binding of PEAK3 to CrkII, it is not sufficient for CrkII

binding and inhibition, suggesting that the C-terminal domain of PEAK3 also plays an essential role in mediating the interaction between PEAK3 and CrkII.

Predicted SHED domain in PEAK3.

Recent crystal structures of the PEAK1 and Pragmin C-terminal domains have revealed identical dimer forms composed of the pseudokinase domains interacting through a set of helical bundles, termed the Split Helical Dimerization (SHED) domain[5,8,9,15]. The SHED domain is unique to NKF3 proteins and is comprised of the helix immediately N-terminal to the pseudokinase domain (αN helix) and three helices C-terminal to the pseudokinase domain (αJ, αK, αL helices) that form an "XL"-shaped helical bundle (lettering of helices reflect Pragmin nomenclature[8]; PEAK1 helices are off-set by one letter from the αJ helix[9])[5,8,9].

In PEAK3, the highest sequence similarity with PEAK1 and Pragmin, apart from the pseudoactive site, falls within the regions corresponding to the SHED domain. In the three C-terminal α-helices, the sequence identity between PEAK3 is 34% and 32% with Pragmin and PEAK1, respectively (**Fig. 3.2; Fig. 3.7**). The most striking conserved sequence motifs include those containing tryptophan residues, such as the EDWLCC sequence in the αK helix, WGP in the loop preceding the αJ helix, and WL in the αJ helix (**Fig. 3.1 A, C; Fig. 3.2**). In the PEAK1 and Pragmin structures, these conserved motifs are involved in the interactions between the pseudokinase domain and the SHED domain[5,8,9]. PEAK3 has remarkably well-conserved sequences in these regions, indicating that the SHED domain is likely present in PEAK3, which we anticipate to adopt a similar structure as the ones in PEAK1 and Pragmin (**Fig. 3.7**). Thus, the SHED domain emerges as a unifying structural feature of the NKF3 family of pseudokinases.

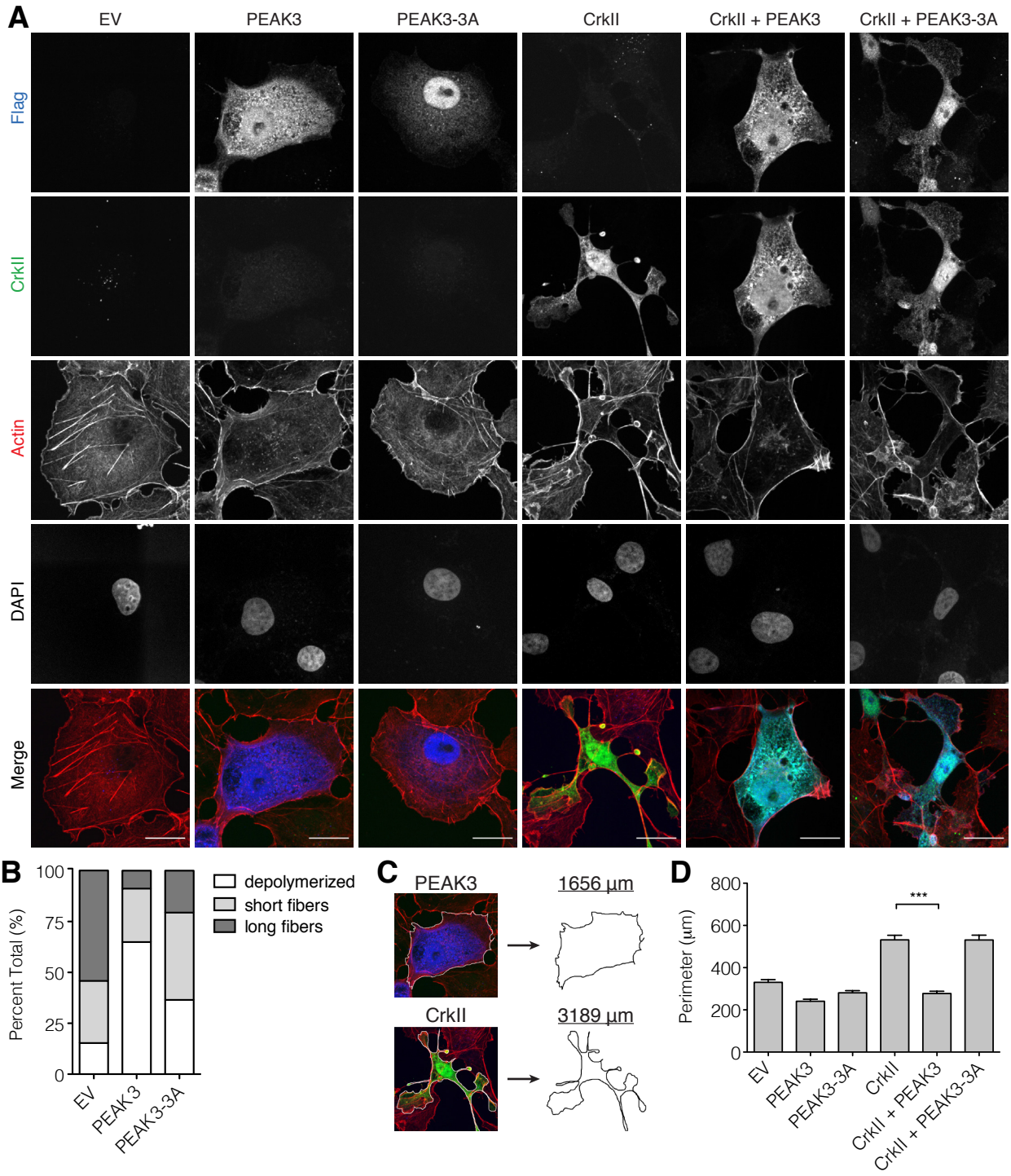


Fig. 3.4. PEAK3 prevents CrkII-dependent membrane ruffling and lamellipodia-like extensions. (A) Confocal microscopy imaging of COS-7 cells transiently co-transfected with wild type PEAK3 or CrkII-binding deficient mutant (PEAK3-3A) and either empty vector (EV) or untagged CrkII. CrkII was detected with an anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin (red). All scale bars correspond to 20 µm. **(B)** Relative percentage of actin fiber phenotypes measured in COS-7

cells transiently transfected with PEAK3 and stained with Alexa Fluor 647-conjugated phalloidin (n = 60 cells per group). Cells were blindly scored and binned based on the extent of visible actin fibers within the cytosol: (1) prominent fibers that traversed over 50% of the cell, (2) short, stochastic fibers, and (3) no significant amount of polymerized actin. **(C)** Schematic illustrating perimeter calculation in representative cells from (A). **(D)** Average perimeter of COS-7 cells imaged in (A), quantified as described in Methods. Data represent the mean ± SEM of three independent experiments (n = 20 cells in each experiment), *** $p < 0.001$.

PEAK3 dimerizes via the SHED domain.

PEAK1 and Pragmin form homo- and hetero-oligomers through two distinct dimer interfaces, one involving the SHED domain and another involving the αG helix/A-loop interface[5,8,15]. Mutation of these interfaces, especially of the hydrophobic interactions between the helices in the SHED domain, impairs the signaling ability of these pseudokinases[8,9]. While the SHED domain in PEAK3 is highly similar to the SHED domains of PEAK1 and Pragmin (**Fig. 3.1 A)**, the αG helix/A-loop interface is not significantly conserved in PEAK3. This lead us to hypothesize that PEAK3 too could dimerize via its putative SHED domain and that this might be critical for its interaction with CrkII.

To assess the ability of PEAK3 to homodimerize, we co-expressed HA- and FLAG-tagged wild type PEAK3 variants in HEK293 cells and assessed their ability to associate by co-immunoprecipitation. These two differentially tagged PEAK3 constructs robustly co-immunoprecipitated (**Fig. 3.8 A, B**). Based on the crystal structures of Pragmin and PEAK1, we designed PEAK3 mutants that carried either individual substitutions of key residues in the predicted dimerization interface or deletion of one of the αN, αJ, or αK helices in the SHED domain (**Fig. 3.8 C**). These mutations/deletions abolished the ability of differentially tagged PEAK3 variants to interact (**Fig. 3.8 D - G**), while control mutations of residues located within the SHED domain helices but distal from the dimer interface had no effect (**Fig. 3.8 H**). These

data suggest that PEAK3 self-associates through the SHED domain in a manner analogous to PEAK1 and Pragmin.

Dimerization of PEAK3 is necessary for CrkII binding and its negative regulation.

While the C-terminal domain of PEAK3 does not contain a CrkII binding motif, it plays an essential role in CrkII binding and inhibition, as demonstrated by our finding that the construct missing this domain, PEAK3 ΔPK, was unable to engage with and antagonize CrkII-induced membrane ruffling (**Fig. 3.6**). To test if this role of the C-terminal domain is linked to its ability to mediate PEAK3 self-association through the SHED domain, we measured how mutations within the SHED domain dimer interface affect CrkII binding. All mutations that compromise PEAK3 dimerization also interfered with the ability of PEAK3 to interact with CrkII (**Fig. 3.9 A - C**). Consequently, the PEAK3 dimerization mutants also failed to inhibit the formation of membrane extensions and membrane ruffles caused by CrkII overexpression (**Fig. 3.9 D, E; Fig. 3.10**).

Our data thus far demonstrate that the N-terminal domain of PEAK3 containing the CrkII binding motif is unable to bind to or interfere with CrkII function when not dimerized by the C-terminal domain, which contains the pseudokinase domain. We wondered if the functional effect of the C-terminal domain-mediated dimerization could be mimicked by substitution of the C-terminal domain with an orthogonal domain that drives constitutive PEAK3 dimerization. To test this, we fused the region N-terminal region of the SHED domain αN helix (residues 1-131) containing the CrkII-binding site, but not the SHED domain or the pseudokinase domain, to a constitutively dimeric coiled coil domain (PEAK3-diCC) (**Fig. 3.11 A**). While the monomeric N-terminal PEAK3 ΔPK construct is unable to bind CrkII (**Fig. 3.6 D**), equivalent levels of the

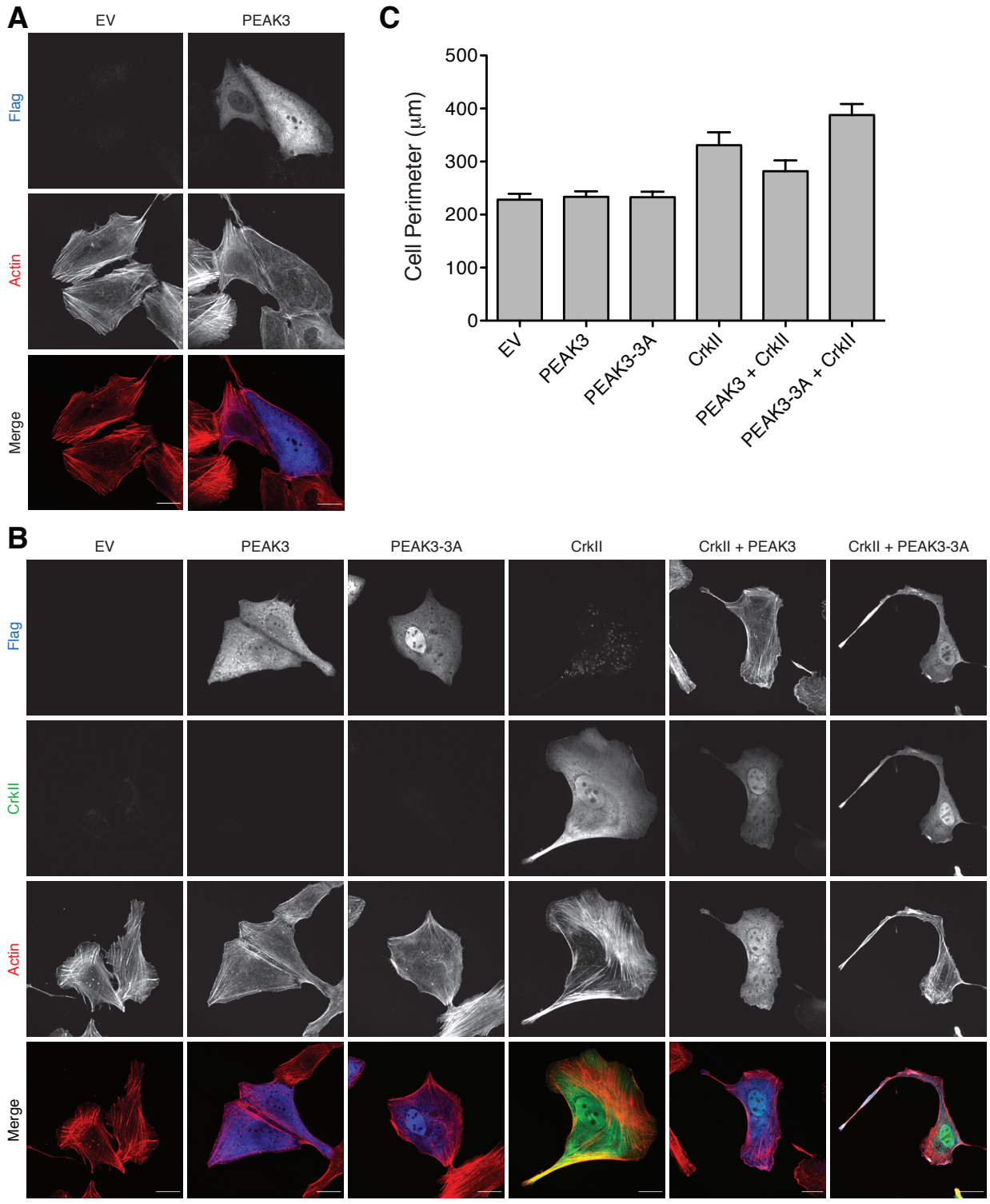


Fig. 3.5. PEAK3 antagonizes CrkII-dependent cell ruffling in U2OS cells. (A) Confocal microscopy imaging of U2OS cells transiently co-transfected with wild type PEAK3. PEAK3 was detected with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 µm. **(B)** Confocal microscopy imaging of U2OS cells transiently co-transfected with wild type PEAK3 or a CrkII-binding deficient

mutant (PEAK3-3A) with either empty vector (EV) or untagged CrkII. CrkII was detected with an anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 μm. **(C)** Average perimeter of COS-7 cells imaged in (B), quantified as described in Methods. Data represent the mean ± SEM of 3 independent experiments (n = 20 cells in each experiment).

immunoprecipitated PEAK3-diCC fusion show notable CrkII binding (**Fig. 13.1 B**). The ability of this minimal construct to restore CrkII binding underscores the importance of dimerization of the CrkII binding motif as a determinant of interaction between CrkII and PEAK3, and possibly also between CrkII and its other interaction partners. To our knowledge such property in known CrkII-binding proteins has not been described.

Despite the ability to interact with CrkII, PEAK3-diCC did not interfere with ability of CrkII to drive formation of membrane ruffling in cells, in contrast with cells expressing wild type PEAK3 (**Fig. 3.11 C, D**). The discrepancy observed between PEAK3-diCC interaction with CrkII and inability to antagonize CrkII activity possibly reflects the weaker binding between CrkII and PEAK3-diCC compared to the wild type PEAK3 (**Fig. 3.11 B**). Alternatively, or concurrently, the C-terminal pseudokinase/SHED module in PEAK3 might play a role in antagonizing CrkII signaling that extends beyond serving as a dimerization domain for the CrkII binding motif. However, additional studies are necessary in order to parse out any further role the SHED domain may play in regulating the activity of NKF3 members.

Mutation of the DFG aspartate impairs CrkII regulation by PEAK3.

NKF3 kinases have evolved multiple sequence alterations within their pseudokinase domains in comparison to active kinases that are predicted to render them catalytically inactive. One notable difference between PEAK3 and PEAK1 and Pragmin is conservation of the DFG

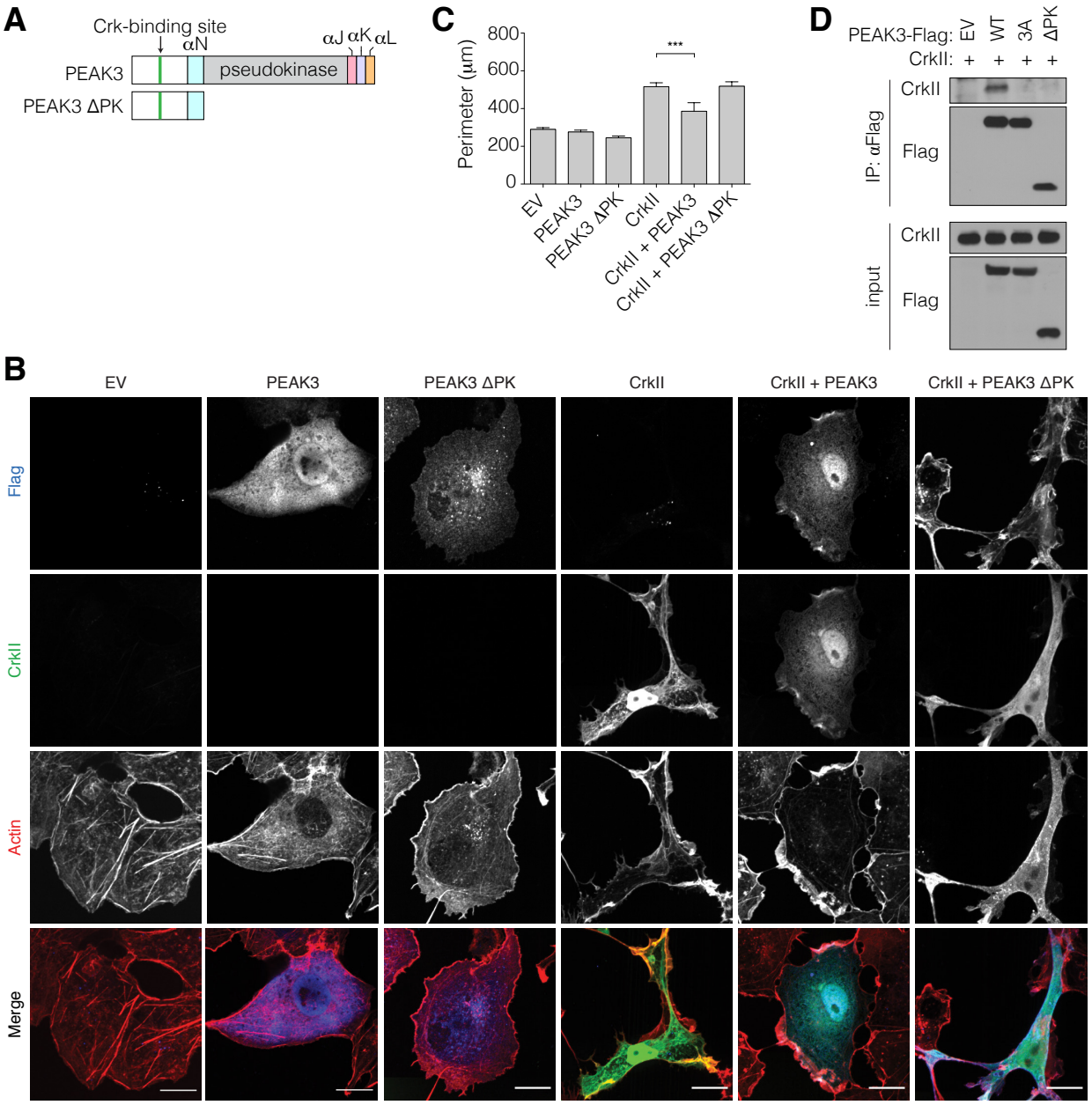


Fig. 3.6. CrkII binding motif is insufficient for PEAK3-dependent inhibition of CrkII. (A) Schematic representation of PEAK3 constructs used. **(B)** Confocal microscopy imaging of COS-7 cells transiently co-transfected with wild type PEAK3 or a variant lacking the predicted pseudokinase domain (PEAK3 ΔPK) with either empty vector (EV) or untagged CrkII. CrkII was detected with an anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 μm. **(C)** Average perimeter of cells imaged in (B), quantified as described in Methods. Data represent the mean ± SEM of three independent experiments (n = 20 cells in each experiment), *** $p < 0.001$. **(D)** Co-immunoprecipitation of Flag-tagged wild type and ΔPK PEAK3 variants with untagged CrkII. Proteins were transiently expressed in HEK293 cells, and protein levels were detected with the indicated antibodies by Western blot analysis.

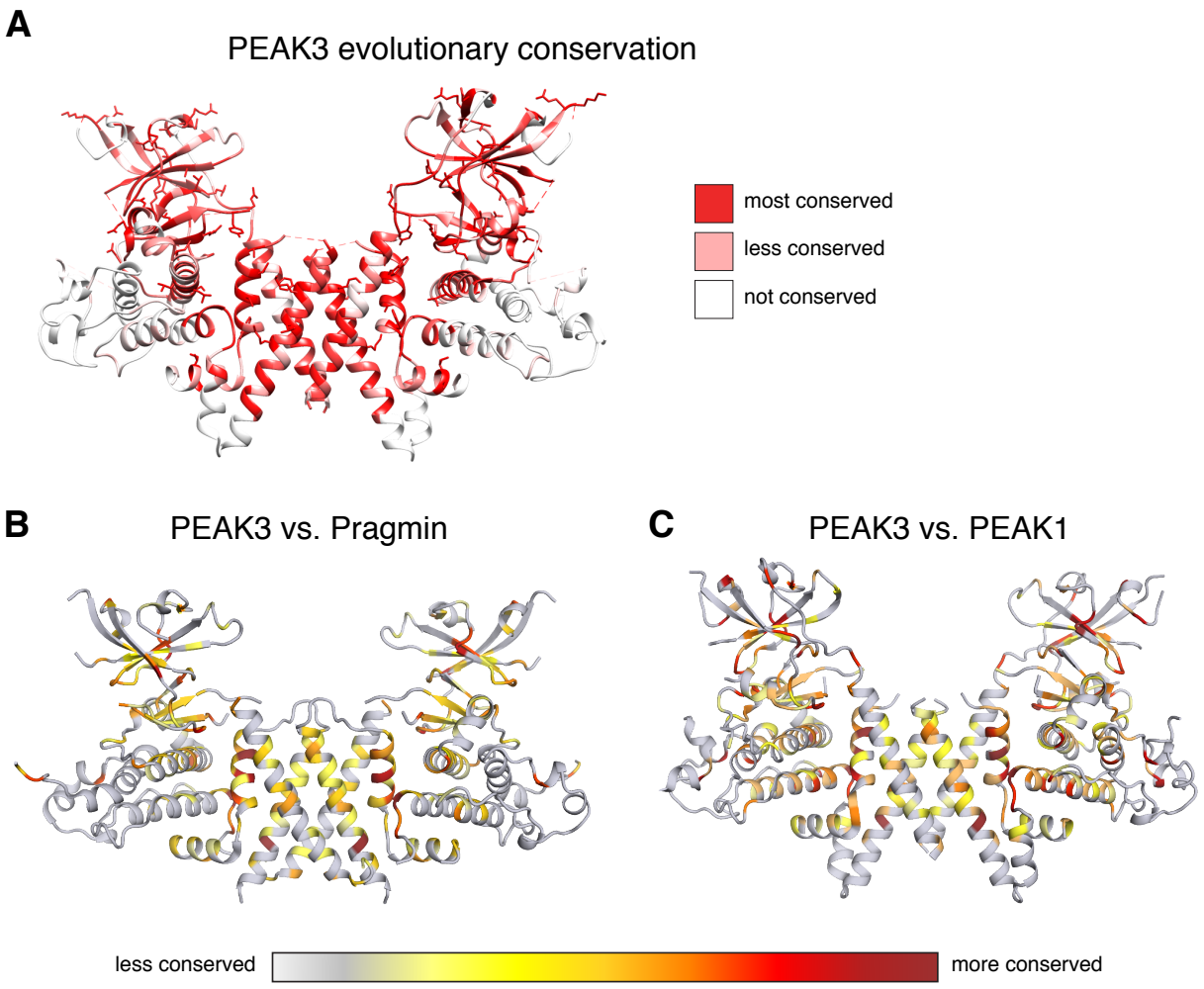


Fig. 3.7. Sequence conservation in the NKF3 family mapped onto 3D structures. (A) Sequence conservation within the identified mammalian PEAK3 homologs across evolution mapped onto the PEAK1 structure. Alignment conservation values are represented by different coloring (lowest conservation = white, highest conservation = red). **(B, C)** Similarity between human Pragmin and PEAK3 (B) or human PEAK1 and PEAK3 (C) mapped onto the three-dimensional structures (B: Pragmin, PDB: 5VE6; C: PEAK1, PDB: 6BHC). Pairwise alignments of PEAK3 with Pragmin (B) or PEAK1 (C) were considered for analysis. Residues either strictly conserved or conservative replacements, as judged by positive BLOSUM62 matrix scores, were rendered as ribbons and colored according to the BLOSUM62 scores (yellow: BLOSUM62 values from 1 to 3; orange: 4-6; red (highest conservation): 7 to 11).

motif in PEAK3, which is degraded to an NFS and NFL motif in PEAK1 and Pragmin, respectively. The DFG motif is also present in a number of human pseudokinases, although its importance for their function is unclear[3]. In active kinases, the aspartate residue within the DFG

motif plays a critical role in catalysis by coordinating Mg^{2+} ions[60]. In a subset of these kinases, the DFG motif is found to adopt two distinct conformations, DFG-in and DFG-out, which are coupled to conformational changes in other regions of the kinase domain[61]. The interactions made by the DFG aspartate are prerequisite for these changes[62]. It is therefore possible that the conformation adopted by the DFG motif in pseudokinases that possess the conserved sequence, such as PEAK3, is coupled to functional conformational changes in the other regions of their pseudokinase domain.

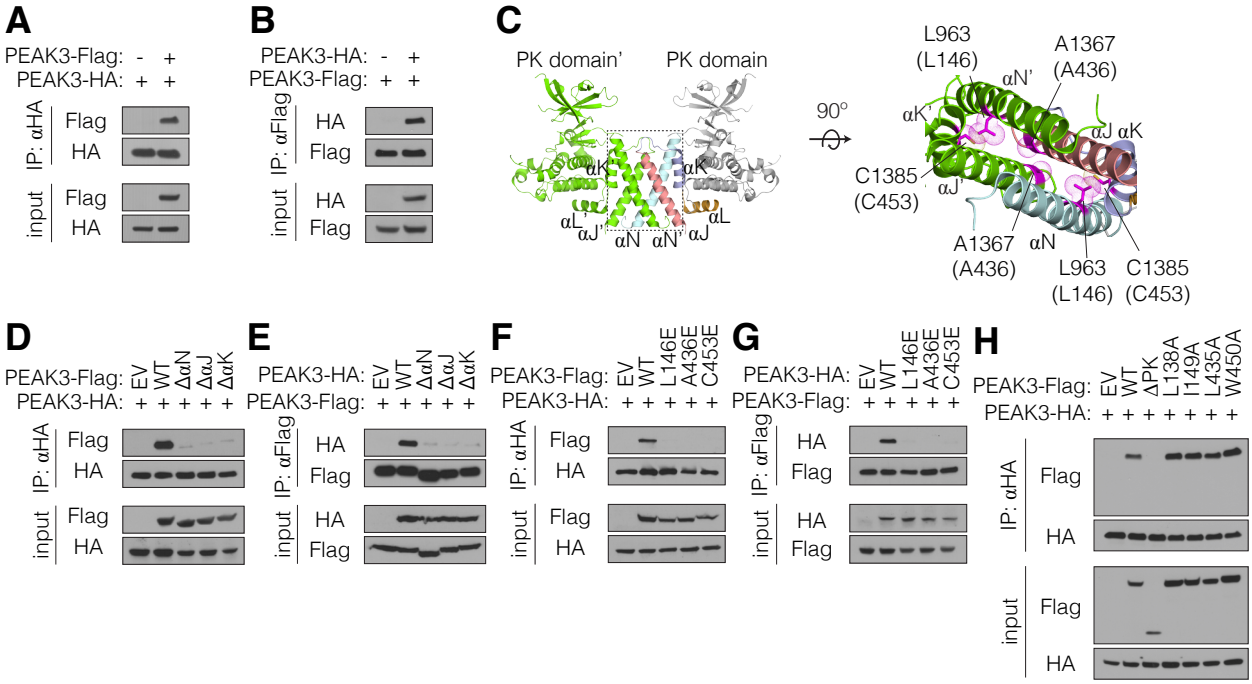


Fig. 3.8. PEAK3 dimerization via the pseudokinase/SHED module is necessary for binding and inhibition of CrkII. (A, B) Co-immunoprecipitation of Flag-tagged and HA-tagged variants of wild type PEAK3. Proteins were transiently expressed in HEK293 cells, and protein levels were detected with the indicated antibodies by Western blot analysis. **(C)** Top panel: Pragmin SHED/pseudokinase domain (PK) dimer structure (PDB ID: 5VE6). Bottom panel: dimerization interface and interactions between helices within the SHED domain. Residues colored in magenta were selected for mutagenesis in PEAK3 based on sequence homology. Top numbering corresponds to Pragmin residues (PDB ID: 5VE6), bottom in parentheses to predicted PEAK3 residues. **(D-H)** Co-immunoprecipitation of wild type PEAK3 and mutants that inhibit dimerization (D-G) or control mutants that are distal to the dimerization interface (H). Proteins were transiently expressed in HEK293 cells, and protein levels were detected with the indicated antibodies by Western blot analysis.

We examined the importance of the DFG motif in PEAK3's function as a negative regulator of CrkII by mutating the DFG aspartate to asparagine (PEAK3 D330N) and testing the ability of this mutant to homodimerize, interact with CrkII, and inhibit CrkII-dependent membrane ruffles in cells. Interestingly, PEAK3 D330N did not dimerize as efficiently as wild type PEAK3 (**Fig. 3.12 A**). Consistent with our observation that PEAK3 dimerization is necessary for CrkII binding, PEAK3 D330N exhibited markedly impaired binding to CrkII (**Fig. 3.12 B**) and did not inhibit CrkII-dependent morphological changes to cell shape to the same extent as wild type PEAK3 (**Fig. 3.12 C, D**). These data show that the integrity of the DFG motif is essential for PEAK3 function as a negative regulator of CrkII.

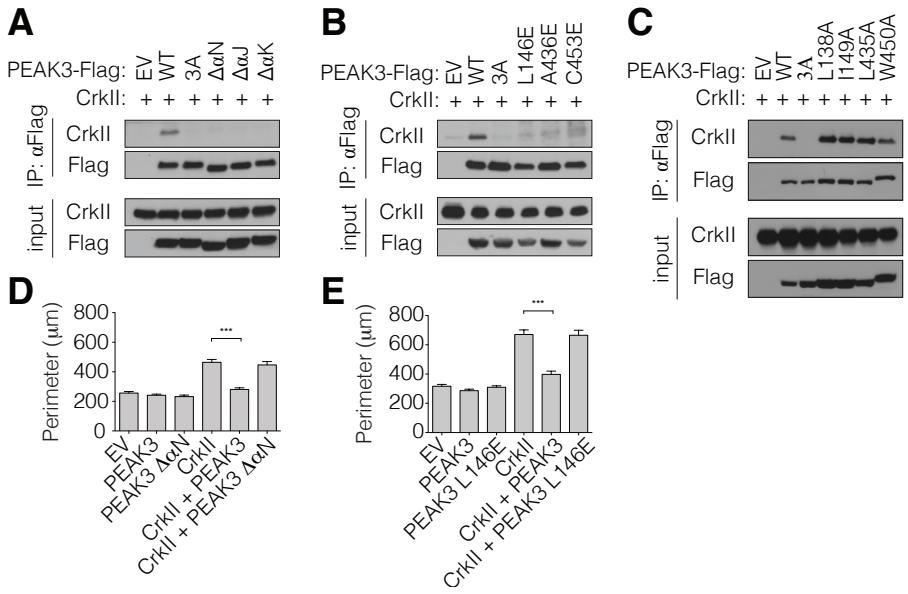


Fig. 3.9. Homotypic association of PEAK3 is required for its interaction with and antagonization of CrkII. (A-C) Co-immunoprecipitation of untagged CrkII with dimerization-deficient mutants (A, B) or control mutants (C) of PEAK3. Proteins were transiently expressed in HEK293 cells, and protein levels were detected with the indicated antibodies by Western blot analysis. **(D, E)** Average perimeter of COS-7 cells expressing wild type PEAK3 or dimerization-deficient variants of PEAK3 with either empty vector (EV) or untagged CrkII. Data represent the mean ± SEM of three independent experiments (n = 20 cells in each experiment), *p < 0.05, *** p < 0.001.

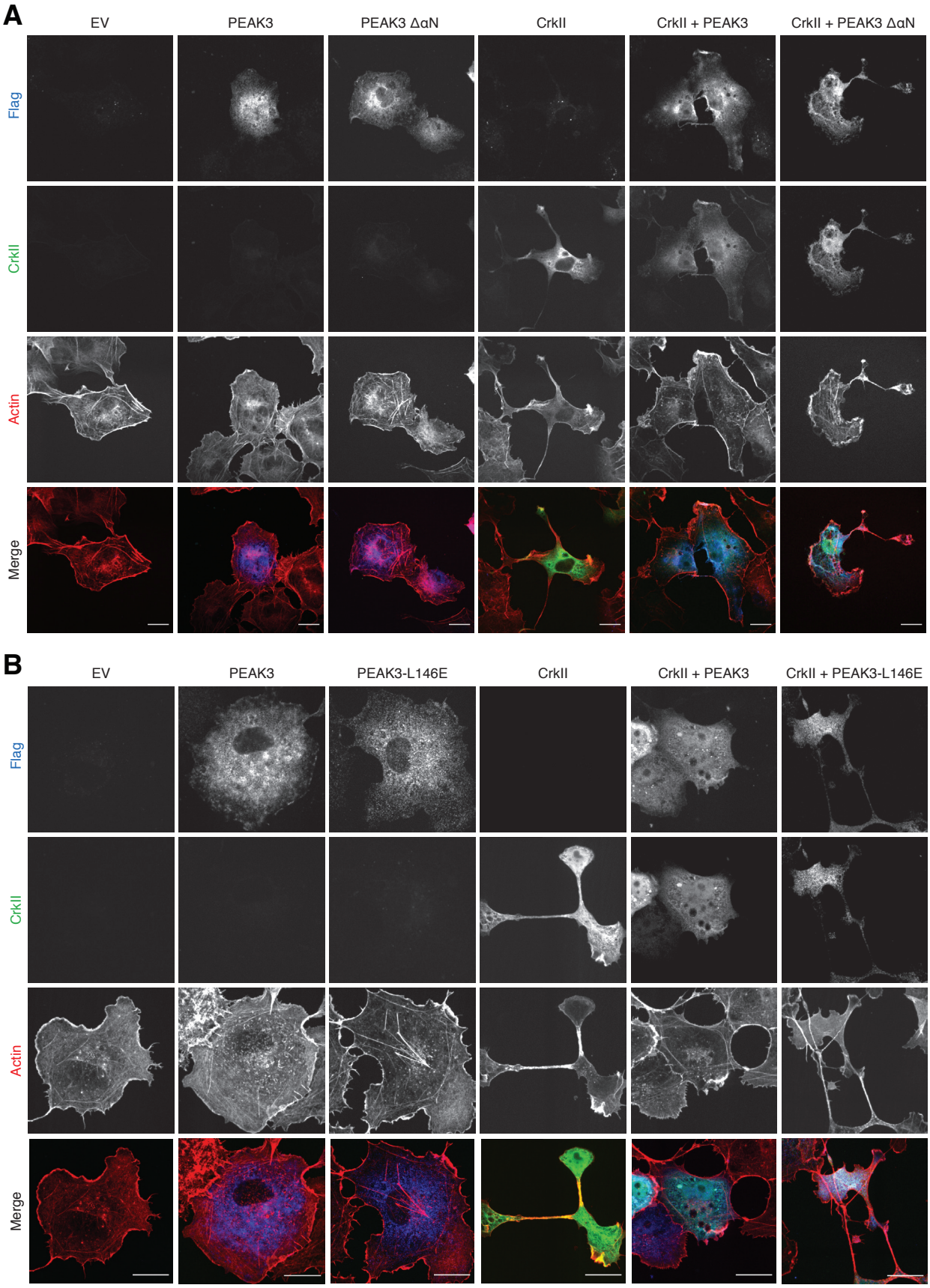

A
EV
PEAK3
PEAK3 ΔαN
CrkII
CrkII + PEAK3
CrkII + PEAK3 ΔαN
Flag
CrkII
Actin
Merge
B
EV
PEAK3
PEAK3-L146E
CrkII
CrkII + PEAK3
CrkII + PEAK3-L146E
Flag
CrkII
Actin
Merge

Fig. 3.10. Homotypic association of PEAK3 is required to antagonize CrkII function. (**A, B**) Confocal microscopy imaging of COS-7 cells transiently co-transfected with wild type PEAK3 or a ΔαN1 (A) or L146E (B) dimerization mutant of PEAK3 with either an empty vector (EV) or untagged CrkII. CrkII was detected with an anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 μm.

Interaction with 14-3-3 may regulate PEAK3 function.

We reasoned that the necessity of the PEAK3 pseudokinase domain in driving its ability to antagonize CrkII might further be linked to its own regulation. Our previous co-immunoprecipitation experiments confirmed that PEAK3 is able to bind to a number of 14-3-3 isoforms (**Fig. 3.3 E**), which have been well characterized as regulatory proteins that modulate signaling through subcellular localization and allosteric interactions[34-36]. We reasoned that the conformation of the pseudokinase domain might be necessary for PEAK3's interaction with 14-3-3. Using co-immunoprecipitation, we confirmed that, as with CrkII, the ability of PEAK3 to bind 14-3-3 is dependent upon the presence of its pseudokinase domain and requires an intact DFG motif (**Fig. 3.13 A**). Furthermore, PEAK3 dimerization, which is driven through the SHED domain, is also critical for its ability to engage 14-3-3 (**Fig. 3.13 A**). However, it remained unclear how 14-3-3 binding to PEAK3 may regulate its interaction with CrkII at the molecular level. We therefore sought to visualize the interaction between PEAK3 and 14-3-3 by obtaining a molecular structure of these proteins in complex. Recombinant PEAK3 co-purifies with endogenous 14-3-3 and elutes on a size exclusion column at a volume consistent with dimeric PEAK3 and 14-3-3, which form obligate dimers (**Fig. 3.13 B - D**). Initial visualization of this complex by negative stain electron microscopy confirms that such a composition is likely the case (**Fig. 3.13 E**), however further experiments using cryo-electron microscopy, which can resolve structures to a higher resolution, are necessary to understand the precise molecular

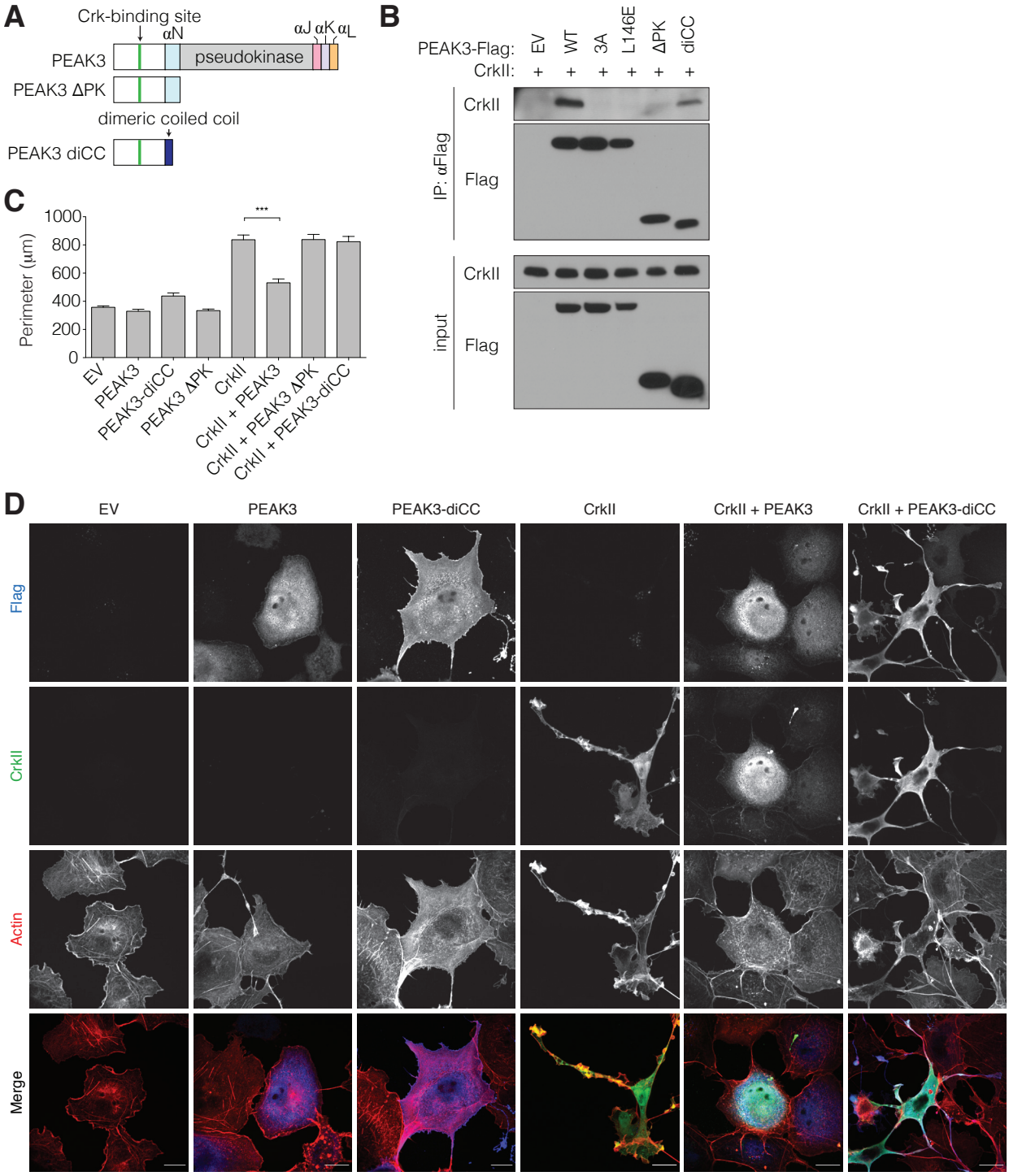


Fig. 3.11. Constitutive dimerization of the CrkII binding site is not sufficient to antagonize CrkII function. (A) Schematic representation of PEAK3 constructs used. **(B)** Co-immunoprecipitation of Flag-tagged wild-type PEAK3 or dimerization-deficient mutants of PEAK3 with untagged CrkII. Proteins were transiently expressed in HEK293 cells, and protein levels were detected with the indicated antibodies by Western blot. **(C, D)** Average perimeter (C) and confocal microscopy (D) of COS-7 cells transiently expressing wild type PEAK3 or a diCC variant of PEAK3 with either empty vector (EV) or untagged CrkII. CrkII was detected with an

anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 μm. Average perimeter data represent the mean ± SEM of three independent experiments (n = 20 cells in each experiment), *** $p < 0.001$.

interaction between 14-3-3 and PEAK3. Furthermore, it would be interesting to examine whether endogenous CrkII is present within these purified complexes, as a molecular structure of the three proteins together would be critical for understanding how 14-3-3 may regulate PEAK3's ability to regulate CrkII.

DISCUSSION

The PEAK1 and Pragmin pseudokinases have grown in prominence due to the key roles they play in the regulation of cellular motility and oncogenesis[18]. Here, we present evidence that PEAK3 is a close homolog of PEAK1 and Pragmin that likely has evaded annotation as a kinase due to the high LCR content within its pseudokinase domain. Despite this difference, PEAK3 retains features that are characteristic of the NKF3 family. These include the residues defined as the inhibitory triad, which occlude the nucleotide-binding pocket. Together with the mutations of several catalytic residues within the kinases domain, these features define PEAK3 as a pseudokinase. Based on sequence analysis and mutagenesis studies, we also demonstrate that PEAK3 self-associates through a conserved SHED domain that flanks its pseudokinase domain, similar to PEAK1 and Pragmin. Hence, the presence of the SHED domain and its role in oligomerization are defining and unique features of NKF3 kinases.

Phylogenetic analysis of the NKF3 family shows that the ancestral NKF3, which likely appeared in the ancestor of Metazoans, was most similar to PEAK1 and was already a pseudokinase. This protein had an NFS motif instead of a DFG although it retained the HxD

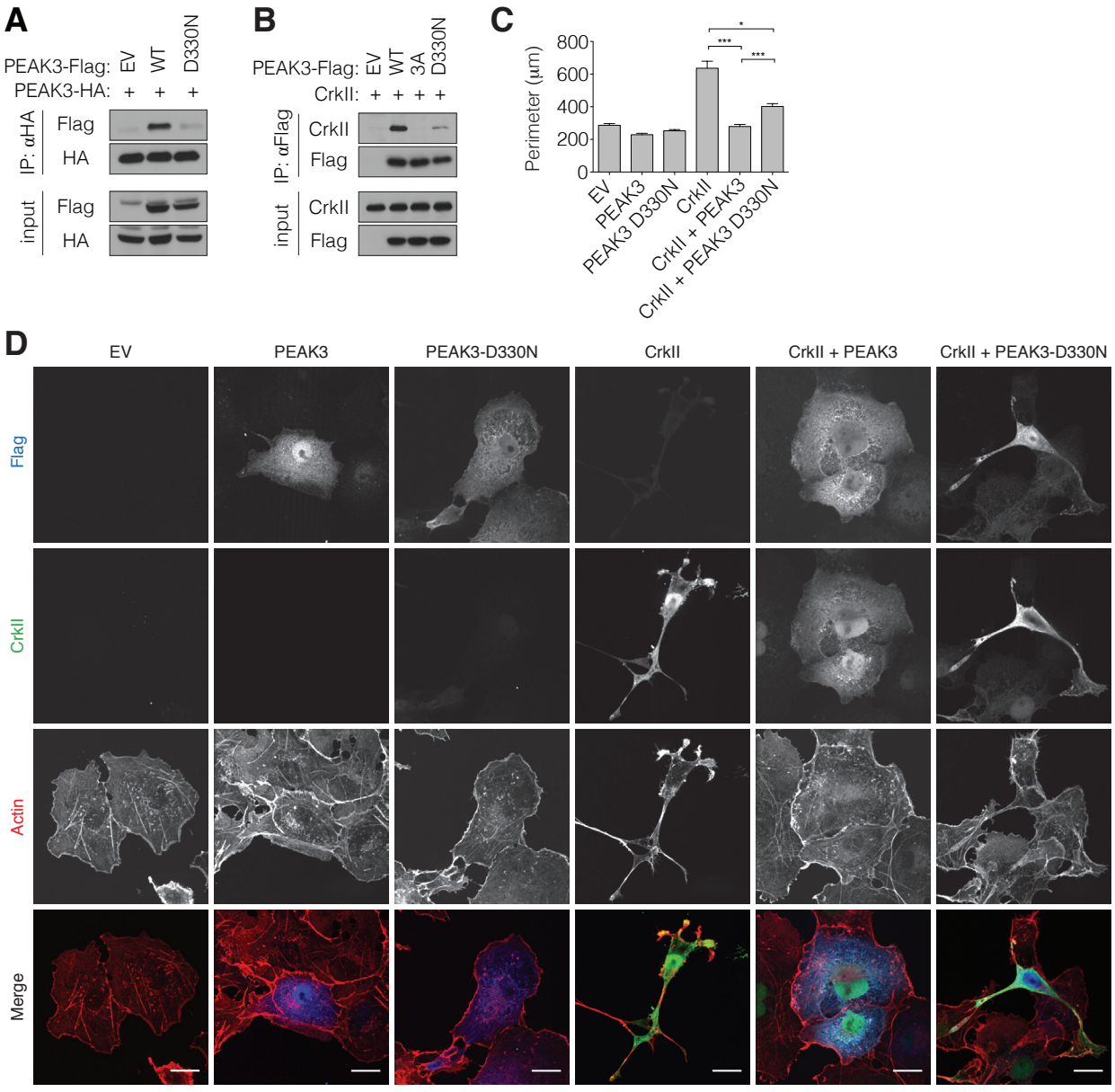


Fig. 3.12. Mutation of the DFG motif in PEAK3 diminishes its ability to rescue CrkII-dependent morphology. (A, B) Co-immunoprecipitation of the PEAK3 D330N variant with wild type PEAK3 (A) or untagged CrkII (B). Proteins were transiently expressed in HEK293 cells, and proteins levels were detected with the indicated antibodies by Western blot. **(C, D)** Average perimeter (C) and confocal microscopy (D) of COS-7 cells transiently expressing wild type PEAK3 or a D330N variant of PEAK3 with either empty vector (EV) or untagged CrkII. CrkII was detected with an anti-CrkII antibody (green), PEAK3 with an anti-Flag antibody (blue), and F-actin with Alexa Fluor 647-conjugated phalloidin staining (red). All scale bars correspond to 20 μm. Average perimeter data represent the mean ± SEM of three independent experiments (n = 20 cells in each experiment), *** $p < 0.001$.

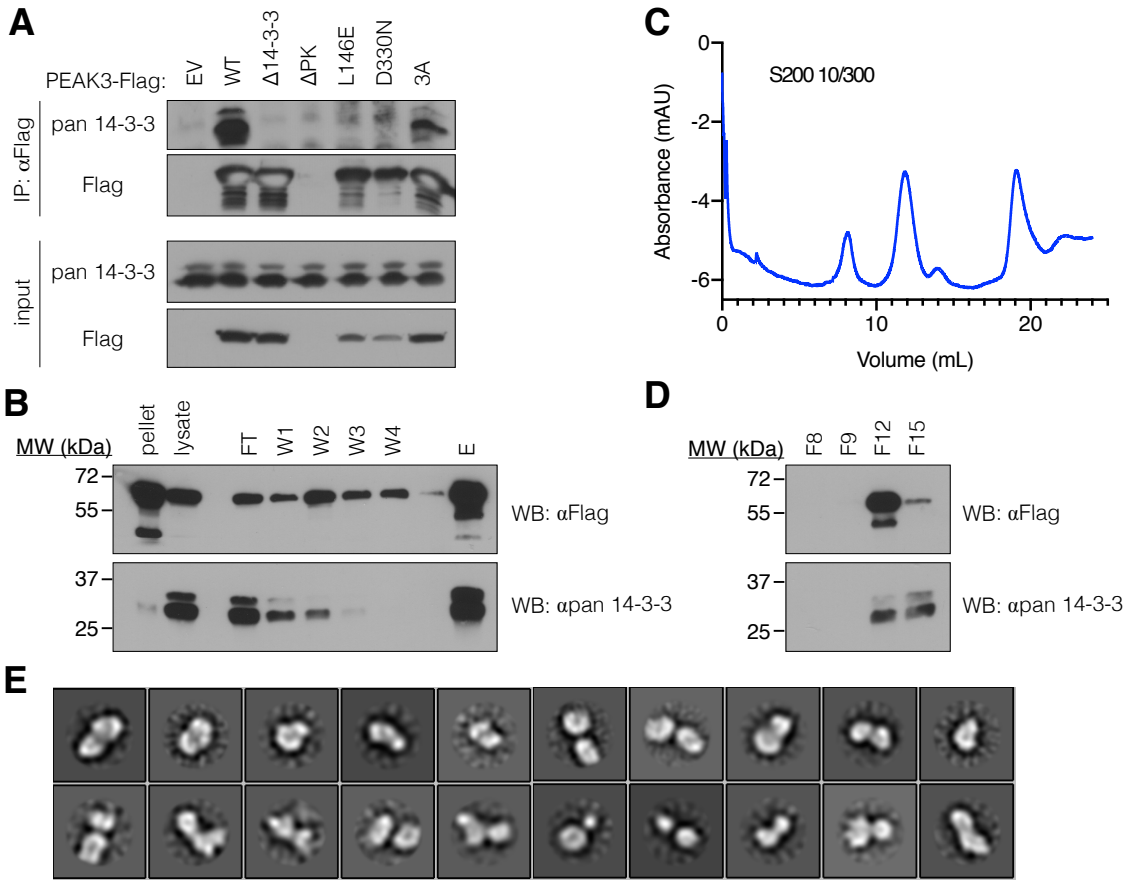


Fig. 3.13. PEAK3 dimerization is required for its interaction with 14-3-3. (A) Co-immunoprecipitation of the PEAK3 variants with endogenous 14-3-3. Proteins were transiently expressed in HEK293 cells, pulled down using an anti-Flag antibody, and proteins levels were detected with the indicated antibodies by Western blot analysis using the indicated antibodies. PEAK3 ΔPK variant was additionally pulled down, however due to issues with film exposure, this mutant was excluded from the blot. **(B)** Western blot analysis of PEAK3 Flag resin purification using the indicated antibodies. **(C)** Size exclusion chromatography trace of recombinant PEAK3 purification. **(D)** Western blot analysis of PEAK3 size exclusion chromatography purification using the indicated antibodies. **(E)** Class averages of negative stain electron microscopy on purified recombinant PEAK3 and co-purified endogenous 14-3-3.

motif, which became HCD in human PEAK1, and in some PEAK1 and Pragmin homologs - a canonical HRD motif. It is intriguing that mammalian (but not avian) PEAK3 proteins have apparently reverted to the canonical DFG motif while their HxD drifted into a degraded LxE motif. Thus at present, our analysis supports a hypothesis that NKF3 family is an example of a

kinase-like family that evolved originally as pseudokinases rather than one that degraded from a family of active kinases, similar to examples discussed by Kannan and co-workers[63].

Our unbiased search for PEAK3 binding partners reveals a propensity for PEAK3 to interact with regulators of cell motility, mirroring documented roles of PEAK1 and Pragmin[11-14]. While both Pragmin and PEAK1 are known to contain proline-rich CrkII binding motifs, and PEAK1 was shown to bind CrkII[14], our study is the first to describe a functional link between CrkII signaling and an NKF3 family member to our knowledge. We show here that PEAK3 regulates CrkII in a manner that likely contrasts that of other NKF3 family members: PEAK3 inhibits CrkII while PEAK1 and Pragmin stimulate pro-motile signaling in cells and would be therefore expected to potentiate CrkII function. The underlying mechanism(s) for these differences is unclear but could reflect the distinct domain structure of PEAK3 compared to other NKF3 family members. Both PEAK1 and Pragmin contain large N-terminal regions that likely encode numerous unique functions that are absent in PEAK3. If PEAK1 and Pragmin do indeed promote CrkII signaling, this function may be encoded within other regions of their N-terminal domain that are absent in PEAK3.

We further demonstrate that the SHED domain-mediated dimerization of PEAK3 is essential for its ability to bind CrkII. The emerging role of dimerization in binding seems to stem from the necessity to dimerize the CrkII-binding motifs themselves, as efficient CrkII binding can be recapitulated when these sites in PEAK3 are brought together by an orthogonal dimerization module. Since the CrkII binding site in PEAK3 represents the canonical proline-rich motif found in other known CrkII binding partners, we anticipate that these proteins might also regulate their interaction with CrkII through dimerization. Some of the known CrkII interactors, such as PEAK1 and Pragmin, are known to exist as dimers. Interestingly, CrkII and

its highly related homolog CrkL, which we also identified as a binding partner of PEAK3 in the IP/MS analysis, have been shown to form dimers[64]. Hence, the inherent property of both CrkII and some of its binding partners to oligomerize might be an essential mechanism for regulation of their mutual interactions and signaling.

While necessary for CrkII binding, dimerization of the CrkII binding motif is not sufficient for CrkII inhibition by PEAK3 in the absence of the SHED and pseudokinase domains. These findings suggest the SHED-pseudokinase domain module plays an important role in CrkII inhibition beyond supporting PEAK3 self-association. One such function could be mediating PEAK3 heterodimerization with PEAK1 and Pragmin, which we predict could occur based on sequence similarities within their SHED domains. Since all NKF3 family members have CrkII-binding sites, all possible NFK3 homo- and heterodimers could efficiently bind CrkII in theory. Given the opposing phenotypic outcomes between PEAK3 and other NKF3 members on cell motility, heterodimerization of PEAK3 with PEAK1 or Pragmin could interfere with their positive signaling properties. The outstandingly shorter length of the N-terminal domain in PEAK3 relative to PEAK1 and Pragmin suggests that PEAK3 may have evolved to act as a dominant negative regulator of PEAK1 and Pragmin, and antagonizes their functions encoded by the N-terminal domains. The regulatory role of PEAK3 for other NKF3 pseudokinases would be consistent with the later appearance of PEAK3 in evolution compared to the other two family members. The regulatory role of PEAK3 for other NKF3 pseudokinases would be consistent with the later appearance of PEAK3 in evolution compared to the other two family members. Furthermore, it is possible that the interaction between PEAK3 and 14-3-3 isoforms plays a role in this regulation by controlling the cellular localization of PEAK3. In this way, monomeric PEAK3, which could potentially heterodimerize with PEAK1 or Pragmin, may have a

subcellular localization distinct to the dimeric, CrkII-interacting pool. However, further studies would be needed to confirm this. Additional structural studies of PEAK3 with its binding partners would additionally aid in parsing out this relationship.

An additional unique feature of PEAK3, which distinguishes it from PEAK1 and Pragmin, is the presence of an intact DFG motif. Our data show that mutation of the DFG motif affects the ability of PEAK3 to homodimerize and subsequently interact with CrkII. The importance of the DFG motif for PEAK3 function is intriguing due to the critical role of this motif in kinase catalysis. While at present we cannot rule out that PEAK3 might be catalytically active, PEAK3 is an unlikely candidate for an active kinase based on its poor conservation of other residues in the putative active site. Rather, the loss-of-function effect of the DFG mutation suggests that it is a resulting conformational change within the pseudoactive site of PEAK3 that influences its dimerization and, consequently, its function. DFG mutations have previously been shown to affect the oligomerization of other kinases and pseudokinases. Notably, mutation of the DFG aspartate (D161N) in the kinase RIPK3 induces oligomerization and assembly of RIPK3 into a multimeric apoptotic complex[65,66]. Other kinase-inactivating mutations, such as those of the β3 lysine or catalytic aspartate, do not affect RIPK3 oligomerization, indicating that the D161N mutation stabilizes a specific conformation of RIPK3 that promotes oligomerization[66]. The opposite effect is observed in the pseudokinase MLKL. In MLKL, the DFG motif is replaced by a GFE motif, and mutation of the GFE glutamate (E351K) prevents MLKL oligomerization. This mutation was proposed to stabilize a conformation that it is not permissive for the release of the adjacent 4-helix bundle (4HB) domain which drives oligomerization[67]. The SHED domain in NKF3 kinases, composed of helices flanking the pseudokinase domain, maintains close contacts with the pseudokinase domain in the crystal structures of PEAK1 and

Pragmin. It is therefore likely that conformational changes within the pseudokinase domain are sensed by the SHED domain and can ultimately affect dimerization.

Potential regulation of PEAK3 dimerization through conformational changes in its pseudokinase domain presents an exciting opportunity for pharmacological modulation of PEAK3 oligomerization, as previously achieved in RIPK3[66,68]. While RIPK3 inhibitors target the ATP-binding site, this will likely not work for PEAK3, whose nucleotide-binding pocket is predicted to be occluded. However, recent studies on another pseudokinase, TRIB1, point to an important role of protein-protein interactions in regulating the conformation of the pseudoactive site. Like in PEAK3, the nucleotide-binding site in TRIB1 is highly occluded and inaccessible to ligands[69]. Binding of the transcription factor C/EBPα to the C-lobe of TRIB1 alters the conformation of the pseudoactive site, including its equivalent DFG motif (SLE in TRIB1), which rotates to a semi-active position upon C/EBPα binding[70]. The TRIB1 studies demonstrate that distal binding events can have global effects on the conformation of the pseudoactive site. Further studies can reveal if such interactions exist in PEAK3 and whether they can be leveraged for the pharmacological manipulation of its function.

The therapeutic relevance of targeting PEAK3 remains to be determined, but there is clear therapeutic potential[68]. While at present there are no known prevalent disease-associated mutations in PEAK3, *PEAK3* mRNA levels are significantly elevated in acute myeloid leukemia (AML) patient samples relative to other cancer types (**Fig. 3.14**)[71,72]. The E3 ubiquitin ligase SIAH1, which is a therapeutic target in AML[39,41], was identified as a PEAK3 binding partner in our IP/MS analysis. SIAH1 targets for degradation the oncogenic protein FMS-like tyrosine kinase 3 with internal tandem duplication mutation (FLT3-ITD), a mutant FLT3 variant detected in 40% of AML cases[73]. It is possible that PEAK3 interferes with this process through direct

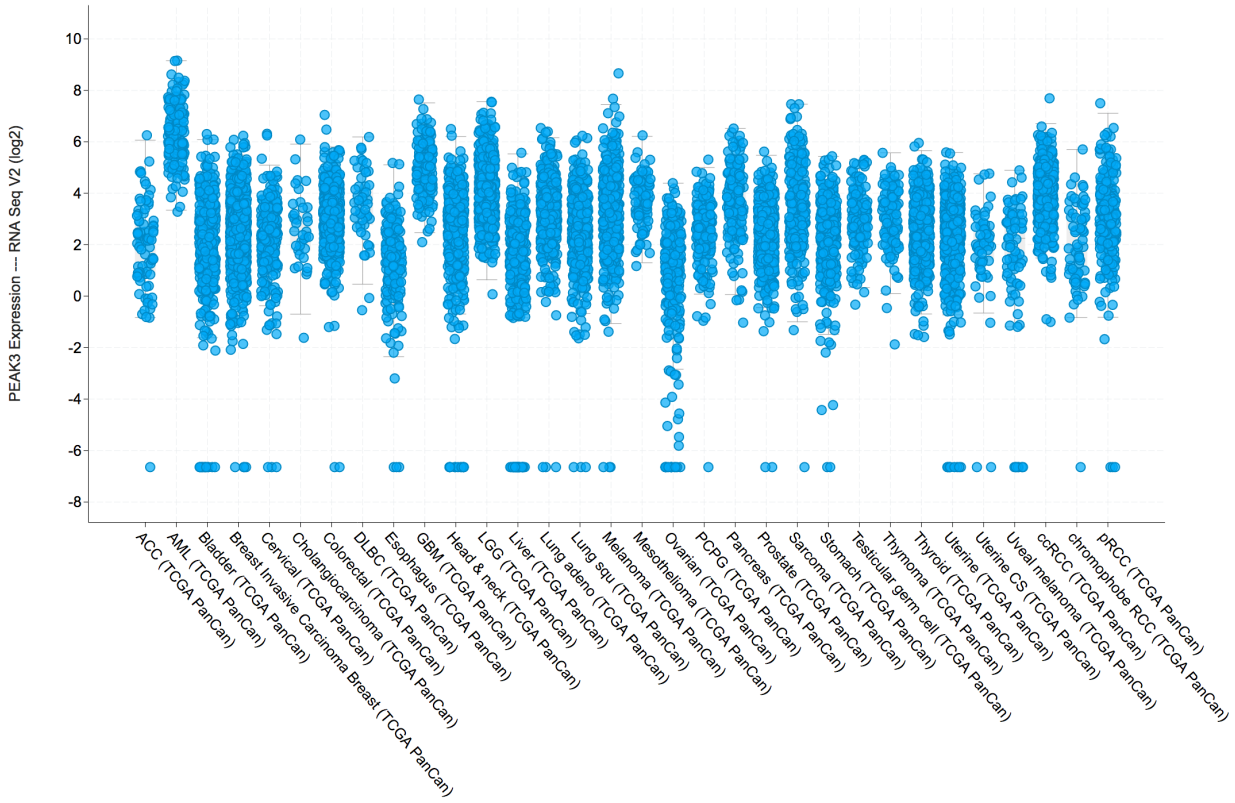


Fig. 3.14. PEAK3 mRNA levels in patient-derived AML cell lines. PEAK3 mRNA levels in patient-derived cancer cell lines from the Cancer Genome Atlas Project.

interaction with SIAH1. CrkL, which too was identified in our IP/MS screen, is one of the downstream effectors of the FLT3 signaling axis that contributes to leukemogenesis[74]. Crk family proteins are known to play key roles in cancer invasion and migration by integrating and amplifying extracellular signals[33]. Genetic knockdown of CrkII specifically decreases the cell migration and malignant potential of multiple human cancer cells including lung, breast, and ovarian cancers[33,49,75]. Mechanistically, inhibition of CrkII leads to reduced or stochastic F-actin networks and reduction in lamellipodia[49,52,75], mirroring the morphological changes we observe when PEAK3 is overexpressed in cells. Therefore, pharmacological targeting of PEAK3 could prove useful in AML and potentially additionally types of cancers that are susceptible to inhibitors that target cellular motility pathways.

ACKNOWLEDGEMENTS

We thank Scott Oakes for the wild-type CrkII plasmid, DeLaine Larsen and Kari Herrington for help with confocal imaging, Jocelyne Lopez, Madeline Ho, and Anna Morrison for their assistance on unpublished biochemical assays, Devan Diwanji for conducting the negative stain and EM experiments, Gwendolyn Jang, John von Dollen, Jeffrey Johnson, and Nevan Krogan for their work on the IP/MS, Mitchell Lopez for his work on confocal imaging and cell tracing, and Malgorzata Dudkiewicz and Krzysztof Pawlowski for the bioinformatic analysis. We would additionally like to thank Jennifer Kung and Kevan Shokat for their critical reading of the manuscript and Isabel Lucet, Roger Daly, and all members of the Jura lab for insightful discussions, especially Christopher Agnew.

MATERIALS AND METHODS

Multiple Sequence Alignment. The alignment was constructed using the Promals3D algorithm and manually adjusted using FFAS03[24] pair-wise alignment results for human PEAK3 and the two homologs PEAK1 and Pragmin. Secondary structure assignments from PDB structure records for Pragmin (PDB ID: 5VE6) and PEAK1 (PDB ID: 6BHC) were added to the final alignment. Low complexity regions in human members of NKF3 family were identified using the SEG server[28].

Phylogenetic tree construction. 29 homologs were selected from vertebrates that possess representatives of the three subfamilies and from four non-vertebrate organisms with single NKF3 homologs (*Trichoplax adherens* (plocozoan)*, Strongylocentrotus purpuratus* (sea urchin), *Branchiostoma belheri* (lancelet), and *Acanthaster planci* (starfish). Multiple sequence alignment of the pseudokinase domains was built using the Promals3D algorithm[76] and was manually

adjusted. The phylogenetic tree for was constructed using the PhyML method with aLRT statistics for calculating significance of branches[77,77]. Branches with bootstrap values better than 70% were marked. Ancestral sequence reconstruction was performed using the Ancescon algorithm[31].

Sequence logos. Homologs were collected from the NCBI database by running the BLAST program with kinase domains of human PEAK1, Pragmin and PEAK3 as queries and maximum expect value of 1E-6. Redundancy in the resulting sequence set was removed using CD-hit algorithm at 95% identity level[78]. Assignment of the resulting representative sequences to the three NKF3 subfamilies was verified by sequence clustering using the fastNJ algorithm. Multiple sequence alignment obtained using the Promals3D algorithm was split into three subfamily alignments with matched column numbering and used to create sequence logos with the WebLogo3 server[79]. The logos included 103 Pragmin homologs, 98 PEAK1 homologs and 35 mammalian PEAK3 sequences. Avian PEAK3 sequences were not included due to substantial heterogeneity of the N-terminal regions.

Plasmids and cell culture. The PEAK3 gene was synthesized by GenScript and subcloned into pcDNA4/TO. The wild-type CrkII plasmid was a gift from Scott Oakes. Mutations and deletions were introduced using Quikchange mutagenesis (Agilent). All constructs were verified via DNA sequencing (Elim Biopharm). HEK293 cells and COS-7 cells were cultured in Dulbecco's modified Eagle media (Life Technologies) supplemented with 10% FBS (Hyclone) and penicillin streptomycin (Life Technologies). Epitope-tagged constructs were transiently transfected into cells using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocols. Cells were transfected for 24 hours prior to imaging or cell lysis.

Immunoprecipitation/mass-spectrometry. FLAG immunoprecipitations were done as previously described[80,81]. Specific details are as follows: 293T cells were transfected with 3xFLAG-tagged PEAK3 expression construct using PolyJet Reagent (SignaGen Laboratories) 20-24 hours after plating $1x10^7$ cells per 14.5 cm dish. 40 hours post- transfection, cells were dissociated and washed with 10 ml PBS +/- 10 mM EDTA, respectively, before centrifugation at ≥ 200xg, at 4°C for 5 minutes. Cell pellets were re- suspended in 1 ml of 0.5% Nonidet P-40 Substitute (Fluka Analytical) in IP buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA) supplemented with cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche), incubated on a tube rotator at 4°C for 30 minutes, and centrifuged at 3,500xg, 4°C for 20 minutes. Cell lysates, 20 ml anti-FLAG M2 magnetic beads (Sigma-Aldrich), and 2 mg 1xFLAG peptide (Sigma-Aldrich) in 0.3 ml IP buffer were incubated on a tube rotator at 4°C for 2 hours. After binding, FLAG beads were washed with 0.05% Nonidet P-40 Substitute in IP buffer (3 x 1 ml) and transferred to a new tube with a final wash in 1 ml IP buffer. Proteins were eluted by gently agitating FLAG beads with 30 ml of 0.05% RapiGest SF Surfactant (Waters Corporation) in IP buffer on a vortex mixer at room temperature for 30 minutes. FLAG-tagged protein expression and protein immunoprecipitation were assessed by western blot and silver stain, respectively, before submitting 10 ml eluate for mass spectrometry. Three independent biological replicates were performed for FLAG immunoprecipitations.

Mass spectrometry analysis. Purified proteins eluates were digested with trypsin for LC-MS/MS analysis. Samples were denatured and reduced in 2M urea, 10 mM NH4HCO3, 2 mM DTT for 30 min at 60°C, then alkylated with 2 mM iodoacetamide for 45 min at room temperature. Trypsin (Promega) was added at a 1:100 enzyme:substrate ratio and digested

overnight at 37°C. Following digestion, samples were concentrated using C18 ZipTips (Millipore) according to the manufacturer's specifications. Digested peptide mixtures were analyzed by LC-MS/MS on a Thermo Scientific Velos Pro dual linear ion trap mass spectrometry system equipped with a Proxeon Easy nLC II high- pressure liquid chromatography and autosampler system. Samples were injected onto a pre-column (2 cm x 100 μm I.D. packed with ReproSil Pur C18 AQ 5 μm particles) in 0.1% formic acid and then separated with a one-hour gradient from 5% to 30% ACN in 0.1% formic acid on an analytical column (10 cm x 75 um I.D. packed with ReproSil Pur C18 AQ 3 μm particles). The mass spectrometer collected data in a data-dependent fashion, collecting one full scan followed by 20 collision-induced dissociation MS/MS scans of the 20 most intense peaks from the full scan. Dynamic exclusion was enabled for 30 seconds with a repeat count of 1. The resulting raw data was matched to protein sequences by the Protein Prospector algorithm[82]. Data were searched against a database containing SwissProt Human protein sequences, concatenated to a decoy database where each sequence was randomized in order to estimate the false positive rate. The searches considered a precursor mass tolerance of 1 Da and fragment ion tolerances of 0.8 Da, and considered variable modifications for protein N-terminal acetylation, protein N-terminal acetylation and oxidation, glutamine to pyroglutamate conversion for peptide N-terminal glutamine residues, protein N-terminal methionine loss, protein N-terminal acetylation and methionine loss, and methionine oxidation, and constant modification for carbamidomethyl cysteine. Prospector data was filtered using a maximum protein expectation value of 0.01 and a maximum peptide expectation value of 0.05. Protein interactions were subsequently scored using the CompPASS algorithm[45].

Co-immunoprecipitation and Western blot analysis. HEK293 cells ($2.5x10^6$ cells) were seeded on 60 mm dishes and transfected the following day. 24 hours post transfection, cells were washed two times on ice with 1xPBS followed by lysis for 30 minutes on ice (0.5% Triton X-100, 0.5% NP-40, 150 mM NaCl, 50 mM Tris pH8.0, 1 mM NaF, 1 mM $Na(VO_4)_3$, 1 mM EDTA, cOmplete mini EDTA-free protease inhibitor cocktail (Roche)). Cells were scraped and clarified by centrifugation for 10 minutes at 15,000 rpm. The whole cell lysates were pre-cleared with Protein A beads (Novex) for 30 minutes at 4°C. The pre-cleared lysates were then incubated with antibody/protein A complexed beads overnight at 4°C (anti-FLAG (mouse, Sigma), anti-HA (mouse, SCBT)). The beads were washed three times with lysis buffer. The bound proteins were eluted from the beads using SDS-loading buffer and were boiled at 95°C for 10 minutes prior to SDS/Page and analysis by Western blot. Samples were run on 12% acrylamide gels, and transferred onto PVDF membranes. Membranes were blocked in 5% milk in 1xTBS and 1% Tween-20 (TBS-T), followed by incubation with primary antibodies diluted in blocking buffer (anti-CrkII (rabbit, Proteintech), anti-FLAG (mouse, Sigma Aldrich; rabbit, CST), and anti-HA (mouse, SCBT)), followed by incubation with secondary antibodies diluted in blocking buffer (anti-mouse IgG Veriblot, Abcam; anti-IgG Veriblot, Abcam). ECL Western blotting detection reagent (GE) or ECL prime (VWR) were used for detection.

Immunofluorescence analysis. COS-7 cells ($7.0x10^4$ per well) were plated onto glass coverslips and transfected the following day. 24 hours post transfection, cells were fixed in 3.7% formaldehyde in PBS for 1 hour at room temperature, permeabilized using 0.1% Triton-X in PBS for 5 minutes, and incubated with blocking buffer (1% BSA in PBS) for 5 minutes. Primary antibodies in the blocking buffer (anti-FLAG (rabbit, CST), anti-CrkII (mouse, SCBT)) were added for 1 hour at 37°C, followed by secondary antibodies (Alexa Fluor 568 anti-rabbit IgG

(Life Technologies) and Alexa Fluor 488 anti-mouse IgG (Life Technologies)) for 1 hour at 37°C. Actin was stained using Alexa Fluor 674 Phalloidin (Life Technologies) for 20 minutes at 37°C. Images were acquired using a Nikon Elipse Ti equipped with a CSU-X1 spinning disc confocal and Andor Clara interline CCD camera with a Nikon Plan Apo 60X oil lens. The effects on actin stress fibers were blindly scored. Data is reported as a percent of the total (n=60 cells per group in each experiment). Cell perimeter calculations were generated by creating an ROI of each cell using Fiji and a Wacom tablet. Each cell was traced by hand using the Wacom tablet and the perimeter of the cell was calculated from the cell-shape vector. Data was analyzed using the Kruskal-Wallis test for the nonparametric comparison of the means followed by Dunn's Multiple Comparison Test for pairwise comparison between groups[83,84].

Mapping sequence conservation onto known NKF3 structures. Structure mapping was performed considering sequence conservation within mammalian subfamily of PEAK3 homologs. Alignment conservation values (Jalview 1.18[85]) were used for residue coloring (lowest conservation value (0) – white, highest conservation value (11) – red). Conservation scores were mapped onto the PEAK1 dimer structure (PDB ID: 6BHC) using UCSF Chimera[86]. Conservation mapping onto the solved crystal structure was performed based on two pairwise alignments: (1) PEAK3 vs PEAK1 and (2) PEAK3 vs Pragmin. Structures for dimers of human Pragmin (PDB ID: 5VE6) and PEAK1 (PDB ID: 6BHC) were rendered using Pymol. Residues conserved in the alignments (strictly conserved or conservative replacements, as judged by positive BLOSUM62 matrix scores[87]), were colored according to BLOSUM62 scores: yellow: BLOSUM62 values from 1 to 3; orange: 4-6; red (highest conservation): 7 to 11.

Recombinant protein expression and purification. PEAK3 constructs were synthesized by Genscript in a pcDNA3.1+ plasmid with a 3xFLAG tag, which was mutated to a 1xFLAG tag

(DYKDDDK) using Quikchange mutagenesis (Agilent). Cloning verification was done by DNA sequencing (Elim biotechnology). Constructs were transfected into 30 mL of Expi293F (ThermoFisher) culture using Expifectamine 293 transfection reagent (ThermoFisher) at 1 μg DNA / 1 mL cell volume. Enhancer was added 24 hours post-transfection, and cells were collected 72 hours post-enhancer addition in an Allegra X-14 centrifuge (Beckman Coulter) at 4,000xg for 10 minutes. Pellets were flash frozen for later purification or resuspended in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 2 mM DTT) with 1 mM NaF, 1 mM $NaVO_4$, DNaseI, and a cOmplete mini EDTA-free protease inhibitor cocktail tablet (Roche) added fresh. Cells were homogenized by pipetting and disrupted using an EmulsiFlex-C5 homogenizer (Avestin). Lysates were clarified in an Avanti centrifuge equipped with an JLA 25.50 rotor at 20,000xg, 40 minutes, 4°C. Clarified lysates were incubated with G1 Flag resin (Genscript) overnight at 4°C with rotating before washing with 50 bead volumes of lysis buffer and eluting in 10 bead volumes of elution buffer (lysis buffer + 0.25mg/ml 1x Flag peptide (SinoBiological)). Elutions were concentrated using an Amicon Ultra-0.5 10k MWCO centrifugal filter (Millipore) before being loaded onto a Supdex 200 10/300 GL column (GE Life Sciences) with an isocratic gradient of buffer.

Negative stain electron microscopy. 3 μl of peak fraction from gel filtration was applied onto glow discharged Cu 400 mesh grids and stained with 0.75% uranyl formate. Images at 52,000x were collected on a Tecnai-FEI T12 microscope equipped with a Gatan CMOS camera and processed with RELION (REgularized LIkelihood OptimizatioN). Briefly, particles were initially manually picked and 2D class averaged from a small subset of micrographs. Representative 2D class averages were then selected for automated particle picking from the entire dataset, extracted, and class averaged.

REFERENCES

1. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
2. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
3. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
4. Kung, J. E. & Jura, N. Structural Basis for the Non-catalytic Functions of Protein Kinases. *Structure/Folding and Design* **24**, 7–24 (2016).
5. Lecointre, C. *et al.* Dimerization of the Pragmin Pseudo-Kinase Regulates Protein Tyrosine Phosphorylation. *Structure* **26**, 545–554.e4 (2018).
6. Patel, O., Roy, M. J., Murphy, J. M. & Lucet, I. S. The PEAK family of pseudokinases, their role in cell signalling and cancer. *FEBS J* **45**, 665 (2019).
7. Murphy, J. M. *et al.* A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties. *Biochem. J.* **457**, 323–334 (2014).
8. Patel, O. *et al.* Structure of SgK223 pseudokinase reveals novel mechanisms of homotypic and heterotypic association. *Nat. Commun.* **8**, 1157 (2017).
9. Ha, B. H. & Boggon, T. J. The crystal structure of pseudokinase PEAK1 (Sugen Kinase 269) reveals an unusual catalytic cleft and a novel mode of kinase fold dimerization. *J. Biol. Chem.* (2017).doi:10.1074/jbc.RA117.000751
10. Tanaka, H., Katoh, H. & Negishi, M. Pragmin, a Novel Effector of Rnd2 GTPase, Stimulates RhoA Activity. *J. Biol. Chem.* **281**, 10355–10364 (2006).
11. Tactacan, C. M. *et al.* The pseudokinase SgK223 promotes invasion of pancreatic ductal epithelial cells through JAK1/Stat3 signaling. *Mol. Cancer* 1–11

(2015).doi:10.1186/s12943-015-0412-3

12. Senda, Y., Murata-Kamiya, N. & Hatakeyama, M. C-terminal Src kinase-mediated EPIYA phosphorylation of Pragmin creates a feed-forward C-terminal Src kinase activation loop that promotes cell motility. *Cancer Sci.* **107**, 972–980 (2016).

13. Safari, F., Murata-Kamiya, N., Saito, Y. & Hatakeyama, M. Mammalian Pragmin regulates Src family kinases via the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif that is exploited by bacterial effectors. *Proc. Natl. Acad. Sci. USA* **108**, 14938–14943 (2011).

14. Wang, Y. *et al.* Pseudopodium-enriched atypical kinase 1 regulates the cytoskeleton and cancer progression [corrected]. *Proc. Natl. Acad. Sci. USA* **107**, 10920–10925 (2010).

15. Liu, L. *et al.* Homo- and Heterotypic Association Regulates Signaling by the SgK269/PEAK1 and SgK223 Pseudokinases. *J. Biol. Chem.* **291**, 21571–21583 (2016).

16. Bristow, J. M., Reno, T. A., Jo, M., Gonias, S. L. & Klemke, R. L. Dynamic phosphorylation of tyrosine 665 in pseudopodium-enriched atypical kinase 1 (PEAK1) is essential for the regulation of cell migration and focal adhesion turnover. *J. Biol. Chem.* **288**, 123–131 (2013).

17. Zheng, Y. *et al.* Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* **499**, 166–171 (2013).

18. O'Rourke, R. L. & Daly, R. J. The pseudokinases SgK269 and SgK223: A novel oncogenic alliance in human cancer. *Cell Adh. Migr.* 1–5 (2017).doi:10.1080/19336918.2017.1394570

19. Leroy, C. *et al.* Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates SRC invasive activity in advanced colon carcinoma cells. *Cancer Res.* **69**, 2279–2286 (2009).

20. Weaver, K. L. *et al.* NACK is an integral component of the Notch transcriptional activation complex and is critical for development and tumorigenesis. *Cancer Res.* **74**, 4741–4751 (2014).

21. Kelber, J. A. *et al.* KRas induces a Src/PEAK1/ErbB2 kinase amplification loop that drives metastatic growth and therapy resistance in pancreatic cancer. *Cancer Res.* **72**, 2554–2564 (2012).

22. Croucher, D. R. *et al.* Involvement of Lyn and the atypical kinase SgK269/PEAK1 in a basal breast cancer signaling pathway. *Cancer Res.* **73**, 1969–1980 (2013).

23. Fujimura, K. *et al.* A Hypusine-eIF5A-PEAK1 Switch Regulates the Pathogenesis of Pancreatic Cancer. *Cancer Res.* **74**, 6671–6681 (2014).

24. Jaroszewski, L., Li, Z., Cai, X.-H., Weber, C. & Godzik, A. FFAS server: novel features and applications. *Nucleic Acids Res.*. **39**, W38–44 (2011).

25. Dudkiewicz, M., Lenart, A. & Pawłowski, K. A novel predicted calcium-regulated kinase family implicated in neurological disorders. *PLoS ONE* **8**, e66427 (2013).

26. Dudkiewicz, M., Szczepińska, T., Grynberg, M. & Pawłowski, K. A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life. *PLoS ONE* **7**, e32138 (2012).

27. Nguyen, K. B. *et al.* Phosphorylation of spore coat proteins by a family of atypical protein kinases. *Proc. Natl. Acad. Sci. USA* **113**, E3482–91 (2016).

28. Wootton, J. C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285 (1994).

29. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, 43 (2010).

30. Knighton, D. R. *et al.* Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–414 (1991).

31. Cai, W., Pei, J. & Grishin, N. V. Reconstruction of ancestral protein sequences and its applications. *BMC Evol. Biol.* **4**, 33 (2004).

32. Park, T.-J. & Curran, T. Essential roles of Crk and CrkL in fibroblast structure and motility. *Oncogene* **33**, 5121–5132 (2014).

33. Birge, R. B., Kalodimos, C., Inagaki, F. & Tanaka, S. Crk and CrkL adaptor proteins: networks for physiological and pathological signaling. *Cell Commun. Signal.* **7**, 272 (2009).

34. Fu, H., Subramanian, R. R. & Masters, S. C. 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.* **40**, 617–647 (2000).

35. Dougherty, M. K. Unlocking the code of 14-3-3. *J. Cell. Sci.* **117**, 1875–1884 (2004).

36. Pennington, K. L., Chan, T. Y., Torres, M. P. & Andersen, J. L. The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein–protein interactions. *Oncogene* **37**, 5587–5604 (2018).

37. Brown, M. T. *et al.* ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates with and is phosphorylated by Src. *Mol. Cell. Biol.* **18**, 7038–7051 (1998).

38. Chen, P.-W. *et al.* The Arf GTPase-activating Protein, ASAP1, Binds Nonmuscle Myosin 2A to Control Remodeling of the Actomyosin Network. *J. Biol. Chem.* **291**, 7517–7526 (2016).

39. Krämer, O. H., Stauber, R. H., Bug, G., Hartkamp, J. & Knauer, S. K. SIAH proteins: critical roles in leukemogenesis. *Leukemia* **27**, 792–802 (2013).

40. Farag, A. K. & Roh, E. J. Death-associated protein kinase (DAPK) family modulators:

Current and future therapeutic outcomes. *Med. Res. Rev.* **39**, 349–385 (2019).

41. Lodygin, D. & Hermeking, H. The role of epigenetic inactivation of 14-3-3sigma in human cancer. *Cell Res.* **15**, 237–246 (2005).

42. Park, E. *et al.* Architecture of autoinhibited and active BRAF–MEK1–14-3-3 complexes. *Nature* **575**, 545–550 (2019).

43. Knudsen, B. S. *et al.* Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *EMBO J.* **14**, 2191–2198 (1995).

44. Kiyokawa, E., Hashimoto, Y., Kurata, T., Sugimura, H. & Matsuda, M. Evidence that DOCK180 up-regulates signals from the CrkII-p130(Cas) complex. *J. Biol. Chem.* **273**, 24479–24484 (1998).

45. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389–403 (2009).

46. Lamorte, L., Royal, I., Naujokas, M. & Park, M. Crk adapter proteins promote an epithelial-mesenchymal-like transition and are required for HGF-mediated cell spreading and breakdown of epithelial adherens junctions. *Mol. Biol. Cell* **13**, 1449–1461 (2002).

47. Ridley, A. J. Rho: theme and variations. *Curr. Biol.* **6**, 1256–1264 (1996).

48. Ridley, A. J. *et al.* Cell migration: integrating signals from front to back. *Science* **302**, 1704–1709 (2003).

49. Rodrigues, S. P. *et al.* CrkI and CrkII function as key signaling integrators for migration and invasion of cancer cells. *Mol. Cancer Res.* **3**, 183–194 (2005).

50. Petit, V. *et al.* Phosphorylation of tyrosine residues 31 and 118 on paxillin regulates cell migration through an association with CRK in NBT-II cells. *J Cell Biol* **148**, 957–970 (2000).

51. Tang, D. D., Zhang, W. & Gunst, S. J. The adapter protein CrkII regulates neuronal Wiskott-Aldrich syndrome protein, actin polymerization, and tension development during contractile stimulation of smooth muscle. *J. Biol. Chem.* **280**, 23380–23389 (2005).

52. Nakashima, N. *et al.* The functional role of CrkII in actin cytoskeleton organization and mitogenesis. *J. Biol. Chem.* **274**, 3001–3008 (1999).

53. Klemke, R. L. *et al.* CAS/Crk coupling serves as a 'molecular switch' for induction of cell migration. *J Cell Biol* **140**, 961–972 (1998).

54. Vallenius, T. Actin stress fibre subtypes in mesenchymal-migrating cells. *Open Biol.* **3**, 130001–130001 (2013).

55. Kobashigawa, Y. *et al.* Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nat. Struct. Mol. Biol.* **14**, 503–510 (2007).

56. Saleh, T. *et al.* Cyclophilin A promotes cell migration via the Abl-Crk signaling pathway. *Nat. Chem. Biol.* **12**, 117–123 (2016).

57. Antoku, S. & Mayer, B. J. Distinct roles for Crk adaptor isoforms in actin reorganization induced by extracellular signals. *J. Cell. Sci.* **122**, 4228–4238 (2009).

58. Kain, K. H. & Klemke, R. L. Inhibition of Cell Migration by Abl Family Tyrosine Kinases through Uncoupling of Crk-CAS Complexes. *J. Biol. Chem.* **276**, 16185–16192 (2001).

59. Jacquemet, G. *et al.* Filopodome Mapping Identifies p130Cas as a Mechanosensitive Regulator of Filopodia Stability. *Curr. Biol.* **29**, 202–216.e7 (2019).

60. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).

61. Jura, N. *et al.* Catalytic Control in the EGF Receptor and Its Connection to General Kinase

Regulatory Mechanisms. *Mol. Cell* **42**, 9–22 (2011).

62. Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. USA* **106**, 139–144 (2009).

63. Kwon, A. *et al.* Tracing the origin and evolution of pseudokinases across the tree of life. *Sci. Signal.* **12**, (2019).

64. Harkiolaki, M., Gilbert, R. J. C., Jones, E. Y. & Feller, S. M. The C-Terminal SH3 Domain of CRKL as a Dynamic Dimerization Module Transiently Exposing a Nuclear Export Signal. *Structure* **14**, 1741–1753 (2006).

65. Newton, K. *et al.* Activity of protein kinase RIPK3 determines whether cells die by necroptosis or apoptosis. *Science* **343**, 1357–1360 (2014).

66. Mandal, P. *et al.* RIP3 Induces Apoptosis Independent of Pronecrotic Kinase Activity. *Mol. Cell* **56**, 481–495 (2014).

67. Petrie, E. J. *et al.* Conformational switching of the pseudokinase domain promotes human MLKL tetramerization and cell death by necroptosis. *Nat. Commun.* **9**, 2422 (2018).

68. Kung, J. E. & Jura, N. Prospects for pharmacological targeting of pseudokinases. *Nat. Rev. Drug Discov.* **18**, 501–526 (2019).

69. Murphy, J. M. *et al.* Molecular Mechanism of CCAAT-Enhancer Binding Protein Recruitment by the TRIB1 Pseudokinase. *Structure/Folding and Design* **23**, 2111–2121 (2015).

70. Jamieson, S. A. *et al.* Substrate binding allosterically relieves autoinhibition of the pseudokinase TRIB1. *Sci. Signal.* **11**, 549 (2018).

71. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, 269 (2013).

72. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

73. Buchwald, M. *et al.* Ubiquitin conjugase UBCH8 targets active FMS-like tyrosine kinase 3 for proteasomal degradation. *Leukemia* **24**, 1412–1421 (2010).

74. Hartman, A. D. *et al.* Constitutive c-jun N-terminal kinase activity in acute myeloid leukemia derives from Flt3 and affects survival and proliferation. *Exp. Hematol.* **34**, 1360–1376 (2006).

75. Linghu, H. *et al.* Involvement of adaptor protein Crk in malignant feature of human ovarian cancer cell line MCAS. *Oncogene* **25**, 3547–3556 (2006).

76. Pei, J., Kim, B.-H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).

77. Guindon, S., Delsuc, F., Dufayard, J.-F. & Gascuel, O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **537**, 113–137 (2009).

78. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).

79. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

80. Jäger, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2011).

81. Jäger, S. *et al.* Purification and characterization of HIV-human protein complexes. *Methods* **53**, 13–19 (2011).

82. Clauser, K. R., Baker, P. & Burlingame, A. L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching.

Anal. Chem. **71**, 2871–2882 (1999).

83. Kruskal, W. H. & Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).

84. Dunn, O. J. Multiple Comparisons among Means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).

85. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

86. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

87. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919 (1992).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
7173AD7A997146E...
Author Signature

9/7/2020
Date