# Structure, Function, and Inhibition of the Oncogenic RhoGEF Trio

by

Sumit Bandekar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2020

Doctoral Committee:

Professor Janet L. Smith, Co-Chair
Professor John J.G. Tesmer, Purdue University, Co-Chair
Assistant Professor Amanda L. Garner
Associate Professor Zaneta Nikolovska-Coleska
Professor Emily E. Scott

Sumit Bandekar

sbandeka@umich.edu

ORCID iD: 0000-0003-2728-7366

# Dedication

This work is dedicated to my family and friends who provided me with unconditional support throughout my Ph.D. First, to my immediate family, mom, dad, and brother, for always providing me a stable, caring, and exciting world to live in. To my extended family, for all the fun we've had together and support we've shared. To all my friends throughout the years, too many to name. I've been so lucky to meet and spent time with each and every one of you. My friends have truly made life worth living. To the wonderful teachers I've had in all aspects of my life who challenged me whilst providing guidance and support. It was this encouragement that allowed me to grow into who I am. To my lab mates, mentors, and colleagues who I truly found another family with, who have helped cultivate in me a profound sense of wonder about protein structure. I'm lucky to have found the laboratory and community that I did. Every person mentioned above had a profound impact on my upbringing, both as a human and scientist. Finally, to all the wonderful people I've had the pleasure to meet in my time in Ann Arbor. This city will always feel like home to me and has been such an important part of my human experience.

.

# Acknowledgements

I would like to acknowledge the following people and organizations: Dr. John Tesmer for training, financial support, and lots of discussion, for providing an incredibly warm and supportive environment for training and development. Drs. Tyler Beyett and Jen Cash for training in structural biology and all manner of science. All members of the Tesmer lab for valuable critiques, discussion, friendship and collegiate-ness. My undergraduate mentees, Brenna Barton, Ena Tully, and Brittany Tang, for providing lots of their time and helping with all aspects of my experimental work. Dr. Janet Smith for taking me in when John chose to move universities and including me in her group's activities, as well as extremely valuable mentorship. Former and current members of the Smith lab for their support, training, and comradeship. All of my thesis committee members for valuable advice and support throughout my time, particularly during my 4th year. Dr. Clay Brown and Jim Delproposto for helpful discussion and advice, and the center for structural biology for crystallography support and discussion. Drs. Jennifer Meagher and David Akey for help with SEC-MALS. The Lin and Ginsburg labs for use of their qPCR machines. Krishnapriya Chinnaswamy for help with ITC. Sheng Li for experimental contributions in HDX-MS. Srinivas Chakravarthy for SEC-SAXS experiments. Silvio Gutkind and Nadia Arang for help envisioning and performing Rhotekin assay experiments, as well as manuscript revisions. Mindy Mackey and Sarah Lloyd for administrative support, PSTP for funding and forging valuable connections

between all the trainees, NIH for funding my F31 and our R01 grants, LS-CAT staff for assistance in remote data collection, Amy and Min from the EM suite for support and valuable discussion, Michael Cianfrocco and the EM community for valuable training and support. Departments of Medicinal Chemistry, Biochemistry, and Pharmacology for many networking opportunities. The LSI administrative staff, especially Danielle Smith, April Miller, Jan Mitchell, and Alyson Carter for help with funding with training grants and the David and Michelle Kroin Family Scholarship. Rackham graduate school for various internal funding opportunities. ASPET for travel awards to attend Experimental Biology conferences. The Signal Transduction Journal club for years of enlightening discussion. This work was supported by NIH Grants 5R01CA221289 (to J.J.G.T.), 5T32GM007767 (to S.J.B.), and 5F31CA224804 (to S.J.B.). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by grant P41GM103622 from the National Institute of General Medical Sciences of the National Institutes of Health. Use of the Pilatus 3 1M detector was provided by grant 1S10OD018090-01 from NIGMS. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of General Medical Sciences or the National Cancer Institute or the National Institutes of Health.

.

## Table of Contents

# List of Tables

# List of Figures

## List of Appendices

# List of Abbreviations

7TM: 7-transmembrane receptor

$A_{280}$: 280 nm absorbance

AF-488: alexafluor 488

ANOVA: analysis of variance

ANS: 8-Anilino-1-naphthalenesulfonic acid

AP-1: activator-protein 1

ASU: asymmetric unit

CaM: calmodulin

CaMK: calmodulin kinase

cDNA: complimentary deoxyribonucleic acid

CCP4: Collaborative Computational Project No. 4

CR: conserved region

CRAL: cellular retinaldehyde binding protein

CV: column volumes

CT: Cral_Trio domain

Dbs: Dbl’s big sister

DDM: dodecyl maltoside

DH: Dbl homology domain

DSF: differential scanning fluorometry

DNA: deoxyribonucleic acid

DTT: dithiothreitol

dTrio: Drosophila melanogaster Trio

E. coli: Escherichia coli

eGFP: enhanced green fluorescent protein

ESI+: positive electrospray ionization

EM: electron microscopy

FnIII: fibronectin type III

FPLC: fast protein liquid chromatography

FRET: Förster resonance energy transfer

GAP: GTPase-activating protein

GDP: Guanosine diphosphate

GEF: guanine nucleotide exchange factor

GFP: green fluorescent protein

GPCR: G protein-coupled receptor

GST: Glutathione-S-transferase

GTP: Guanosine triphosphate

GTP$\gamma$S: guanosine 5′-O-[gamma-thio]triphosphate

GTPase: Guanosine triphosphatase

HDX-MS: hydrogen deuterium exchange-mass spectrometry

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hTrio: human Trio

hTrioFL: full-length human Trio

$IC_{50}$: half maximal inhibitory concentration

Ig: Immunoglobulin domain

IPTG: isopropyl β-D-1-thiogalactopyranoside

ITC: isothermal titration calorimetry

KD: kinase domain

LARG: leukemia-associated RhoGEF

LIC: ligation-independent cloning

LC-MS: liquid chromatography-mass spectrometry

MALS: multiangle light scattering

MANT-GTP: 2´/3´-O-(N-Methyl-anthraniloyl)-guanosine-5´- triphosphate

MAPK: mitogen-activated protein kinase

MBP: maltose-binding protein

MLCK: myosin light chain kinase

MR: molecular replacement

MS: mass spectrometry

MyBP: myelin basic protein

NCS: non-crystallographic symmetry

NHS: N-Hydroxysuccinimide

Ni-NTA: nickel-nitriloacetic acid resin

NMR: nuclear magnetic resonance spectroscopy

$OD_{600}$: 600nm optical density

PCR: polymerase chain reaction

PDB: protein data bank

PEG: polyethylene glycol

PH: pleckstrin homology domain

PLCβ: phospholipase Cβ

PM: plasma membrane

PMSF: phenylmethanesulfonyl fluoride

qPCR: real-time quantitative polymerase chain reaction

RNA: ribonucleic acid

$R_g$: radius of gyration

RH: regulator of G protein signaling homology domain

RMSD: root mean square deviation

RNAi: Ribonucleic Acid Interference

S/B: signal to background ratio

SAXS: Small-angle X-ray scattering

SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SEC: size-exclusion/gel filtration chromatography

SH3: Src Homology 3 domain

SO: sypro orange

SUMIT: Selumetinib uveal melanoma investigator trial

TEM: transmission electron microscope

TEV: tobacco etch virus

TEAD: TEA domain

Titin: TTN

Tgat: transforming gene in adult T-cell leukemia

TLS: translation-libation-screw

$T_m$: thermal denaturation temperature (melting temperature)

TRIPα: Trio inhibitory peptide α

UM: uveal melanoma

WT: wild type

YAP: yes-associated protein

# Abstract

The guanine exchange factor Trio links $G\alpha_{q/11}$-coupled seven-transmembrane receptors to the activation of small GTPases in the Rac and Rho subfamilies, thereby regulating the processes of cytoskeletal rearrangements and gene transcription. This signal transduction pathway underlies the development of uveal melanoma, a fatal malignancy. Trio is a feasible drug target for uveal melanoma; RNAi knockdown of Trio is efficacious in a xenograft model of metastatic uveal melanoma. Trio is also found overexpressed in a variety of cancers and thus may serve as a target in these. Additionally, Trio is thought to primarily be involved in developmental processes and does not play many roles in adult humans. The path to targeting Trio for therapeutic development is not clear as we still do not understand the structural and functional roles of ~80% of the protein, and the portions we do understand are protein-protein interaction modules that interact with their targets using broad, relatively flat interfaces. My work has focused on determining structure-function relationships within Trio to lay groundwork for future inhibitor development. I have defined the autoregulatory mechanism employed by the C-terminal guanine exchange factor module of Trio (TrioC) by determining the crystal structure of TrioC in its basal state, which revealed how its pleckstrin homology domain and preceding linker region blocks the RhoA binding site on the Dbl homology domain. I demonstrated that missense and nonsense mutations found in cancer patients are able to hyperactivate TrioC, demonstrating how these mutations may be involved in cancer

growth or development. My later work focused on determining structure/function relationships within larger fragments of Trio, because the full-length protein is expressed in cancer. I have biochemically characterized these larger fragments and shown that they have higher Rac1 GEF activity than Trio's Rac1-selective GEF module alone. These constructs can be used to interrogate the function of additional Trio domains and I have begun characterizing these domains using electron microscopy and biochemical assays. Through my efforts, the materials and conditions necessary to determine the structure of a larger construct of Trio using electron microscopy have been assembled. My work has opened new avenues toward targeting Trio: inhibitor molecules could either stabilize the autoinhibited conformation of TrioC I report here or disrupt the region of Trio producing extra Rac1 GEF activity.

# Chapter 1 **Introduction**

## 1.1 Background and Introduction

This dissertation concerns the structure and function of the RhoGEF Trio, a $G\alpha_{q/11}$ target implicated in uveal melanoma. The following introductory Chapter will provide background in GPCR signaling and the link between Trio and uveal melanoma. Following that, a description of Trio homologs, splice variants, and information gained about Trio via genetic experiments are summarized. The Chapter ends with a domain by domain breakdown of what is known in the literature regarding the molecular determinants of Trio function and regulation.

### 1.1.1 GPCR and G Protein Signaling

G protein-coupled receptors (GPCRs) constitute the largest family of extracellular signaling receptors in eukaryotes. They are responsible for conveying signals from the extracellular milieu to the controlled intracellular environment and are hugely important for the physiological processes of sight, taste, smell, neurotransmission, and other processes induced by hormones, such as heart contractility. GPCRs are flexible membrane proteins that respond to extracellular inputs, such as a photon of light or the hormone epinephrine. The binding of these extracellular inputs changes GPCR conformation such that they can productively interact with intracellular heterotrimeric G proteins, which transmit these signals within the cell.[1] Heterotrimeric G proteins are composed of an α subunit (~45 kD) which binds to guanine nucleotides, and a βγ heterodimer (~45 kD) which both regulates the α subunit and signals on its own. The phosphorylation state of the bound nucleotide on α can change the conformation of the α subunit and allow this protein to bind to other proteins in the cell. G protein α subunits typically cycle between a guanosine diphosphate (GDP) bound state, in which they associate with the βγ dimer (and this complex can associate with the GPCR), and a guanosine triphosphate (GTP) bound state, in which the affinity for βγ dimer is decreased and the α subunit can interact with its downstream signaling partners. G protein α subunits (Gα) were classified first by their physiological effects and later by sequence homology and are outlined in Figure 1.1. There are four Gα classes, including the stimulatory (s) class, the inhibitory (i) class, the 12 class, and the q/11 class.[2-4]

| **Class** | **$G\alpha_s$** | **$G\alpha_i$** | **$G\alpha_{12}$** | **$G\alpha_q$** |
|---|---|---|---|---|
| **Members** | **$G\alpha_s$, $G\alpha_{olf}$** | **$G\alpha_i$, $G\alpha_t$, $G\alpha_o$, $G\alpha_z$, $G\alpha_{gust}$** | **$G\alpha_{12}$, $G\alpha_{13}$** | **$G\alpha_q$, $G\alpha_{11}$, $G\alpha_{14}$, $G\alpha_{16}$** |
| **Major Effectors** | **Adenylyl Cyclase (AC) Isoforms** | **AC Isoforms, cGMP-PDE*** | **RH#-RhoGEFs** | **PLCβ Isoforms<br>TrioC RhoGEFs** |
| **Associated Cancers** | **Thyroid/ Pituitary Tumors** | **Ovarian/ Adrenal Tumors** | **Burkitt's Lymphoma<br>Diffuse large-B-cell Lymphoma** | **Uveal Melanoma<br>Other Melanomas** |

***cGMP-PDE: cyclic guanosine monophosphate-phosphodiesterase**
**#RH: Regulator of G-protein Signaling Homology Domain**

Figure 1.1 Gα proteins are associated with various cancers and cancerous signaling likely occurs through their major effectors. Major Gα classes are shown, with class members listed below, followed by major effectors, and associated cancers.

Gα subunits are regulated by other proteins which can cycle bound nucleotide into the "off" (GDP-bound) state or the "on" (GTP-bound) state. Gα subunits are typically found in the GDP-bound state in an animal cell and are cycled to a GTP-bound state through interaction with activated GPCRs. The receptor acts as a guanine nucleotide exchange factor (GEF) in that it selectively binds to the nucleotide-free state of Gα, thereby forcing Gα to release the nucleotide it has bound. Eukaryotic cells maintain a higher intracellular concentration of GTP compared to GDP, and thus GTP enters Gα, triggering release from the GPCR and Gβγ. For its GTP-bound lifetime, the α subunit will interact with downstream effectors, proteins that carry out downstream signaling functions. α subunits have intrinsic GTPase activity that hydrolyzes the bound GTP to GDP, although this process does not occur at a relevant timescale to most signaling cascades, with the exception of $G\alpha_s$.[4,5] The GTPase activity of many α subunits can be enhanced by proteins called GTPase-activating proteins (GAPs). These proteins act by stabilizing the Gα catalytic site in a transition state conformation and thereby speeding up hydrolysis of the terminal phosphate of GTP. The largest family of GAP proteins for Gα subunits is the regulator of G protein signaling (RGS) family. Phospholipase Cβ enzymes also have a GAP domain selective for $G\alpha_{q/11}$ proteins [6] [7,8] The cycle of a heterotrimeric G protein is depicted in Figure 1.2.

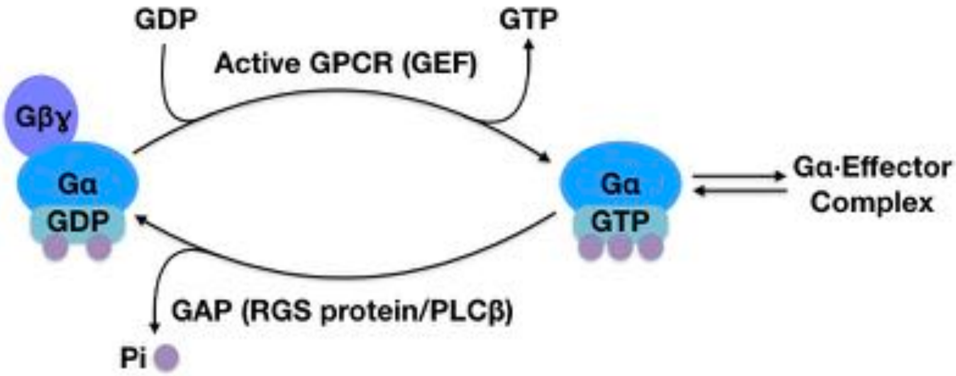


Figure 1.2 The life cycle of a heterotrimeric G protein is regulated by other proteins and the kinetics of this cycling determines the timescale and magnitude of signaling. The Gα subunit primarily associates with the βγ heterodimer until it is cycled to the GTP state by its cognate GEF molecule, usually an activated GPCR. GTP-bound Gα can freely associate with effector molecules to coordinate a series of downstream functions. Gα is cycled back to the GDP state by a GAP protein. G protein subunits are shown as cartoon ovals, with guanine nucleotides shown as a rectangles and phosphate groups as small purple circles.

### 1.1.2 G$\alpha_{q/11}$ Mutations in Uveal Melanoma and Signaling Implications

Recently it has been found that α subunits harboring mutations in their active site residues can lead to various malignancies due to their inability to return to the GDP-bound "off" state (Figure 1.1).[9] Most commonly, a variant in the "catalytic" glutamine or arginine residues (Figure 1.3) leads to a large decrease in rate in the GTP hydrolysis reaction. In the transition state for hydrolysis, Arg183 (G$\alpha_q$ numbering) contacts the β and γ phosphates and a water molecule in the $Mg^{2+}$ coordination sphere. Q209 also contacts the γ phosphate, and most importantly is in position to act as the base to deprotonate the nucleophilic water molecule. A mutation in either of these residues leads to inefficient

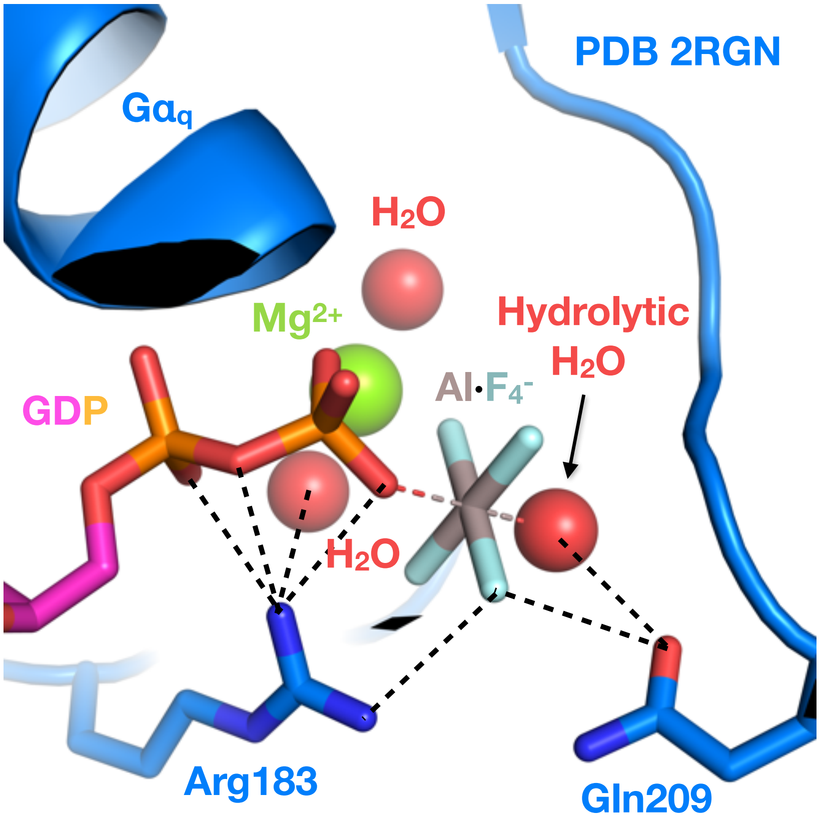


Figure 1.3 Heterotrimeric G protein active site shown with GDP·$ALF_4^-$ bound, showing the transition state for γ phosphate hydrolysis and highlighting the role of Arg183 and Gln209 in catalyzing hydrolysis reaction. G$\alpha_q$ is shown as blue cartoon representation, with key active site residues Arg183 and Glu209 shown as sticks. The GDP-$ALF_4^-$ conjugate is shown as sticks, where $ALF_4^-$ represents the transition state of γ phosphate hydrolysis. The $Mg^{2+}$ ion and waters are shown as spheres. Polar contacts <3.5 Å are indicated as black dotted lines and show how phosphates are coordinated to align the active site for γ phosphate hydrolysis.

GTP hydrolysis even in the presence of a cognate GAP, which is the key cancer driving event that occurs in uveal melanoma (UM).[10,11]

UM is the most common ocular cancer in the United States, with ~1500 new cases diagnosed per year and a high metastasis rate (~50 %). The disease is fatal in all but a few metastatic patients, and no efficacious treatments for metastatic UM exist. UM tumors originate in the pigmented cell types of the eye, and commonly spread to the liver. In UM, the $G\alpha_{q/11}$ is mutated in ~90 % of cases, with the most common variations occurring in residues Q209 and R183. The $G\alpha_{q/11}$ variation most often occurs early in disease progression and is considered the driving oncogene in most UM cases. There are also rare cases of mutations in upstream or downstream proteins in the $G\alpha_{q/11}$ signaling pathway leading to UM, including the cysteinyl leukotriene receptor 2 and phospholipase Cβ4. The Q209 and R183 variations in $G\alpha_{q/11}$ are hypothesized to lead to the constitutive activation of $G\alpha_{q/11}$ effectors, including phospholipase Cβ-family enzymes (PLCβ) and the Rho-family GEFs (RhoGEFs) Trio, Kalirin, and p63RhoGEF.[12-25]

Because $G\alpha_{q/11}$ regulates a host of other physiological processes including platelet activation, cardiac contractility, and neurotransmission, it would be challenging to target the G protein systemically using a dosed drug molecule, although $G\alpha_{q/11}$- selective probes such as FR900359 are currently being used to further dissect the progression of UM signaling events throughout the stages of tumor development.[26-29] Given the difficulties in inhibiting $G\alpha_{q/11}$, attention has largely focused around targeting downstream molecules.

Early attempts to define UM signaling events centered around the study of the canonical $G\alpha_{q/11}$ signaling pathway through PLCβ, which activates $Ca^{2+}$-dependent signaling, ultimately impinging on the mitogen-activated protein kinase (MAPK) pathway.[10,30-32] These studies led to clinical trials targeting MAPKs as potential targets for the treatment of UM, because well-defined roles for these kinases exist in cancer, and drugs targeting MAPKs already are on the market. The "Selumetinib Uveal Melanoma Investigator Trial" (SUMIT) dosed Selumetinib, an allosteric MEK1/2 inhibitor, to metastatic UM patients and advanced to a phase III clinical trial vs. the standard of care decarbazine, a cytotoxic DNA-alkylating drug. Selumetinib did not enhance progression-free survival vs. the control condition and also led to several adverse effects in patients.[33-35] This in combination with other evidence from models of UM suggests that the aberrant signaling events in UM involve other pathways than the canonical downstream target of $G\alpha_{q/11}$. Although PLCβ could still play a role in UM, it may be limited to the early stages of disease and dispensable in fully established UM. A comparison of $G\alpha_{q/11}$-mutant UM cell lines to those in which $G\alpha_{q/11}$ is not the causative oncogene shows no difference in MAPK activation between the two, suggesting that $G\alpha_{q/11}$ signaling to MAPK is dispensable in UM, providing a potential explanation for the results of these clinical trials.[36] The SUMIT clinical trial stands as the strongest evidence that targeting MAPKs alone is ineffective for UM treatment.

In 2014, an RNA interference-based knockdown study performed using a model of UM in *D. melanogaster* cells identified the nucleotide exchange factor Trio as a potential target for UM. In a mouse xenograft model of metastatic $G\alpha_q$-dependent UM, Trio knockdown

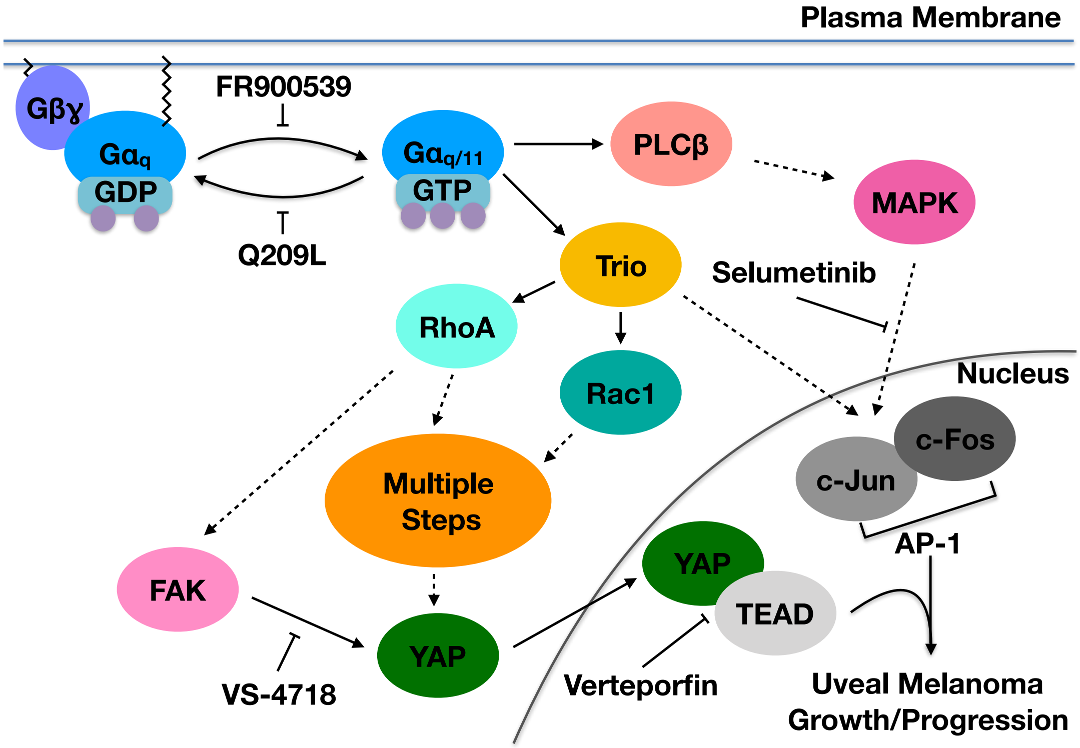


Figure 1.4 Signaling events dependent on $G\alpha_q$ in uveal melanoma occur through the effector proteins PLCβ and Trio, involve several signaling proteins, and converge on two families of transcriptions factors. The diagram is incomplete due to steps in the pathway that are not yet defined. Proteins are shown as ovals, with arrows indicating activation. Dashed arrows represent multiple steps not shown. Inhibitor molecules are shown in the diagram pointing to the step they inhibit using blocking symbols.

reduced both the size and weight of grafted tumors.[37] In-depth studies of Trio-mediated signaling revealed that Trio signaled to both the activator protein-1 (AP-1) and TEA domain (TEAD) transcription factors, both implicated in cancer.[38-41] A simplified signaling diagram is shown in Figure 1.4, where Trio signals to AP-1 and TEAD depending on RhoA and Rac1 nucleotide exchange activity.[42] Knockdown of Trio significantly reduces signaling to both AP-1 and TEAD in luciferase systems. Trio knockdown also reduces the mRNA levels of common genes regulated by TEAD in UM cell lines as measured by quantitative PCR. Knockdown of variant $G\alpha_q$-Q209L ($G\alpha_q$-QL), the variant most commonly found in UM, or the downstream GTPases RhoA or Rac1 replicate these transcriptional effects. Small molecule strategies for UM proposed include direct inhibition of $G\alpha_q$ using FR900359, inhibition of Focal Adhesion Kinase (downstream of Trio and

RhoA) using VS-4718, and use of verteporfin to inhibit the yes-associated protein (YAP)/TEAD interaction. These avenues are still being explored for feasibility. [28,37,42,43]

### 1.1.3 Trio Domain Architecture and Implications for Cancer

Trio is named for its three enzymatic domains, which are comprised of two closely related Dbl family GEF domains with different GTPase selectivity and a kinase domain (KD) with unknown activity. The N-terminal GEF module is termed TrioN and is specific to Rac1 and RhoG, and the C-terminal GEF module is called TrioC and is specific to the GTPase RhoA. The Trio KD is of the calmodulin kinase family.[44] G$\alpha_q$ directly binds to TrioC and stimulates nucleotide exchange on RhoA[45,46]. G$\alpha_q$ activation translocates Trio to the plasma membrane (PM)[37], which may have additional signaling implications because the GTPase substrates of Trio's GEF domains are also anchored at the plasma membrane through lipid modifications (Figure 1.5).[47] In addition to increased RhoA activation, overexpression of G$\alpha_q$-Q209L in 293 cells also stimulates Rac1 exchange in a Trio-dependent manner.[42] KD activity, in theory, would thus also be enhanced on PM-localized

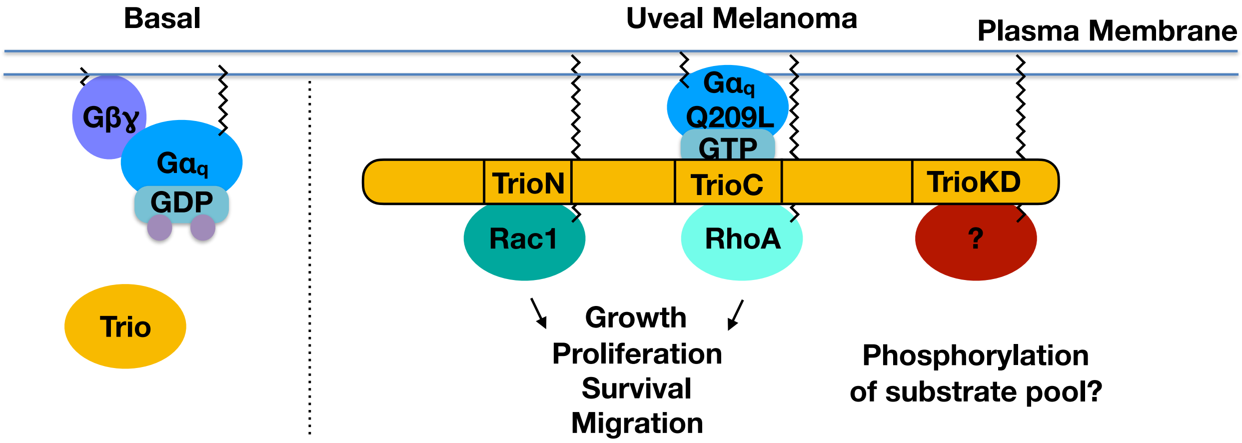


Figure 1.5 Trio signaling in uveal melanoma involves membrane translocation by G$\alpha_q$-Q209L and results in the activation of small GTPases and possible kinase signaling. In the basal state (left), no signaling is expected to occur through Trio. In the UM state (right), Trio is recruited to the membrane by G$\alpha_q$-Q209L and signaling can occur through Rac1, RhoA, and potentially the kinase domain. Proteins are shown as ovals, with arrows indicating activation. Serrated lines represent lipid modification.

substrates. Thus, in UM, Trio-dependent events through $G\alpha_q$ likely include Rac1 activation, RhoA activation, and possible phosphorylation events. Although not the focus of this document, Trio is found upregulated or overexpressed in a variety of cancers including glioblastoma, cervical and head and neck carcinomas, and other melanomas.[37,44]

### 1.1.4 Trio Homologs and Splice Forms

Mammalian Trio is ubiquitously expressed and has two homologs, Kalirin and p63RhoGEF (Figure 1.6). Unless otherwise mentioned, this document refers to the largest isoform of human Trio (hTrio), isoform A in Figure 1.7. Kalirin is largely expressed in the nervous system and shares Trio's overall domain architecture and 61% amino acid identity. Kalirin includes a fibronectin type III (FnIII) domain near the C-terminus which is not found in Trio. Trio knockout is embryonic lethal, wherein mice display large defects in skeletal muscle formation and minor defects in brain organization.[48] In contrast, Kalirin

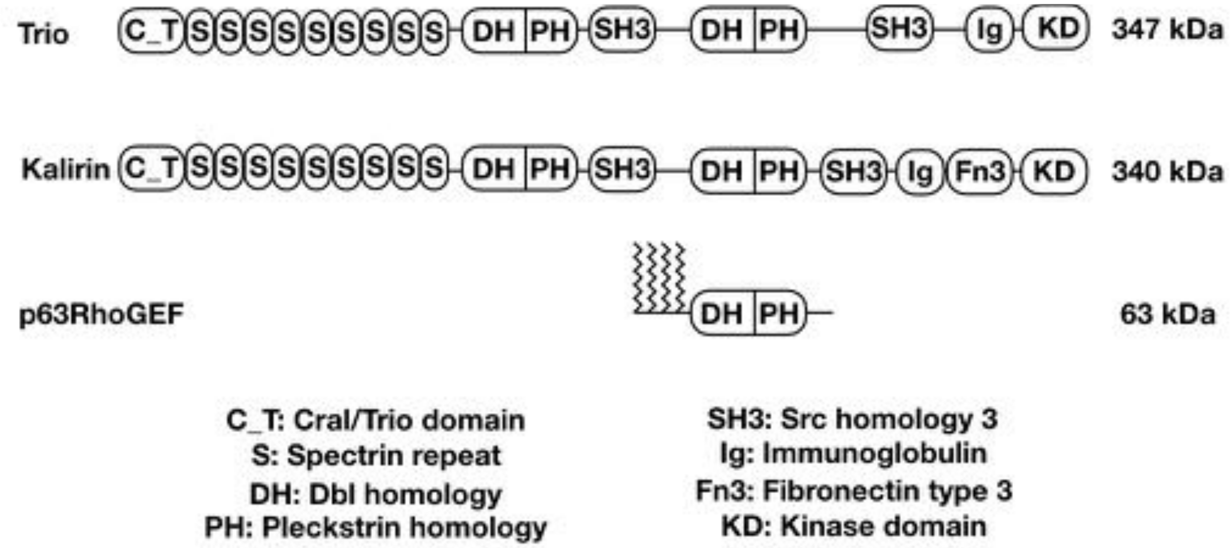


Figure 1.6 Domain architecture of Trio and its mammalian homologs highlight commonalities and differences between the proteins. Domain architecture is shown, with the backbone of the protein shown as a horizontal line and approximate domain boundaries shown as ovals. Approximate molecular masses are shown to the right of each protein. Lipid modifications are shown as serrated lines. The human versions of each protein were used to generate this diagram. C_T, Cral/Trio domain; S, Spectrin repeat; DH, Dbl homology; PH, Pleckstrin homology; SH3, Src homology 3; Ig, Immunoglobulin; Fn3, Fibronectin type 3; KD, Kinase domain.

knockout mice are viable, yet display reduced dendritic spine formation specifically in the cortex, and associated defects in working memory. Thus, although the two are highly related, Trio has developmental roles that Kalirin is unable to compensate for, likely in part due to its limited expression profile compared to Trio.[44,49] Both Trio and Kalirin are involved in coordinating neurite outgrowth, the projection of a nerve outside of its main cell body typically used to take in signals from a neighboring neuron. This process is deeply tied with coordination of the actin cytoskeleton and is primarily dependent on the GEF modules of Trio and Kalirin.

p63RhoGEF is a smaller RhoGEF involved in heart contractility and shares 65% identity to TrioC within its GEF module. [44,45] Outside of this GEF module, p63RhoGEF does not share Trio's overall architecture. p63RhoGEF contains a 148-residue N-terminal extension with several N-terminal cysteines that are palmitoylated, and an 88-residue C-terminal extension. Trio, when expressed as a fusion with green fluorescent protein (GFP) slowly recruits to the plasma membrane on the timescale of minutes in response to $G\alpha_q$-coupled receptor activation. [37] In contrast, a GFP fusion of p63RhoGEF is constitutively membrane localized, with $G\alpha_q$ playing a major role in allosterically activating the GEF module.[50] Membrane localization of p63RhoGEF is dependent on palmitoylation, and deletion of the N-terminus or substitution of the modified cysteines to serines can shift the protein to the cytoplasm and blunt its RhoGEF activity as a result. Artificially tethering p63RhoGEF to the plasma membrane using a PH domain from another protein restores its normal localization and activity. $G\alpha_q$ is also palmitoylated, and as such can rescue p63RhoGEF's ability to activate RhoA, but non-palmitoylated $G\alpha_q$ cannot replicate this activity. p63RhoGEF is therefore optimized for fast signaling kinetics through RhoA in accordance with its role in mediating angiotensin-II receptor signals in heart contractility, in contrast to Trio's slow and sustained kinetics, which are more relevant to development and cancer. It is possible that the immediate availability of p63RhoGEF at the membrane can act as a $G\alpha_q$ sponge and influence the kinetics of available $G\alpha_q$ to activating other effectors, including PLCβ isozymes. [51-53]

Trio and Kalirin are both large proteins whose genes an encode for a variety of proteins of variable length based on alternative splicing (Figure 1.7). Experiments profiling these splice variants found in developing rat brains indicate that both GEF modules of these proteins are required for development but come into play at different times.

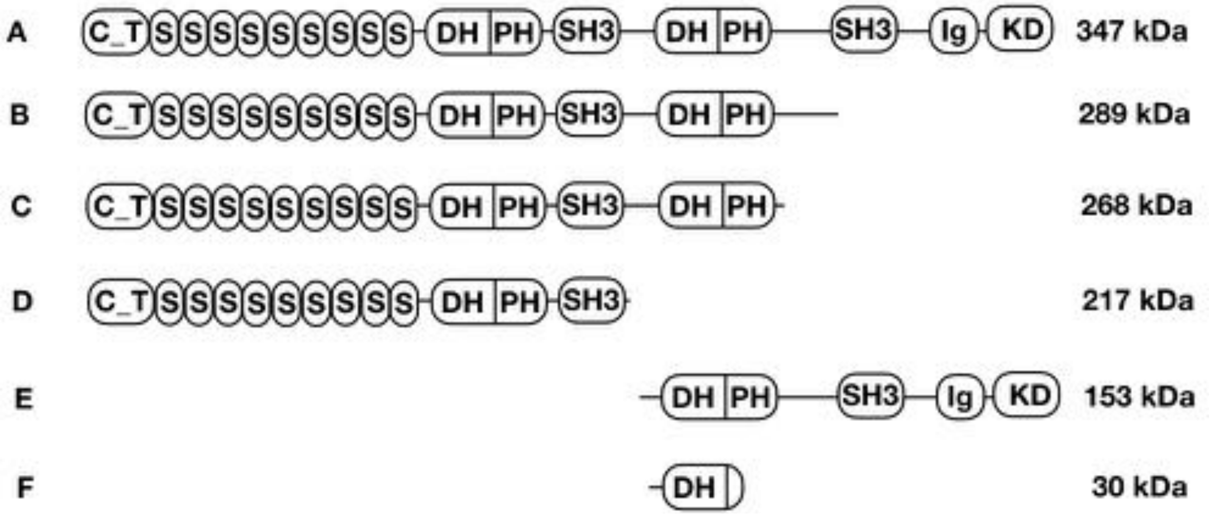


Figure 1.7 Domain organization of major splice variants of Trio to highlight diversity of domains present in different splice variants. Splice variants are designated A-F and approximate molecular masses are listed on the right. Kalirin also expresses similar splice variants to A-E of Trio. Horizontal lines show protein backbone and approximate domain boundaries are shown above as ovals. C_T, Cral/Trio domain; S, Spectrin repeat; DH, Dbl homology; PH, Pleckstrin homology; SH3, Src homology 3; Ig, Immunoglobulin; Fn3, Fibronectin type 3; KD, Kinase domain.

Earlier in mammalian development, isoforms of Kalirin encoding both GEF domains are present (A-C in Figure 1.7), giving way to an isoform containing only KalirinN later in development and throughout adulthood (D in Figure 1.7).[54] Trio isoform expression is more complicated. Full length Trio (TrioFL, A in Figure 1.7) is highly expressed in skeletal muscle and expressed at a low level in the brain during development. Expression of TrioFL tapers to a low level both in the brain and skeletal muscle fractions in adulthood.[55] Compared to other isoforms, TrioFL uniquely contains the Ig and kinase domains and thus these domains may play a role in skeletal muscle development. Two isoforms of Trio which terminate at slightly different endpoints after TrioC (B, C in Figure 1.7) are found in the brain and are barely detectable in skeletal muscle. These isoforms show little change in expression level from development through to adulthood. An isoform (D in Figure 1.7) of Trio terminating after the first SH3 domain increases in expression through development, with highest expression levels occurring during adulthood. A splice form found in humans is named "Duet" (E in Figure 1.7) and both Trio and Kalirin express a version of this. This splice variant does not change in expression significantly throughout development and is expressed in the brain. Similar to Kalirin, Trio also produces a splice variant terminating closely after TrioN (D in Figure 1.7) which increases in expression through development.

A final Trio splice variant (F in Figure 1.7) is a transforming gene in adult T-cell leukemia (Tgat). Tgat encodes the enzymatic portion of TrioC's GEF module followed by a unique 15 residue extension, not including the C-terminal pleckstrin homology domain normally present. Tgat is capable of transforming NIH 3T3 cells in culture, and these transformed cells have the ability to induce tumors when grafted into mice. The unique C-terminal extension on Tgat may play roles in membrane recruitment through its basic and hydrophobic character, and potential cysteine lipid modification sites. [56,57]

### 1.1.5 Trio Throughout Evolution

Human Trio was originally discovered in a screen to discover interactors with the LAR tyrosine phosphatase, and LAR was found to interact with a peptide encoding the Ig-KD domain tandem of Trio.[58] Shortly afterwards, UNC-73 was identified as a Trio/Kalirin homolog (29% identity to hTrio) in *Caenorhabditis elegans* [59] and *Drosophila melanogaster* Trio (dTrio, 44% identity to hTrio) was also discovered. [60] These homologs have been used to study the effects of Trio knockout or mutation on conserved signaling pathways. Like mammalian Trio (Figure 1.7), these invertebrate Trios also are expressed as various splice forms starting from a similar N-terminus. A domain pileup of Trio throughout the animal kingdom (Figure 1.8) shows that the overall domain architecture is found in every phylum within the animal kingdom except bryozoa, rotifers, and sponges.

Figure 1.8 Domain organization of Trio throughout the animal kingdom reveals a core conserved domain architecture with variable C-termini. Phyla are listed on the left with domain layout of Trio orthologs shown on the right. Horizontal lines show protein backbone and approximate domain boundaries are shown above as ovals. C_T, Cral/Trio domain; S, Spectrin repeat; DH, Dbl homology; PH, Pleckstrin homology; SH3, Src homology 3; Ig, Immunoglobulin; Fn3, Fibronectin type 3; KD, Kinase domain.

Trio is not found outside of the animal kingdom. Chordates all contain a domain architecture similar to hTrio. In cnidarians, mollusks, and echinoderms all domains are found, and the protein contains a FnIII domain similar to that of Kalirin (Figure 1.8). Nematodes lack the kinase domain and also have an FnIII domain. Annelids, arthropods, and Platyhelminthes terminate after TrioC's pleckstrin homology domain. A core portion of Trio, containing the Cral_Trio lipid binding domain, spectrin repeats, both Dbl and pleckstrin homology domain tandems, and one SH3 domain is thus highly conserved within the animal kingdom.

### 1.1.6 Dbl Family Overview

Trio's two GEF domains are part of the Dbl family, the largest family of RhoGEFs found in mammals. Both Dbl proteins and their substrates, the Rho GTPases, are conserved as far as the last eukaryotic common ancestor, underscoring their importance in regulating the actin cytoskeleton, cell growth, and migration.[61] The Dbl family is named for the founding member Dbl, originally identified as an oncogene in diffuse B-cell lymphoma and the family includes at least 69 members in humans.[62,63-65] The Dbl homology (DH) domain is a ~200 amino acid sequence and is the minimal region necessary for GEF activity. It is almost always followed by a regulatory pleckstrin homology (PH) domain of about ~150 residues. The DH domain is responsible for catalyzing the nucleotide exchange reaction by selectively binding a nucleotide-free state of the substrate GTPase. This occurs through disruption of the $Mg^{2+}$ binding site on the GTPase, as well as a change in conformation of certain residues interacting with the nucleotide. DH domains can range from being very specific to being promiscuous regarding the GTPases they bind to. The adjacent PH domain is typically involved in regulating the DH domain, either enhancing or inhibiting GEF activity on specific substrates by directing subcellular localization or through direct interactions or competition with the bound GTPase. PH domains exert these functions typically by binding the headgroups of phosphoinositides or to protein binding partners in the cell. [65]

### 1.1.7 Regulation of Dbl Family RhoGEFs by G Proteins

Many Dbl family GEFs are regulated by G proteins, with examples of regulation by both heterotrimeric and small Ras-like G proteins.[45] A subfamily of GEFs containing a domain with homology to the regulator of G protein signaling domain (RH) use this domain to bind to activated $G\alpha_{12/13}$ subunits. These include leukemia-associated RhoGEF (LARG), PDZ-RhoGEF, and p115RhoGEF. Binding of the Gα results in membrane localization of these GEFs and is speculated to also relieve intramolecular autoinhibition in the GEF protein through concurrent interaction with other domains. [45] $G\alpha_{q/11}$ regulate a subfamily of RhoGEFs consisting of TrioC, p63RhoGEF, and KalirinC. Instead of an RH domain, $G\alpha_{q/11}$ primarily bind to a uniquely conserved extension of the C-terminal helix of the PH domain but use additional contacts on the PH domain and the DH domain to bridge the domain tandem and confine the relative orientation of the two domains. This results in the relief of PH domain mediated autoinhibition, and possibly membrane localization in the case of Trio and Kalirin.[37] Gβγ subunits regulate a subfamily of RhoGEFs including P-Rex1, p114RhoGEF, Tim, and Clg. However, the molecular basis of how Gβγ interacts with and activates these proteins remains to be determined, although it likely involves membrane localization and is likely to be different in each case. Ras-like small G proteins have also been shown to regulate Dbl RhoGEFs. GTP-bound RhoA has been shown to directly interact with the PH domain of PDZ-RhoGEF and to the Ig domain of Trio. [66,67] Because RhoA is localized to the plasma membrane, this provides a possible avenue for feed-forward regulation of Dbl family members by stabilizing them at the membrane. GTP-bound Ras has also been shown to stimulate nucleotide exchange on Rac1, and this activity occurs both directly through Tiam1 and indirectly through SOS. [68,69]

### 1.1.8 Structure/Function Relationships in the DH/PH Tandem

The Dbl homology domain consists of a helical bundle comprised of 6 helical segments formed into a shape representing a chaise longue, with a concave face for the bound GTPase to “sit” in. DH domains were initially defined to have three highly conserved regions (CR). [70] CR1 and 3 come together to form a binding interface or “seat” for the GTPase, whereas CR2 is on the opposite face of the DH domain and is hypothesized to maintain the structural integrity of the domain. Residues from every helical segment except α2 contribute to the GTPase binding surface. (Figure 1.9).[71,46,72-76].

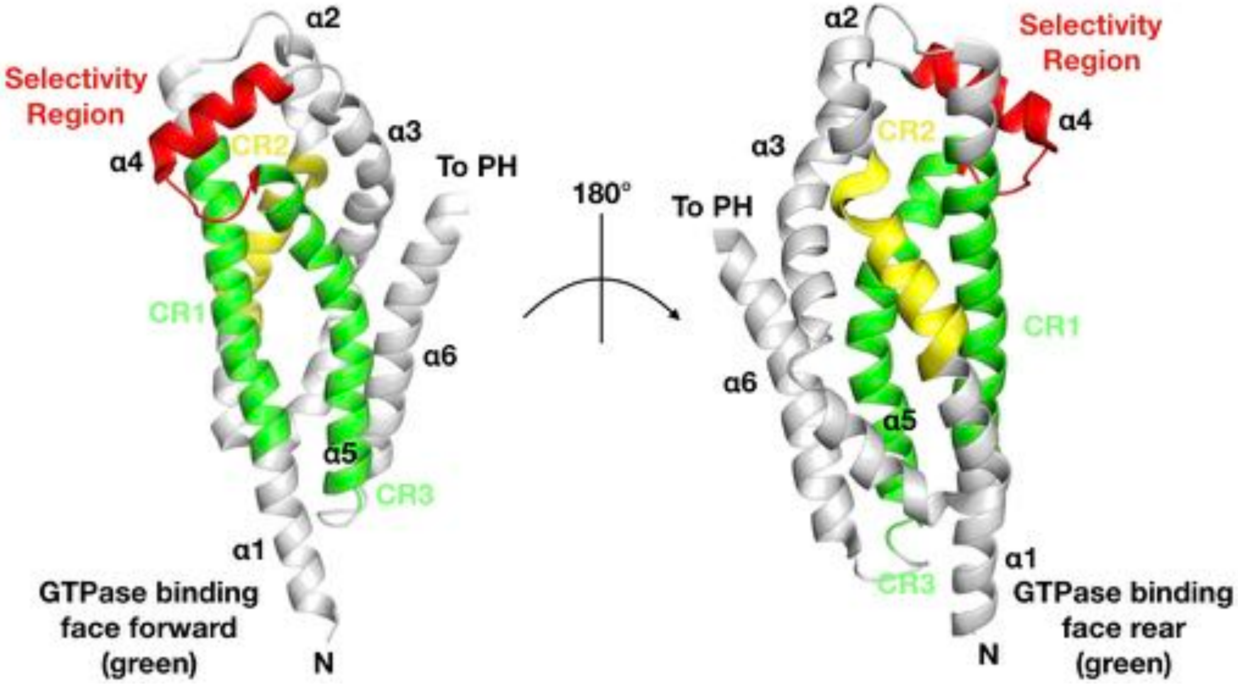


Figure 1.9 Conserved structural elements of the Dbl homology domain using p63RhoGEF as a model (PDB 2RGN). Conserved regions important in GTPase interaction are shown in green (CR1, 3), conserved region important for structural integrity of the domain in yellow (CR2), and selectivity region in red. Two views are shown for clarity. The DH domain of p63RhoGEF is shown in cartoon representation with secondary structure labeled.

Binding of a compatible GEF-GTPase pair produces a rearrangement of switches 1 and 2 in the GTPase. A comparison of RhoA in its GTP$\gamma$S form[77,78] with its GEF-bound state[46] can be used to understand the structural implications of GEF binding (Figure 1.10). In switch 1, Thr37 (human RhoA numbering) is displaced from the site where it coordinates the $Mg^{2+}$ ion. Val35 is placed about 2 Å from the 2' hydroxyl of the ribose. Switch 1 movement exposes the nucleotide binding pocket to solvent, most notably seen in the movement of Tyr34 away from its position as a "lid" and Phe30 from interacting with the nucleotide base, increasing the likelihood of nucleotides diffusing in and out of the pocket. In switch 2, major changes include Glu64 inserting into the nucleotide pocket ~2 Å from where the $\gamma$-phosphate would sit, and insertion of Ala61 to about ~2 Å from the $Mg^{2+}$ site.

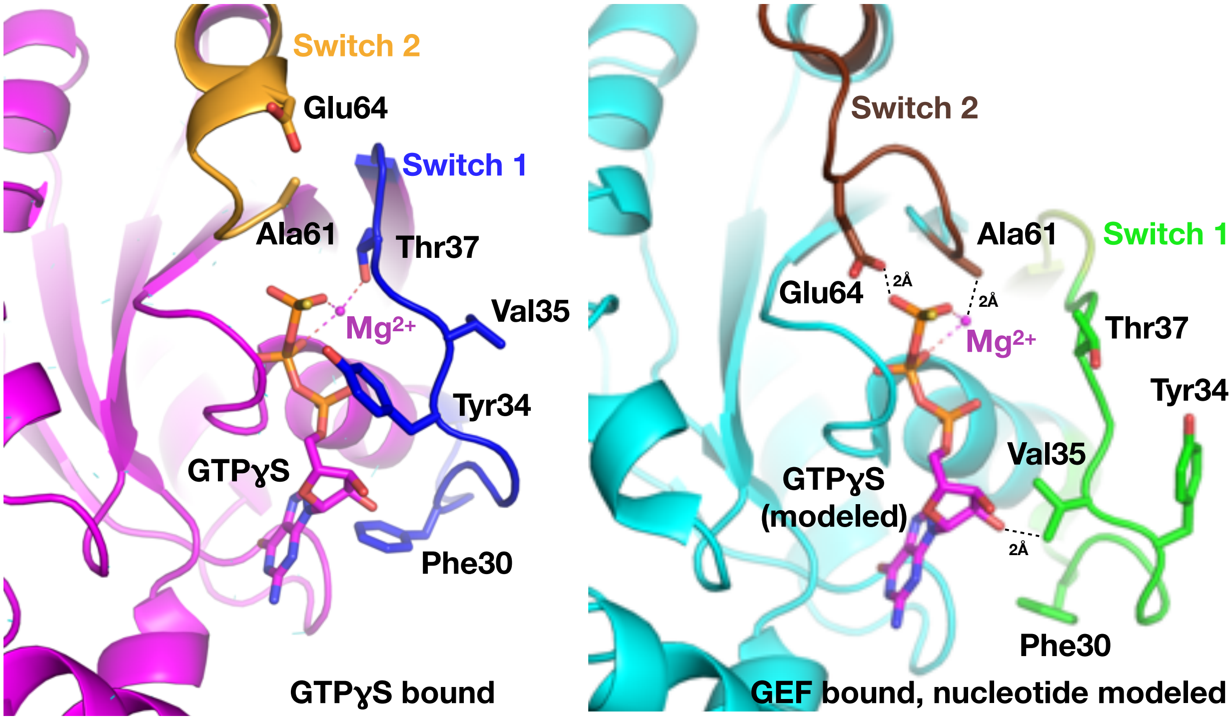


Figure 1.10 Comparison of GTP (left, PDB 1A2B) and GEF bound (PDB 2RGN with nucleotide modeled from 1A2B) states of RhoA to show structural elements which lead to nucleotide release in the GEF-bound state. On the left, RhoA in its GTP$\gamma$S bound state is shown as a magenta cartoon with switch 1 region highlighted in blue and switch 2 in orange. Key residues are shown as sticks. On the right, the GEF-bound (nucleotide-free) state of RhoA is shown as a cyan cartoon with switch 1 shown in green and switch 2 shown in brown. GTP$\gamma$S from the left structure is modeled in the nucleotide free state to show clashes, which are shown as black dotted lines with distance measurements shown. $Mg^{2+}$ is shown as a sphere and polar contacts with the $Mg^{2+}$ ion are shown as purple dotted lines.

On the GEF side, important residues include Glu1976 (human Trio numbering) which forms contacts with Thr37 and Tyr34, and Lys2106 which interacts with the backbone of Ala61. Both of these residues are highly conserved in the Dbl family. [63]

Substrate specificity for GEFs is determined by residues on both the enzyme and its substrate GTPase. On the GTPase side, a series of residues in the β2 and β3 strands usually dictate specificity for the given GEF.[79] On the GEF side, residues in α4, α5 and the loop connecting them contribute to specificity. (Figure 1.9) Specificity for RhoA is dictated primarily by a Lys/Arg in the α4-α5 loop (Arg2090 in Trio) interacting with Asp45 and Glu54 of RhoA.[76,79] Trio has two highly related DH domains (39% identity) yet each domain has distinct GTPase specificity. TrioN prefers Rac1 and RhoG, and TrioC is selective for RhoA. The specificity determinants for the TrioN/Rac1 pair are unclear as the two regions described above do not contact each other in the crystal structure, and the above polar residues are instead neutrally charged Gln and Asn. [58,74,80]

The DH domain is linked to the PH domain by a helix of variable length termed the α6-αN linker helix. Residues in α6 are formally considered part of the DH domain whereas residues in αN are considered part of the PH domain. Some GEFs have a βN secondary structure as opposed to αN.[81] α6 contains key residues which contact switch II of the bound GTPase. In addition, the α6-αN linker helix plays an important role in confining the relative orientation of the DH and PH domains and can thereby play an important role in the regulation of DH domain activity given binding events that occur in the bracketing DH and PH domains.

PH domains consist of a series of seven antiparallel β strands which form a sandwich-like structure (Figure 1.11). At one end of the sandwich, basic residues, if present, are thought to interact with the headgroups of phospholipids, sometimes in a highly specific manner.[82] The other end of the β-sandwich is capped by an extended C-terminal helix which marks the end of the PH domain. This terminal helix can form contacts with upstream or downstream GTPases. The loops connecting the β-strands in the PH domain opposite to the C-terminal helix often form part of a phospholipid interaction site, whereas lateral (at the membrane) protein-protein contacts are formed by the sides of the sandwich.[46,72-74,82] Low-affinity (and possibly non-specific) lipid binding may serve to

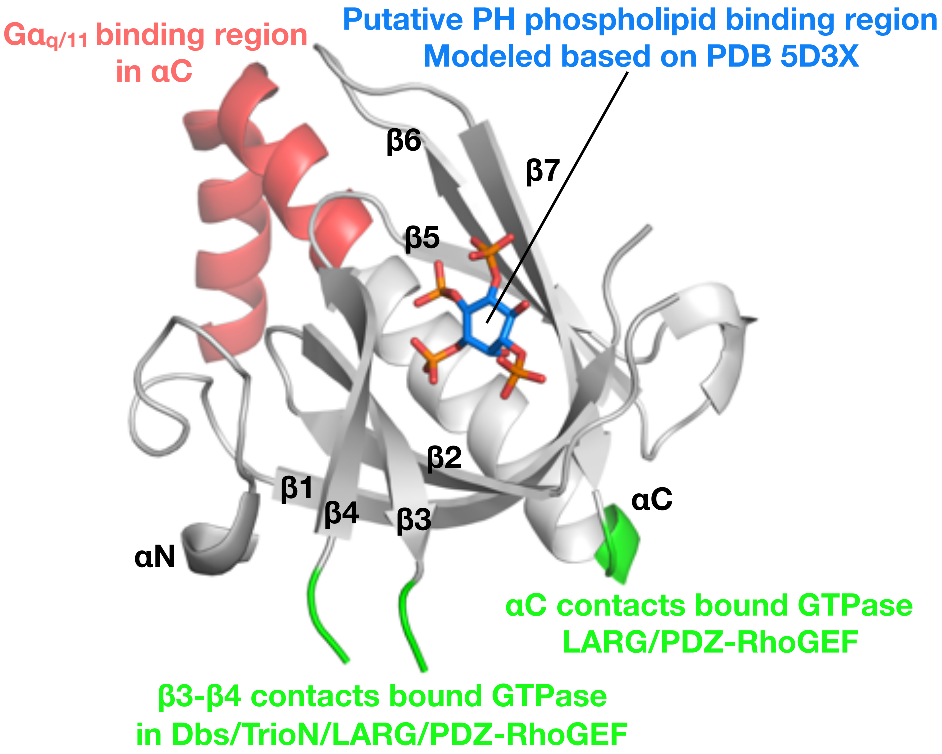


Figure 1.11 Key structural elements of the Pleckstrin homology domain shown using the PH domain of p63RhoGEF (PDB 2RGN) as a model. Regions important for GTPase contact are shown in green, region for Gα$_{q/11}$ interaction shown in red, and the head group of a phospholipid is modeled into the concave surface of the sandwich to show the putative phospholipid binding region. PH domain is shown in cartoon representation with secondary structure labeled.

orient the PH domain, and through α6-αN, the DH domain in order to facilitate productive substrate contact.[45,61,64,67,75-81]

Several examples of PH domain mediated regulation exist in the Dbl family. The PH domains of TrioN and Dbs enhance nucleotide exchange activity of the DH domain.[72,74,83,84] Specifically, residues from the β3-β4 loop in the PH domain orient residues in α6 of the DH domain to optimize contacts with switch II of the GTPase. There are also contacts seen directly between the β3-β4 loops of these GEFs and the GTPase, but mutation of these contacts does not produce any effect in vitro. Structures of TrioN and Dbs show a similar DH/PH relative orientation either in the presence or absence of bound GTPase, indicating these GEF modules are prearranged for catalysis. These PH domains also may provide correct localization through the binding of lipids or protein partners.[80,85,86]

The PH domains of the RH-RhoGEFs seem to play a similar role to those in TrioN and Dbs. Contacts between the PH domain proper and α6 of the DH domain contribute positively to nucleotide exchange. Structures of LARG and PDZ-RhoGEF in complex with RhoA show additional contacts formed between αC in the PH domain and RhoA, and mutations to these residues affect activity in LARG. [73] These same experiments have not been performed with PDZ-RhoGEF.

The PH domains in TrioC, KalirinC, and p63RhoGEF are autoinhibitory towards the RhoGEF activity of their DH domains, although the molecular basis of this autoinhibition is unknown.[46,80,87,88] A uniquely conserved amphipathic helix-turn-helix extension of the PH domain on these GEFs has been shown to be the primary binding determinant of Gα$_q$ (αC, Figure 1.12). The structure of the Gα$_q$·GDP·AlF$_4^-$-p63RhoGEF-RhoA complex revealed a broad range of additional contacts and the first instance of a Gα subunit interacting with the DH domain. Gα$_q$ makes contacts with several loops on the PH domain and extensive contact with the back face of the DH domain, including CR2. This binding mode, bridging the two domains, specifically constrains their relative orientation to one compatible with binding of nucleotide-free RhoA. (Figure 1.12) The conformation of the DH/PH tandem in this subfamily in the basal state is not known, but extensive functional

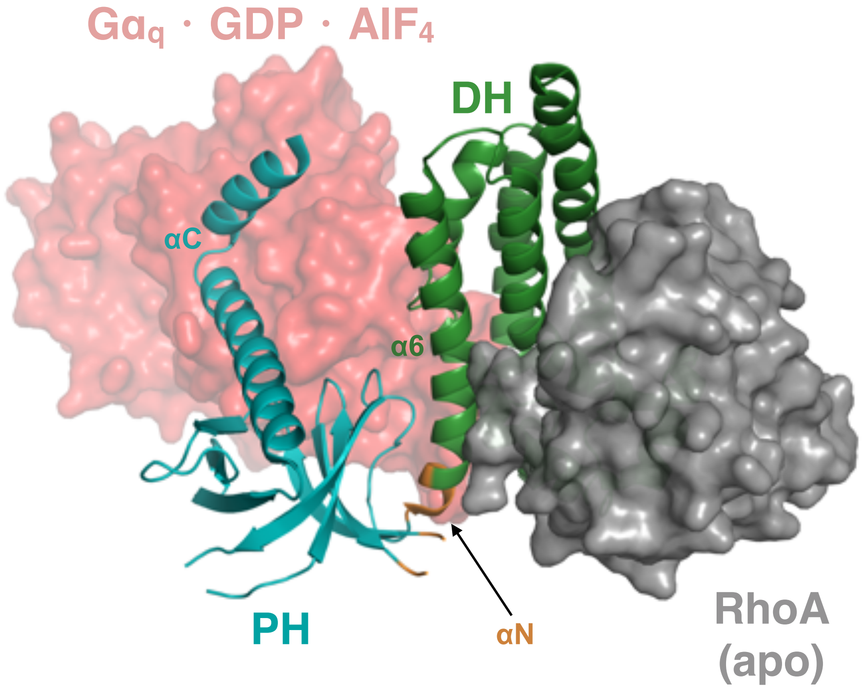


Figure 1.12 The structure of the Gα$_q$·GDP·AlF$_4^-$-p63RhoGEF-RhoA complex highlights the unique structural interface Gα$_q$ uses to activate TrioC subfamily members, comprised of surfaces on both the DH and PH domains. Gα$_q$·GDP·AlF$_4^-$ is shown using surface representation in red, and nucleotide-free RhoA (apo) is shown as a gray surface representation. The DH domain of p63RhoGEF is shown as a green cartoon, with αN shown as an orange cartoon, and the PH domain as a blue cartoon.

analysis of p63RhoGEF has localized the autoinhibitory constraint to the α6-αN region, yet the specific residues that mediate autoinhibition and the mechanism of autoinhibition remain unclear. The most likely candidates include residues in α6-αN present only in the $G\alpha_q$-regulated subfamily, although mutation of these residues in p63RhoGEF, including Gly2149 (Trio numbering) and Arg2150, only confers a modest level of activation compared to the full PH domain deletion.[88]

### 1.1.9 Cral_Trio Domain

Trio's N-terminal domain belongs to the CRAL_Trio (CT) family, named for its discovery in the cellular retinaldehyde binding protein (CRAL) and Trio. CT domains share homology with the Sec14 family of lipid transferases.[89,90,91] Sec14 domains have the ability to extract a hydrophobic ligand from one membrane and transport it into another. Common Sec14 substrates are inositol phospholipids and other hydrophobic molecules such as vitamins A and E. Of the Dbl family, Trio, Kalirin, Dbl, and Dbl's big sister (Dbs) contain N-terminal CT domains. Curiously, the CT domains found in Dbl proteins lack an N-terminal helical region present in most Sec14 domains. It has been speculated that the lack of this N-terminal region leads to a loss of lipid transferase activity while retaining lipid binding activity. As such, CT domains in Dbl proteins are hypothesized to have allosteric effects upon binding to their ligand.[92] Alternatively, they may direct subcellular localization through binding a specific type of lipid. Ligands of Dbl family CT domains are unknown in vivo, although in vitro lipid dot blot experiments have shown that both the Kalirin and Dbs CT domains bind to phosphoinositides.[93,94,95] Deletion of Kalirin's CT domain has been attempted but results are inconclusive. [95,96,94] Dbs, when overexpressed as a GFP fusion, localizes to the Golgi apparatus in 3T3 cells. However, when its CT domain is deleted (DbsΔCT), DbsΔCT is found at the plasma membrane, co-localizing with cytoskeletal structures not present when Dbs WT is transfected. DbsΔCT can transform 3T3 cells in vitro (Dbs WT does not) yet does not have altered GEF activity, leading to the hypothesis that the localization is Dbs is highly important in maintaining its normal cellular function. Dbl increases its transforming potential about ten-fold upon deletion of its CT domain.[97] Finally, a single report has shown binding between the CT domains of Dbl, Dbs, and Kalirin with $G\beta\gamma$ subunits.[98] Thus, as seen with other signaling domains, it is possible that the CT domain which originally evolved to bind lipids has been co-opted to bind G proteins. Crystal structures of several Sec14 domain proteins have been solved. The overall domain consists of an N-terminal α-helical region, followed by a β-sheet enveloped by α-helices which form the deep lipid-binding cavity. Residues that contact the bind lipid typically are housed on the central β-strand and opposing helices. At one end of the cavity is a lid helix which is thought to open and close to allow ligand access in response to membrane association. [99] Trio does retain all these major structural features based on sequence conservation, but the physical consequence of the lid helix opening and closing in Trio remains an open question.

### 1.1.10 Spectrin Repeats

C-terminal to the CT domain, Trio contains a series of spectrin repeats. The spectrin repeat is a ~100 residue elongated 3-helix bundle of about 50Å in length found in many proteins associated with the actin cytoskeleton. There are typically >4 repeats per protein.[100] Trio and Kalirin both contain 9 spectrin repeats, whereas Dbl and Dbs have only one. One hypothesized function of spectrin is to maintain linear distance between the two domains on either end of the repeat region. The structure of $\alpha$-actinin, a muscle protein which has 4 spectrin repeats, shows these repeats dimerizing in a linear antiparallel fashion in order to maintain distance between functional domains on either end of the dimer.[101] Crystal structures of non-dimeric spectrin repeat regions are also linear, although these may be influenced by crystal packing.[102-104] Assuming the spectrin repeat region in Trio is also linear, the N-terminal CT domain is kept at a maximum of about 5 nm away from the TrioN GEF module. Since this Rac1/RhoG-specific GEF module is heavily involved in regulating neurite outgrowth, this 5 nm distance may have an important physiological role. However, if Trio does not dimerize, these spectrin repeats may have more flexibility than actinin and thus mediate intramolecular interactions. Spectrin domains are also thought to act as scaffolding domains for signaling complexes. The myosin regulatory protein Supervillin has been reported to directly interact with the spectrin repeat region of Trio and Kalirin.[105] Deletion of Kalirin's spectrin repeat region and biophysical characterization has been attempted.[94,95,106]

### 1.1.11 Src Homology 3 Domain

Greater than 20 Dbl family members contain a Src homology 3 (SH3) domain in close proximity in primary sequence to their GEF module. Trio and Kalirin contain two SH3 domains (SH3N, SH3C) found C-terminal to each GEF module.[65] SH3 domains typically bind to polyproline motifs of the form PXXP, where X is typically a hydrophobic residue.[107] SH3 domains mediate protein-protein interactions, and as such they might bind to intramolecular sequences to mediate conformational change or steric occlusion, or they might bind to other proteins. SH3 domains are small globular domains of about 50-60 amino acids that form a series of two β-sheets connected by variable loops. Conserved hydrophobic residues in β2 and β3 and residues on the largest loop mediate substrate recognition and specificity. The peptide ligand must assume a left-handed polyproline helix 2 conformation, and two turns per helix sit in the SH3 binding pocket. A basic residue upstream of the prolines is also required to direct proper substrate orientation. SH3 domains typically have low binding affinity and poor specificity for given substrates.[108] The two SH3 domains in Trio share 33% identity and they are found different distances from their upstream GEF modules. SH3N is 60 residues C-terminal to TrioN, and SH3C is 300 residues C-terminal to TrioC. SH3C is closer in primary sequence to the Ig domain following it in the primary sequence (60 residues). It is not known what the function of SH3 domains are in Dbl family proteins. Solution nuclear magnetic resonance (NMR) structures of rat Kalirin SH3N (68% identity to human Trio SH3N) and mouse Kalirin SH3C (61% identity to human trio SH3C) are deposited in the protein data bank (PDB). Intramolecular ligands of Kalirin SH3N have been proposed, with the isolated ligand having 0.5 mM affinity for isolated SH3N via NMR. The SH3-ligand interaction is hypothesized to be autoinhibitory towards KalirinN GEF activity.[109]

### 1.1.12 Immunoglobulin Domain

Immediately before the kinase domain, Trio has a 100 residue Immunoglobulin (Ig) domain of the C2 type, which refers to its similarity to the constant region of immunoglobulin.[110] Ig domains are commonly thought to mediate protein-protein or protein-small molecule interactions, with a focus on the actin cytoskeleton. Trio's Ig domain is most similar to those in myosin light chain kinase (MLCK) (38% identity) and Titin (TTN) (37% identity) who also have calmodulin kinase domains. Ig domains in MLCK and TTN are involved in kinase domain substrate recognition in these molecules.[111] Both of these proteins are involved in muscle development. Trio knockout mice have defects in muscle development, and as such Trio's Ig domain may also play a role in recognizing kinase domain substrates. [48] The structure of the Ig domain is a conserved core of 4 β-strands surrounded by 3-5 variable strands with at least one disulfide linkage.[110] Apart from this general information, the molecular details of Ig domain structure-function are not understood. One report has demonstrated binding of the Ig domain of Trio to the prenylated form of RhoA-GTP. [67] In this model, RhoA-GTP generated through Trio-dependent nucleotide exchange can anchor Trio to the plasma membrane where its substrate GTPases are located, thus feeding forward into further Trio activation.

### 1.1.13 Calmodulin Kinase Domain

At its C-terminus, Trio has a kinase domain (KD) that belongs to the calmodulin kinase (CaMK) family of about 250 residues. CaMK domains are serine/threonine specific kinase domains and are typically regulated by the calcium binding protein calmodulin (CaM). Trio's KD shares highest homology to death-associated protein kinases (42% identity to death-associated protein kinase 1), myosin light chain kinase (40% identity), calmodulin-dependent protein kinase (35% identity), and titin kinase (34% identity). These kinases all share the hallmark of a ~50 residue sequence C-terminal to the KD involved in CaM regulation. Typically, this C-terminal sequence binds to both the peptide binding region and the nucleotide binding site of the kinase domain to autoinhibit kinase activity. CaM binding to this C-terminal sequence relieves autoinhibition and restores kinase activity. [112-115] Trio shares this C-terminal regulatory sequence, and it is thus possible Trio is regulated in the same fashion. Trio's KD substrate pool and biological function are not known, although as discussed above it may play a role in skeletal muscle development in line with the Trio knockout phenotype and data with Trio splice variants. [55] [48]

### 1.1.14 Research Summary and Goals

Trio proves a promising target for the treatment of uveal melanoma, yet drugging this protein is still challenging as many structure/function relationships within Trio remain unclear. I plan to determine the molecular basis for function in Trio's domains in order to provide an avenue for inhibitor development. I had two major hypotheses when approaching this work; first that TrioC regulation is determined by residues in the linker region between DH and PH domains; second that domains outside of the two GEF modules of Trio regulate one or both of its GEF activities. In Chapter two, I describe my work in determining the molecular basis for autoinhibition in the TrioC GEF module through a combination of structural biology, biophysical, and biochemical techniques. This work defines the mode of regulation for the TrioC subfamily. In Chapter three, I summarize preliminary attempts to characterize the only known TrioC inhibitor and develop screening assays against TrioC. In Chapter four, I report on my attempts to purify and characterize individual uncharacterized domains of Trio. I have developed constructs for both SH3 domains and the Ig domain that can be used in future experiments to interrogate the function of these domains. I also detail my efforts in characterizing larger constructs of human Trio. I have developed constructs which could be used in future efforts to determine structures of larger fragments of Trio in complex with binding partners. To this end, I have optimized purification and preliminary freezing conditions for a future cryoEM structure of a fragment of Trio hyperactive in Rac1 GEF activity in complex with Rac1.

# Chapter 2 **TrioC Structure and Function**

## 2.1 Introduction

The following Chapter concerns the molecular basis of TrioC autoinhibition, and the majority of this work has been published in *Science Signaling*. The figures represented here are modified from the open access version of the article.[116] This work set out to test the hypothesis that TrioC autoinhibition is conferred primarily through residues in the αN region of the PH domain. A brief introduction to DH/PH structure and function and regulation of TrioC subfamily members by $G\alpha_{q/11}$ precedes the methods and results sections of this Chapter concluding with a discussion and conclusion section summarizing my model for TrioC autoregulation.

Rho guanine exchange factors (RhoGEFs) are signaling modules that activate Rho-family small molecular weight GTPases.[65] These enzymes stabilize a nucleotide-free state of their cognate GTPases, thereby accelerating the process of nucleotide exchange. The C-terminal RhoGEF module of Trio (TrioC) regulates developmental and growth processes by influencing the actin cytoskeleton and gene transcription through activation of RhoA.[87] TrioC and the related RhoGEF modules p63RhoGEF and KalirinC are downstream effectors of $G\alpha_{q/11}$ and thereby give rise to a chain of phospholipase C-β-independent events upon activation of $G\alpha_{q/11}$-coupled GPCRs. [45,46] In >80% of cases, a constitutively active mutation in $G\alpha_{q/11}$ drives the progression of uveal melanoma (UM) in a Trio-dependent fashion. [15,37,42]

The prototypic RhoGEF module is composed of a Dbl homology (DH) and pleckstrin homology (PH) domain tandem linked by a flexible helix of variable length. [45,88] The DH domain is responsible for binding the nucleotide-free state of substrate GTPases, whereas the PH domain plays various roles ranging from enhancement of GEF activity, such as in Dbl's big sister (Dbs) and the N-terminal DH/PH module of Trio (TrioN), to suppression of GEF activity as in the TrioC subfamily. Regulation mediated by the PH domain is known to occur by one of several mechanisms, including protein-protein or lipid-protein interactions. [72,73,82,83] [46,74,88]

Structural and functional studies of p63RhoGEF, a close homolog of TrioC, show that $G\alpha_{q/11}$·GTP binds to both the DH and PH domains and thereby constrains the DH/PH module in a manner that optimizes the RhoA binding site. However, the structural basis for how the PH domain mediates autoinhibition in the TrioC subfamily remains unclear.[46,88] Such information would enable a better understanding of how Trio contributes to cancer progression and pave the way for future therapeutics that could stabilize the less active, basal form of TrioC. There are currently no effective approved therapies for the treatment of UM.[35]

In this work, I used X-ray crystallography to show that the TrioC PH domain inhibits GEF activity by forming an interface with the DH domain that blocks the binding site for switch II of RhoA. Using biochemical assays, I have demonstrated the importance of residues unique to TrioC, as well as subfamily members p63RhoGEF and KalirinC, in the N-terminal α-helix of the PH domain (αN) that contribute to the interface. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) supports a model wherein the RhoA binding site on the DH domain is occluded by the PH domain through contacts made by the observed DH-PH interface. Furthermore, we demonstrated that mutations found in the TrioC αN region in cancer patients not only activate the TrioC fragment in GEF assays, but also full-length Trio in human cells, allowing for sustained signaling through RhoA. [117-119]

## 2.2 Methods

### 2.2.1 TrioC Cloning and Site Directed Mutagenesis

All residue numbering in this manuscript is with respect to human Trio isoform 1, which is 3097 amino acids in length (UniprotKB O75962). Human TrioC complimentary deoxyribonucleic acid (cDNA) (hTrio residues 1960-2290) was a gift from the Gutkind laboratory and was inserted into a modified pMAL expression vector (pMalC2H10T)[73] with restriction cloning using 5′ EcoRI and 3′ XbaI sites. “TrioCΔC” (hTrio 1960-2275) was also inserted into pMalC2H10T using the same restriction cloning scheme. TrioCΔC was designed based on the prediction that the conserved C-terminal extension of the PH domain (residues 2275-2290) is likely disordered in the absence of $G\alpha_q$. [87] The RhoA construct was described previously and consists of human RhoA cDNA inserted into pMalC2H10T.[73] $G\alpha_q\Delta N$ was used in the $G\alpha_q$ activation assays and corresponds to residues 35-359 of murine $G\alpha_{q.}$[120] Site-directed mutants and deletions were generated on the TrioC vector described above using an inverse polymerase chain reaction (PCR) or the Quikchange protocol from Agilent. TrioC cDNA was set in an inverse PCR reaction with Q5 polymerase (New England Biolabs (NEB)), and the PCR reaction was digested with Dpn1 (NEB), phosphorylated with T4 polynucleotide kinase (NEB), and ligated with T4 DNA ligase (NEB). The ligated plasmid was then transformed into XL1-Blue cells (Agilent), and the mutation was confirmed using Sanger sequencing of plasmid DNA purified using a Mini-prep kit (Qiagen). The Quikchange protocol was similar except the PCR reaction is DpnI digested and the resultant reaction was directly used to transform XL1-Blue cells. Sequencing was performed through the University of Michigan DNA sequencing core. (Below cloning performed by Nadia Arang - Gutkind Laboratory) To generate variants in TrioFL-pEGFP (containing Trio residues 61-3097), a 1000 base pair fragment was PCR amplified from TrioC mutant constructs in pMalC2H10T. Gibson assembly was used to join fragments to a fragment from TrioFL containing an additional restriction site. Final assembled fragments and TrioFL were then digested with SpeI and FseI and ligated to generate the full-length gene variants (reagents from NEB).

### 2.2.2 Protein Expression and Purification

Plasmids encoding TrioC variants and RhoA were transformed into Rosetta (DE3) pLysS *Escherichia coli* (*E. coli*) cells (Novagen) and grown in Terrific Broth (EMD Millipore Sigma) with 100 µg/mL ampicillin or carbenicillin at 37 ˚C with 200 rpm shaking. Once a 600 nm optical density (OD600) of 0.6-0.8 was reached, expression of N-terminally tagged maltose binding protein (MBP) fusion proteins was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside and cells were further allowed to grow at 20 ˚C with 200 rpm shaking for 20 h. E. coli were then harvested at 5000 × g for 15 min and cell pellets were flash-frozen or prepared. Cell pellets were resuspended using a Dounce homogenizer in an ice-cold "lysis buffer" containing 20 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) HEPES pH 8.0, 200 mM NaCl, 2 mM dithiothreitol (DTT), 2 mM $MgCl_2$, 5 % glycerol, 0.001 mM leupeptin, 1 mM lima bean trypsin inhibitor, and 0.1 mM phenylmethylsulfonyl fluoride. Resuspended cell solution was then lysed using an EmulsiFlex C3 homogenizer (Avestin), and deoxyribonuclease I (DNase I) or benzonase was added to hydrolyze excess nucleic acid. Lysate was centrifuged at 32000 × g in an Avanti J-20 centrifuge (Beckman-Coulter) to remove insoluble material. The soluble fraction was then filtered through a glass fiber filter and loaded onto nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) equilibrated with lysis buffer. Next, 10 column volumes (CV) of lysis buffer were used to wash the column, followed by a 10 CV wash of lysis buffer containing 20 mM imidazole and 500 mM NaCl. A final wash step consisted of 10 CV of lysis buffer plus 20 mM imidazole. The recombinant protein was then eluted using lysis buffer plus 200 mM imidazole. The elution fractions were then incubated with 5 % (w/w) tobacco etch virus protease in order to cleave the N-terminal MBP expression tag and the mixture was dialyzed against a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT in order to remove imidazole. MBP was then removed using another round of Ni-NTA purification, TEV protease reaction mixture was flowed over the Ni-NTA column, and the column was washed (20 mM imidazole) and eluted (200 mM imidazole) using 10CV each of dialysis buffer with imidazole to check for MBP and TEV protease removal. Proteins were then checked for purity using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and pure fractions were concentrated in a 30kD cutoff Amicon Ultracel concentrator (Millipore), flash frozen in liquid nitrogen, and stored at –80 ˚C. For DSF and GEF assays, protein was thawed from -80 ˚C, 0.2 µm filtered, and used for experiments.  For crystallography and hydrogen deuterium exchange experiments, fractions containing TrioC were thawed from –80 ˚C on ice, and polished using gel filtration chromatography on a Superdex 75 10/300GL column (General Electric healthcare) in a buffer containing 20 mM HEPES pH 8.0, 100 mM NaCl, and 2 mM DTT. Fractions containing TrioC were then concentrated using an

Amicon Ultracel concentrator before using. Human RhoA was purified in a similar manner to that described above with the addition of 3 mM $MgCl_2$ and 10 µM GDP in all buffer solutions.[73] $G\alpha_q\Delta N$ was purified as described previously[120]. Protein concentrations were evaluated using 280 nm absorbance ($A_{280}$) on an ND-1000 (NanoDrop Technologies) and standardized using theoretical extinction coefficients generated by the ProtParam ExPASy webserver.[121]

### 2.2.3 TrioCΔC Crystallization

TrioCΔC fractions containing pure protein (~37 kD) were concentrated using a 30 kD cutoff centrifugal concentrator and sitting drop screening trays were set using commercial screens from Qiagen and Hampton Research: IndexHT, Protein Complex, Classics Lite, and JCSG+. 96 well plates from Hampton Research (HR3-185) were used and incubated in 20 °C and 4 °C rooms. Hits from original screens were optimized in 24-well hanging drop VDX plates (Hampton Research) in order to optimize crystal size and quality. The final TrioCΔC crystal form was grown at 20 °C using hanging drop vapor diffusion. The drop contained 2 µL of 2 mg/mL protein mixed with 2 µL of well solution containing 100 mM HEPES pH 7.5 and 14% polyethylene glycol 3350 and 1 µL of double distilled $H_2O$. Crystals grew over a course of 2 days and took the form of 3 dimensional plates of about 25×100×300 µm. Crystals were harvested from the drop using nylon cryoloops and transferred into a solution of freshly prepared mother liquor plus various cryoprotectants: 20% glycerol, 20% ethylene glycol, 20% PEG 400, and no cryoprotectant. Crystals were then plunged into liquid $N_2$ and shipped to the Advanced Photon Source for remote data collection. The crystal used for the final crystal structure was harvested from the drop it grew in and plunged directly into $N_2$.

### 2.2.4 Data Collection and Processing

Diffraction data were collected remotely at 110 K on the Life Sciences Collaborative Access Team beamline 21-ID-G at the Advanced Photon Source at Argonne National Laboratory and reduced, indexed, integrated, and scaled using the HKL2000 software package.[122] Initial attempts to solve ~6 Å datasets of TrioC using various Dbl family members as models (DH/PH structures of p63RhoGEF, Dbl, Dbs, TrioN) with molecular replacement did not yield any solutions. Statistics for the final TrioC dataset are shown in Table 2.1. Initial phases for the dataset which yielded the final TrioC structure were provided by molecular replacement in Phaser[123] with Chainsaw[124] processed versions of the individual DH and PH domains of p63RhoGEF[46] (PDB 2RGN) as sequential search models. Initial rounds of refinement consisted of manual model building and real space refinement in Coot[125] alternating with reciprocal space refinement in the PHENIX software package[126] using simulated annealing and torsional non-crystallographic symmetry (NCS) to restrain the three copies of TrioC in the asymmetric unit. Final rounds of refinement consisted of manual model building and real-space refinement in Coot, and refinement using REFMAC5[127] with manual X-ray weighting and translation-libration-screw refinement. NCS restraints were omitted in the final rounds to allow for minor variations between chains. Structure validation was performed using built in validation tools in Coot, the MolProbity server[128], and the PDB_REDO webserver[129]. Refinement statistics are shown in Table 2.1. The TrioCΔC coordinates have been deposited into the PDB under accession code 6D8Z. [116]

### 2.2.5 Structural Comparisons

Unless stated otherwise, Chain A of the TrioC structure was used for comparisons. Chain B of PDB entry 2RGN was used for models of p63RhoGEF. Chain A of PDB entry 1RJ2 was used as a model for Dbs, and Chain A of PDB entry 1LB1 was used for RhoA-bound Dbs. Comparisons were made by visual inspection using the PyMOL Molecular Graphics System, Version 1.8.6.2, Schrodinger, LLC. The script Rotation_Axis.py was used to calculate angles of rotation.

### 2.2.6 Differential Scanning Fluorimetry

DSF experiments were performed first on a ThermoFluor plate reader (Johnson & Johnson) using 8-anilinonaphthalene-1-sulfonic acid as the fluorescent dye (WT TrioC $T_m$= 44.8 °C with a 95% CI of [44.4, 45.3], N=11 independent experiments in at least duplicate). Data were also collected on a 7900HT Fast Real-Time PCR system (Applied Biosystems) or a QuantStudio 6 Real-Time PCR system (Applied BioSystems) using Sypro Orange (SO) dye (ThermoFisher) (WT TrioC $T_m$= 49.4 °C with a 95% CI of [49.2, 49.7], N=15 independent experiments in at least duplicate). Although the absolute $T_m$ for WT TrioC was different for each dye, we tested TrioC WT and R2150E and found $\Delta T_m$ for R2150E ($T_m$ R2150E - $T_m$ WT) was similar: +0.9 °C on the ThermoFluor instrument, and +0.6 °C on the PCR instruments (N=3 independent experiments performed in at least duplicate). Purified TrioC variants were incubated at 0.2 mg/mL in a buffer containing 20 mM HEPES pH 8.0, 200mM NaCl, and 2mM DTT with dye. Black 384 well PCR plates (Applied Biosystems) were used for the ThermoFlour instrument and wells were covered with silicon oil. For the QPCR instruments, white 384 well PCR plates (Applied Biosystems) were used and covered with sealing tape. These plates were exposed to a temperature gradient of 20–60 °C. Fluorescence was monitored as a function of temperature, and the $T_m$ was determined by fitting the fluorescence data to a sigmoidal curve and calculating the inflection point in GraphPad Prism.[82,130]

### 2.2.7 Förster Resonance Energy Transfer (FRET) Nucleotide Exchange Assay

FRET was used to assess the nucleotide exchange activity of TrioC variants. [131] [73] First, 3 μM RhoA·GDP was incubated with 500 nM TrioC variants for 5 min at room temperature in freshly prepared nucleotide exchange buffer: 20 mM HEPES pH 8.0, 100 mM NaCl, 2 mM DTT, and 10 mM $MgCl_2$. Immediately before measurement, 1 μM 2′/3′-O-(N-Methyl-anthraniloyl)-guanosine-5′-triphosphate (MANT-GTP) (Jena Biosciences), was injected to a final assay volume of 100 μL. The mixture was then excited at 280 nm, and fluorescence intensity at 450 nm was read in 2 s intervals on a Flexstation 3 plate reader until a plateau was reached (30 min). Fluorescence curves were fit to a one-phase exponential association model using GraphPad Prism to derive the observed kinetic constant $k_{obs}$. The resulting $k_{obs}$ were then compared to that of matched TrioC WT rates with a representative rate constant of (0.004 ± 0.0006 $s^{-1}$, N=3 independent experiments in triplicate). The 2152Δ and 2147Δ variants displayed an exchange rate too fast to be measured under the initial assay conditions. Thus, the exchange rate of these variants was measured using 50 nM GEF, and all relative rates are therefore reported as specific activities. Prior literature suggests that under our assay conditions the nucleotide exchange reaction rate is linear with respect to GEF concentration.[46] For $G\alpha_q$ activation assays displayed in Table 2, TrioC WT and variants were added at 250 nM GEF concentration (in order to lower the observed basal rate and better capture activation), G$\alpha_q$ at 550nM, and the assay was run with and without the addition of 30 μM $AlCl_3$ and 10 mM NaF in the reaction buffer to generate $AlF_4^-$ in solution.

### 2.2.8 Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS)

The following was performed by Sheng Li, Ph.D. at the University of California San Diego. First, 1 μL of 2.6 mg/ml TrioC in 20 mM HEPES pH 7.5, 50 mM NaCl, and 2 mM DTT was mixed with 15 μL of a buffer containing 8.3 mM tris(hydroxymethyl)aminomethane (Tris) and 50 mM NaCl at pH 7.2 at 0 °C, and 24 μL of ice cold quench buffers containing 0.1 M glycine pH 2.4, 16.6 % (v/v) glycerol and various concentrations of guanidinium HCl (0.08, 0.8, 1.6 and 3.2 M) were then added. The quenched samples were then subjected to an immobilized pepsin column (16 μL bed volume) on ice at a flow rate of 20 μL/min for inline digestion. Proteolytic products were collected on a trap column for desalting and liquid chromatography-mass spectrometry (LC-MS) analysis was performed using an Agilent Poroshell C18 column (EC-C18, 35×0.3 mm, 2.7 μm) with a linear acetonitrile gradient (6.4 %-38.4 % over 30 min). Both trap and C18 columns were kept at 0 °C. MS analysis was done using an OrbiTrap Elite Mass Spectrometer (Thermo Fisher Scientific), and MS/MS data were searched against TrioC sequence by Proteome Discoverer (Thermo Fisher Scientific). Coverage maps of identified peptides were compared with each other and the 1.6 M GuHCl quench buffer was selected. This enabled us to determine the proper quench buffer concentration for the HDX experiments.

All exchange stock solutions were kept on ice and contained 1.0 mg/mL of each TrioC variant, 8.3 mM Tris pH 7.2 and 50 mM NaCl. HDX-MS experiments were initiated by adding 48 μL of exchange stock solutions to 144 μL of $D_2O$ buffer (8.3 mM Tris, 50 mM

NaCl, pD 7.2) and quenching the reaction at varied time points (10, 100, 1000, 10,000 and 100,000 s) at 0 °C. At the indicated time, 16 μL of exchange reaction solution was taken out and mixed with 24 μL of ice-cold quench buffer (0.1 M glycine pH 2.4, 16.6 % (v/v) glycerol, 1.6 M GuHCl) and immediately frozen on dry ice. Non-deuterated and equilibrium deuterated control samples were also prepared for back exchange correction. All frozen samples were thawed at 4 °C and subjected to the above system for enzymatic digestion, LC separation and MS analysis. All the columns were kept at 0 °C to minimize back exchange. The extent of deuterium incorporation of deuterated peptides was determined using HDXaminer (Sierra Analytics, LLC, Modesto, CA), which calculates centroid deuterium incorporation values for each peptide. Ribbon maps were generated with in-house Excel macro and MatLab scripts.

### 2.2.9 Rhotekin Pulldown Assay

Below experiments were performed by Nadia Arang in the Gutkind Lab at the University of California San Diego. HEK293 cells cultured in Dulbecco's modified eagle medium plus 10 % fetal bovine serum (Sigma) in 15 cm dishes were cultured to 70% confluency and transfected with 15 μg deoxyribonucleic acid (DNA) encoding TrioFL wild type and variants using Turbofect transfection reagent (Thermo Fisher) at a 1:2 ratio of DNA to Turbofect. Active Rho levels were measured using the RhoA Pull-Down Activation Assay Biochem Kit (bead pull-down format) following manufacturer's instructions (Cytoskeleton, Inc.). Briefly, 24 h after transfection, cells were serum starved for an additional 24 h and lysed using provided lysis buffer. Protein concentrations were quantified using DC Protein Assay (BioRad), samples were adjusted to the same concentration using provided lysis buffer and then snap frozen in liquid $N_2$. Lysate with

600 μg of protein was added to 30 μL GST-tagged Rhotekin-RBD protein bound to sepharose beads. Samples were incubated while rocking at 4 °C for 1.5 h. Beads were then washed, eluted in Laemmli sample buffer, and analyzed by western blot using a mouse monoclonal anti-RhoA antibody (Cytoskeleton, Inc. Cat # ARH04) to measure the ratio of active to total RhoA. Bands were quantified using ImageJ software (NIH).

### 2.2.10 Statistical Analysis

DSF and FRET nucleotide exchange experiments described above were performed in 3 experiments in at least duplicate. For DSF assays, average $T_m$ values from each experiment for each variant were subtracted from compiled WT values obtained on the same instrument to obtain $\Delta T_m$. Statistical significance was assessed using a one-way analysis of variance (ANOVA) test with a post-hoc Dunnett's test for multiple comparisons to compare N=3 experiments of each variant with N=11 WT experiments (J&J Thermofluor) or N=15 experiments (qPCR instruments). For FRET activity assays comparing TrioC variants to WT, $k_{obs}$ values for each variant were normalized to matched WT $k_{obs}$ for each experimental N to generate fold GEF activation values for each variant. To capture the statistical spread in WT measurements, each experimental WT $k_{obs}$ was normalized to the average WT $k_{obs}$ of the N=3 experiments. Statistical significance was assessed using a one-way ANOVA test with a post-hoc Dunnett's test for multiple comparisons to compare N=3 experimental fold GEF activation values of each variant with N=51 WT fold GEF activation values. For $G\alpha_q{\cdot}GDP{\cdot}AlF_4^-$ activation assays, $k_{obs}$ in the presence of $G\alpha_q{\cdot}GDP{\cdot}AlF_4^-$ for each variant was normalized to $k_{obs}$ in the presence of $G\alpha_q{\cdot}GDP$ to generate fold activation values. The fold $G\alpha_q{\cdot}GDP{\cdot}AlF_4^-$ activation value for TrioC WT was compared to cancer variants using a one-way ANOVA test with a post-hoc Dunnett's test for multiple comparisons. Errors are provided as 95% confidence intervals. Rhotekin assay data for Trio variants profiled in N=3 were normalized to the average of N=3 Trio WT values and then compared to the average Trio WT value using a one-way ANOVA test with a post-hoc Dunnett's test. Analysis was performed using GraphPad Prism version 7.00.

### 2.2.11 Small Angle X-ray Scattering (SAXS)

SAXS was performed at BioCAT by Srinivas Chakravarthy (beamline 18ID at the Advanced Photon Source, Argonne National Lab) with in-line size exclusion chromatography (SEC-SAXS) to separate the sample from aggregates and other contaminants thus ensuring optimal sample quality. TrioC WT and R2150W at 3 mg/mL were loaded onto a Superdex 200 Increase 10/300 GL column, which was run at 1.0 mL/min by an AKTA Pure FPLC (GE Healthcare Life Sciences, Chicago, IL) with a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT. The eluate was passed through a UV monitor and flown through the SAXS flow cell, which consists of a 1.5 mm ID quartz capillary with 10 μm walls. Scattering intensity was recorded using a Pilatus 3 1M detector (Dectris, Baden-Dättwil, Switzerland) which was placed 3.5 m from the sample giving access to a q-range of 0.004 $Å^{-1}$ to 0.4 $Å^{-1}$. Exposures of 0.5 s were acquired every 2 s during elution and the data were reduced using BioXTAS RAW 1.4.0.[132] Additional experimental parameters can be found in Table 2.3. Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the I(q) vs q curves used for subsequent analyses. The ATSAS software package was used to further process data and generate data plots.[133]

## 2.3 Results

### 2.3.1 TrioC Protein Expression and Purification

TrioC and TrioCΔC were expressed and purified as described in the methods section. Briefly, TrioC constructs were expressed in E. coli as MBP fusions. Proteins were purified using Ni-NTA resin and subjected to TEV protease cleavage overnight in tandem with dialysis to remove imidazole. Originally, a Source 15S column was used to remove MBP as TrioC sticks to the S column and MBP flows through. Complex formation with RhoA indicated that TrioCΔC was properly folded (Figure 2.1). After further optimization TrioCΔC was instead subjected to Ni-NTA resin twice to remove MBP for ease of use.

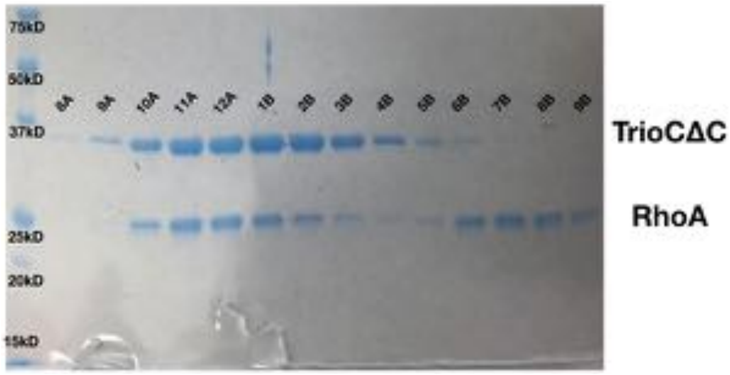


Figure 2.1 Complex formation between TrioCΔC (37 kD) and RhoA (22 kD) via size exclusion chromatography suggests that TrioC purified from E. coli is properly folded. This experiment was run on a single 24 mL analytical Superdex 75 column, however the chromatogram was lost due to computer failure. A Coomassie stained 12 % gel of fractions from the SEC run is shown to illustrate complex formation and an excess of RhoA eluting after the complex.

This protein appeared homogenous on a gel and was used for biochemical assays. TrioC WT and variants were purified in the same manner (Figure 2.2).

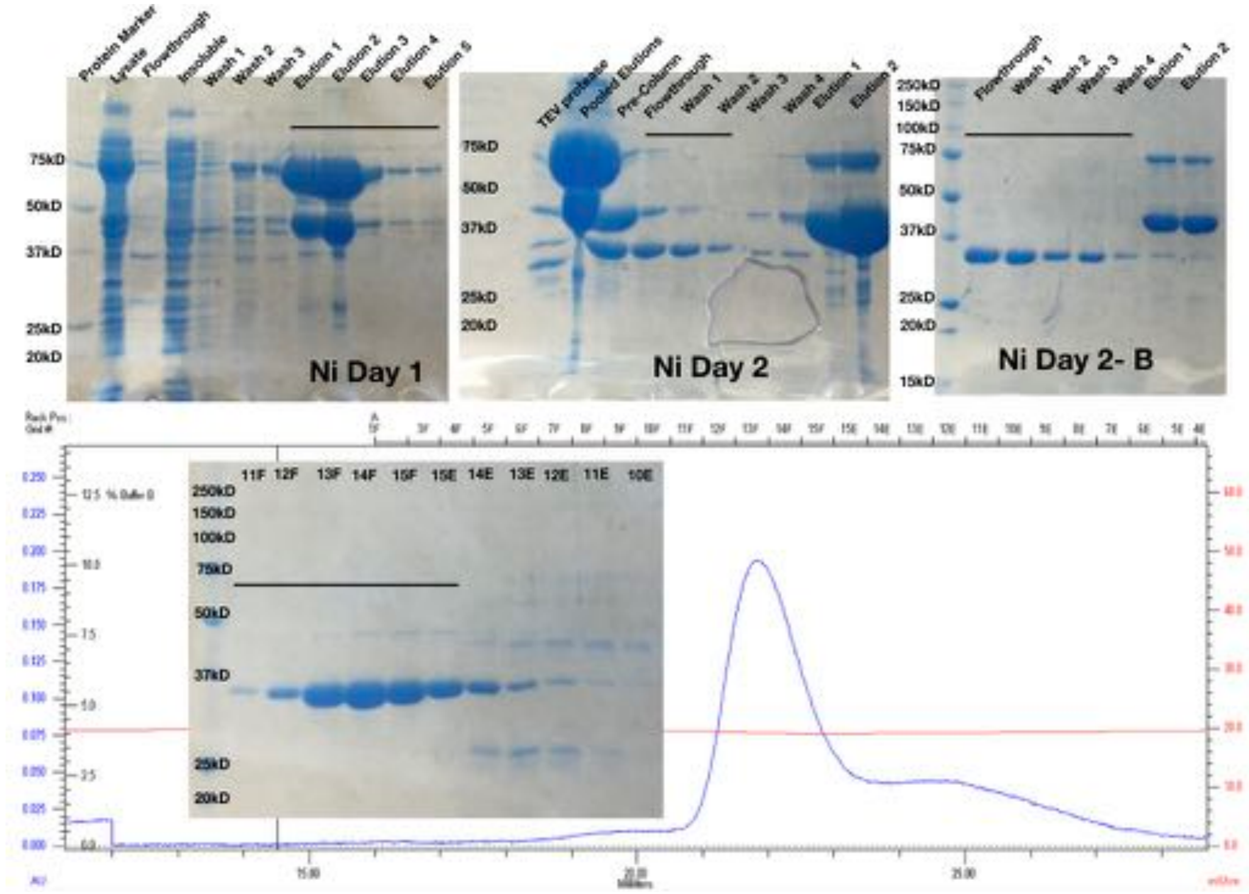


Figure 2.2 TrioCΔC Purification Scheme showing sequential steps starting from the top left and results in homogenous TrioCΔC. Top left, Coomassie stained 12 % gel showing initial Ni-NTA purification. Top middle, Coomassie stained 12 % gel showing re-IMAC purification. Top right, Coomassie stained 12 % gel showing continued re-IMAC purification. Bottom, chromatogram showing $A_{280}$ trace (blue line) from SEC run of TrioCΔC fractions on tandem 24 mL analytical Superdex 75 columns. Coomassie stained 12 % gel showing sequential fractions from SEC trace. Black bars indicate fractions pooled to the next step, or for the last step, fractions pooled for crystallography or other experiments.

Protein used for crystallography, SAXS, and HDX-MS was further polished using size exclusion chromatography. TrioC from SEC appeared as a monomer of ~37 kD (TrioCΔC expected MW 37.4 kD) and final yield was 25-100 mg of pure TrioC per liter of cell culture. Gels and chromatograms representing a typical TrioC purification using the final standardized protocol are shown in Figure 2.2.

### 2.3.2 TrioC Crystallization

TrioCΔC was purified using the scheme described above in a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT. Protein was concentrated to 15, 30, and 60 mg/mL in an attempt to reproduce the method used to crystallize TrioN. [81] Crystals formed after overnight incubation of TrioCΔC at 15 and 30 mg/mL at 20 °C (Figure 2.3 top left), however these crystals melted back into solution after 2 days and could not be reproduced. Screening trays were set in a 1 μL+1 μL sitting drop format using commercial screens: Classics Lite, Index HT, JCSG+, and ProComplex. Hits were obtained in Tris-HCl pH 8.5, 200 mM $MgCl_2$, and polyethylene glycols (PEGs) 4000 and 8000 at 13 and 16 mg/mL. Further optimization of protein concentration, PEG type, and

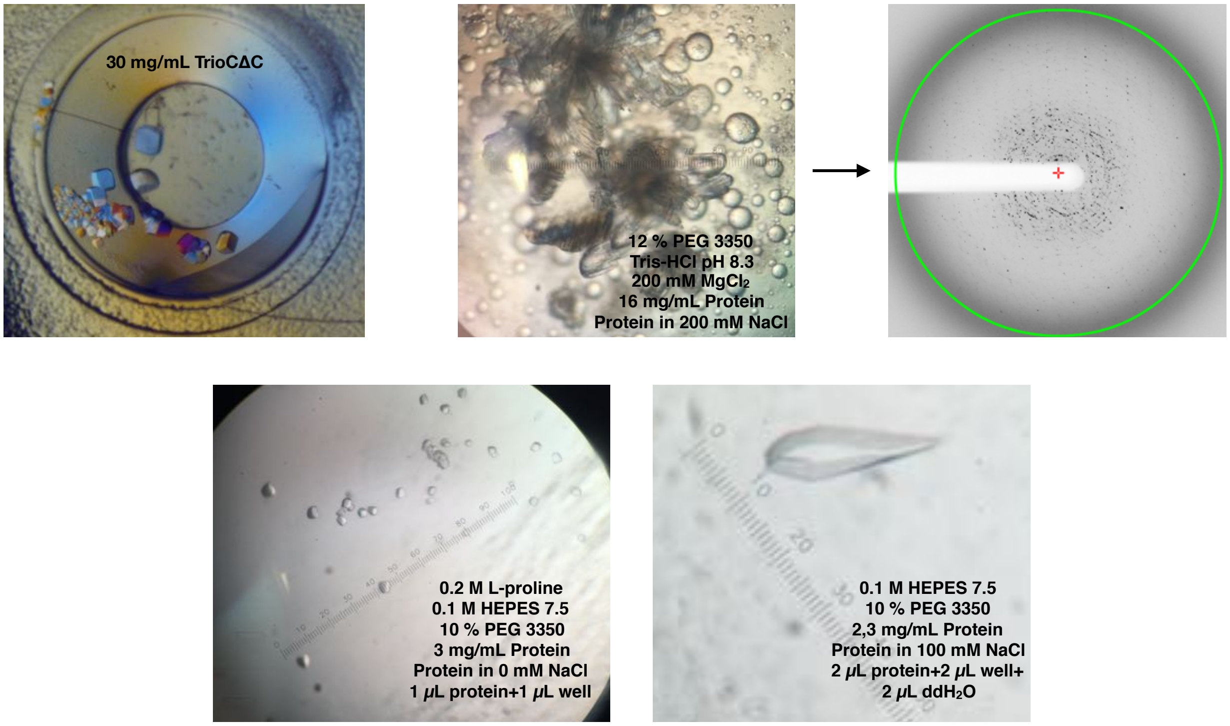


Figure 2.3 Early TrioCΔC crystallization attempts resulted in two crystal forms which were intractable and one which could be optimized into large single crystals. Top left, crystals grown using TrioCΔC incubated in high concentration in buffer. Top middle, TrioCΔC crystal form consisting of stacks of inseparable plates. Top right, example diffraction from top middle crystals harvested using ethylene glycol as the cryoprotectant. Bottom left, TrioCΔC 3-dimensional single crystals grown using no salt in buffer. Bottom right, optimized TrioCΔC crystal ~200 μm in length. Crystallization conditions are provided in each case.

PEG concentration in hanging drop trays led to a crystal form which presented itself as inseparable stacks of plates and diffracted to a very low resolution (Figure 2.3 top middle and top right). Attempts to refine this crystal form using the additive screen (Hampton) led to two crystal forms which both turned out to be small molecule crystals: one dextran sulfate, another cesium chloride. Attempts to improve the stacked plate crystal form through micro and macroseeding did not yield any improvement. Next, lower NaCl concentration was used in the protein buffer to lower solubility and potentially slow down crystal growth. TrioCΔC was run on SEC in a buffer containing 20 mM HEPES pH 8.0 and 2 mM DTT. Protein in this buffer had a lower retention volume on an S75 column and much of the sample was found in the void fraction of the column. Setting trays of TrioCΔC purified in this manner yielded several hits, some of which resembled the earlier stacks of plates seen. However, a new crystal form was seen in three different conditions which consisted of 3-dimensional non-birefringent single crystals (Figure 2.3 bottom left). Two hits contained a buffer pH ~7, 6-10 % PEG, and one of the two also had 200 mM L-Proline. The third hit was 12.5 % ethylene glycol. These crystals were stained with Izit crystal dye (Hampton Research) and absorbed the blue stain, consistent with them being protein. Subsequent attempts focused on improving crystal size by varying protein concentration, well solution composition, and introducing NaCl back into the protein buffer. Shifting the equilibrium away from nucleation and towards crystal growth was necessary in order to improve crystal size. A series of SEC runs determined that 50mM was the lowest NaCl concentration in which TrioCΔC behaved as a monomer. Below 50mM NaCl, TrioCΔC formed multimers and shifted towards the void volume of the S75 column. Trays set using 50 mM NaCl in the SEC buffer yielded crystals of a similar size to seen before. However, trays set at 100 mM NaCl yielded larger 3-dimensional crystals of about 200 µm in the longest dimension. (Figure 2.3 bottom right) Adding a layer of 200 µL of silicon oil over the well solution and adding 1 equivalent of $ddH_2O$ into the drop solution (1 µL well solution + 1 µL protein solution + 1 µL $ddH_2O$, or 2+2+2, etc.) were also attempted in order to further slow crystal growth. Several of these crystals of 100-300 µm in the largest dimension were harvested in mother liquor including 20% ethylene glycol as a cryoprotectant. Data collection yielded diffraction that at best extended to 6 Å spacings. This data was also streaky with poorly defined spots (Figure 2.4 top). The 6 Å dataset could not be solved using molecular replacement with the DH/PH tandems of p63RhoGEF, TrioN, or Dbs, and as such, concern arose that a contaminant was in fact being crystallized. Crystals were washed and run on a gel to ensure that TrioC was in fact the major protein component in the gels (Figure 2.4 top right). Further optimization

**10 % PEG 3350**
**100 mM HEPES pH 7.5**
**TrioC @ 3 mg/mL**
**0.2 M L-Proline**
**Drop: 1 µL well+1 µL protein**
**200 µL silicon oil over well**
**Cryoprotectants: PEG 400, glycerol, ethylene glycol**

**14 % PEG 3350**
**100 mM HEPES pH 7.5**
**TrioC @ 2 mg/mL**
**Drop: 2 µL well+2 µL protein+1 µL $ddH_2O$**
**No Cryoprotectant**

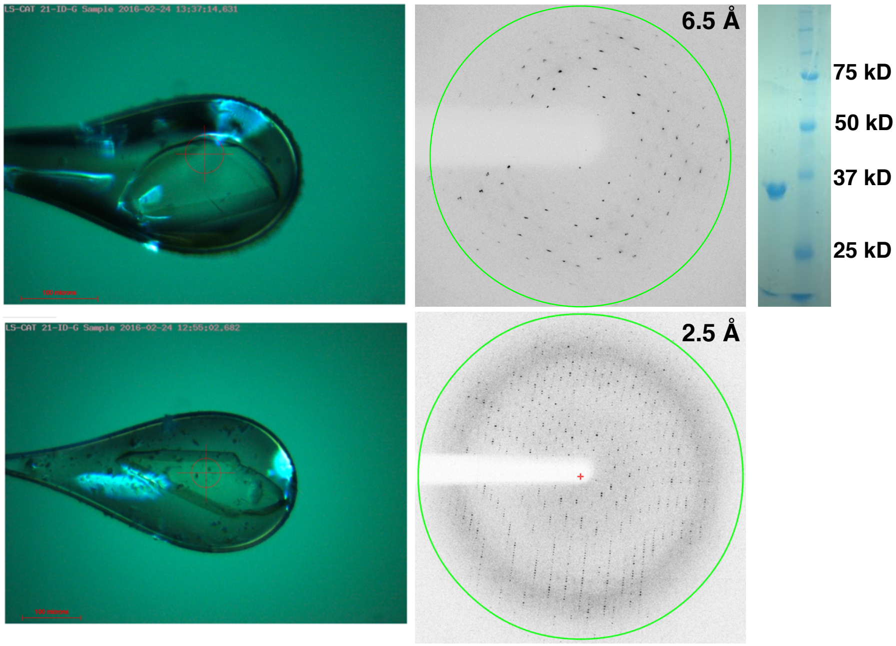


Figure 2.4 Optimization of diffraction resolution for final TrioCΔC crystal form resulted in data that extended to 2.65Å spacings. Top left, crystallization conditions. Top middle, crystal shown frozen in nylon loop and resultant diffraction pattern. Top right, washed TrioCΔC crystals were run on a 12% Coomassie stained gel to determine if these crystals consisted of TrioCΔC protein. Bottom left, crystallization conditions. Bottom middle, crystal shown frozen in nylon loop and resultant diffraction pattern. The green rings indicate diffraction extending to the designated resolution limit.

trials focused on improving resolution by screening cryoprotectants. Addition of 25 % ethylene glycol, glycerol, or PEG 400 to the mother liquor did not improve resolution. Lack of cryoprotectant improved diffraction from 6 Å to 2.7 Å (Figure 2.4 bottom). Two datasets were collected at this improved resolution, with one second exposures and a two second exposure set, which improved resolution to 2.65 Å. This two second exposure set of frames was used to solve the final TrioC structure.

### 2.3.3 TrioCΔC Structure Solution, Processing, Model Building, and Refinement

Final diffraction data collected with 2.65 Å spacings were indexed, integrated, and scaled using the HKL 2000 software package. [122] The presence of systematic absences in diffraction maxima in all three dimensions allowed confident assignment of the $P2_12_12_1$ space group. The data was integrated and scaled in $P2_12_12_1$ and initial phases were determined using molecular replacement with the Phaser program within the Collaborative Computational Project No. 4 (CCP4) software package. [123,134] Input models included the DH/PH tandems of TrioN, p63RhoGEF, and Dbs, all which have >30 % sequence identity to TrioC. None of these models yielded any solutions of high quality. Solution was then attempted using DH and PH domains of p63RhoGEF with the linker region deleted as sequential search models. These input models were processed using the CHAINSAW package within CCP4 in order to remove side chains. [124] This method of structure solution yielded a solution with high log likelihood gain and translation function Z-score values, markers of a likely correct solution. Structure solution inserted the bulk of the DH and PH domains into density. Three copies of the TrioC DH/PH tandem were found in the asymmetric unit. Initial rounds of structure building included manual model building and real space refinement in Coot and simulated annealing and refinement with torsional NCS restraints in PHENIX.refine.[125,126] Once refinement statistics stopped improving, REFMAC5 in CCP4 was used, in combination with translation-libation-screw (TLS) refinement and manual X-ray weighting, with non-crystallographic symmetry turned off. [127] Six TLS groups were predicted using the TLS motion determination server and corresponded to the DH and PH domains in each of the three chains in the ASU.[135,136] These changes accounted for the medium resolution of the structure and allowed for refinement against ideal bond lengths and angles, in addition to taking into account differences between the 3 protein copies within the asymmetric unit. Final considerations for structure building incorporated changes suggested by Molprobity[128], PDBRedo[129], and Coot validation metrics for rotamers, geometry, and Ramachandran restraints. Statistics for processed data and the final model are shown in Table 2.1.

|  | TrioCΔC (human Trio 1960-2275) |
|---|---|
| *Data collection statistics* |  |
| Synchrotron source | LS-CAT beamline 21-ID-G, Advanced Photon Source |
| Wavelength (Å) | 0.97856 |
| Resolution range (Å) | 50-2.65 (2.70-2.65) |
| Space group | $P2_12_12_1$ |
| Unit cell (a, b, c) (Å) | 59.2, 85.8, 182.4 |
| Unique reflections | 30511 (1371) |
| Multiplicity | 5.0 (3.8) |
| Completeness (%) | 98.7 (90.1) |
| Mean I/σI | 18.9 (1.2) |
| $CC_{1/2}$ | ND (0.699) |
| $R_{sym}$ | 0.085 (0.746) |
|  |  |
| *Refinement statistics* |  |
| Resolution limits (Å) | 15-2.65 (2.70-2.65) |
| Number of test reflections | 28787 (1862) |
| $R_{work}$ | 0.23 (0.34) |
| $R_{free}$ | 0.27 (0.37) |
| Number of nonhydrogen atoms | 7525 |
| Macromolecule | 7499 |
| Ligand | 0 |
| Water | 26 |
| Protein Residues | 907 |
| RMS bonds (Å) | 0.008 |
| RMS angles (°) | 1.2 |
| Ramachandran favored (%) | 97.8 |
| Ramachandran outliers (%) | 0 |
| Clashscore calculated from MolProbity | 1.33 |
| Average B-factor | 76.0 |
| Macromolecule | 76.0 |
| Ligand | N/A |
| Water | 48.0 |

Table 2.1 Data collection and refinement statistics for the TrioCΔC crystal structure. Parentheses indicate values for highest resolution shell. ND, not determined.

### 2.3.4 Structural Comparison to other Dbl Family GEF Modules

As in previously determined Dbl family DH/PH structures, the TrioC module begins with an α-helical DH domain, consisting of 6 helical spans (α1-α6; A in Figure 2.5). The α6 helix is continuous with the short first helix of the PH domain, αN, and their junction serves as a flexible hinge between the DH and PH domains. The remainder of the PH domain is composed of a seven-stranded (designated β1-β7) antiparallel β-sandwich capped on one end by a C-terminal helix (αC). Three copies of the TrioC module are found in each asymmetric unit (A in Figure 2.6). They are similar in overall conformation with a root mean square deviation (RMSD) of 0.8 Å for Cα atoms (B in Figure 2.6), with minor differences arising from unique crystal contacts. The only other reported structure of a TrioC subfamily RhoGEF module is that of p63RhoGEF (65% sequence identity) in complex with $G\alpha_q$ bound to guanosine diphosphate and aluminum tetrafluoride ($G\alpha_q \cdot GDP \cdot AlF_4^-$) and nucleotide-free RhoA [Protein Data Bank (PDB) entry 2RGN], representing a TrioC subfamily member in its activated, signal competent state. In Dbs, a closely related RhoGEF with a PH domain that positively contributes to nucleotide exchange, the DH and PH domains adopt a conformation more similar to activated p63RhoGEF than to autoinhibited TrioC (B and C in Figure 2.5). The region encompassing the α6/αN junction (Trio residues 2139-2150) in the TrioC structure

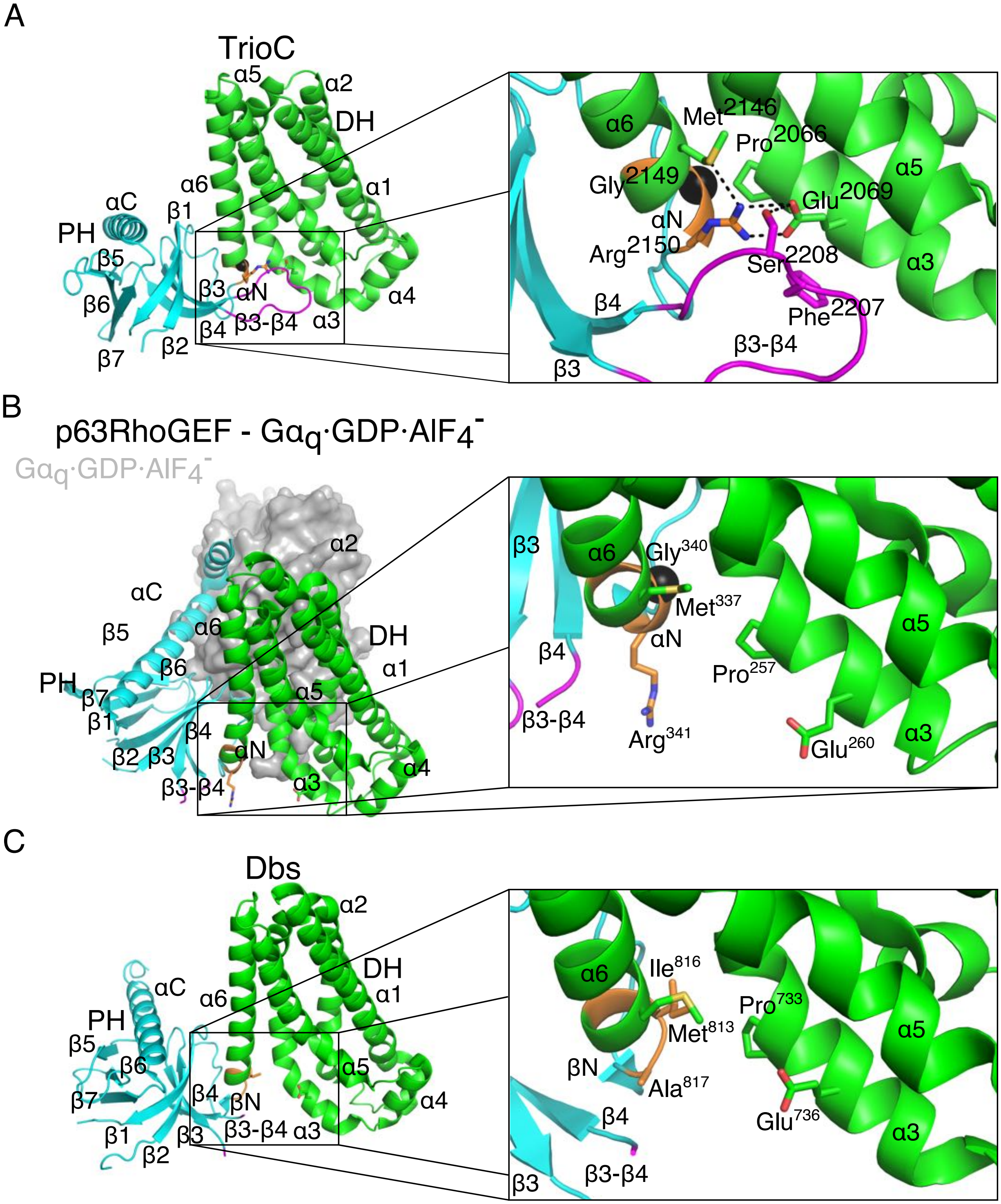


Figure 2.5 Comparison of the TrioC structure (A) with that of Gα$_q$·GDP·AlF$_4^-$ activated p63RhoGEF (B) and Dbs (C). Proteins are shown in cartoon representation with secondary structure elements labeled. DH domains are colored green, αN helices are orange, PH domains are blue, and β3-β4 loops are magenta. Gα$_q$·GDP·AlF$_4^-$ is shown as a gray surface representation. On the right sides of each structure are close-up images of the interaction region between DH and PH domains, with key residues shown as sticks and hydrophilic contacts shown as black dotted lines. Glycine residues are depicted as black spheres.

adopts a less bent α-helical conformation, with the αN helix forming the bulk of the

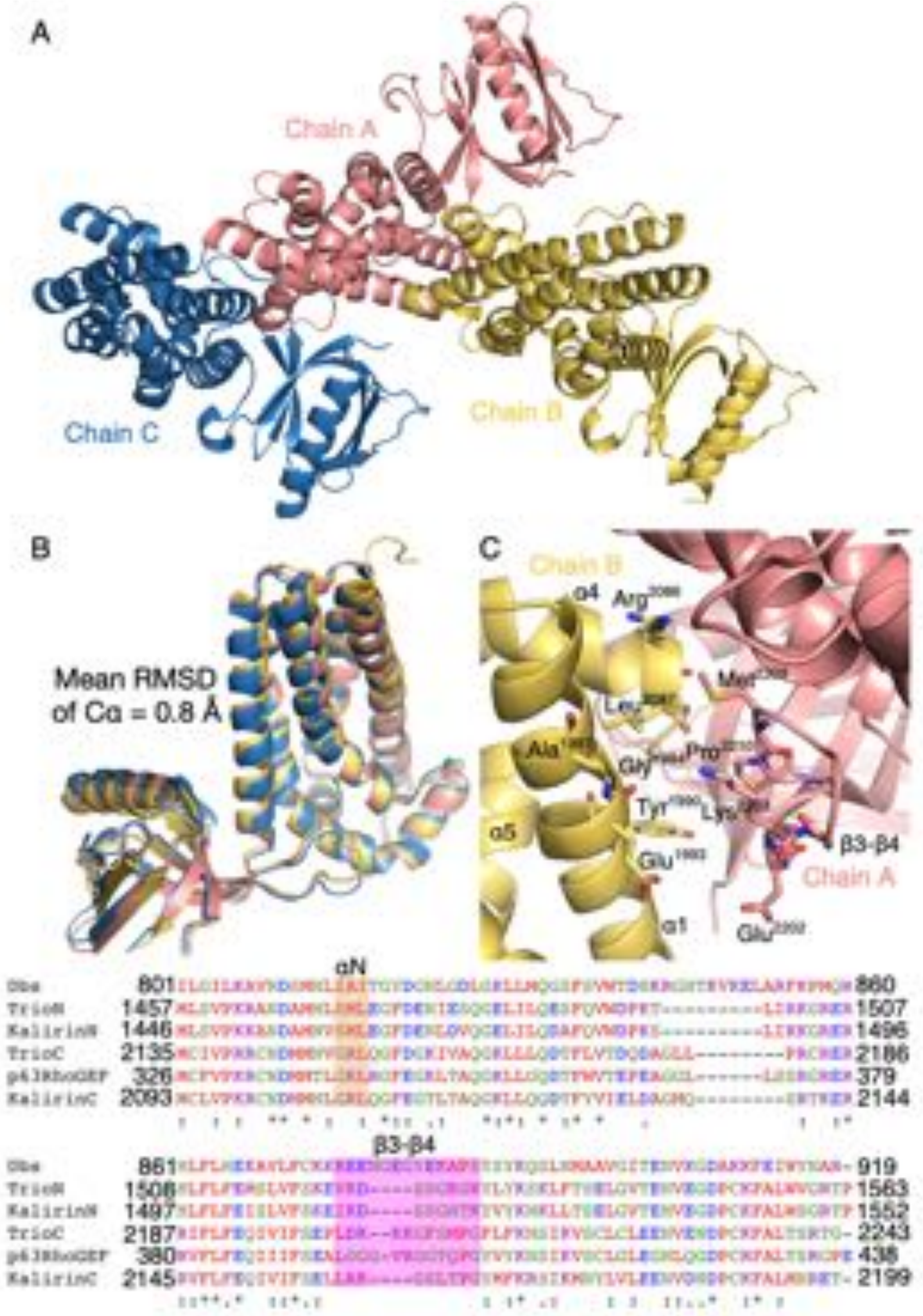


Figure 2.6 Details within the TrioC crystal structure. Three copies of TrioC are found in the asymmetric unit and are depicted in cartoon representation (A). Alignment of the three TrioC copies within the asymmetric unit using LSQ superpose command in Coot. RMSD of Cα atoms between chains are shown (B). Crystal packing likely explains the position of the β3-β4 loop within the autoinhibited interface (C). Side chains involved in packing between the DH domain of chain B and the β3-β4 loop of Chain A are shown as sticks. Bottom, multiple sequence alignments showing conservation in αN and β3-β4 regions among related RhoGEF modules.

interactions with α3 in the DH domain (A in Figure 2.5). The analogous αN elements in activated p63RhoGEF (B in Figure 2.5) and Dbs (C in Figure 2.5) are instead displaced from the DH domain, leaving space for switch II of RhoA to bind.

In TrioC, Gly2149 (Ile816 in Dbs, see Figure 2.5) packs against the side chain of Pro2066 in α3 of the DH domain, allowing for closer approach of the TrioC DH and PH domains (Figure 2.5). Meanwhile, TrioC-Arg2150 forms an interdomain salt bridge with Glu2069 in α3, which is not possible with the analogous Dbs-Ala817. The side chain of Met2146 also bridges the DH and PH domains by forming a hydrogen bond with the side chain of Arg2150 and hydrophobic contacts with the DH domain. These interactions are broken upon the binding of $G\alpha_q$, as seen in the active p63RhoGEF structure (Figure 2.5). Glu2069, Gly2149, and Arg2150 are all invariant in the TrioC subfamily, and are not conserved in Dbs or the related N-terminal DH/PH modules of Trio and Kalirin (See αN; bottom of Figure 2.6). Met2146, however, is conserved as a hydrophobic residue in most RhoGEFs and instead forms direct contacts with Switch II of bound GTPases in activated structures of p63RhoGEF and Dbs. Although not well conserved among TrioC subfamily members (See β3-β4; bottom of Figure 2.6), residues 2204-2212 in the β3-β4 loop of the PH domain bury the Arg2150-Glu2069 salt bridge and form additional interactions with α3 in the DH domain; the hydroxyl of Ser2208 forms a hydrogen bond with Glu2069. In

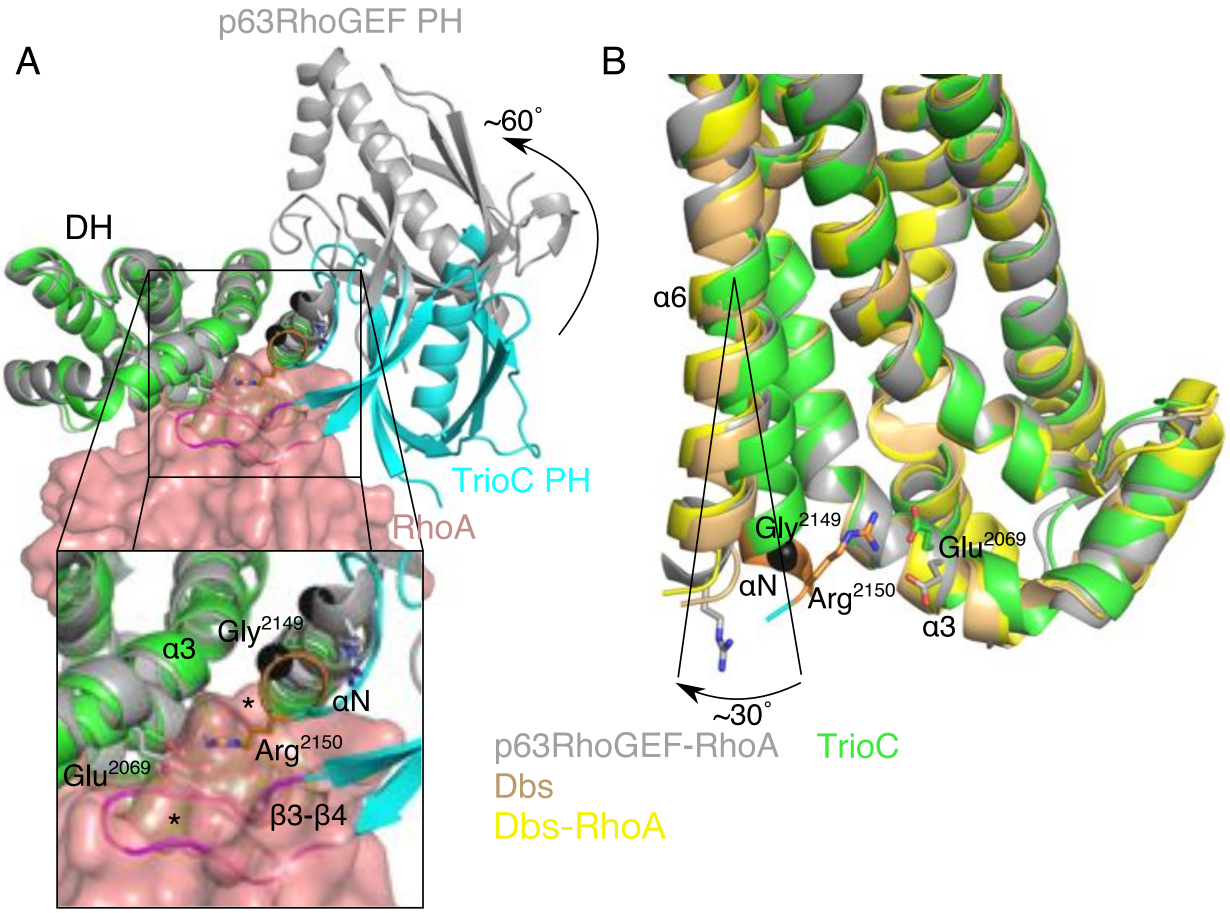


Figure 2.7 Conformational changes in the α6-αN region in RhoGEF structures. The autoinhibited TrioC structure and activated p63RhoGEF structure are shown as cartoons and aligned by their DH domains to highlight conformational change in α6-αN and the PH domains (A). The difference in relative orientation of the PH domain is highlighted with a curved arrow. TrioC is shown with the DH domain green, the PH domain cyan, and the linker region orange, and p63RhoGEF is colored gray. RhoA is shown in red as a surface representation. Key autoinhibitory residues are shown as sticks and spheres to orient the reader. Regions of TrioC which overlap with the RhoA binding site (based on the p63RhoGEF structure) are demarcated with asterisks (inset). DH domains, including α6-αN linker regions, of GTPase free and bound structures are shown as cartoons, with key residues in the autoinhibited interface shown as sticks. Dbs in its GTPase free state is colored tan and in its RhoA bound state is colored yellow. Changes in the conformation of the linker helix are outlined using two lines and a curved arrow to indicate the angle formed between their principal axes.

the active p63RhoGEF structure, β3-β4 is disordered (Figure 2.5). In Dbs, the equivalent loop is disordered without bound GTPase (Figure 2.5).

Aligning the core of the DH domain of TrioC (Trio residues 1967-2136, Chain A) with activated p63RhoGEF revealed a 60° relative rotation of their PH domains around the axis of the α6-αN hinge (A in Figure 2.7). This leads to the formation of an interface between the DH and PH domains of TrioC which directly overlaps with the RhoA binding face of the DH domain, providing a likely mechanism for autoinhibition (A inset in Figure 2.7). In particular, the more extended conformation of αN in autoinhibited TrioC occupies the RhoA Switch II binding site of the DH domain, and Glu2069 and Met2146, which interact directly with RhoA in the p63RhoGEF structure, are instead directly engaged with Arg2150.

Comparison of the α6-αN hinge regions suggests that TrioC undergoes a 30° rotation in this region when transitioning from inactive to active states (B in Figure 2.7). Whether in GTPase bound or free states, the hinge region in Dbs adopts a conformation most similar to that of activated p63RhoGEF (B in Figure 2.7). Thus, the Dbs DH/PH tandem is prearranged for competent nucleotide exchange, whereas TrioC is found in the autoinhibited conformation until $G\alpha_q$·GTP binds and changes the helical track of α6-αN to that seen in the active p63RhoGEF structure.

### 2.3.5 Site Directed Mutations in the DH-PH Interface Destabilize and Activate TrioC

To test the inhibitory role of the observed interface, we altered positions in α6-αN, the β3-β4 loop, and α3 using site-directed mutagenesis. We hypothesized that variants which disrupted important contacts in the DH-PH interface would be more sensitive to thermal denaturation (lower melting temperature ($T_m$) relative to wild type (WT)), as measured by differential scanning fluorimetry (DSF), and display increased GEF activity. E2069A, M2146A, and S2208A exhibited 3-7 °C lower melting points, whereas R2150A and F2207A (β3-β4 loop) were similar to WT TrioC (Table 2.2). Most of the alanine variants had a similar exchange rate to WT TrioC, with the exception of R2150A, and

S2208A, which were ~1.5-fold more active than WT, a slightly elevated but not significantly

| Variant | ΔTm (°C) | 95% CI for Δ$T_m$ | GEF activation (Fold/WT) | 95% CI for GEF activation | Fold activation by $G\alpha_q$·GDP·$AlF_4^-$ | 95% CI for Fold activation by $G\alpha_q$·GDP·$AlF_4^-$ |
|---|---|---|---|---|---|---|
| WT | 0.0 | [-0.5,0.5]$^{\#}$ | 1.0 | [0.9, 1.1] | 3.1 | [0.4, 5.9] |
| E2069A | -6.0*** | [-7.3, -4.7] | 0.7 | [0.1, 1.3] | ND | ND |
| M2146A | -3*** | [-3.6, -2.4] | 0.8 | [-0.5, 2.2] | ND | ND |
| S2208A | -3.2*** | [-4.8, -1.7] | 1.4 | [0.1, 2.7] | ND | ND |
| R2150A | 0.9* | [-0.2, 1.9] | 0.9 | [0.03, 1.8] | ND | ND |
| F2207A | 1.1 | [-0.6, 2.8] | 0.7 | [-0.1, 1.6] | ND | ND |
| Δ2204-2208 | -0.4 | [-1.1, 0.2] | 1.5 | [0.2, 2.8] | ND | ND |
| Δ2203-2209 | -0.5 | [-0.9, -0.1] | 1.0 | [-0.5, 2.7] | ND | ND |
| G2149I | -6.3*** | [-6.7, -5.9] | 1.9 | [-0.7, 4.6] | ND | ND |
| E2069R/R2150E | -7.9*** | [-14, -2.2] | 3.2** | [1.6, 4.7] | ND | ND |
| R2150E | 0.9 | [0.7, 1.1] | 1.0 | [-0.4, 2.4] | ND | ND |
| 2153Δ | ND | ND | 3.0* | [0.7, 5.3] | ND | ND |
| 2152Δ | ND | ND | 11*** | [4.5, 18] | ND | ND |
| 2147Δ | ND | ND | 14*** | [4.8, 22] | ND | ND |
| 2143Δ | ND | ND | 0.2 | [-0.2, 0.6] | ND | ND |
| G2149W | -3.6*** | [-3.9, -3.3] | 4.3*** | [2.4, 6.4] | 1.5* | [0.7, 2.3] |
| R2150Q | -1.6*** | [-3.3, 0.2] | 4.5*** | [2.1, 6.9] | 2.2 | [1.2, 3.1] |
| R2150W | -3.1*** | [-5.1, -1.2] | 9.3*** | [-0.4, 18] | 0.9** | [0.7, 1.0] |

Table 2.2 Table depicting DSF and GEF activity data for TrioC variants. ΔTm= $T_m$(variant)-$T_m$(WT). Fold GEF activation=average $k_{obs}$(variant)/$k_{obs}$(WT). Fold activation by $G\alpha_q$·GDP·$AlF_4^-$=average kobs(variant+$G\alpha_q$·GDP·$AlF_4^-$)/kobs(variant+$G\alpha_q$·GDP). Each variant was profiled in N=3 experiments performed in at least duplicate. ND, not determined. A one-way ANOVA with post Dunnett's test was used to test for significance for ΔTm, Fold GEF activation, and $G\alpha_q$·GDP·$AlF_4^-$ activation for each variant in comparison to WT. * refers to $p<0.05$, ** to $p<0.01$, *** to $p<0.005$. #, DSF data was collected using two experimental setups, the larger 95% CI was chosen here. See methods section for further explanation of methods/statistics used in this table.

different rate (Table 2.2). β3-β4 loop deletions Δ2204-2208 and Δ2203-2209 were 4 and 1 °C destabilized relative to wild type, although neither was more active than wild type. A larger loop deletion, Δ2201-2211, yielded insoluble protein and could not be assayed. The G2149I variant, replacing the position with the cognate residue in Dbs, reduced the $T_m$ over 7 ˚C and displayed 2-fold higher exchange relative to WT. The E2069R/R2150E double mutant, designed to test the importance of the salt bridge, destabilized protein 9 °C and enhanced GEF activity 4-fold. Thus, the electrostatic complementarity of these residues is not as important as their contributions to the local structure. The single R2150E mutant was not significantly different from WT in nucleotide exchange activity, whereas E2069R yielded insoluble protein and thus could not be assayed. Thus, the

E2069R substitution is most likely responsible for the activation exhibited by the E2069R/R2150E double variant protein is only stable in the context of a salt bridge swap. Ultimately, mutations which disrupted contacts made in the closed interface (G2149I and E2069R/R2150E), as opposed to mutations that removed contacts (such as R2150A) had the greatest ability to destabilize and activate the DH/PH module.

### 2.3.6 Mutations Found in Cancer Patients also Destabilize and Activate the TrioC Module

Analysis of the cBioPortal database revealed that truncations 2152Δ and 2153Δ occur in cancer patients. [137] The 2153Δ variant removes the bulk of the PH domain yet leaves αN intact. This variant activated the DH/PH module 3-fold relative to wild type TrioC (Table 2.2). The 2152Δ variant activated the module 12-fold over WT. To assess the consequences of further truncation, we profiled 2147Δ, which removes all of αN and found this variant 16-fold hyperactive. A final truncation, 2143Δ, which also removes a portion of α6, had 5-fold lower GEF activity than WT, likely due to loss of RhoA binding residues. The cBioPortal database also contains the G2149W, R2150Q, and R2150W variations, which we hypothesized would be activating due to steric or electrostatic disruption of the autoinhibited DH-PH interface. Indeed, we found that all had lower $T_m$ values and were >4-fold more active than WT (Table 2.2), consistent with a model wherein the PH domain must be dislodged from the DH domain in order to facilitate RhoA binding. We also found that TrioC WT, G2149W, and R2150Q were activated by $G\alpha_q$ in an $AlF_4^-$-dependent manner in the same assay format, although the G2149W and R2150Q variants were activated to a lesser extent (~1.5 fold) than TrioC WT (~3-fold). Our most active TrioC variant, R2150W, was not activated by $G\alpha_q \cdot GDP \cdot AlF_4^-$ (Table 2.2).

### 2.3.7 HDX-MS Shows Higher Dynamic Behavior in the DH-PH Interface in Constitutively Activated TrioC

We subjected R2150W and WT TrioC to HDX-MS experiments to study activated and basal forms of the enzyme in solution. The exchange maps for each variant agreed with relative levels of solvent exposure seen in the crystal structure, with buried residues exchanging much more slowly than residues found on the exterior of the protein. By examining a difference map between R2150W and WT, we hoped to understand the conformational changes taking place upon activation of TrioC. In the HDX-MS difference map R2150W-WT (Figures 2.8 and 2.9), the α6 region in close proximity to the R2150W

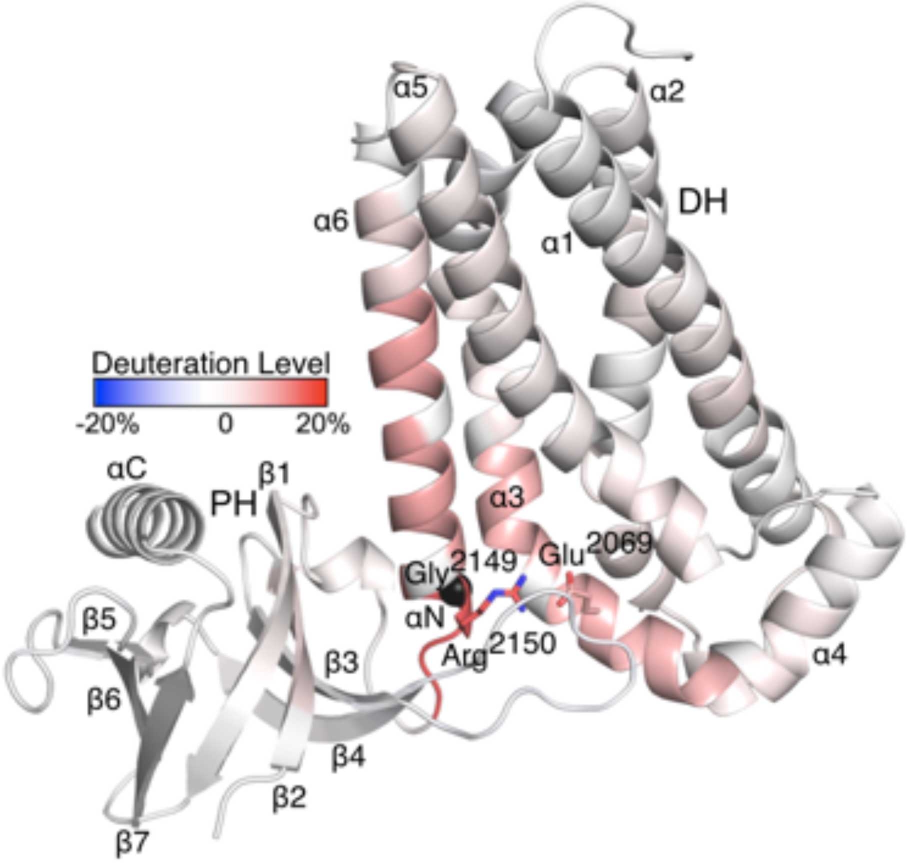


Figure 2.8 Difference HDX-MS values for TrioC (R2150W minus WT) plotted onto the TrioC crystal structure. Difference hydrogen-deuterium exchange values (R2150W-WT) plotted onto the TrioC structure shown as a cartoon with secondary structure elements labeled. Data shown represents an average of N=2 experiments from separate protein preparations. Darker red indicates higher exchange rates in R2150W, and off-white represents no change in R2150W vs. WT. A scale bar at the top left indicates relative exchange rates. Key residues involved in the autoinhibitory interface are shown as sticks and spheres to orient the reader. Experiments performed by Sheng Li, Ph.D.

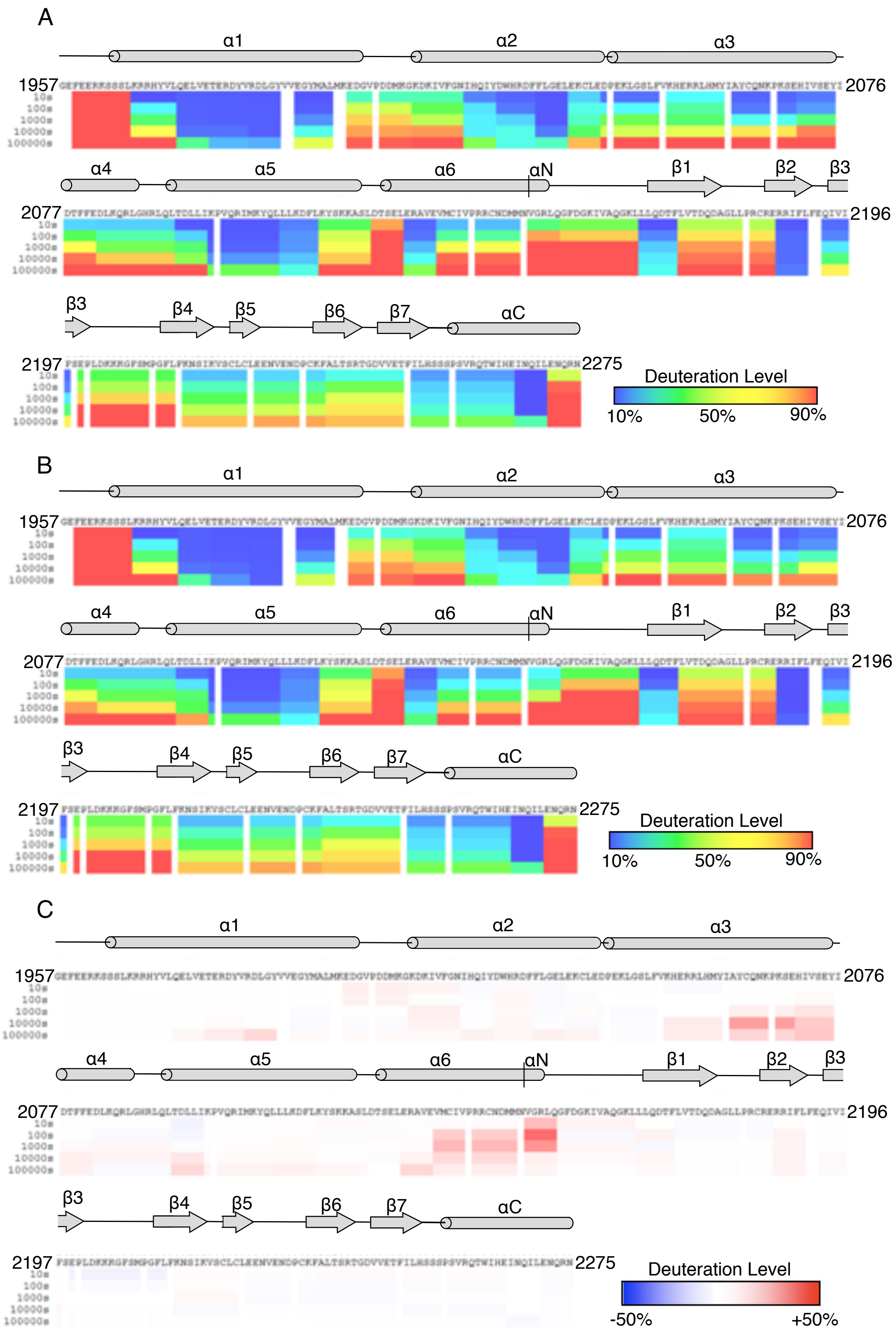
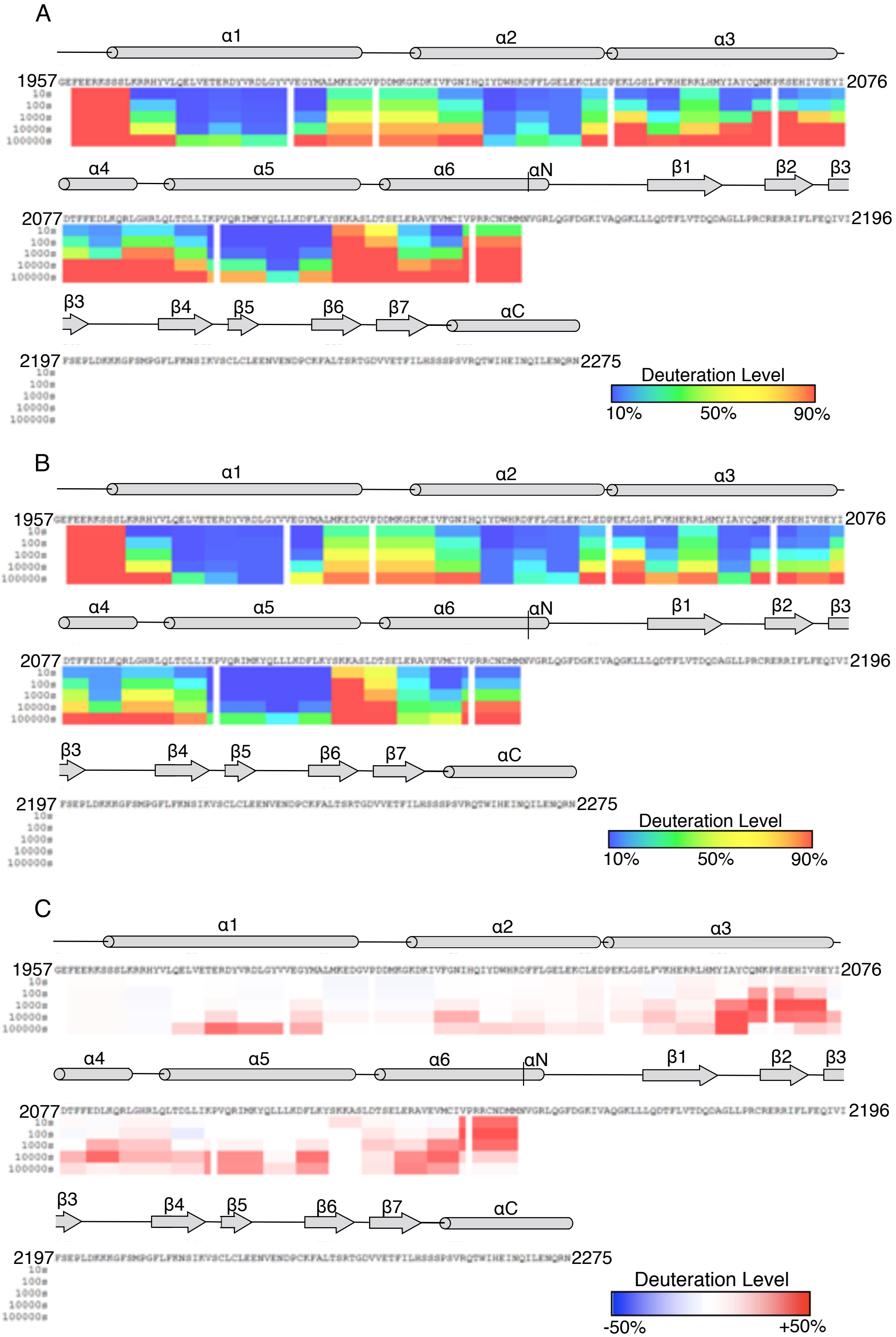

A
α1
α2
α3
1957
2076
10s
100s
1000s
10000s
100000s
α4
α5
α6
αN
β1
β2
β3
2077
2196
β3
β4
β5
β6
β7
αC
2197
2275
Deuteration Level
10%
50%
90%
B
α1
α2
α3
1957
2076
α4
α5
α6
αN
β1
β2
β3
2077
2196
β3
β4
β5
β6
β7
αC
2197
2275
Deuteration Level
10%
50%
90%
C
α1
α2
α3
1957
2076
α4
α5
α6
αN
β1
β2
β3
2077
2196
β3
β4
β5
β6
β7
αC
2197
2275
Deuteration Level
-50%
+50%

Figure 2.9 (Previous page) Primary HDX-MS data corresponding to the Figure 2.8. The TrioC primary sequence is shown left to right with starting and ending residues on each line numbered. Above the sequence the secondary structure is shown as cylinders (alpha helices) or arrows (beta strands). Below the primary sequence the HDX-MS values are plotted over various time points shown on the left. Blue indicates low exchange and red indicates high exchange, see scale bars at bottom right of each section for detail. TrioC R2150W is shown in (A), TrioC WT in (B), and the difference map (R2150W-WT) in (C). This data is the average of N=2 experiments performed by Sheng Li, Ph.D.

mutation is more solvent accessible compared to WT and residues in α3 also display a marked increase in solvent exposure, supporting the notion that these two regions directly interact in the autoinhibited, basal state. Thus, in the R2150W variant, residues from both α3 and α6 are more solvent exposed and more available for interaction with RhoA. We hypothesize that this occurs by breaking the interactions between these two regions seen in the autoinhibited crystal structure. The β3-β4 loop exhibits no difference in exchange dynamics between the two states, and in conjunction with our biochemical data we believe the position of this poorly conserved loop as a part of the DH-PH interface is likely due to crystal contacts. The HDX-MS difference map comparing the truncation 2152Δ to WT TrioC also showed deprotection in the α6-αN region and in α3 (Figure 2.10). These results are consistent with α6-αN losing helicity in the context of 2152Δ and thus explains its hyperactivity.

### 2.3.8 DH-PH Interfacial Mutants Activate Full-Length Trio in Mammalian Cells

Full-length Trio is the primary splice variant transcribed, and thus likely the predominant variant expressed, in UM cell lines. [37] Thus we compared the activity of the cancer-associated variants, G2149W, R2150Q, and R2150W, to WT in the context of human Trio (residues 61-3097) under serum starved conditions to detect inherent Trio activity. Proteins were expressed as C-terminal enhanced green fluorescent protein (eGFP) fusions using transient transfection in human embryonic kidney 293 (HEK293) cells.



Because 2152Δ was the most active cancer truncation variant in vitro, it was also

Figure 2.10 (Previous page) Primary HDX-MS data for TrioC 2152Δ and matched WT data. The TrioC primary sequence is shown left to right with starting and ending residues on each line numbered. Above the sequence the secondary structure is shown as cylinders (alpha helices) or arrows (beta strands). Below the primary sequence the HDX-MS values are plotted over various time points shown on the left. Blue indicates low exchange and red indicates high exchange, see scale bars at bottom right of each section for detail. TrioC 2152Δ is shown in (A), TrioC WT in (B), and the difference map (2152Δ-WT) in (C). This data is the result of N=1 experiment performed by Sheng Li, Ph.D.

profiled. A pulldown assay using the Rho-binding domain of the RhoA effector Rhotekin (Cytoskeleton, Inc.) [37] was employed to determine the relative ratio of active Rho (RhoA·GTP) to total RhoA content in response to expression of WT and mutant Trio. Expression of R2150W, R2150Q, G2149W, and G2152Δ Trio mutants all led to a >2-fold increase in levels of RhoA·GTP as compared to WT Trio (Figure 2.11). All Trio variants expressed to a similar level (C in Figure 2.11).

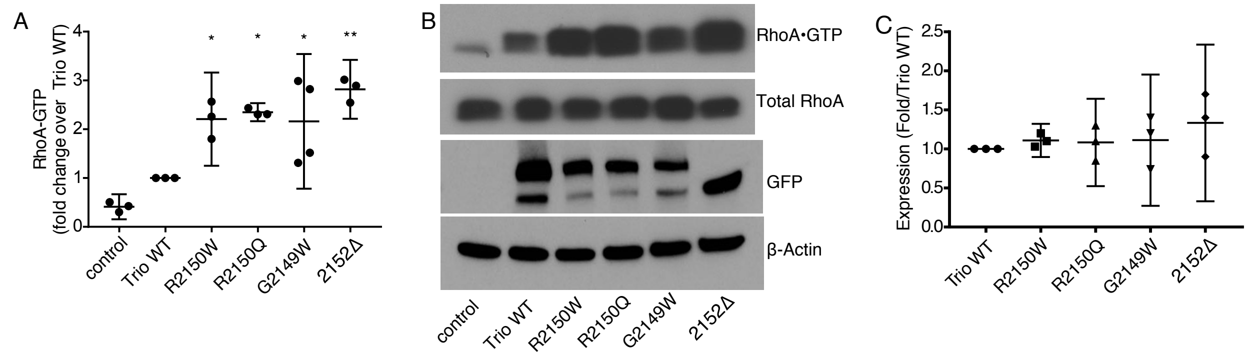


Figure 2.11 Rhotekin pulldown assay for pEGFP-TrioFL variants in 293 cells. Performed by Nadia Arang in the Gutkind Laboratory at UCSD. The quantified results of N=3 experiments normalized to WT (A). Fold change in RhoA·GTP content normalized to the Trio WT condition shown on Y axis and different variants shown on the X axis. Representative western blot of the experiments used to derive panel A (B). On top is RhoA·GTP blot, below is the blot for total RhoA, below that the GFP blot for Trio variant expression, and finally a β-Actin blot is used as a loading control for the GFP blot. A quantification of N=3 blots for Trio variant expression is shown in (C).

### 2.3.9 Small Angle X-ray Scattering on TrioC WT and R2150W

Size exclusion chromatography-coupled small angle x-ray scattering (SEC-SAXS) experiments were performed by Srinivas Chakravarthy, a beamline scientist at the biological collaborative access team at the advanced photon source. SAXS data collection parameters are detailed in Table 2.3. TrioC WT and R2150W proteins were

| SAXS data collection parameters | |
|---|---|
| Instrument | BioCAT facility at the Advanced Photon Source beamline 18ID with Pilatus3 1M (Dectris) detector |
| Wavelength (Å) | 1.033 |
| Beam size (μm$^2$) | 150 (h) x 25 (v) |
| Camera length (m) | 3.5 |
| $q$-measurement range (Å$^{-1}$) | 0.004-0.4 |
| Absolute scaling method | N/A |
| Basis for normalization to constant counts | To incident intensity, by ion chamber counter |
| Method for monitoring radiation damage | Automated frame-by-frame comparison of relevant regions |
| Exposure time, number of exposures | 0.5 s exposure time with a 2 s total exposure period (0.5 s on, 1.5 s off) of entire SEC elution |
| Sample configuration | SEC-SAXS. Size separation by an AKTA Pure with a Superdex 200 Increase 10/300 GL column. SAXS data measured in a 1.5 mm ID quartz capillary. |
| Sample temperature (ºC) | 20 |
| **Software employed for SAXS data reduction, analysis and interpretation** | |
| SAXS data reduction | Radial averaging; frame comparison, averaging, and subtraction done using BioXTAS RAW 1.4.0 |
| Basic analysis: Guinier, P(r) | Guinier fit and P(r) using ATSAS |

Table 2.3 SAXS data collection parameters and data reduction method used by Srinivas Chakravarthy to collect and process SAXS data at BIO-CAT 18-ID-D.

shipped to Argonne national laboratories and run on the SEC-SAXS setup. Scattering profiles were collected and normalized to buffer and blank controls and then used to derive paired distance distribution functions (P(r)) in order to determine the radius of gyration ($R_g$), maximum paired distance vector ($D_{max}$), and Porod volume (Table 2.4).

| | TrioC WT | TrioC R2150W |
|---|---|---|
| Concentration (mg/mL) | 3.0 | 3.0 |
| Guinier I(0) | 14.9 | 12.3 |
| Guinier $R_g$ (Å) | 25.3 | 24.5 |
| Guinier Pearson CC | 0.69 | 0.56 |
| P(r) I(0) | 15.0 | 12.7 |
| P(r) $R_g$ (Å) | 26.4 | 27.2 |
| P(r) Quality of Fit | 0.91 | 0.85 |
| $D_{max}$ (Å) | 90.0 | 94 |
| Porod Volume (Å$^3$) | 50735.0 | 52624.8 |

Table 2.4 SEC-SAXS data displayed for TrioC WT and R2150W variant reveals no major differences between them. Zero angle intensity [I(0)], radius of gyration ($R_g$), and quality of fit/Pearson correlation coefficient derived from Guinier and Pairwise-distance distribution function [P(r)] analyses are displayed. The maximum particle size $D_{max}$ was determined from the largest radius present in the P(r) plot.

TrioC WT had a $R_g$ of 26.4 Å, a $D_{max}$ of 90 Å, and a Porod volume of 50735 Å$^3$. The R2150W variant had a $R_g$ of 27.2 Å, a $D_{max}$ of 94 Å, and a Porod volume of 52625 Å$^3$. Although the R2150W variant displays larger values in each individual measure, it remains unclear whether these differences are meaningful without more experimental replicates. The low Guinier Pearson CC also indicates these samples need further optimization before further analysis. Envelopes generated using the DAMAVER software suite showed no visible differences for WT vs. R2150W. [138]

## 2.4 Discussion

The definition of the molecular events underpinning carcinogenic signaling in UM has showcased the role of Trio in transmitting signals from constitutively active $G\alpha_{q/11}$ subunits to the nucleus. [37,42] $G\alpha_{q/11}\cdot GTP$ activates Trio by relieving an autoinhibitory constraint on TrioC, which leads directly to nucleotide exchange of RhoA into its GTP-bound form. In this work, we have defined the structural basis of this autoinhibitory constraint using X-ray crystallography accompanied by HDX-MS and validated our model with biochemical and cell-based assays. The crystal structure of the TrioC module revealed the structure of a DH/PH tandem in a conformation incapable of binding GTPase. All three copies of the protein in the asymmetric unit adopt a highly similar conformation, supporting that the observed configuration of the DH and PH domains is not an artifact of crystal packing (Figure 1.7). In other published crystal structures of Dbl family members, the DH and PH domains exhibit distinct relative orientations[73,82,84], consistent with the α6-αN connecting region serving as a hinge. The crux of our structure is a novel interface formed between the α6-αN hinge region and the DH domain, wherein the most important contacts are made by residues in αN. It showed that Gly2149 and Arg2150 both make extensive contacts with the DH domain, with Gly2149 enabling closer proximity of αN to the DH domain than in other DH/PH modules, and Arg2150 sequestering DH domain residues Glu2069 and Met2146, which both make contact with RhoA in the bound state. These residues are conserved among TrioC homologs, but not in other RhoGEF modules including the N-terminal modules of Trio and Kalirin. The presence of Gly2149 and Arg2150 is likely a prerequisite for αN to follow a standard helical track at the end of α6 in order to block the switch II binding site of RhoA on the

DH domain. Accordingly, mutations of Gly2149 and Arg2150 were generally activating. Mutation of the analogous residues in p63RhoGEF (Gly340 and Arg341) were also activating[88], and we predict this trend would hold true for the KalirinC module. In contrast, the closely related PH domains of Dbs/TrioN instead enhance GEF activity by positioning residues in the α6-αN hinge region to form beneficial interactions with the GTPase substrate. [74,83] To the best of our knowledge, Sos1 is the only other example of a Dbl family structure in which the PH domain directly binds to the DH domain. In Sos1, the DH-PH interface involves the GTPase binding site of the DH domain and the αC helix of the PH domain. Given the poor conservation of the residues involved across the Dbl family, this interaction surface is likely unique to Sos1. [68,139]

The TrioC 2153Δ, 2152Δ, and 2147Δ truncations were all activated relative to WT TrioC (Table 2.2). Thus, the basal activity of the DH domain is highly sensitive to whether the PH domain is present. The bulk of the PH domain, and in particular its extended C-terminal helical region, is also required for binding and activation by $G\alpha_q$-GTP. [88]

Activation by truncation is a likely explanation for the transforming activity of Tgat, a splice variant of Trio encoding Trio residues 1921-2160 followed by a unique 15 residue extension. [56] Although the α6-αN region in Tgat is entirely present, the lack of the core PH domain fold in this variant could mean that the αN helix is disordered and thus cannot confer full autoinhibition as we hypothesize for the 2153Δ and 2152Δ variants. Alternatively, the 15-residue extension could drive an activated conformation of the α6-αN hinge region by making unique contacts within the protein, with other signaling partners, or with the cell membrane. [57,140,141]

Alanine scanning mutations throughout the DH-PH interface did not affect TrioC activity. In contrast, substitutions that introduced bulk into the interface such as G2149I/W, E2069R/ R2150E, and R2150W were able to activate TrioC >2-fold. These data suggest that in the absence of the interactions formed by one side chain, as in the R2150A variant, the remaining residues in the DH-PH interface can still contact each other and stabilize the autoinhibited conformation. In contrast, variants which insert steric bulk into the interface will disrupt the majority of DH-PH interfacial contacts from forming.

Although the β3-β4 loop is resolved as part of the closed DH-PH interface in our crystal structure, this poorly conserved loop likely plays little to no role in the autoinhibition of TrioC family members. Its removal did not affect TrioC activity or thermostability, and our HDX-MS profiling of the active mutant R2150W showed no difference in exchange rates in the β3-β4 loop, which we would expect to see if this loop formed a part of the DH-PH interface. Instead, the crystal contacts that β3-β4 makes with neighboring TrioC monomers are likely responsible for its ordered conformation in the crystal structure. (Figure 2.7). Truncation of the analogous loop in p63RhoGEF (Δ397-402) has no effect on activity. [88] The β3-β4 loops in structures of TrioN and Dbs make contacts with the bound GTPase, but mutation of the loop has no effect in vitro. [74,83]

Our data support a model wherein TrioC exists in an equilibrium of autoinhibited and active conformations (Figure 2.16). The existence of an equilibrium is supported by the measurable basal GEF activity of TrioC and full-length Trio. The basal autoinhibited state is represented by the crystal structure we have reported here (Figure 2.16, top left quadrant). TrioC can sample a conformation which is active in the absence of $G\alpha_q$·GTP, and the cancer point variants we profiled are able to shift the equilibrium towards this state (Figure 2.16, bottom left quadrant). This is accomplished through displacement of αN from its contacts with α3 (confirmed by HDX-MS). $G\alpha_q$·GTP binds primarily to an extension of αC in the PH domain (Figure 2.16, top right quadrant) yet does not activate TrioC until it binds to both DH and PH domains and displaces αN from α3 (Figure 2.16, bottom right quadrant, represented by PDB entry 2RGN). Our SAXS results hint that activated TrioC may occupy a larger size in solution, although this is a small change that could be due to the inherent limitations of the SAXS method itself. Our results suggest that because the cancer-associated point variants favor a conformation similar to that produced by $G\alpha_q$·GTP binding, they synergistically enhanced $G\alpha_q$·GTP binding and a maximum activation rate. The exception is R2150W, whose activity was not further enhanced by saturating $G\alpha_q$·GDP·$AlF_4^-$. This variant may be present in a fully activated conformation without need for $G\alpha_q$·GDP·$AlF_4^-$.

We posit that activation of TrioC by $G\alpha_q$·GTP, by point mutation, and by truncation all depend on the same biophysical mechanism: displacement of αN from the contacts made with α3 seen in our crystal structure. As is seen in our cell based RhoA activation assay, this paradigm held true in an overexpression model of full-length Trio in human cells. Thus, in human cancer, Trio has the potential to bypass regulation by $G\alpha_{q/11}$ by truncation or point variation, which would lead to the activation of RhoA and downstream proliferative signaling through the AP-1 and YAP-TEAD axes. [38,42] A small molecule stabilizer of the autoinhibited TrioC conformation reported here could prove fruitful in halting proliferative signaling through Trio. In combination with therapies targeting the other arms of $G\alpha_q$ signaling, a TrioC inhibitor could be an effective part of a combination therapy for UM.

## 2.5 Conclusions and Future Directions

Within this Chapter, I set out to evaluate the hypothesis that TrioC autoregulation is mediated by residues in the αN region of the PH domain. Through X-ray crystallography and a panel of biochemical and biophysical techniques, I have shown that the primary determinants of TrioC autoinhibition do indeed reside within αN, namely Gly 2149 and Arg2150, thereby defining the mode of regulation for this subfamily. In the R2150W variant, which is found in human cancer, I have demonstrated that a TrioC point variant can represent a fully activated conformation of TrioC. My structural model for TrioC regulation is supported by both crystallographic and solution-based data (HDX-MS), and this model seems to hold true for full-length Trio. Thus, I have demonstrated the molecular basis of TrioC subfamily autoinhibition conferred by αN. Future work in this project will evaluate the biological consequences of TrioC activation through use of the point variant R2150W. Experiments to be performed include the evaluation of downstream signaling cascades, including the evaluation of RhoA-dependent transcriptional activity. Phenotypes driven by the R2150W variant could also be profiled using focus formation assays, migration assays, transcriptomics, and xenograft experiments.

# Chapter 3 **Structure and Function of Understudied Trio domains**

## 3.1 Introduction

The following Chapter describes my efforts to understand structure/function relationships in understudied domains in Trio. Only the two GEF modules of Trio are characterized and thus ~80% of the protein is still unexplored. My hypothesis going into this work were that nonenzymatic portions of Trio regulate its GEF activities through allostery. I also considered the hypothesis that $G\alpha_q$ functions as a master regulator of Trio by inducing global conformational change, based on the fact that $G\alpha_q$ stimulates not only RhoA exchange through Trio but also Rac1 exchange in cells. I start with purification and crystallization trials of individual domains of Trio, including the CT domain, SH3 domains, Ig domain, and kinase domain, and test a hypothesis from the literature that the Ig domain interacts with GTP-bound RhoA. I then shift to my efforts to purify full-length Trio from mammalian cells. I close with characterization of various constructs of Trio using EM and biochemical assays. I have generated constructs and laid the groundwork for future work in studying Trio function using larger portions of the molecule using electron microscopy.

## 3.2 Methods

### 3.2.1 Cloning

hTrioFL (hTrio residues 61-3097) cDNA inserted into peGFP and *Drosophila melanogaster* Trio (dTrio) cDNA in pRmHA-3 were gifts from the Gutkind Laboratory at UCSD. The hTrioFL construct was mutagenized via inverse PCR using Q5 DNA polymerase in order to install an N-terminal 10xHis tag followed by a TEV protease cleavage sequence (N-10xHis-TEV-eGFP-hTrioFL-C). Construct was then DpnI digested, subjected to a PCR cleanup kit, phosphorylated, ligated, and transformed into XL-1 Blue E. coli cells. Other Trio constructs were amplified out of the hTrio plasmid using PCR with ligation-independent cloning (LIC) handles. [142] For shorter constructs, Q5 DNA polymerase was used. For longer constructs, KOD polymerase was necessary in order to get productive and high-fidelity amplification and was used. PCR reactions were run on agarose gels to confirm the presence and correct size of amplicons. Reactions were then DpnI digested, subjected to a PCR cleanup kit (Qiagen 28106), and processed with T4 DNA polymerase to generate sticky ends. T4 processed inserts were mixed with linearized and T4 processed vectors for E. coli, insect, or mammalian expression from the University of Michigan high-throughput protein production core, and reconstituted plasmids were transformed into DH5$\alpha$ or XL1-Blue E. coli cells. Constructs were screened for the presence of the insert in the correct reading frame using Sanger sequencing at the University of Michigan DNA sequencing core facility. Primers were directed reading towards the 3’ direction starting prior to the LIC site and reading towards 5′ immediately following the LIC insertion site. Longer Trio constructs were sequenced using primers directed at the internal Trio sequence itself. Vectors used included those with MBP and Glutathione-S-transferase (GST), and all contained 6xHis sequences N-terminal to the fusion tag (N-6xHis-MBP/GST/-TEV-hTrio-C).

### 3.2.2 Protein Expression and Purification

RhoA expression and purification is described in Chapter two. Rac1 was expressed and purified using the same protocol as used for RhoA. Plasmids encoding Trio constructs were transformed into Rosetta (DE3) pLysS Escherichia coli (E. coli) cells (Novagen) and grown in Terrific Broth (EMD Millipore Sigma) with 100 μg/mL ampicillin or carbenicillin plus 50 μg/mL chloramphenicol at 37 °C with 200 rpm shaking. Once a 600 nm optical density ($OD_{600}$) of 0.6-0.8 was reached, expression of N-terminally tagged fusion proteins was induced using 0.5 mM isopropyl β-D-1-thiogalactopyranoside and cells were further allowed to grow at 20 °C with 200 rpm shaking for 20-24 h. E. coli were then harvested at 5000 × g for 15 min and cell pellets were flash-frozen or prepared. Cell pellets were vortexed and resuspended using a Dounce homogenizer in an ice-cold "lysis buffer" containing 20 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) HEPES pH 8.0, 200 mM NaCl, 2 mM dithiothreitol (DTT), 1 mM $MgCl_2$, 5 % glycerol, 0.001 mM leupeptin, 1 mM lima bean trypsin inhibitor, and 0.1 mM phenylmethylsulfonyl fluoride. Individual Trio domain construct purifications also included 10 mM imidazole. Longer Trio constructs included 5mM EDTA in the lysis buffer and no $MgCl_2$. Resuspended cell solution was then lysed using a handheld VirSonic 100 sonicator (Boston Laboratory Equipment) for five 30-second pulses at 18 watts on ice. Lysate was then centrifuged at 40000 rpm in a Beckman Optima L-90K ultracentrifuge (Beckman-Coulter) to remove insoluble material. The soluble fraction was then filtered through a 0.45 μm filter and loaded onto nickel- nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) equilibrated with lysis buffer. Larger Trio constructs were prepared using Roche cOmplete His-tag resin (Roche 5893682001) equilibrated with lysis buffer. Two aliquots of 10 column volumes (CV) of lysis buffer containing 20 mM imidazole were used to wash the column (10 mM for cOmplete resin). The recombinant protein was then eluted using lysis buffer plus 200 mM imidazole. The elution fractions containing desired protein were then incubated with

5 % (w/w) tobacco etch virus protease in order to cleave the N-terminal expression tag and the mixture was dialyzed against a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT in order to remove imidazole. The fusion tag was then removed using another round of Ni-NTA purification for individual domain constructs. Proteins were then checked for purity using sodium dodecyl sulfate-polyacrylamide gel electrophoresis and pure fractions were concentrated in an Amicon Ultracel concentrator (Millipore), flash frozen in liquid nitrogen, and stored at –80 ˚C. For GEF assays, protein was thawed from -80 ˚C on ice, 0.2 μm filtered, and used for experiments. For crystallography, fractions were thawed from –80 ˚C on ice, and polished using gel filtration chromatography on a Superdex 75 or 200 10/300GL column (General Electric healthcare) in a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT. After purification using cOmplete resin, larger Trio constructs were dialyzed against 20 mM HEPES pH 8.0, 10 mM NaCl, and 2 mM DTT with 1 % (w/w) TEV protease overnight at 4 ºC. These solutions were subjected to anion exchange chromatography using a 5 mL HiTrap Q HP column (GE 17115401) using a buffer of 20 mM HEPES pH 8.0 and 2 mM DTT with a gradient of 10 mM to 1000 mM NaCl over 100 mL. Desired fractions were concentrated in a 30 or 50 kD cutoff concentrator and loaded on either a 24 mL Superose 6 or Superdex 200 column (GE) equilibrated with 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT. Desired fractions from SEC were concentrated to 1-2 mg/mL and flash frozen. Protein aliquots for biophysical experiments were concentrated further and frozen as required. For GTPase complex formation, Trio constructs were mixed with >2-fold molar excess of GTPase in a buffer containing 0.1 mM EDTA to drive complex formation. Complexes were incubated at 4 ºC for >30 minutes and loaded on an S200

column equilibrated with 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT and 0.1 mM EDTA. Fractions containing complex were used for biophysical experiments.

Human Embryonic Kidney 293 Freestyle cells (293F-ThermoFisher) were cultured in suspension in Freestyle 293 expression media (ThermoFisher) and transfected with a 1:3 (w/w) mixture of hTrioFL-pEGFP DNA:Polyethylenimine (see eukaryotic cell culture). After 48 hours, cells were harvested and pelleted at 4 °C at 500 x g for 10 minutes and either flash frozen or carried forward for preparation. A mixture of 2.5 mL/50 mL cell culture Cell-Lytic M (Sigma C2978) was supplemented with 200 mM NaCl, 2 mM DTT, 5 % glycerol, 10 mM EDTA, and one protease inhibitor cocktail tablet (Sigma S8830) and added to cells. The cell pellet was gently rocked in lysis buffer at 4 °C for 15 minutes in order to resuspend the 293F pellet. The cell suspension was brought up to 80 mL in a lysis buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, 10 mM EDTA, and 2.5 % glycerol and subjected to ultracentrifugation at 40000 rpm for 45 minutes. Supernatant was run through a 0.4 µm filter and loaded onto cOmplete His-tag purification resin (Roche 5893682001) equilibrated with “wash buffer”: 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, 5 % glycerol, 10 mM EDTA. Column was washed twice with 5 CV of wash buffer including 10mM Imidazole, and protein was eluted using 5 x 1 mL aliquots of wash buffer + 200 mM imidazole. Samples from the purification were run on a 4-15 % gradient SDS-PAGE gel (Bio-Rad) and imaged on a Typhoon (GE Healthcare) to measure eGFP fluorescence (488 nm excitation, 532 nm emission) before being stained using Coomassie (Bio-Rad). Purification on HiTrap Q HP and Superose 6 columns was attempted similarly to described above.

Rac1/RhoA affinity columns were generated using Affi-Gel 10 and further purification of hTrioFL was attempted (Bio-Rad). One mL of Affi-Gel 10 slurry was washed with 10 CV of water and 10 CV of 20 mM HEPES pH 7.5, 200 mM NaCl, 2 mM DTT, and 1 mM $MgCl_2$. GTPase, fresh from an SEC purification, was incubated with the Affi-Gel resin at 4 °C for four hours. Flowthrough and three 10 CV washes were collected and then the loaded resin was used for hTrioFL purification attempts. For purification, the column was equilibrated with 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, and 5 % glycerol. Elutions from the His-tag purification were added to the column and rocked for 15 minutes at 4 °C. Ten mM EDTA was added into the reaction mixture, encouraging complex formation. Column was washed four times with 10 CV of 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, and 5 % glycerol with 10 mM EDTA. The complex was then eluted with 5 x 0.5 mL of 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, and 5 % glycerol with 1 mM $MgCl_2$ and 20 μM GDP. Samples were judged for purity using a fluorescence scan of the SDS-PAGE gel and Coomassie stain for total protein.

### 3.2.3 Crystallography

Sitting drop screening trays were set using commercial screens from Qiagen and Hampton Research: IndexHT, Protein Complex, Classics Lite, and JCSG+. Ninety-six well plates (Art Robbins Instruments 102-0001-03) were set using a Gryphon Crystal Robot (Art Robbins Instruments) and incubated at 20 ˚C. Hits from original screens were optimized in 24-well hanging drop VDX plates (Hampton Research) in order to optimize crystal size and quality.

### 3.2.4 GEF assays

FRET was used to assess the nucleotide exchange activity of Trio variants in a Corning Black 384 well low-volume round bottom micro plate (Corning 4514). [131] [73] First, 2 µM RhoA·GDP (or Rac1) was incubated with 50 nM GEF for 5 min at room temperature in freshly prepared nucleotide exchange buffer: 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, and 10 mM $MgCl_2$. Immediately before measurement, 1 µM 2´/3´-O-(N-Methyl-anthraniloyl)-guanosine-5´- triphosphate (MANT-GTP) (Jena Biosciences), was added to a final assay volume of 20 µL. The mixture was then excited at 280 nm, and fluorescence intensity at 450 nm was read in 5 s intervals on a SpectraMax M5 plate reader (Molecular Devices) for 5-10 min. Fluorescence curves were fit to a linear regression model using GraphPad Prism to derive the observed kinetic constant $k_{obs}$. The resulting $k_{obs}$ of the experimental condition was then compared to that of matched rates of GTPase alone and GTPase + control GEF (TrioN or TrioC). For $G\alpha_q$ activation assays, $G\alpha_q$ was added at 200nM, and the assay was run with and without the addition of 30 µM $AlCl_3$ and 10 mM NaF in the reaction buffer to generate $AlF_4^-$ in solution.

### 3.2.5 Kinase Assays

Radiometric $^{32}P$ assays were run in 96-well PCR plates (Fisherbrand 14230237) at 10 µL reaction volumes. Reaction mixtures were set up with 50 nM kinase enzyme, 500 nM substrate, and 100 µM ATP (2-5% $\gamma$-$^{32}P$ ATP- PerkinElmer NEG502H250UC) in a reaction buffer containing 20 mM HEPES, pH 8.0, 2 mM DTT, and 10 mM $MgCl_2$. To test activation of Trio kinase constructs by $Ca^{2+}$•CaM, 1 µM of CaM and/or 1 mM $Ca^{2+}$ were used. Control kinases used were GRK5 and TBK1 and were purified as described.[143,144] Substrates used were bovine brain Tubulin (Cytoskeleton, Inc. HTS02), bovine Myelin

Basic Protein (Sigma-Aldrich M1891) and recombinant human Myosin light chain 2 (Novus Biologicals NBC1-18536). The ADP-Glo assay was also used (Promega V6930) in a 384-well white polystyrene assay plate (Corning 3572) to a final volume of 20 µL using the same reaction conditions as above. An ADP dilution curve was used to show that the kit reagents were functioning properly and could achieve a S/B value of ~6 with an ADP concentration of 100 nM. At the end of the kinase reaction, 10 µL of the ADP-Glo reagent was added to the reaction mixture and incubated for 40 minutes at room temperature. 20 µL of kinase detection reagent was then added and luminescence was measured on the Spectramax M5 plate reader after a 30-minute incubation at room temperature.

### 3.2.6 SEC-MALS

Proteins were run on a SEC-MALS system configured with an AKTAmicro system (GE) with a UV detector, a Protein-KW804 Silica column (Shodex), a Dawn Heleos II Multi-Angle Laser Light Scattering (Wyatt), and an Optilab T-rEX differential refractometer (Wyatt). Superose 6 (GE) and WTC-050S5 (Wyatt) columns were also tried in Trio-G$\alpha_q$ complexing experiments. The system was equilibrated in a base buffer of 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT with the addition of $MgCl_2$, $AlF_4$, and guanosine nucleotides as needed. The system was first standardized using a control injection of 100 µg bovine serum albumin (BSA) (Sigma) dissolved in water. Following the correct determination of both the BSA monomer and dimer molecular masses, 50-100 µg of proteins were injected, and their scattering and absorbance profiles were analyzed to determine their molecular weight in solution. The output signals were imported into Astra 6 software (Wyatt Technology Corporation).

### 3.2.7 Isothermal Titration Calorimetry

ITC experiments were performed by Krishnapriya Chinnaswamy in the University of Michigan Center for Structural Biology. Ig domain and RhoA were dialyzed against the same buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 1 mM $MgCl_2$. ITC experiments were performed in the Nano-ITC Low volume (TA Instruments). 400 µL of RhoA in the above buffer was added to the cell and 50 µL of Ig domain in above buffer was drawn into the syringe. The experiments were performed at 25 C with 33 x 1.5 µL of binding partner injected into RhoA every 250 seconds with stirring speed of 250 rpm. Buffer alone was used as a blank. The $K_a$ and ΔH of the reactions were calculated using the Launch NanoAnalyze software (TA Instruments).

### 3.2.8 Differential Scanning Fluorimetry

Data were collected on a QuantStudio 6 Real-Time PCR system (Applied BioSystems) using Sypro Orange (SO) dye (ThermoFisher). Purified Trio constructs were incubated at 1-5 µM in a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT with 1x dye. White 384 well PCR plates (Applied Biosystems) were used and covered with sealing tape. These plates were exposed to a temperature gradient of 20–60 °C. Fluorescence was monitored as a function of temperature, and the $T_m$ was determined by fitting the fluorescence data to a sigmoidal curve and calculating the inflection point in GraphPad Prism.

### 3.2.9 Negative Stain Electron Microscopy

Negative staining of Trio samples was done as previously described.[145] Purified protein samples were applied to a glow discharged formvar coated copper grid (Electron Microscopy Sciences FCF400-Cu-50) for 1 minute. Grids were blotted against filter paper and dipped in $ddH_2O$ quickly and again blotted. Grids were again dipped in $ddH_2O$ and again blotted. Grids were then quickly dipped into a 0.75 % Uranyl Formate (UF) solution made in $ddH_2O$ and then blotted. Finally, grids were dipped into another drop of UF solution and soaked for 1 minute. Afterwards, grids were blotted dry and then dried using a vacuum line. Stained grids were evaluated for stain quality, contrast, particle quality, and particle spread using the Morgagni 100 kV transmission electron microscope (TEM) (FEI). Grids which displayed high stain quality, high contrast, well spread and well defined particles were taken for further imaging on the Tecnai T12 120 kV transmission electron microscope (FEI) operated using the Leginon automated data collection system.[146] Particles were picked and 2-dimensional class averaging was done using the cisTEM software suite [147].

### 3.2.10 Cryogenic Electron Microscopy

The following was done with extensive help from Jennifer Cash, Ph.D. Protein samples were frozen using the Vitrobot automated grid freezing system (FEI) on Quantifoil 300 mesh Copper grids (Electron Microscopy Sciences Q3100-CR1.3). Liquid ethane was prepared, 4 µL of protein sample was added onto the grid, the sample was blotted using filter paper to remove excess liquid, and the grid was flash plunged into liquid ethane. Grids were clipped using C-clips (Electron Microscopy Sciences) and samples were loaded into the 200 kV Talos Artcica TEM for sample screening using Leginon. Segments of the grid containing thin ice with good particle contrast and spread were collected and motion corrected using the MotionCor2 software package within Leginon. [148] Motion-corrected micrographs were used to generate preliminary 2D class averages using cisTEM.

### 3.2.11 Eukaryotic Cell Culture and Transfection Procedure

HEK293F Freestyle (ThermoFisher R79007) cells were maintained in suspension culture at a 25 mL volume in Freestyle expression media (ThermoFisher 12338-026) at 110 rpm shaking in 5 % $CO_2$ at 37 ˚C in a Multitron incubator (INFORS HT). Cells were cultured between $0.3x10^6$ and $5x10^6$ cells/mL as determined by trypan blue (Sigma) counting with a hemocytometer. DNA for transfection was transformed into E. coli DH5α and prepared using a Midiprep kit (Qiagen). Prior to transfection, cells were expanded to 50 mL or larger as needed. At ~$1x10^6$ cells/mL, cells were transfected with a 1:3 w/w mixture of prepared DNA:PEI using the Opti-MEM transfection reagent, warmed to 37 ºC (Thermofisher 31985062). Transfection cultures were returned to the shaking incubator for ~48 hours before cells were harvested and frozen or freshly prepared.

## 3.3 Results

### 3.3.1 Cral_Trio (CT) Domain

Constructs encoding the Trio CT domain with different domain boundaries were designed using a combination of manual comparison to homolog structures and the XtalPred prediction server.[149] These inserts were cloned into pMCSG9 using the LIC protocol and transformed into Rosetta E. coli cells. Constructs were expressed as MBP fusions at 20 ºC using 0.5 mM IPTG and purified using Ni-NTA resin, processed with TEV protease, and run on Ni-NTA resin once more to remove MBP. Proteins purified this way were pure by SDS-PAGE with the exception of a minor contaminant at ~70 kD. CT constructs displayed abnormal behavior on SEC, running as an extremely broad peak. (Figure 3.1) Samples from this preparation were subjected to the lab suite of four crystal screens at 10 mg/mL, but no hits were obtained. Because the CT domains bind to lipid

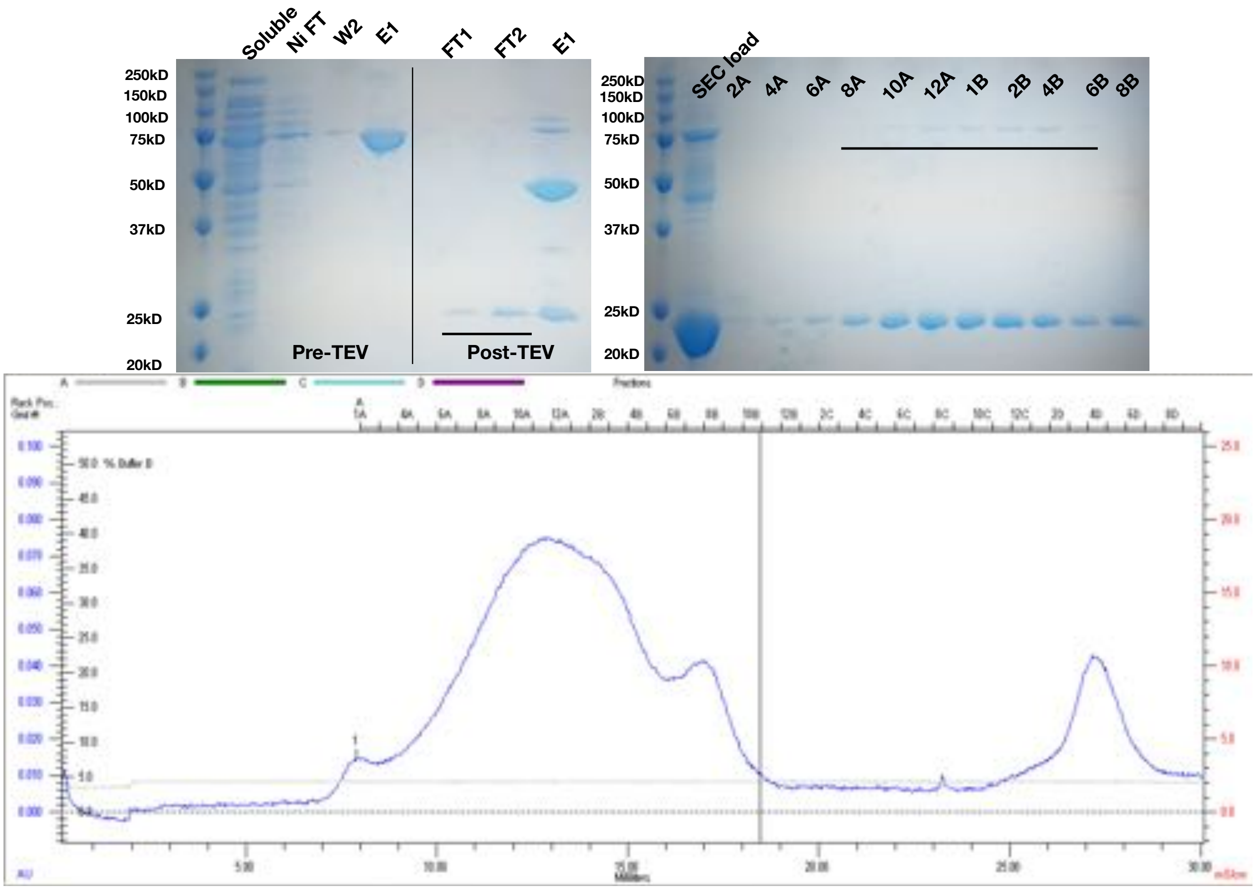


Figure 3.1 Purification of the Trio CT domain shows aberrant SEC behavior. Top left, Coomassie stained 12 % polyacrylamide gel showing Ni-NTA purification of CT domain before and after TEV cleavage. Top right, Coomassie stained 12 % polyacrylamide gel of fractions from below SEC trace. Black lines indicate fractions taken to the next step or fractions concentrated and used for crystallography. Bottom, $A_{280}$ trace (blue line) from single analytical 24 mL Superdex 200 run of CT domain.

molecules, detergent extraction of potential bound lipids was attempted using Triton X-100 as described for a homolog.[99] CT samples were washed with 0.01 % Triton X-100 (critical micellar concentration ~0.02 %) while immobilized to the nickel column, and then eluted in buffer with no Triton. This seemed to slightly reduce the presence of larger species, but the protein still eluted over a broad curve. (Figure 3.2 top vs. bottom) Next, in addition to the steps described above, 0.01 % Triton was included in the SEC buffer to try to eliminate protein aggregation. This SEC run produced a much narrower peak, however the interpretation of this result is not simple as Triton has significant $A_{280}$. This narrow peak had a significantly higher absorbance at peak height ~0.5 AU compared to that seen with broader peaks ~0.06 AU even though similar protein amounts were loaded (by $A_{280}$). This was far more absorbance than could be accounted for by the peak narrowing. Thus, we hypothesized this new narrower peak was not actually a protein peak, but rather a micelle of concentrated Triton. This is supported by a Coomassie stained gel showing extremely small amounts of protein present. (Figure 3.3) The small amount of protein remaining did elute at the right elution volume for a 22 kD monomer.

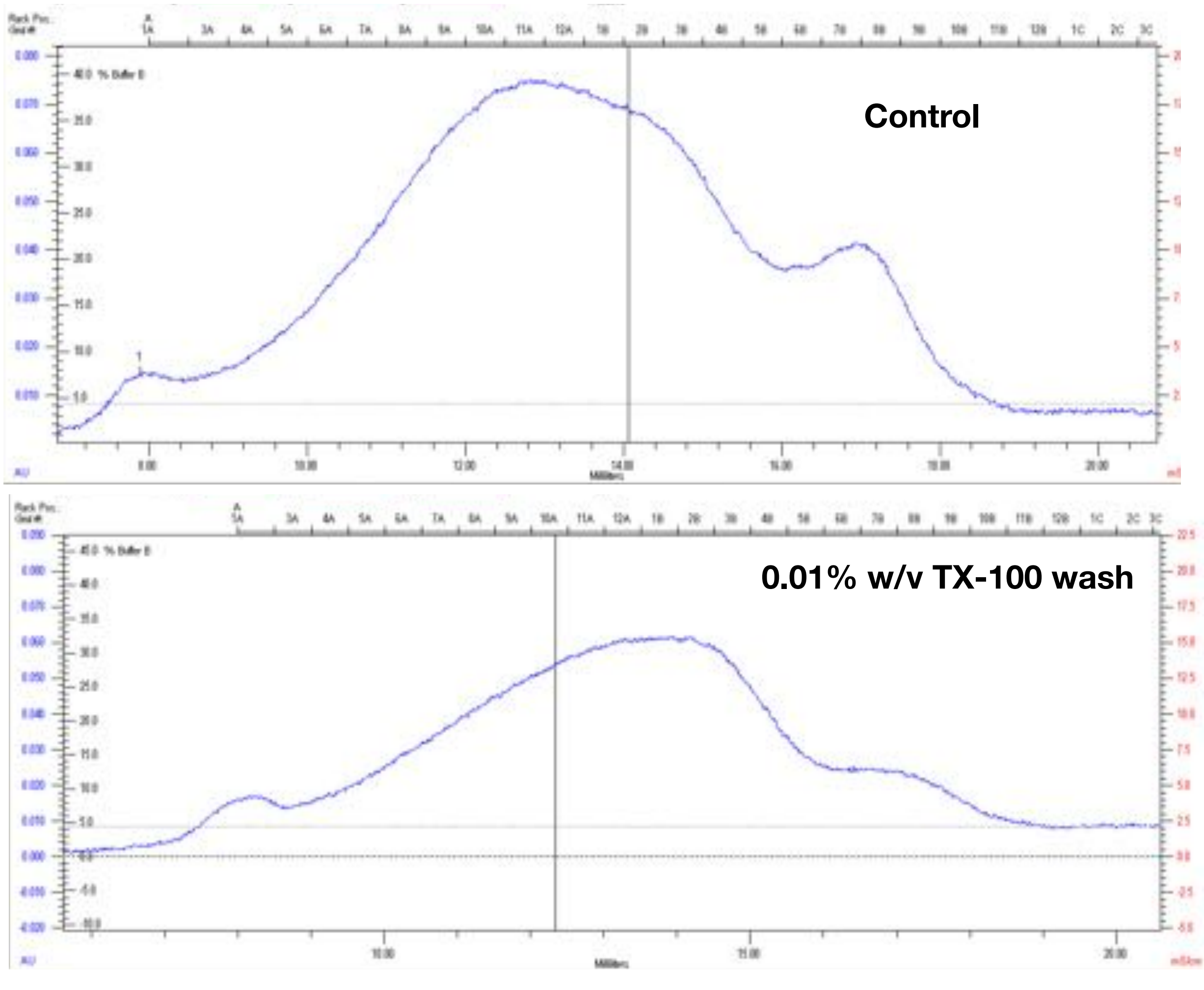


Figure 3.2 $A_{280}$ traces (blue line) from single analytical 24 mL Superdex 200 run of CT domains show that CT domain behavior is not improved by TX-100 washing. Control conditions (top), or a 0.01 % (w/v) Triton X-100 wash while CT domain is immobilized to the Ni-NTA column. A protein monomer of 22 kD is expected to elute at 15 mL based on protein standards run in the same buffer.

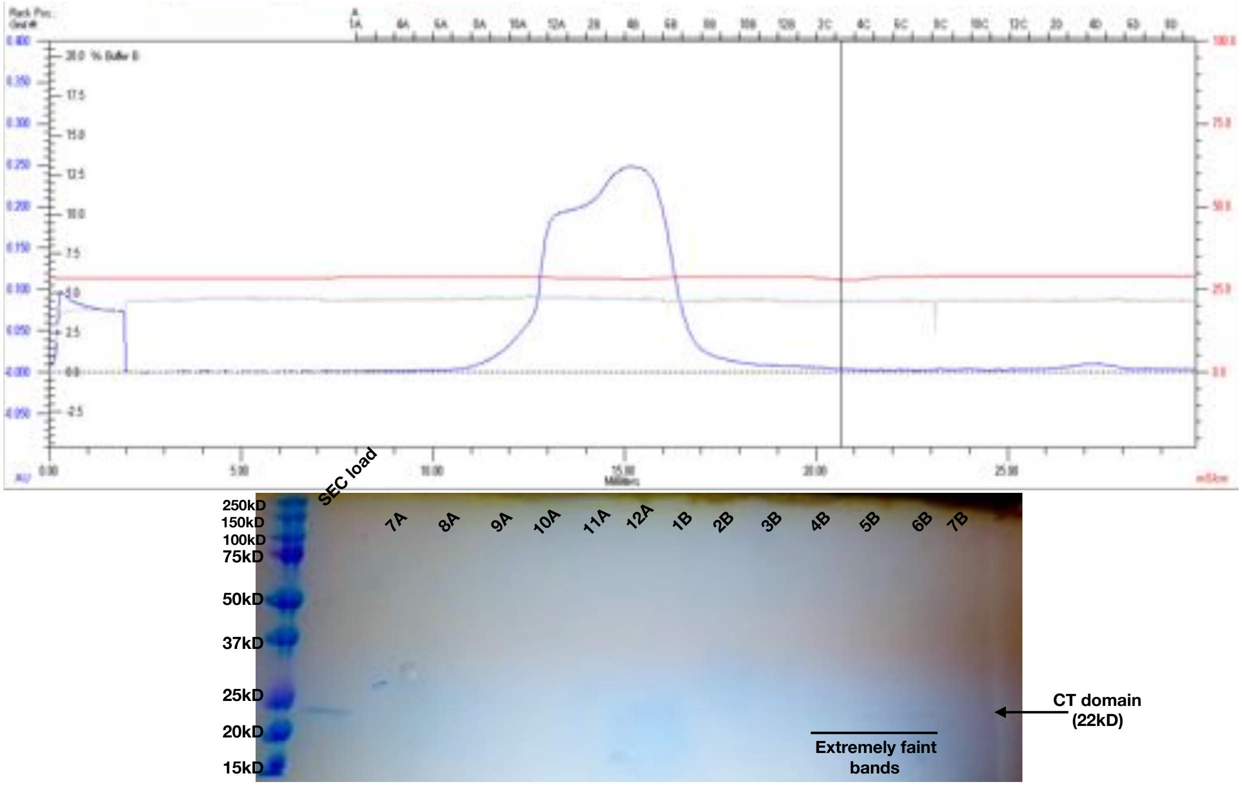


Figure 3.3 Single Superdex 200 SEC run using 0.01 % Triton X-100 in all purification steps and the SEC buffer likely leads to formation of triton micelles which complicates interpretation of purification results. Top, $A_{280}$ trace of SEC run (blue line). Bottom, Coomassie stained 12 % gel of fractions from SEC trace. Extremely faint bands are visible in fractions 4-6B.

Remaining protein may have been denatured by the Triton and removed during one of several purification steps. The incubation of protein samples with bio-beads (Bio-Rad) to remove Triton before SEC was also attempted. Samples purified this way also ran as an extremely broad peak on SEC. CT domain samples co-expressed with GroEL/ES or DNAJ/K did not fare any better.

### 3.3.2 SH3_N Domain

A construct encoding Trio's SH3_N domain was designed based on the rat Kalirin SH3_N NMR structure (PDB: 1U3O)[109], which has 68% sequence identity to the human Trio sequence. The XtalPred server was also used to slightly change domain boundaries without extending into regions of predicted disorder. This construct was expressed as

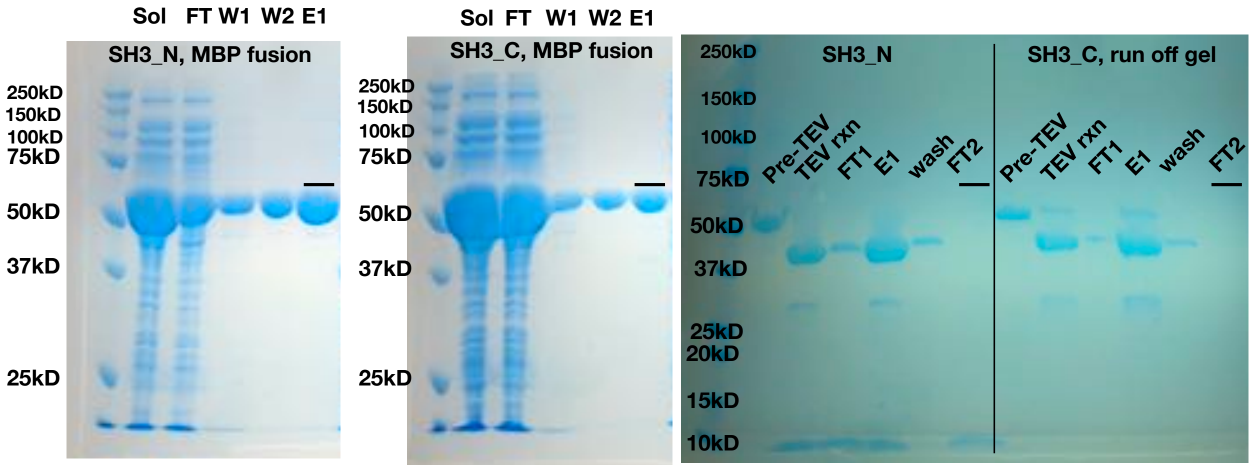


Figure 3.4 Ni-NTA purification of SH3_N and SH3_C as MBP fusions show that both SH3 domains express and purify well. Left, Coomassie stained 12 % gel showing Ni-NTA purification of SH3_N as an MBP fusion. Middle, Coomassie stained 12 % gel showing Ni-NTA purification of SH3_C as an MBP fusion. Following TEV cleavage, both constructs were subjected to reverse IMAC (Right) and resultant fractions are shown on a Coomassie stained 12 % gel. Horizontal black lines indicate fractions that were taken on to the next steps.

an MBP fusion in Rosetta E. coli cells and purified using Ni-NTA resin, TEV cleavage, reverse IMAC, and a final preparative grade S200 SEC run. (Figures 3.4 and 3.5) Screening trays were set with SH3_N at 9.2 mg/mL, yet they yielded no hits and the drops in the tray remained largely clear. Further studies

of this domain are hindered due to the lack of information about what ligands it interacts with.

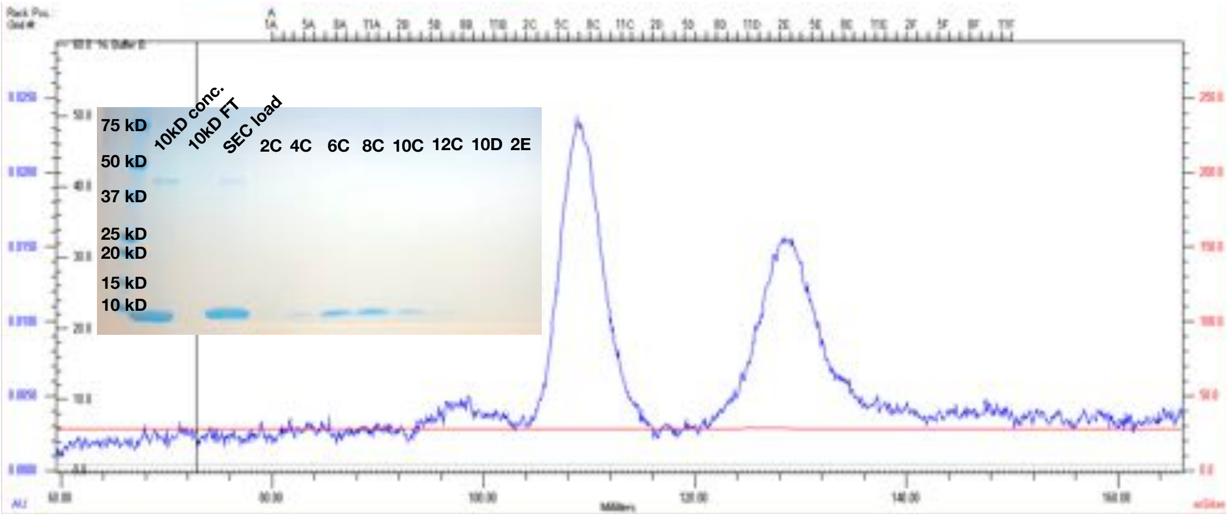


Figure 3.5 SEC purification of SH3_N shows homogenous monomeric protein. $A_{280}$ trace (blue line) from preparative 120 mL Superdex 200 run of SH3_N and Coomassie stained 4-15 % gradient gel of fractions from SEC showing pure SH3_N.

### 3.3.3 SH3_C Domain

A construct encoding Trio's SH3_C domain was designed based on the mouse Kalirin SH3_C NMR structure (PDB: 1WFW), which has 61% identity. The XtalPred server was also used to slightly change domain boundaries without extending into regions of predicted disorder. This construct was expressed as an MBP fusion in Rosetta E. coli cells and purified using Ni-NTA resin, TEV cleavage, reverse IMAC, and a final preparative grade S200 SEC run. (Figures 3.4, 3.6) Crystal screens were set with SH3_C at 10 mg/mL: JCSG+, Classics Lite, Index HT, and ProComplex. A preliminary crystal form was obtained and manifested as sprays of fine needles around a central nucleation point. This crystal form, grown against a grid centered around 3 M NaCl including either 0.1 M HEPES pH 7.5 or 0.1 M Tris pH 8.5, reproduced in a 24-well hanging drop tray. Similar crystals were observed in a condition with 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 5.5, and 45 % methylpentanediol. Some wells contained fewer nucleation points, however the crystals had similar morphology. (Figure 3.6) Further optimization in hanging

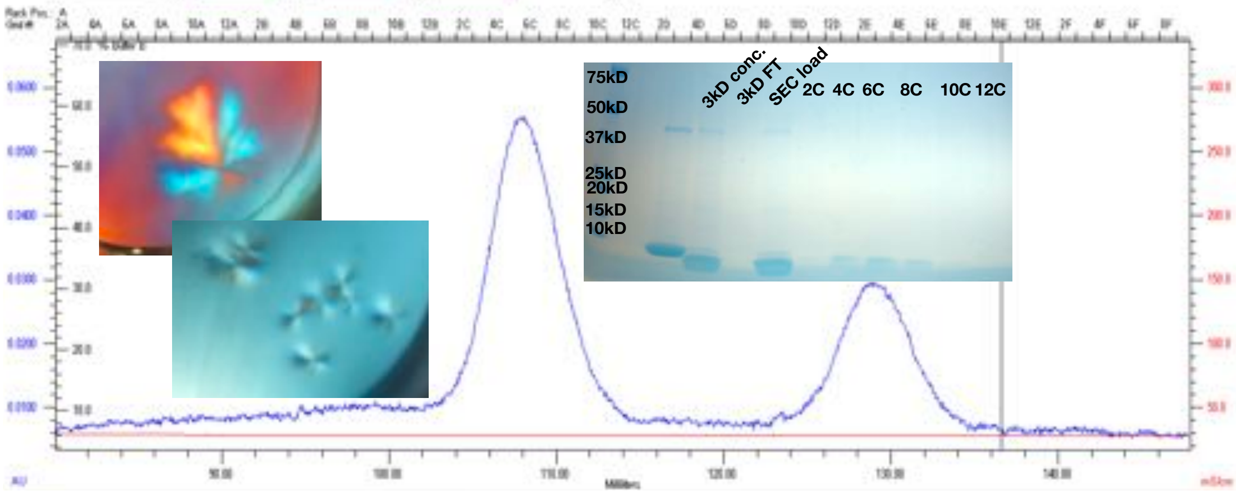


Figure 3.6 SEC purification of SH3_C shows homogenous monomeric protein and a preliminary crystal form is shown. $A_{280}$ trace from preparative 120 mL Superdex 200 run of SH3_C and Coomassie stained 4-15 % gradient gel of fractions from SEC showing pure SH3_C. Left, crystal form obtained for SH3_C in the following conditions: 0.1 M HEPES pH 7.5, 3.0 M Sodium chloride or 0.1 M Tris pH 8.5, 3.0 M Sodium chloride and 0.2 M Ammonium acetate,

drop trays, screening protein, precipitant, and additives, will be required to improve this form into a single crystal, or a new crystal form will need to be found. Seeding could also be employed. Similar to the SH3_N domain, studies of this domain in isolation will be enhanced by definition of its ligand or ligands, if any, which remain unknown.

### 3.3.4 Ig Domain

A construct encoding Trio's Ig domain was designed using a combination of manual comparison to homolog structures and the XtalPred prediction server. This construct was expressed as an MBP fusion in Rosetta E. coli cells and purified using Ni-NTA resin, TEV cleavage, reverse IMAC, and a final preparative grade S200 SEC run. (Figure 3.7)

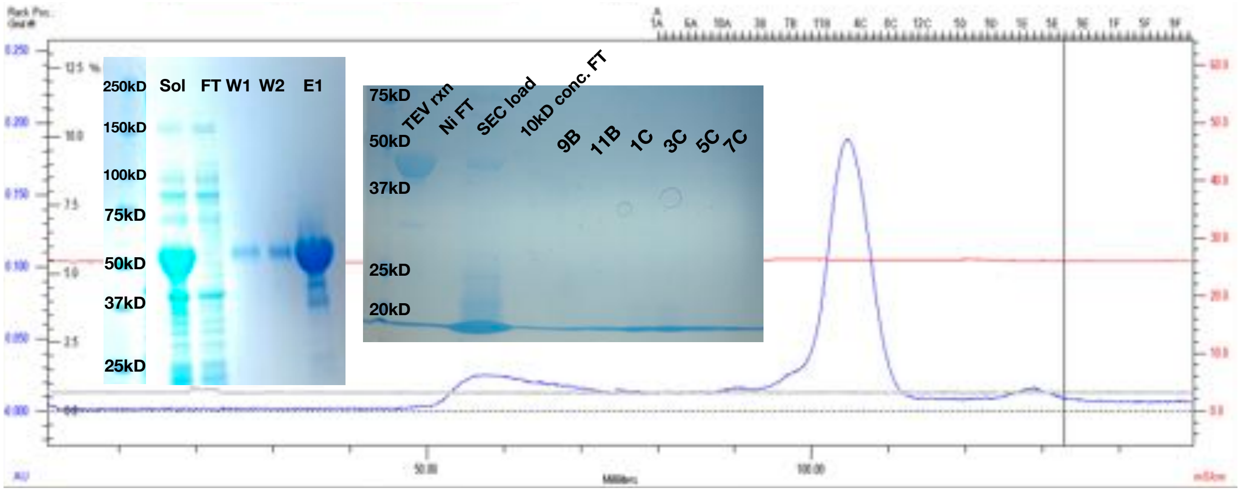


Figure 3.7 Ig domain purification using Ni-NTA and SEC yields homogenous monomeric protein. Left, Coomassie stained 4-15 % gradient gel showing Ni-NTA purification of Ig domain as MBP fusion. Middle, Coomassie stained 4-15 % gradient gel showing reverse IMAC purification and fractions from 120 mL preparative grade Superdex 200 column. Back, $A_{280}$ trace (blue) from SEC.

This construct was set in screening trays using JCSG+, Classics Lite, Index HT, and ProComplex @ 9.75 mg/mL, no crystalline hits were obtained. This protein was also used to evaluate the hypothesized interaction between the Ig domain and the GTP-bound conformation of RhoA.

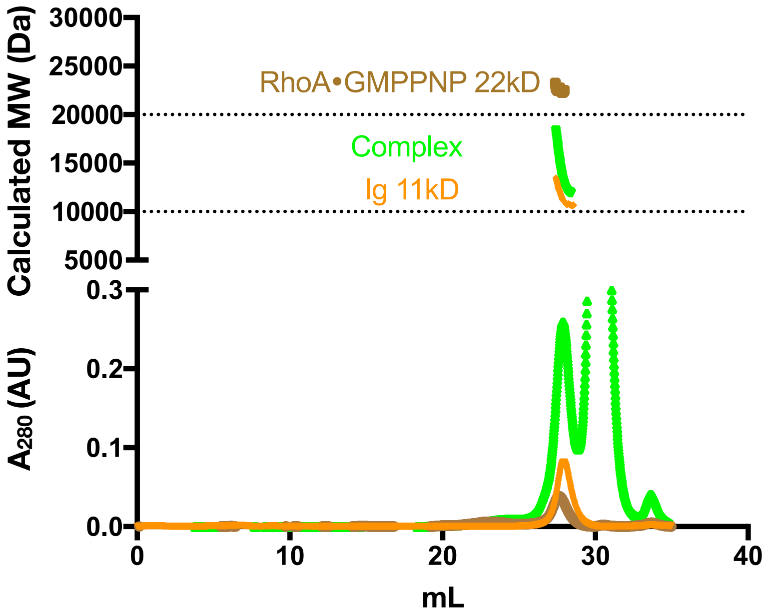


Figure 3.8 SEC-MALS data for the Ig·RhoA complex indicates lack of complex formation under tested conditions. Annotated UV traces show $A_{280}$ on the bottom portion of the Y axis and calculated molecular weight on the top part of the Y-axis. Dotted lines indicate where molecular weight lines for 20 kD and 10 kD would run. Complex formation attempted in the presence of GMP-PNP and control samples of Ig and RhoA·GMP-PNP run alone on the same curve. The green molecular weight trace decays along the peak from 22 to 11 kD indicating that RhoA and Ig co-elute rather than forming a complex.

A prior report suggested that the Ig domain interacts with the GTP-bound conformation of RhoA.[67] GTP-bound RhoA was prepared by incubating 90 μM RhoA with 10 mM EDTA and 5 mM GMP-PNP in the following buffer: 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, 3 mM $MgCl_2$, and 10 μM GMP-PNP. This exchange reaction was left overnight at 4 °C similar to that previously described.[73] The next morning, Ig domain was added to the exchange reaction to a final concentration of 180 μM, excess EDTA was removed using a desalting column and the complex was loaded onto SEC-MALS. This complex was then subjected to SEC-MALS analysis in comparison to either protein alone run in the same buffer using GMP-PNP. The two proteins run separately on SEC-MALS and do not appear to form a complex (Figure 3.8). These two proteins were then subjected to ITC in order to determine whether there was potential for weak affinity interactions. No interaction was seen using either the GDP or GMP-PNP bound states of RhoA, where

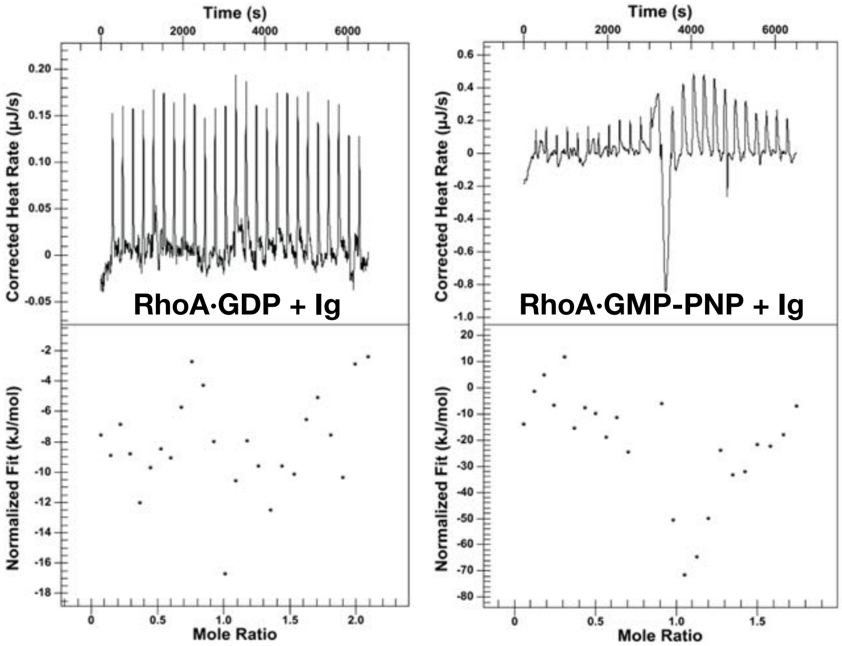


Figure 3.9 ITC analysis of the RhoA-Ig domain interaction shows no interaction between the two proteins in either nucleotide state of RhoA. Graphs of ΔH over time are shown on the top panel and fits of the heat per molar ratio are shown on the bottom panel. Left, ITC data using GDP loaded RhoA. Right, ITC data using GMP-PNP loaded RhoA. Experiments performed by Krishnapriya Chinnaswamy .

RhoA was added to the cell and excess Ig domain was injected using the syringe. (Figure 3.9)

### 3.3.5 Kinase Domain

Trio’s kinase domain contains a putative C-terminal autoregulatory sequence, and as such constructs were designed containing different truncations at the C-terminal end. This construct was expressed as an MBP fusion in Rosetta E. coli cells and purified using Ni-NTA resin and subjected to TEV cleavage. Further attempts at removing excess MBP via reverse IMAC, ion exchange, or SEC were unsuccessful. In each case, the small amount of kinase domain present co-migrated with MBP on these resins. These constructs were abandoned.

### 3.3.6 hTrioFL 293F Expression and Purification

hTrioFL (hTrio residues 61-3097) were mutagenized via inverse PCR to include an N-terminal 10x Histidine tag followed by a TEV protease cleavage site. This construct was prepared via a midiprep kit and transfected into HEK 293F cells grown in suspension at a 1:3 ratio of DNA:PEI (w/w). 50 ug of DNA was transfected for a 50 mL 293F culture. After a transfection time of 48 hours, cells were spun down and lysed using detergent, resuspended in lysis buffer containing 10 mM EDTA, and purified using Roche cOmplete His-tag resin. A panel of detergents and buffer systems were screened below their critical micellar concentrations (CMC) to see if they would help solubilize hTrioFL. However, none of the detergents tested produced a better result than using no detergent and a HEPES pH 8.0 based buffer system. The cOmplete resin's nickel coupling chemistry gives it the unique ability to stand high EDTA concentration. Sans EDTA, protein isolated in this manner ran as a smear on a gel of degradation products from a top band of estimated size 370 kD. After initial nickel purification in the presence of EDTA, samples did not seem to degrade further at 4 °C. Presence of eGFP-tagged Trio was determined by running a fluorescence scan on SDS-PAGE gels with an excitation

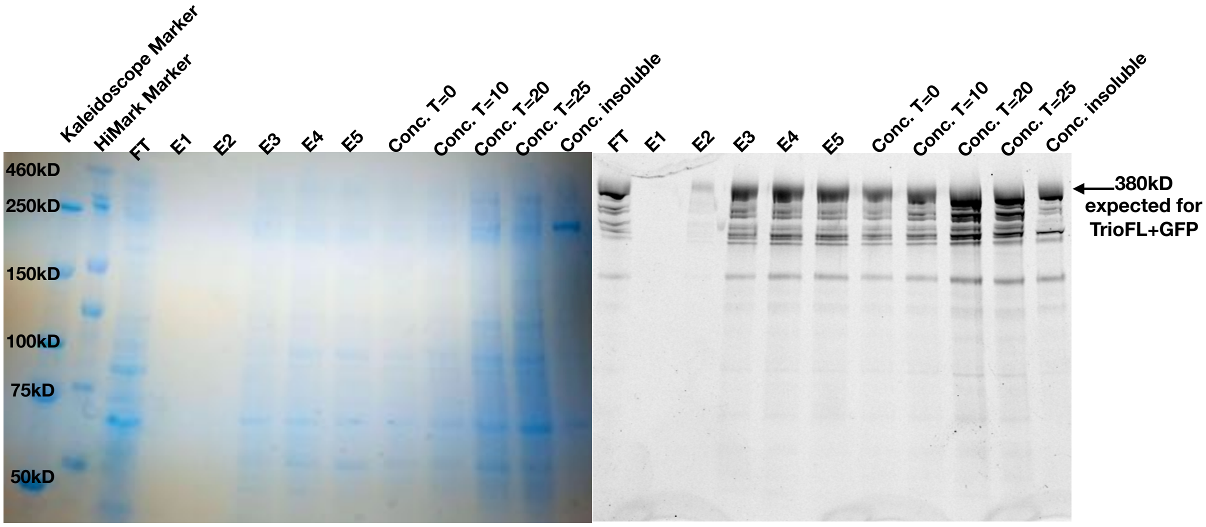


Figure 3.10 Ni purification of hTrioFL from 293F cells using Roche cOmplete resin shows that protein expresses and partially purifies on nickel resin. Left, 4-15 % Coomassie stained gradient gel showing his-tag purification and concentration using a 100 kD centrifugal concentrator. Right, Fluorescence scan of the same gel showing eGFP specific bands. hTrioFL at 380kD was difficult to visualize via Coomassie but could easily be seen by fluorescence scanning.

of 488 nm and emission of 532 nm (Figure 3.10). Figure Samples from nickel affinity purification were not very pure, and ion exchange, size exclusion, and a Rac1 affinity column were generated to try to purify hTrioFL further. Both anion exchange and size exclusion, attempted via FPLC, resulted in miniscule amounts of hTrioFL seen on resulting SDS-PAGE gels. Trio products are seen in the gel accompanying analytical

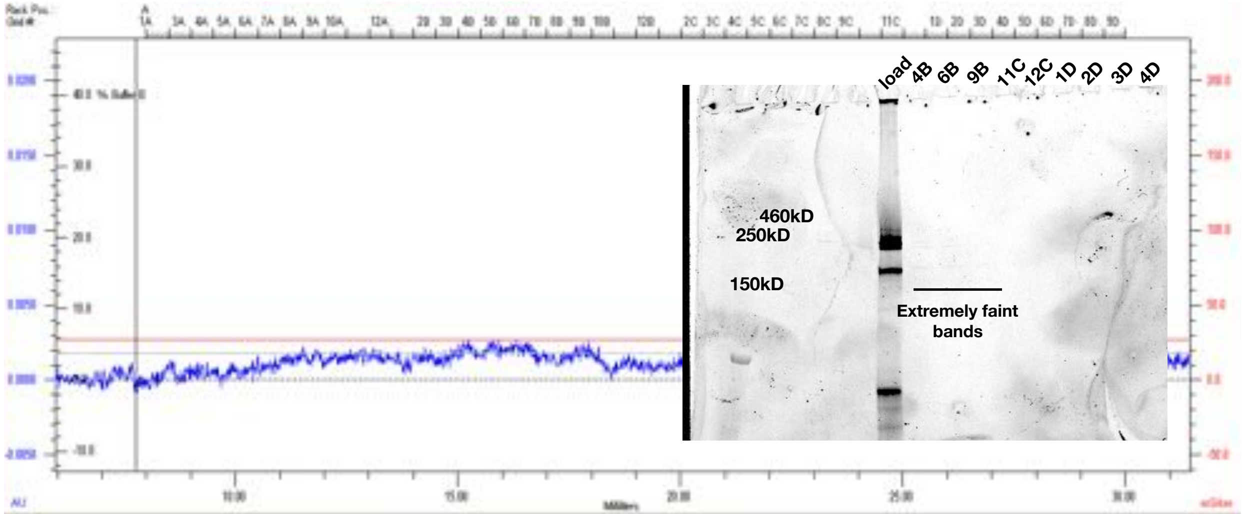


Figure 3.11 SEC purification of hTrioFL from 293F cells reveals little protein remaining after the run. The $A_{280}$ trace (blue line) of protein loaded onto a single 24 mL analytical Superose 6 column and fluorescence scan of 4-15 % gradient gel of fractions from SEC run. Extremely faint bands ~250 kD and ~175 kD are visible in fractions 4B-9B and likely represent Trio degradation products. Coomassie stained gel not shown as it is blank, no bands.

SEC, however they are extremely faint and run over about 4-5 mL on the sizing column, not indicative of a well-behaved protein (Figure 3.11). This behavior is similar to that seen for another full-length RhoGEF in our laboratory, P-Rex1.[150] It may be that the protein aggregated on one of the FPLC's inline filters or the column's pre-filter itself. Another possibility is that the protein interacts with the resin such that it does not run as a discrete peak. A Rac1 affinity column was generated similar to that previously described[150], however Rac1 did not couple to the column in my attempts.

### 3.3.7 Larger Trio Construct Screening

Initial constructs encoding hTrioFL, hTrioΔ2290, and dTrioFL were cloned into E. coli expression vectors for expression as MBP fusions in Rosetta cells, and GST fusion vectors for trial expression in insect and 293F cell lines. Baculovirus generation failed

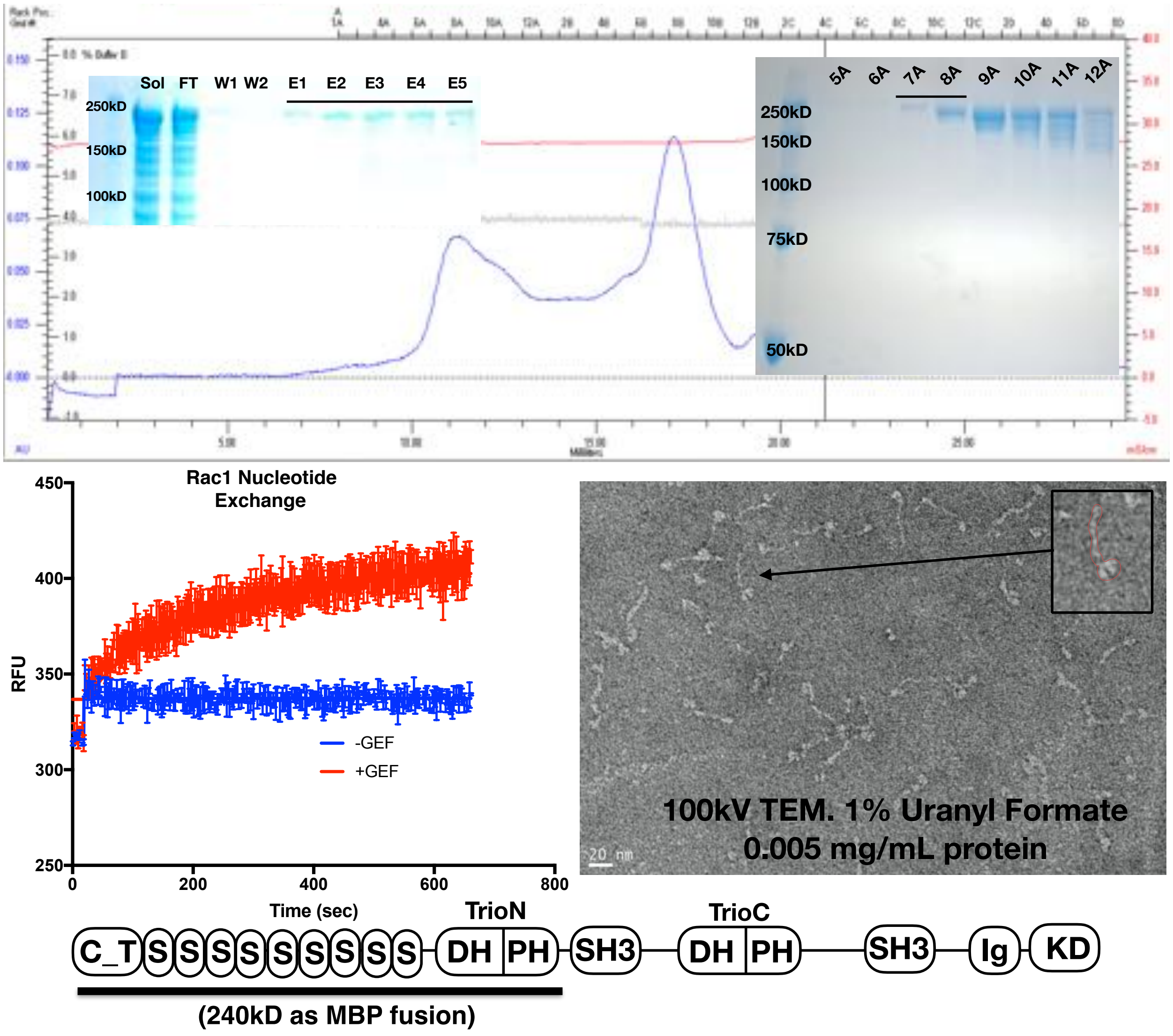


Figure 3.12 Purification and Characterization of hTrio Δ1594 shows folded, active protein which was characterized using negative stain EM. Top left, his-tag purification of Δ1594 with corresponding fractions run on Coomassie stained 4-15 % gradient gel. Top, $A_{280}$ trace (blue line) from analytical 24 mL Superose 6 run of Δ1594. Right, corresponding fractions from SEC are run on a Coomassie stained 4-15 % gradient gel. Middle left, control nucleotide exchange experiment showing that Δ1594 is capable of catalyzing Rac1exchange. Error shown is the standard deviation between three experimental replicates. Red shows condition including enzyme and blue shows Rac1 exchange rate alone. Middle right, negative stain TEM micrograph showing single particles of Δ1594 with a single particle highlighted in a box and outlined in red to show particle structure. Bottom, Trio domain architecture with black bar showing domain boundaries.

and was abandoned due to the long time required for virus generation. 293F expression of hTrioFL and hTrioΔ2290 were tested as 6xHis-GST fusions at a 48 hour time point in a 50 mL 293F culture using 50 µg:150 µg DNA/PEI mixture, but no protein was seen using a nickel pulldown. hTrioFL, hTrioΔ2290, and dTrioFL were also tried as MBP fusion constructs in Rosetta E. coli cells. These constructs yielded smears upon nickel purification using the Roche resin and EDTA, but the smears contained a sharp band at ~220 kD. The buildup of this 220 kD product may be due to degradation or early termination of translation. Based on approximate annotated Trio domain boundaries two further constructs were designed to mimic this 220 kD product, hTrioΔ1718 and hTrioΔ1594, both ending at the residue indicated. hTrioFL was also tried using coexpression of GroEL/ES and DNAJ/K but with no reduction in truncation products.

### 3.3.8 Characterization of hTrioΔ1594

hTrioΔ1718 and hTrioΔ1594 both expressed in Rosetta E. coli cells and were purified using Roche nickel resin and size exclusion chromatography. TEV cleavage was attempted but samples aggregated after TEV cleavage. Δ1594 was used for following experiments as it displayed a lesser degree of degradation than the longer construct. Δ1594 was assayed using the GEF assay and found to have nucleotide exchange activity on Rac1, indicating the construct was likely folded properly. This construct was then subjected to negative stain electron microscopy screening. Micrographs taken using 1 % uranyl formate stain using the Morgagni 100 kV TEM revealed an extended particle ~50 nm in length at maximum (Figure 3.12). The Δ1594 particle (Inset, Middle right of Figure 3.12) has a globular head, likely representing the MBP fusion partner and perhaps the CT domain which is also predicted to be globular. The globular head is linked to an extended tail region which is thin and somewhat linear. This tail region is likely the spectrin repeat domains of Trio, linear 3-helical bundles arranged in tandem, with the overall region similar in length (50 nm) with an estimate based on the solution structure of the spectrin repeat (PDB 1AJ3) with a length of ~60 Å (0.6 nm x 9 repeats = 54 nm).[151] These particles were not all found in an extended conformation, with the tail in various states from those shaped like an "S" to those more like a "C" where the tail curves back towards the head of the particle. This prevents further data processing as particles need to be in a highly similar conformation for averaging to work. In addition, linear particles are challenging to tackle using particle picking software as most are trained to search for globular particles. As such, the size of the picking radius would have to be inflated to take this into account. Because no binding partners of the CT/spectrin repeat region of Trio are known, and there was no obvious path forward to improving this sample for structural analysis, the sample was de-prioritized.

### 3.3.9 Characterization of hTrio 1284-2638

Since the N-terminus of Trio was extended and not globular on negative stain EM, we hypothesized that the remainder of the protein might prove easier to work with. A series of constructs starting with the TrioN GEF module (residue 1284) and ending at different portions of the kinase domain C-terminal extension (3053, 3063, 3072, 3097) were inserted into pMCSG9 and purification experiments were performed using Roche Ni resin. The “1284-3072” construct expressed well and was further pursued as it included all three of Trio’s enzymatic domains (TrioN/TrioC/Kinase) as well as the three protein-protein interaction domains (SH3_N/SH3_C/Ig). This construct was purified initially using IMAC, TEV cleavage, and SEC. “1284-3072” was active in the GEF experiment on both Rac1 and RhoA, indicating it was likely folded. Initial negative stain experiments showed globular particles (Figure 3.13). Finally, this construct showed complex formation with

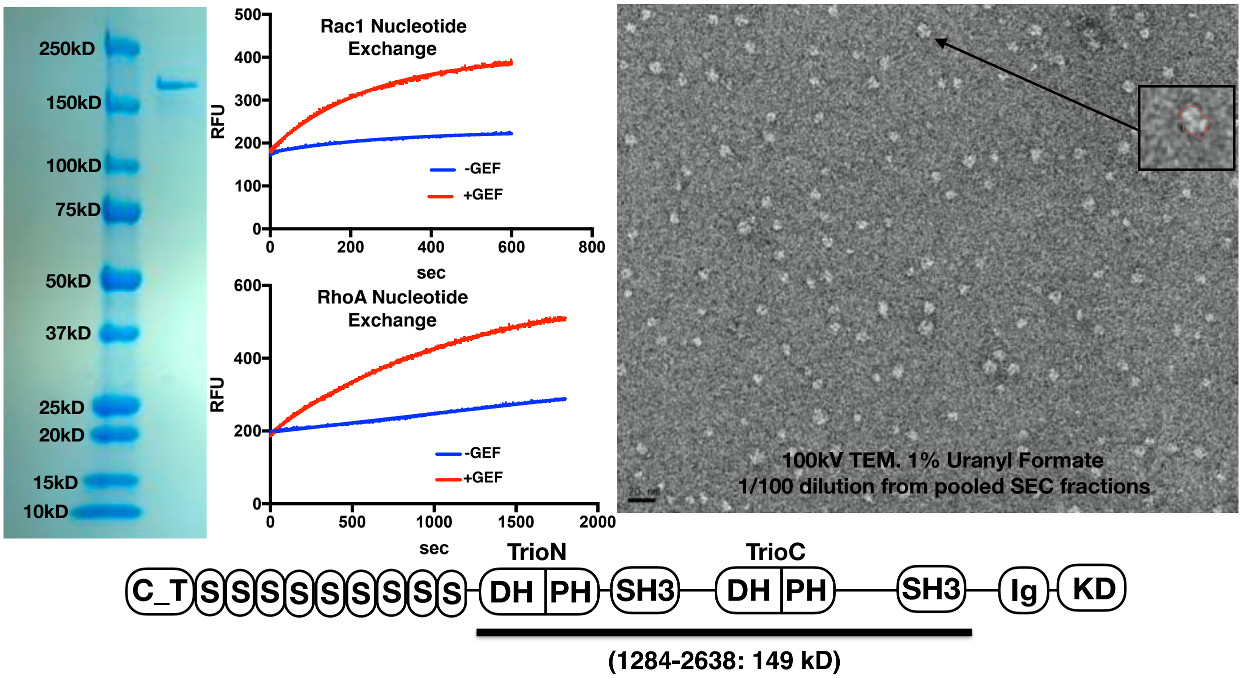


Figure 3.13 Characterization of hTrio1284-2638 shows folded, active protein and was characterized using negative stain EM. Left, Coomassie stained 4-15 % gradient gel showing pure protein. Middle, control Rac1 and RhoA nucleotide exchange experiments showing that the construct is active on both GTPases. Red, plus GEF. Blue, GTPase alone. Right, negative stain TEM micrograph showing single particles with a single particle highlighted in a box and outlined in red to show particle structure. Bottom, Trio domain architecture with black bar showing domain boundaries.

$G\alpha_q$, further indication it was likely folded properly. Following control experiments, the purification protocol was modified to include IMAC, TEV cleavage, anion exchange, and a final SEC step. Curiously, although this Trio sample was estimated to be 197 kD, a large amount flowed through a 100 kD cutoff spin concentrator, and a 50 kD concentrator had to be used to capture all protein. I faced this same problem with all Trio constructs with similar or smaller domain boundaries. Radiometric kinase assays were attempted with this construct on three model substrates, tubulin, myelin basic protein, and myosin light chain, with no detectable kinase activity. Mass spec coverage data from our collaborator (Sheng Li, UCSD) combined with DNA sequencing confirmed that this construct actually ended at residue 2638 (1284-2638), shortly after the C-terminus of SH3_C, due to an unintended nonsense mutation present in the construct. Therefore, the kinase reactions were run without the kinase domain present. SEC-MALS was attempted with this construct with the intent of determining whether the binding of $G\alpha_q$ results in a global conformational change in Trio. The 1284-2638 construct did not yield a usable scattering profile on SEC-MALS. The UV absorbance peak did not match with the elution volume of the peaks from the light scattering and refractive index

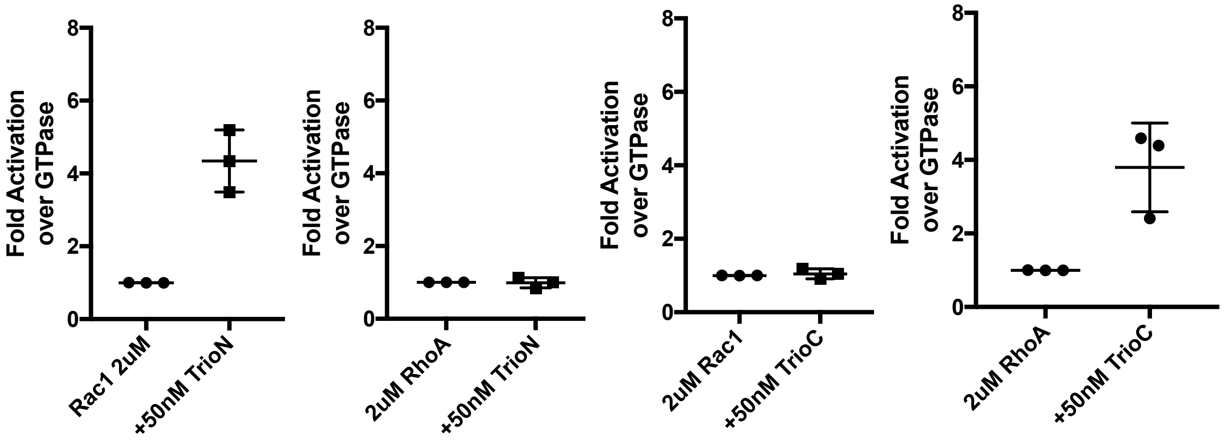


Figure 3.14 Control experiments showing that each Trio GEF module is selective for its given GTPase. Box and whisker plots of GEF assays normalized to rate of GTPase alone, N=3 experiments in at least duplicate. Fold activation is shown on the Y axis and conditions on the X axis. Error is represented as standard deviation.

detectors. Several columns were tried to see if this was a unique effect of the sample with the KW-804 column. Neither the Wyatt WTC-0305S or the Superose 6 columns yielded usable scattering profiles from this Trio construct. It is unclear what the problem with SEC-MALS is because all similar Trio constructs (starting with 1284) purified run as sharp peaks of expected elution volume on analytical 24 mL Superose 6 or Superdex 200 columns. Finally, the relative GEF activity of the 1284-2638 construct was tested

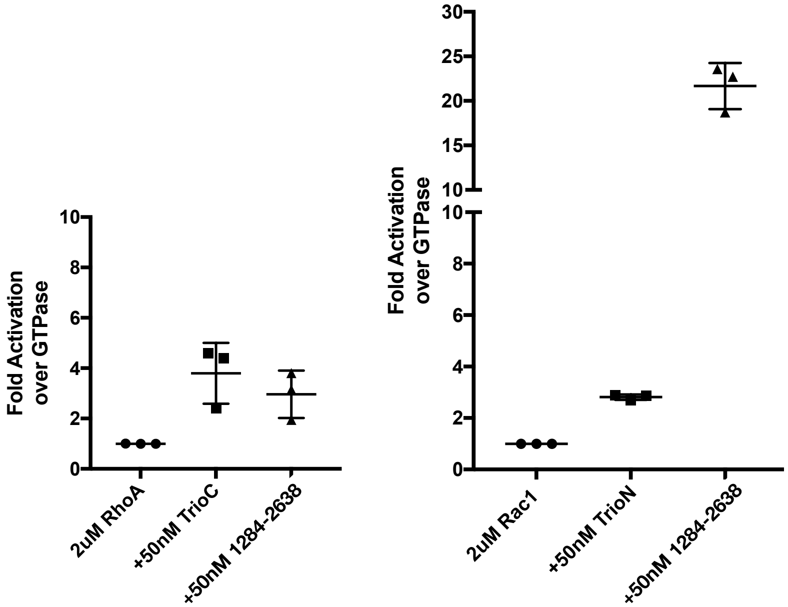


Figure 3.15 Basal nucleotide exchange activities of hTrio 1284-2638 reveals enhanced Rac1 exchanged rate. Left, RhoA nucleotide exchange assay. Right, Rac1 nucleotide exchange assay. Box and whisker plots of GEF assays normalized to rate of GTPase alone, N=3 experiments in at least duplicate. Fold activation is shown on the Y axis and conditions on the X axis. Error is represented as standard deviation

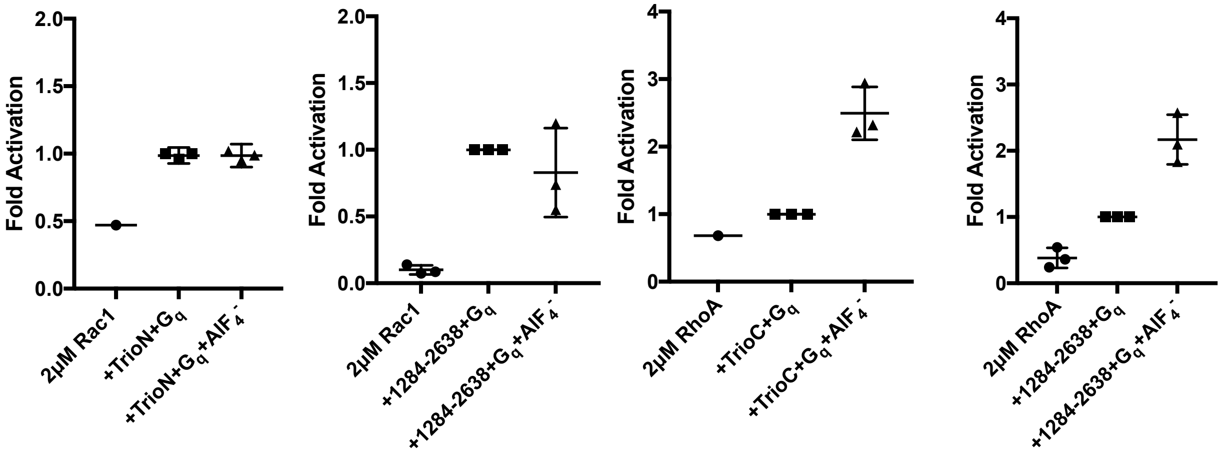


Figure 3.16 $G\alpha_q$ activation assays on hTrio 1284-2638 in comparison to each isolated GEF module shows that $G\alpha_q$ activates RhoA exchange but not Rac1 exchange. Left two plots depict Rac1 exchange experiments and right two plots depict RhoA exchange experiments. Box and whisker plots of GEF assays normalized to rate of GTPase + GEF + $G\alpha_q$, No $AlF_4^-$ , N=3 experiments in at least duplicate. Fold activation is shown on the Y axis and conditions on the X axis. Error is represented as standard deviation.

relative to each of TrioN and TrioC constructs alone. First controls were run to show TrioN and TrioC had activity on only their cognate substrates of about 3-4 fold rate enhancement at concentrations tested (Figure 3.14). 1284-2638 had similar GEF activity on RhoA as TrioC did, yet basal activity of Rac1 exchange by the larger construct was 3-fold higher than TrioN alone (3.15). RhoA exchange by 1284-2638 was activated by $G\alpha_q \cdot AlF_4^-$ to the same extent as TrioC, ~2.5-fold. Rac1 GEF activity of 1284-2638 was not affected by $G\alpha_q \cdot AlF_4^-$ addition (Figure 3.16).

### 3.3.10 Production of Truncation Constructs and GEF Activity Assays

In order to determine the source of the observed extra Rac1 activation, a series of truncation constructs were designed in order to determine the domain or region responsible for this activation. Each construct started with 1284 at the N-terminus of the TrioN GEF module and ended at 1718 (including through SH3_N), 1959 (through 200 residue conserved linker region), 2290 (through the TrioC GEF module), 2616 (through SH3_C), and 2780 (through Ig domain). All of these constructs expressed in Rosetta *E. coli*, although shorter constructs expressed to a much higher extent and did not include

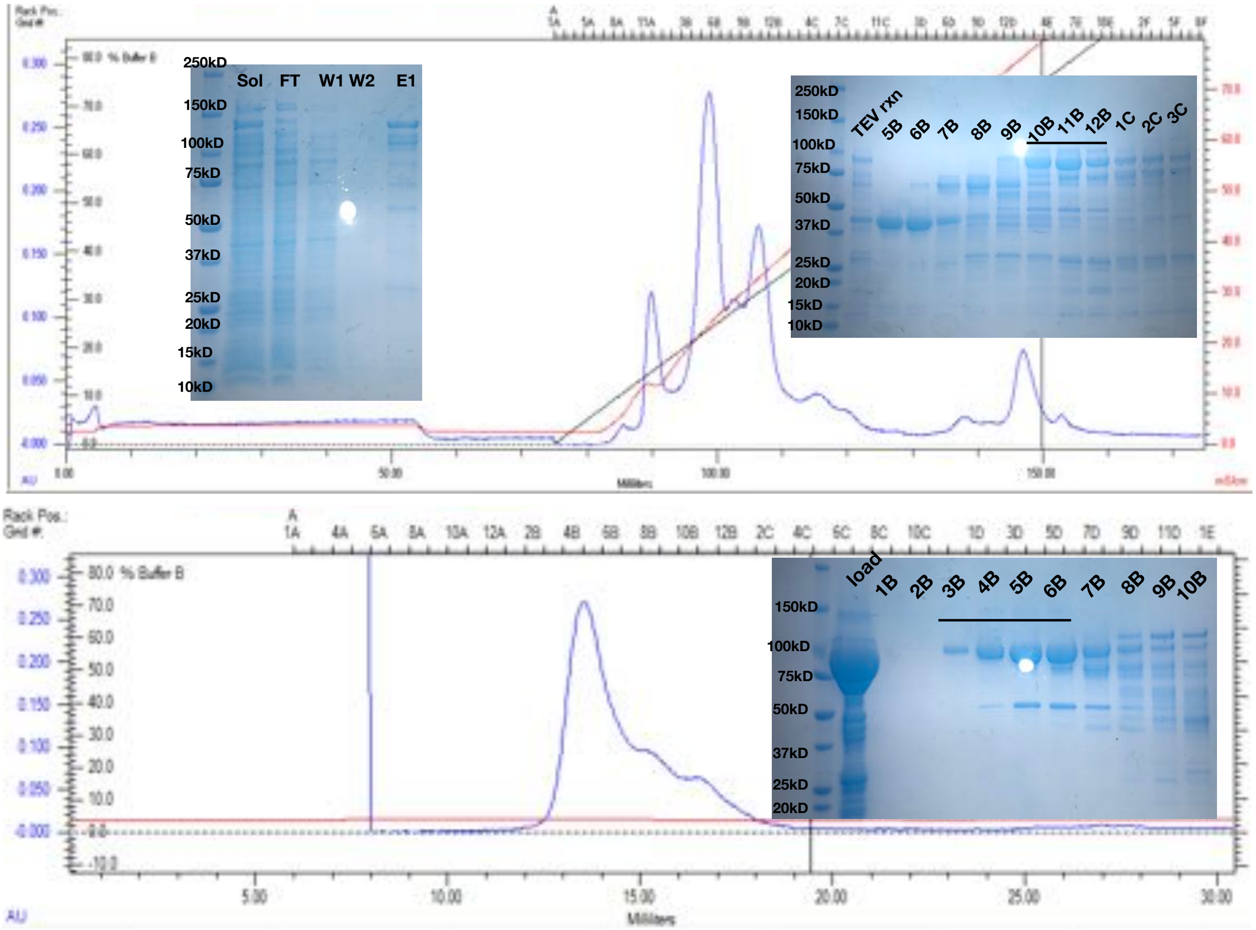


Figure 3.17 Purification of Trio Truncation constructs yields homogenous protein outside of a minor 50 kD contaminant. Purification of Trio 1284-1959 depicted above. Top left, his-tag purification with corresponding fractions run on Coomassie stained 4-15 % gradient gel. Top, $A_{280}$ trace (blue line) from HiTrap Q anion exchange column with corresponding fractions run on a Coomassie stained 4-15 % gradient gel. Bottom, analytical 24 mL Superdex 200 run. Bottom right, corresponding fractions from SEC are run on a Coomassie stained 4-15 % gradient gel. In each case, fractions taken to the next step or pooled to aliquot protein are indicated by horizontal black bars.

as many truncation products. All were purified using IMAC followed by TEV cleavage, anion exchange, and a final SEC step (Figure 3.17).

Each construct was tested for activity vs. both TrioN and TrioC in paired experiments. All of the constructs including the TrioC GEF domain had similar activity on RhoA to TrioC. These included all listed above except for 1718 and 1959, which did not include the TrioC GEF module and had no RhoA GEF activity. Constructs including 1959 and larger displayed an extra 3-6 fold exchange activity on Rac1 over TrioN, similar to the higher rate observed previously (Figure 3.18). We hypothesize that the 200-residue region between SH3_N and TrioC (1718-1959), which has >90% sequence identity conserved across representative species from the animal kingdom, likely contributes this extra activation, or perhaps stabilizes SH3_N so that the SH3 domain can enhance

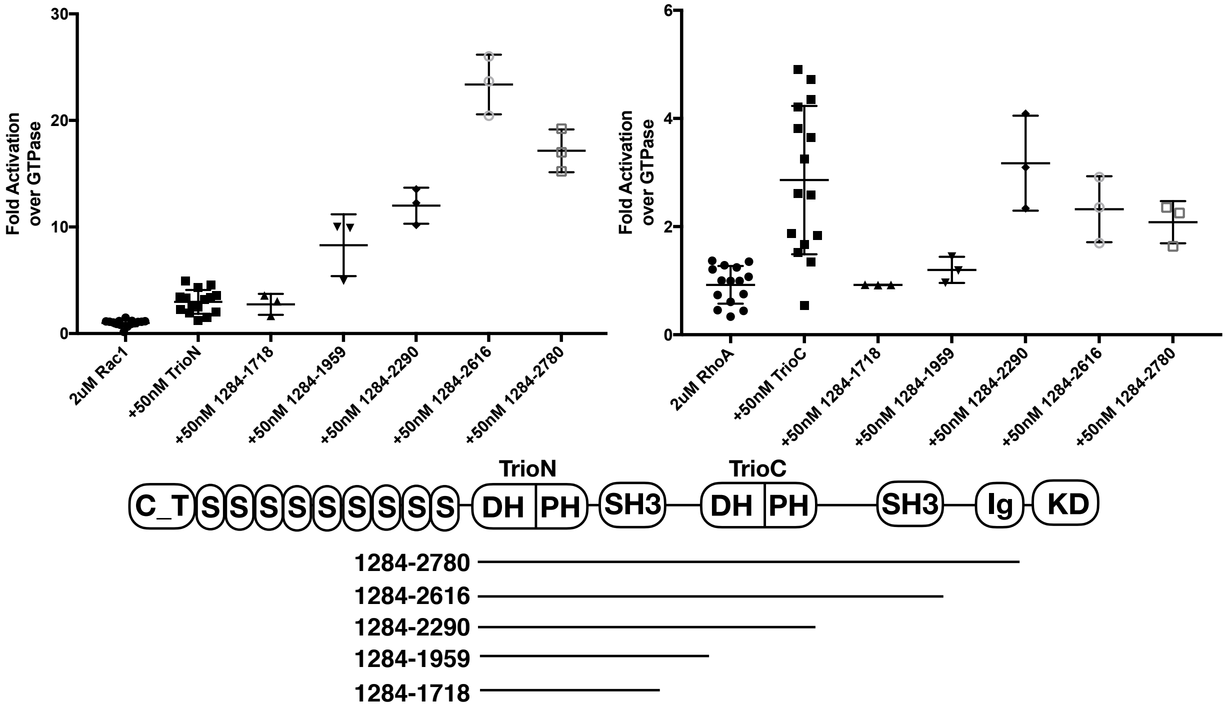


Figure 3.18 Summarized GEF activity assay data for Trio truncation constructs shows no effect of additional domains on RhoA activity, and enhancement of Rac1 activity by additional domains. Left plot depicts Rac1 exchange experiments and right plot depicts RhoA exchange experiments. Box and whisker plots of GEF assays normalized to rate of GTPase alone, N=3 experiments in at least duplicate. Fold activation is shown on the Y axis and conditions on the X axis. Error is represented as standard deviation. Below, Domain layout of hTrio showing approximate domain boundaries corresponding to each truncation construct.

activity on Rac1. I isolated the 1284-1959 construct in complex with Rac1 in order to determine the molecular basis of this activation (Figure 3.19).

### 3.3.11 Negative Stain Electron Microscopy of Trio Truncation Constructs

Negative stain EM was used to visualize the truncation constructs described in the previous section. All constructs were poorly behaved on micrographs and it was difficult to obtain micrographs with well spread particles with high contrast. All of the samples tested had low contrast relative to literature samples. Only the preliminary construct described, 1284-2638 in its less pure form was observed well-spread and discrete on micrographs. Averaging of this particle produced a globular image of about the right assumed size but yielded no discernible features (Figure 3.20). The 1284-1959·Rac1 complex, separated from excess Rac1 using SEC, yielded well dispersed, higher contrast, and better ordered particles than seen before (Figure 3.21). Averaging of ~15000 particles yielded globular averages of about the correct size, yet features were still not yet discernable. Improvements suggested by EM site staff included using darker, thicker stain, and using more careful focusing technique. The current negative stain data

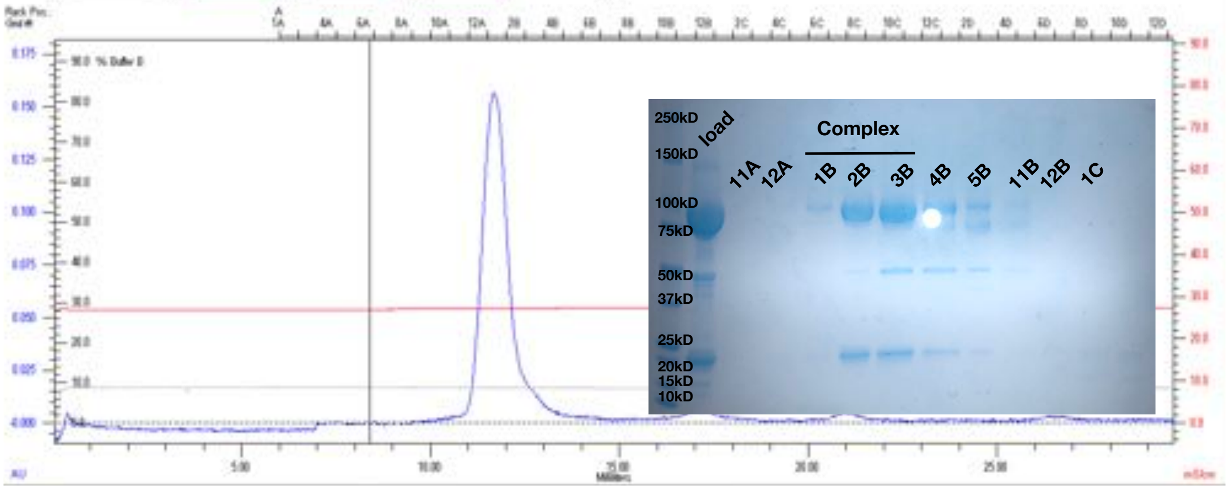
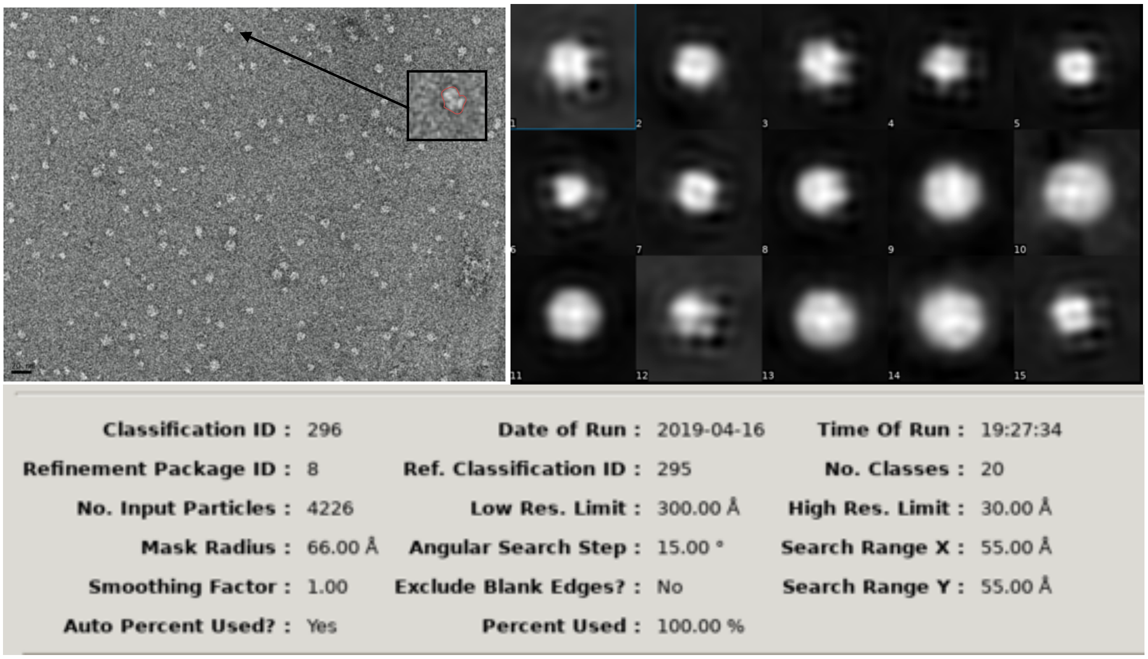


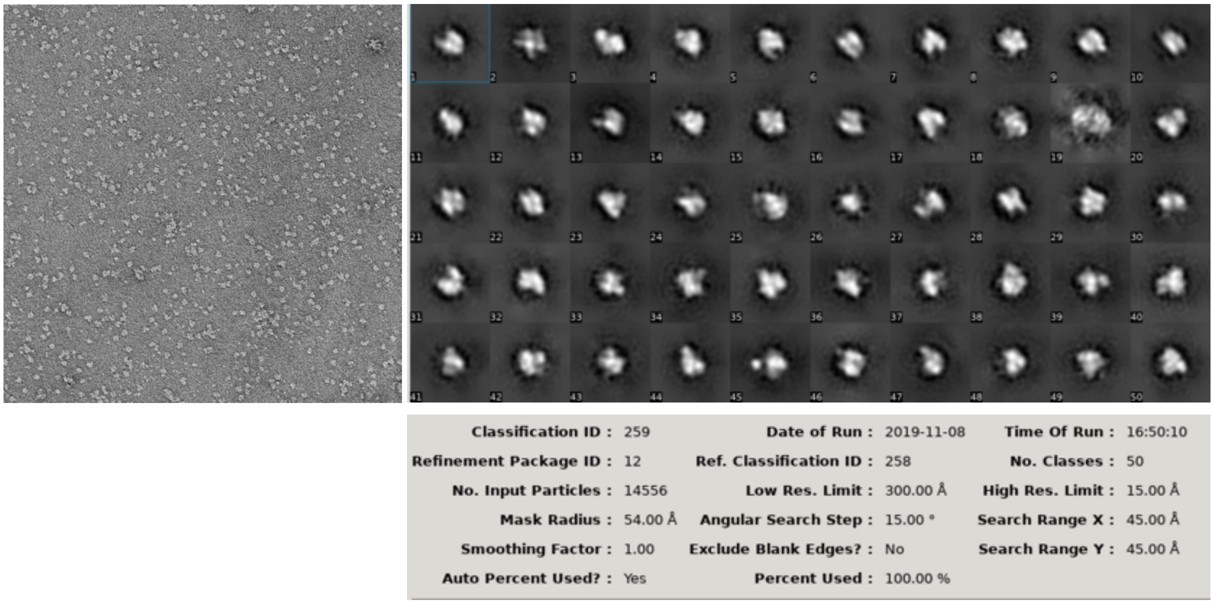
Figure 3.19 Isolation of the Trio 1284-1959·Rac1 complex via SEC yields homogenous complex with a minor contaminant at 50 kD. Analytical 24 mL Superdex 200 Increase run. Right, corresponding fractions from SEC are run on a Coomassie stained 4-15 % gradient gel. Fractions pooled to aliquot complex are indicated by horizontal black bars. A minor contaminant is present at ~50 kD.



Classification ID : 296 Date of Run : 2019-04-16 Time Of Run : 19:27:34
Refinement Package ID : 8 Ref. Classification ID : 295 No. Classes : 20
No. Input Particles : 4226 Low Res. Limit : 300.00 Å High Res. Limit : 30.00 Å
Mask Radius : 66.00 Å Angular Search Step : 15.00 ° Search Range X : 55.00 Å
Smoothing Factor : 1.00 Exclude Blank Edges? : No Search Range Y : 55.00 Å
Auto Percent Used? : Yes Percent Used : 100.00 %

Figure 3.20 Negative stain EM averaging on hTrio 1284-2638 using cisTEM shows a globular particle with no discernible features. Top left, raw micrograph of particles with one particle highlighted in the insert. Top right, the result of 2D classification with classification parameters listed on the bottom portion of figure.

collection procedure collects one image for focus, and then collects a series of desired exposures (5-10 in this case), which may have slightly different focus points. Therefore, the data could be improved by collecting frames one at a time, with one focus point



Classification ID : 259 Date of Run : 2019-11-08 Time Of Run : 16:50:10
Refinement Package ID : 12 Ref. Classification ID : 258 No. Classes : 50
No. Input Particles : 14556 Low Res. Limit : 300.00 Å High Res. Limit : 15.00 Å
Mask Radius : 54.00 Å Angular Search Step : 15.00 ° Search Range X : 45.00 Å
Smoothing Factor : 1.00 Exclude Blank Edges? : No Search Range Y : 45.00 Å
Auto Percent Used? : Yes Percent Used : 100.00 %

Figure 3.21 Negative stain EM averaging on Trio 1284-1959·Rac1 complex using cisTEM shows an ordered particle. Top left, raw micrograph of particles with one particle highlighted in the insert. Top right, the result of 2D classification with classification parameters listed on the bottom portion of figure.

taken near the desired one exposure frame, although this method is more time-consuming. The 1284-1959·Rac1 complex sample itself may not stain well or respond poorly to the stain environment and is also at the lower end of molecular weight range for samples imaged by negative stain at 97 kD.

### 3.3.12 Cryogenic Electron Microscopy on Trio Truncation Constructs

First, Trio 1284-2638 was subjected to cryoEM trials. Using concentrations of 3-4 μM protein with 0.08 mM dodecylmaltoside (DDM) produced visible particles in both carbon and gold grids. However, it remained very difficult to see what a discrete particle looked like even when good ice was found, and thus this sample was abandoned. Perhaps the protein samples a lot of conformations in solution and is thus low contrast on an EM grid, similar to that seen in negative stain (Figure 3.22). A complex of 1284-1959 with Rac1 was subjected to cryoEM trials in carbon grids using 2-4 μM protein with 0.08 mM DDM. Particles were well-resolved in this instance indicating that complex formation leads to the formation of a more discrete Trio particle. Preliminary 2D averages

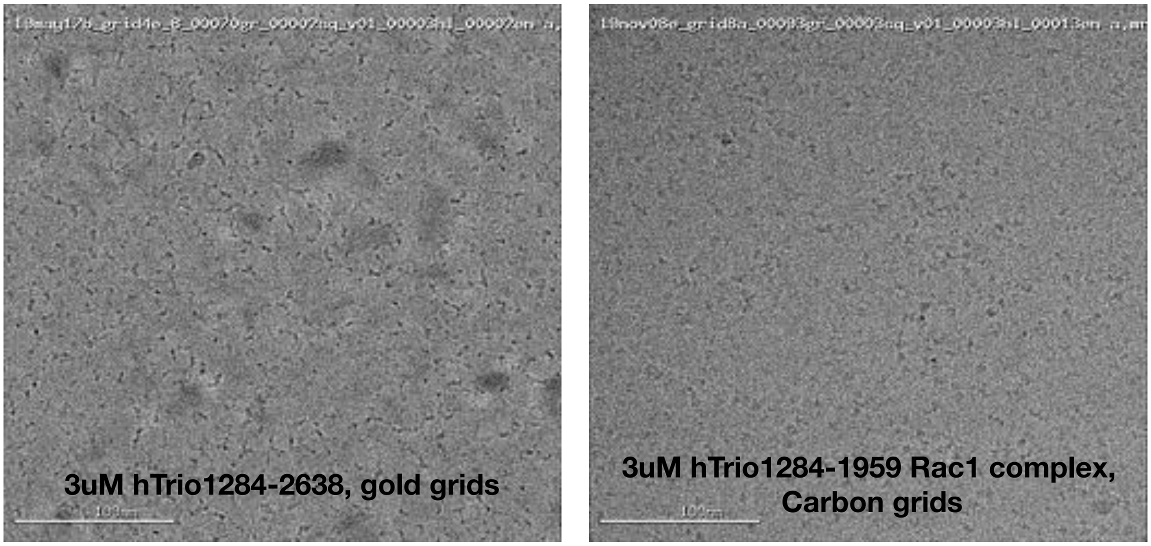
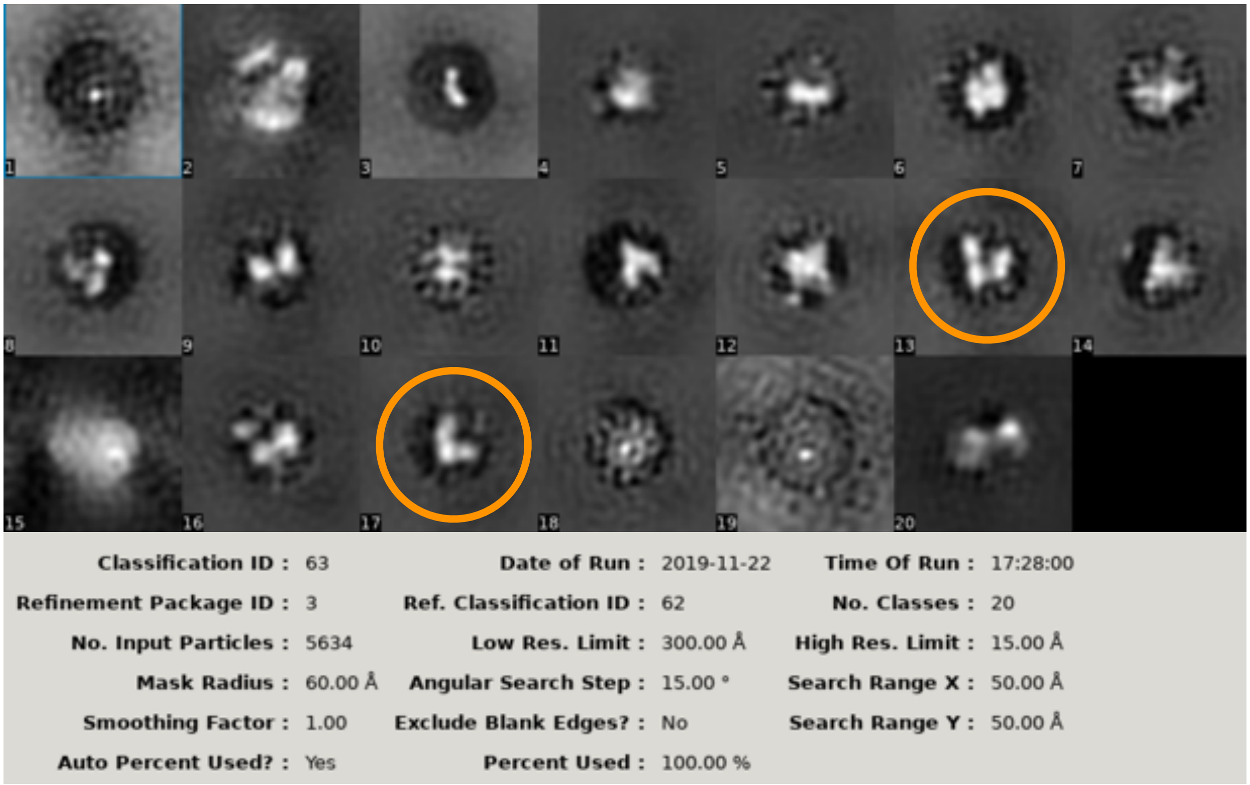


Figure 3.22 CryoEM screening of larger Trio constructs with raw micrographs shown reveals that the Rac1 complex is more suited to averaging and data processing. Left, raw micrograph of Trio 1284-2638 shows dark particles, but they do not have a discrete shape. Right, raw micrograph of the 1284-1959·Rac1 complex shows discrete particles that can be used for averaging. Data collected with Jennifer Cash, Ph.D.

generated using cisTEM are presented in Figure 3.23. The averages represented in classes 13 and 17 (highlighted in orange, Figure 3.23) show 4 discrete domains (accounting for DH, PH, SH3_N in Trio, and Rac1) and this sample thus is promising to pursue to high resolution. Averages in classes 6, 9, and 11 may represent different orientations of the particle. Due to limited microscope time, I was unable to collect a large dataset and these averages are the result of ~5000 particles. More data needs to be collected before this sample can yield higher resolution structural information.



Classification ID : 63
Date of Run : 2019-11-22
Time Of Run : 17:28:00
Refinement Package ID : 3
Ref. Classification ID : 62
No. Classes : 20
No. Input Particles : 5634
Low Res. Limit : 300.00 Å
High Res. Limit : 15.00 Å
Mask Radius : 60.00 Å
Angular Search Step : 15.00 °
Search Range X : 50.00 Å
Smoothing Factor : 1.00
Exclude Blank Edges? : No
Search Range Y : 50.00 Å
Auto Percent Used? : Yes
Percent Used : 100.00 %

Figure 3.23 Preliminary class averages of the 1284-1959·Rac1 complex processed using cisTEM reveals an ordered particle with four observed domains. Bottom, 2D averaging parameters. Classes that show 4 discrete domains are highlighted using orange circles. Data collected and processed with Jennifer Cash, Ph.D.

### 3.3.13 Kinase Assays using hTrio 1284-3053

I also was able to produce a construct encoding residues 1284-3053, confirmed both by sequencing and via Coomassie gel, where it runs larger than 1284-2780. This construct was used to perform the ADP-GLO assay on 3 different kinase substrates in an attempt to recover kinase activity. The ADP-GLO assay was checked using concentrations of ADP to show that it did have a measurable signal/background as low as 100 nM ADP, to show the reagents in the kit were working. Although TBK1 was able to phosphorylate all control substrates, Trio showed no activity against tubulin, myosin light chain, or myelin basic protein, even when the kinase reaction was allowed to proceed overnight.

### 3.4 Discussion:

A variety of structure/function studies on Trio were conducted in this section. Individual Trio domains were expressed as MBP fusions in E. coli to varying success. The CT domain displays a recurring aggregation problem that needs to be tackled before this domain can be crystallized or studied further. The addition of Triton X-100 throughout purification seemed to shift the protein towards monomeric state, yet the high $A_{280}$ of Triton X-100 complicates interpretation of results as our FPLC systems only have the capability to read at 280 nm. This experiment should be retried using a detergent system that does not have $A_{280}$. Both SH3 domains and the Ig domain were successfully expressed and purified. A potential interaction between the Ig domain and RhoA was explored, and the two proteins were not found to interact on SEC-MALS or ITC. The literature shows this interaction between RhoA and Ig domains requires prenylated RhoA, which was not used in my experiment. *E. coli* lacks the prenyltransferase required to lipidate small GTPases. It may be that the Ig domain interacts with some complicated epitope including this lipid modification including some protein contacts on RhoA. The kinase domain was expressed but did not behave well in solution and was thus deprioritized. Overall, the study of these domains is hindered by a lack of knowledge of binding partners or ligands that they bind to. Even if we were able to obtain diffracting crystals, these domains likely would look like any number of homologs available in the PDB.

Larger constructs of Trio were explored. hTrioFL expressed well as an eGFP fusion in 293F cells but could not be purified past the initial Ni-resin step. A C-terminal tag would make the biggest difference in this purification, and further downstream purification methods will need to be optimized in order to produce soluble, well behaved protein. A future protein production scheme could include his-tag purification, then a secondary purification via a C-terminal tag to ensure only the full-length protein was obtained, and one or two polishing steps done via affinity columns using Rac1 and RhoA.

Constructs encoding the N-terminal half of hTrio were purified from E. coli and behaved reasonably well, although negative stain data show that this protein adopts many conformations. A binding partner or other method of conformational restriction will be needed in order to produce a protein that could yield a high-resolution structure of this region. Regardless, new information was obtained: the spectrin repeat region of Trio is extended in conformation. The extended structure of this region likely plays a role in its function, as this region is extremely conserved throughout evolution.

Next, a series of constructs exploring Trio's C-terminal region, with N-termini coinciding with the TrioN GEF module were produced and profiled using EM and GEF assays. Kinase activity was evaluated and not detected, although the number of substrates used was not exhaustive. The C-terminal region of Trio was found to be more globular in nature than the N-terminus, yet likely flexible in solution, as negative stain EM was not able to yield domain-level 2D information. Through GEF assays, several key pieces of information were obtained. Larger constructs of Trio (larger than 1284-1959) have up to a 10-fold higher rate of Rac1 nucleotide exchange than that of TrioN alone. Thus, in the context of full-length Trio, additional domains may contact this GTPase or the DH/PH core to further assist nucleotide exchange. In addition, I showed that TrioC behaves and is regulated as if other domains are not present because basal and $G\alpha_q$-enhanced RhoA nucleotide exchange in context of larger constructs is similar to that of TrioC. Finally, $G\alpha_q$ addition does not affect Rac1 nucleotide exchange, opposing the hypothesis that $G\alpha_q$ induces global conformational change to activate Trio. Efforts to study $G\alpha_q$ induced conformational change on Trio using SEC-MALS were hindered by technical difficulties.

## 3.5 Conclusion and Future Directions

Structure/function studies on the understudied parts of Trio were performed in order to determine the structure of these domains and how they may contribute to nucleotide exchange. I have developed several protein constructs for isolated Trio domains and larger constructs of Trio that in the future can be used to study this protein. I determined that the region immediately C-terminal to the TrioN GEF module enhances Rac1 nucleotide exchange. The atomic details of this rate enhancement will be determined by the solution of the 1294-1959·Rac1 complex via CryoEM, a structure for which I have laid the groundwork for by producing preliminary 2D averages. Perhaps the interface which assists in Rac1 rate enhancement can be targeted in the future by a small molecule in order to reduce Trio-dependent Rac1 GEF activity likely important for cell migration. Biochemical experiments hint that $G\alpha_q$ regulates TrioC independently of the rest of the protein. However, this is done in a soluble system and other factors could be in play when the membrane is included, along with GTPase lipid modifications. Whether $G\alpha_q$ induces conformational change in Trio will be explored in the future using CryoEM and HDX-MS to study the Trio-G$\alpha_q$ complex which can be readily formed. Another loose end that will be followed in the future is the kinase domain. If characterized, Trio's kinase domain represents the most canonically druggable portion of the molecule. However, its substrate scope and biological role remain unknown. Although I was unable to demonstrate kinase activity by Trio's kinase domain, it seems unlikely that this domain is inactive. It contains almost all highly conserved serine/threonine kinase domain residues [58], and the presence of this domain is conserved throughout much of the animal kingdom. It may be that this domain requires eukaryotic folding or post-translational machinery in order to be produced properly. Future attempts should produce this protein in isolation and attempt to reconstitute activity. Alternatively, a cell-based chemical biology approach could be used to identify Trio KD substrates in cells. [152] This work has laid foundation for the study of uncharacterized portions of Trio and provides groundwork for CryoEM structures of larger constructs of Trio in complex with binding partners.

# Chapter 4 **Conclusions and Future Directions**

## 4.1 Conclusions

This dissertation has primarily covered the structure and function of TrioC, with the primary result of defining the autoinhibitory mechanism of this GEF module. The TrioC DH/PH GEF module was expressed, purified, and crystallized and the structure compared to that of G$\alpha_q$-bound p63RhoGEF, a close homolog. Upon comparison, the conformation of the α6-αN linker region was hypothesized to be the key determinant of the autoinhibited conformation. We attempted to break this autoinhibition by changing residues in the interface formed between DH and PH domains. Variation of residues in this region were initially difficult to interpret, as single point variation did not activate the protein over WT. However, the cancer-sourced TrioC variants in the α6-αN region, most notably R2150W, led to hyperactivation of TrioC. By inserting steric bulk into the autoinhibitory interface, we hypothesized that the contacts were not able to form efficiently, and the substrate binding face of the DH domain was free to interact with RhoA. Truncations of the PH domain were also able to accomplish this same task, perhaps by reducing order in α6-αN. This hypothesis was supported by HDX-MS data, which showed that the two helical regions contacting each other to form the autoinhibitory interface, α3 and α6-αN, were more solvent exposed in the active variant R2150W, whereas the rest of the protein surface remained relatively unchanged. The same was true for the truncation variant 2152Δ. In addition, the R2150W variant was not further activated by G$\alpha_q$, suggesting we have generated a variant capable of fully escaping autoinhibition by the PH domain. We tested both our point variations and truncations in the context of full-length Trio an overexpression model and found that the model for PH domain mediated autoinhibition seems to hold. We have proposed a four-phase conformational equilibrium model for how TrioC is regulated (Figure 4.1), wherein

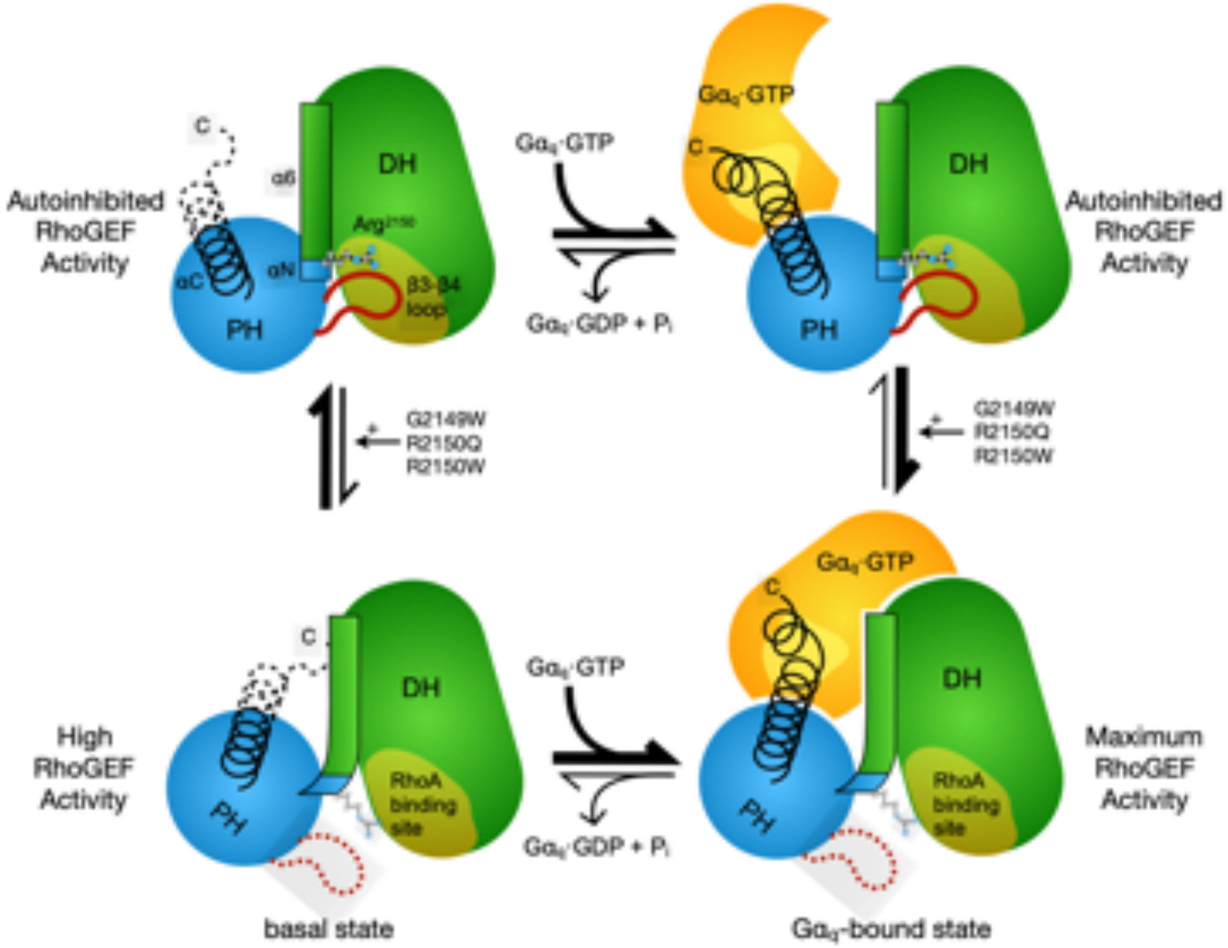


Figure 4.1 Model for TrioC activation. TrioC exists in a conformational equilibrium between inactive and active states that can be biased towards the active state by either active G$\alpha_q$ or mutations in the DH-PH interface. The thicker half-arrows represent the favored direction in each equilibrium. The DH domain is represented by a green oval with its RhoA binding site highlighted in yellow. α6 is shown as a green rectangle that forms a continuous helix with αN, represented as a blue rectangle. The PH domain is represented as a blue circle, with its C-terminal αC helix as a black helix. Arg2150 is shown as a ball-and-stick model, and the β3-β4 loop as a red cartoon loop. G$\alpha_q$ is shown as a gold shape with its effector binding region in light yellow. Disorder is indicated with dashed lines and a blurring of Arg2150. The autoinhibited conformation in the top left quadrant is represented by PDB 6D8Z and the maximum activity state is represented by PDB 2RGN, where G$\alpha_q$ binds to both the PH and DH domains and constrains them in a more open configuration that features a bent α6-αN helix.

point mutation can lead to a partial activation in the module (bottom left, Figure 4.1) and $G\alpha_q$ binding to both DH and PH domains leads to full activation (bottom right, Figure 4.1). The TRIPα peptide was explored as a possible lead to inhibit Trio and was found to be both active and specific for TrioC even over its close homolog p63RhoGEF. Regrettably, this peptide proved difficult to work with. If the stability of TRIPα could be improved, likely through removal of hydrophobics and cysteines by mutagenesis, it could be a valuable tool to use to discover new leads. Screening via FP and looking for displacement of labeled TRIPα would be a simple and effective manner to find compounds binding in the same region, and thus likely acting by the same mechanism. In addition, structural work using this peptide could define the molecular basis for how it induces RhoGEF inhibition, which remains unknown. The most likely mode of action would be binding to the face of the DH domain to block GTPase access. An inhibitor directed against TrioC could also act by stabilizing the closed interface and decreasing affinity for RhoA. If future crystallography attempts are not successful, this problem could be tackled through HDX-MS and site-directed mutagenesis. Since HDX-MS has the resolution to identify ~5 AA peptides which would presumably be protected from deuterium exchange by interaction with TRIPα, these regions could be mutagenized and resulting TRIPα variants could be profiled using DSF and the GEF activity assay.

RhoGEF inhibition has been attempted for roughly two decades and no significant progress has been made. Most discovered molecules are directed against TrioN and LARG and have overtly toxic or reactive groups attached to them, and others have not progressed since the initial report.[153,154] The best success story in GEF inhibition is the natural product Brefeldin A, an inhibitor for the ArfGEF domain family Sec7.[155,156]

Brefeldin A is a complex heterocyclic natural product, and as such I hypothesize that an effective RhoGEF inhibitor will likely be a large complex molecule capable of mimicking the relatively flat yet nuanced surface in the RhoGEF·GTPase interaction. Unfortunately, Brefeldin A is not bioavailable and as such has failed to progress through the development pipeline. I began optimization of a few methods in order to run a high-throughput screening campaign against the TrioC GEF module. None of these presented promising screening statistics in early optimization attempts and were thus deprioritized. If TRIPα could be improved such that exhibits reproducible purification and activity, a displacement screen against TRIPα would be the strongest screening strategy to use as it is already selective for TrioC. Since that is currently not the case, I propose the best method to discover a TrioC inhibitor would be to screen using a complex scaffold such as a peptide, aptamer, or antibody, then try to understand the determinants of the interaction and ultimately convert this large molecule into a smaller molecule. A focus on natural products could discover a small molecule with enough complexity to bind specifically to the TrioC GEF module (or TrioN, for that matter).

With the path to direct inhibition of GEF activity remaining unclear, basic science structure/function work on Trio is warranted to understand how other domains contribute to GEF activity and if they could instead be targeted for therapeutic benefit. The remainder of my dissertation consisted of structure/function work on isolated domains and larger fragments of Trio. Full length Trio purification was not successful; after loading detectable amounts of pEGFP-TrioFL into either size exclusion or anion exchange resin, the protein was not detectable by eGFP fluorescence in the elution fractions in each case, similar to that seen for another RhoGEF in our laboratory, P-Rex1. The protein may be aggregating on the column or in the FPLC system. hTrioFL purification could be improved using a C-terminal tag, as the protein degrades even when overexpressed in 293F cells. The Trio cDNA is difficult to work with due to length and high GC content and I was unable to clone such a tag. Affinity steps using sequential Rac1 and RhoA columns could also help isolate homogenous protein. If my preliminary electron microscopy imaging on Trio truncations is any guide, this full-length sample would likely be too flexible to conduct a high-resolution 3D reconstruction, so a GTPase complex would be most practical to pursue. For solving a structure of full-length Trio, multiple binding partners may need to be employed, such as the GTPases Rac1, RhoA, and G$\alpha_q$. It is also possible that the full-length protein purified from mammalian cells adopts a wrapped up and ordered conformation due to PTM or folding assistance uniquely provided by this system or due to the presence of all domains.

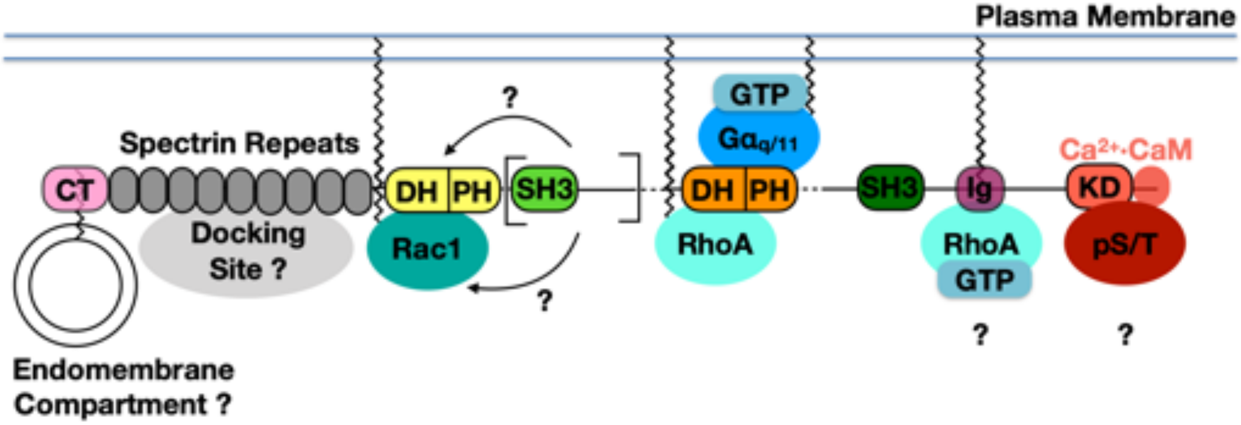


Figure 4.2 Model summarizing hypotheses for Trio function. From left to right, the CT domain may interact with specific lipids in endomembrane compartments and cooperate with the spectrin region to spatially regulate Rac1 signaling. The spectrin repeat region may serve as a docking site for other signaling proteins or complexes. The SH3-N and region following (bracketed) is hypothesized to interact with Rac1 or the TrioN GEF module to increase nucleotide exchange. Dotted lines show unstructured regions to the N- and C terminus of the TrioC DH/PH module. The Ig domain may interact with a complex epitope including the prenyl modification on RhoA and protein-protein contacts. Ca2+·CaM binding to the C-terminal extension of the KD may be required for kinase activity. The plasma membrane is shown at the top as two horizontal lines. Trio domains are shown as ovals on a line representing the protein backbone. Other proteins are shown as ovals. Serrated lines represent lipid modifications and dotted lines represent putative unstructured regions. The CT and Ig domains are shown as transparent ovals with lipids inside of them to indicate they may interact with lipids. Question marks indicate hypothesized steps or mechanisms.

Figure 4.2 accompanies the following paragraphs summarizing what has been learned from studying portions of Trio outside of the GEF modules. Nonenzymatic domains of Trio, including CT, both SH3 domains, and the Ig domain were purified and subjected to crystallization trials, to no major success. A preliminary crystal form for SH3_C was obtained and can be optimized in the future. It will likely be more relevant to identify binding partners or ligands of these domains before further structural studies are warranted, as apo crystal structures will not be very informative as to their underlying function. A possible interaction between the Ig domain of Trio and RhoA·GTP was explored via SEC-MALS and ITC, however no interaction was found to occur. This interaction was previously found to depend on the prenylation status of RhoA, and as such this experiment can be repeated in the future with lipidated RhoA·GTP. [67]

The N-terminal half of Trio was purified from E. coli and found active in Rac1 nucleotide exchange. This sample was characterized by negative stain EM and consisted primarily of an extended region likely corresponding to Trio's nine spectrin repeats. Trio's N-terminal region thus consists of a putative lipid binding domain followed by these spectrin repeats before the TrioN GEF module. Most particles adopt an extended conformation and thus this region may be responsible for maintaining linear distance between the CT domain and the TrioN GEF module, perhaps to spatially regulate Rac1 nucleotide exchange around the presence of a specific type of membrane lipid. Alternatively, the extended spectrin region could provide a docking surface for other proteins or complexes to assemble and perhaps be phosphorylated by Trio's kinase domain. As this region of Trio is implicated in neural development, I posit that this specific domain architecture is key to this complex, spatially regulated process. [44]

Trio's C-terminal region, including its 2 GEF modules and the SH3_N, SH3_C, and Ig domains, seems to form a more globular structure as seen by electron microscopy, although averaging indicates that this structure is likely flexible in solution, with formation of the Rac1 complex leading to a more ordered particle. As seen in my biochemical experiments, this construct has elevated Rac1 activity over TrioN alone, and similar RhoA exchange activity, both basal and G$\alpha_q$-stimulated. Through a series of truncation constructs, the determinant of additional Rac1 GEF activity was localized to the region C-terminal of the SH3_N domain, within residues 1718-1959. This region may wrap around to productively contact another surface of Rac1 or the GEF module itself in order to increase affinity or turnover rate for this GTPase. The Trio (1284-1959)-Rac1 complex forms a stable structure as determined by EM and as such would be a good target to try and capture more of Trio.

Using a construct encoding from TrioN through the putative kinase domain proper (1284-3053), I ran kinase assays using the ADP-GLO protocol and found that Trio was inactive against 3 model substrates. It is possible that we have not identified a substrate that Trio will phosphorylate, or that this kinase domain needs to be produced in a eukaryotic system in order to be fully functional. Trio's kinase domain contains a putative calmodulin regulatory sequence C-terminal to the kinase domain, so perhaps $Ca^{2+}\cdot$CaM is required to bind to this region to change KD conformation to promote full kinase activity. The constructs I tested did not contain this regulatory region.

### 4.2 Future Directions

Trio has been defined as the focal point for the malevolent signaling leading to the development of uveal melanoma. This signaling is thought to depend on Trio's dual GEF activities, with prior experiments showing that elimination of either of Trio's GEF activities reduces cancerous signaling.[37] In addition, Trio has been found to be upregulated in several types of cancer including glioblastoma, carcinomas, and leukemias, and is also highly mutated in cancer. [37,44] Important biochemical events remain to be defined in the regulation of Trio, starting with how this highly complex molecule is translocated to the PM where its GTPase substrates are. Collaborators have shown that Trio is translocated to the membrane upon activation of $G\alpha_q$-coupled receptors. $G\alpha_q$ directly binds to Trio and is palmitoylated and could thus translocate the complex to the membrane, but that has not been directly shown. The localization question becomes more complex with shorter isoforms of Trio that do not contain the $G\alpha_q$ binding site (Figure 1.7). There must be an alternative mechanism for these isoforms to become co-localized with their substrates at the cell surface or other intracellular compartment.

$G\alpha_q$ is thought to be a master regulator of Trio, yet it is not known whether the binding of this protein to Trio results in a global conformational change, or if it simply regulates TrioC alone. My biochemical data suggests that $G\alpha_q$ regulates TrioC by activating RhoA exchange ~2.5-fold while being insensitive to the presence of other Trio domains. This hypothesis is strengthened by primary sequence analysis and secondary structure prediction. On either end of the TrioC module there are large unstructured regions, with a ~200 residue unstructured region N-terminal to this GEF module (~1718-1959), and a ~30 residue glycine-rich region immediately C-terminal (~2290-2310). However, the answer to this question could be different in the context of the true full-length enzyme, which I was not able to purify.

Trio remains a hugely multifunctional molecule, with protein-protein binding effects. The specific functional roles of the understudied Trio domains must be identified before future mechanistic study can be done. As of now, the cellular roles of the CT, SH3_N, SH3_C, Ig, and kinase domains remain ambiguous. Each domain has a putative role we can identify based on homology with their respective protein families, yet the cellular ligands or binding partners that make each role specific are not known. A lipid binding screen could be used to identify CT ligands, and it would be fortuitous to use a phosphoproteomic screen or other cell-based method to define Trio's KD substrate profile. SH3 ligands are notoriously low affinity, and as such it could be difficult to define ligands. One method could be an NMR screen using labeled intramolecular Trio polyproline candidate sequences, as NMR can detect low affinity interactions. The Ig domain may bind to small molecules, lipids, or proteins, and as such a pulldown mass spectrometry approach might be best to identify binding partners. As stated before, this domain may even interact with a complicated epitope consisting of a prenyl group on RhoA including protein contacts. With the ultimate goal of inhibiting Trio GEF activity in cancers, primarily uveal melanoma, much more basic science is needed to determine the best strategy for effective modulation of Trio function. Efficacious Trio probes could further unravel the complex signaling networks this enzyme is at the core of, and provide valuable leads for cancer therapeutics.

# Appendix A: TrioC Inhibitor Discovery

## Introduction

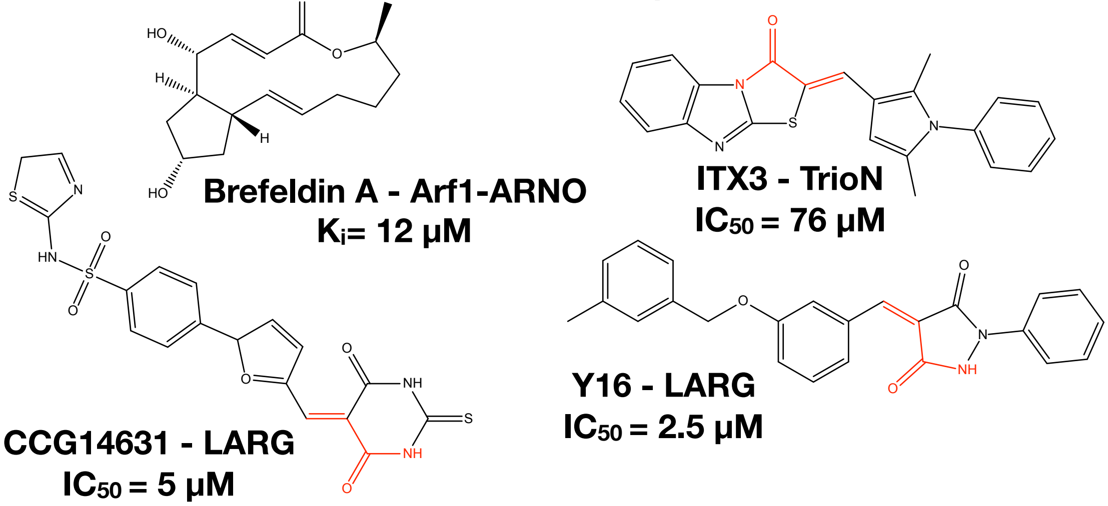
This Chapter covers preliminary inhibitor discovery efforts against TrioC. I hypothesized that the Trio inhibitory peptide α (TRIPα) peptide, the only known inhibitor of TrioC, would provide a valuable lead to characterize structurally, and to use as a tool to find new TrioC binders. The Chapter begins with a background on RhoGEF inhibition and TRIPα and continues to describe efforts made to purify TRIPα in order to discover the molecular basis of its function and use the peptide as a tool for inhibitor discovery in high-throughput assays. I also attempted to optimize high-throughput assays to use in inhibitor discovery campaigns. The Chapter closes with a discussion on avenues to move forward with TrioC inhibitor discovery.

The first RhoGEF inhibitor discovered was TRIPα and it was discovered using a yeast two-hybrid screen designed to discover molecules binding to TrioC.[157] The original TRIPα peptide was part of a library of peptide fragments called aptamers displayed as a loop on the protein thioredoxin. This peptide was shown to have activity as an excised 42-residue peptide using the in vitro GEF assay and was mutagenized to improve activity (TRIPα E32G shown in Figure A.1). [158] When overexpressed, this peptide is effective at reducing tumor volume and weight in a xenograft model of Tgat-dependent ATL. TRIPα is an especially valuable lead as it is selective for TrioC even over its close homolog p63RhoGEF. Thus, the structural basis of TRIPα inhibition will not only reveal the molecular mechanism of RhoGEF inhibition, but also how selectivity between related

RhoGEFs can be achieved, phenomena not yet defined. The large nature of TRIPα indicates it may have the complexity to productively interact with the relatively flat surface of the DH/PH module of TrioC. Determining the basis of this inhibition may allow “conversion” of TRIPα into a bioavailable molecule that captures enough complexity to be efficacious.

**TRIPα E32G: AREGADGAICGYNLATLVMLVMLGPSERVFCPLCGPCSSDIYELM**

**$K_i$ = 7.4 μM**



**Brefeldin A - Arf1-ARNO**
**$K_i$= 12 μM**

**ITX3 - TrioN**
**$IC_{50}$ = 76 μM**

**CCG14631 - LARG**
**$IC_{50}$ = 5 μM**

**Y16 - LARG**
**$IC_{50}$ = 2.5 μM**

Figure A.1 Described RhoGEF inhibitors from the literature show toxic motifs common to discovered RhoGEF inhibitors. TRIPα is shown as the primary peptide sequence with hydrophobic residues highlighted in blue, and cysteines highlighted in red. Regions deemed indispensable for activity of the peptide are underlined. Shown below in stick models are described GEF inhibitors listed next to their molecular targets and their inhibitory values against those targets. Michael acceptor regions are highlighted in red.

Small molecule GEF inhibitors have also been described in the literature (Figure A.1). Brefeldin A is a natural product ArfGEF inhibitor and binds to the GEF·GTPase·GDP ternary complex in order to prevent productive nucleotide exchange. This is the only GEF inhibitor for which a crystal structure and thus the mechanism of action is known. Brefeldin A does not target a member of the Dbl family so conversion to Trio inhibition is not possible. [155,156] ITX3 is a described small molecule inhibitor of TrioN originally discovered through a yeast gene reporter assay. [159,160] This molecule reportedly inhibits TrioN GEF activity in vitro and in cells, although it is not very potent ($IC_{50}$ = 76 μM for in vitro GEF assay). A series of 5 compounds have been described which inhibit LARG GEF activity in both fluorescent and radiometric in vitro assays, some with ~5 μM $IC_{50}$ values.[153] A competing group discovered the molecule Y16 using a virtual screen directed against the crystal structure of the DH/PH tandem of LARG.[161] At micromolar amounts, Y16 effectively inhibits LARG and subfamily members p115RhoGEF and PDZRhoGEF. This compound inhibits growth and migration of breast cancer cells in vitro. However, all three of these described compounds and derivatives contain known toxic or pan-assay interference motifs (Figure A.1).[154] For example, the above compounds contain Michael acceptor motifs (acrylamide groups) which could have off target effects. No follow-up reports of these compounds have been published.

Brefeldin A is a natural product GEF inhibitor and the only one whose mechanism of action is understood (Figure A.1). This molecule inhibits several members of the Sec7 family, GEFs for the Arf family of GTPases. Brefeldin A binds to the complex of the Sec7 domain with the GDP-bound form of Arf, trapping this state and preventing nucleotide release.[155,156] Brefeldin A is a complex heterocyclic natural product and interacts with the relatively flat interface between GEF and GTPase. I hypothesize a successful GEF inhibitor will also need to be similarly structurally complex and a relatively large molecule.

In this Chapter, I describe purification of TRIPα from *E. coli* and preliminary characterization of its activity. I also describe preliminary assay optimization for high throughput screens directed against TrioC using TRIPα as a control for TrioC binding and inhibition.

## Methods

### TRIPα Expression and Purification

Codon-optimized TRIPα cDNA was ordered from GenScript and was inserted in pMCSG9 for bacterial expression. TRIPα was expressed and purified similarly to TrioC, as an MBP fusion in E. coli. After MBP cleavage by TEV protease, the reaction mixture was concentrated using a 30 kD cutoff Amicon Ultracel concentrator and the flowthrough was collected. Flowthrough from the 30 kD concentrator was concentrated in a 3kD cutoff concentrator to collect TRIPα at ~5 kD. Silver-staining of SDS-PAGE indicated the sample was pure. Coupled liquid chromatography-mass spectrometry (done by Finn Maloney, Ph.D.) confirmed the peptide mass as 5099.34 D (predicted mass of 5101.89 D, difference of 2.5 D). Since TRIPα contains several cysteines, 2 mM DTT was included in all buffer steps and the purification was performed at 4 °C.

### GEF Activity Assays and DSF Assays

These assays were performed as described in sections 2.2.6 and 2.2.7 with the addition of varying amounts of purified TRIPα in buffer.

### High-Throughput Screening Optimization

Attempts were made to optimize a high-throughput screen against the TrioC DH/PH module. TRIPα peptide was tried as a positive control in order to optimize a high-throughput DSF assay. DSF assays were performed as described above with the addition of varying amounts of TRIPα into the reaction mixture and varying the protein construct in order to improve signal to background ratio (S/B) where $S/B = \frac{\mu_+}{\mu_-}$ and the Z′ assay statistic. Z′ was calculated as previously described where $Z' = 1 - \frac{3(\sigma_+ + \sigma_-)}{|\mu_+ - \mu_-|}$.[162] In order to perform a fluorescence polarization (FP) assay using amine labeled TRIPα, amine labeling was attempted via lysine using a peptide mutagenized via to include a C-terminal lysine-tryptophan. Alexafluor-488 (AF-488) was used as an N-Hydroxysuccinimide ester (NHS ester) (A20000 ThermoFisher Scientific) in a 1:1 molar ratio to TRIPα. The reaction took place at ambient temperature for 1 hour in a buffer composed of 20 mM HEPES pH 8.0, 200 mM NaCl, and 2 mM DTT. The reaction mixture was buffer exchanged using a 3kD centrifugal concentrator thrice into the above buffer in order to remove excess labeling reagent. Peptide was assumed to be 100% labeled and concentration of the labeled peptide was estimated using $A_{280}$. The concentration of AF-488-TRIPα and TrioC were varied in order to optimize the FP signal for the described parameters. An FP version of the nucleotide exchange activity assay described above was also attempted.[153] In a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, and 10 mM $MgCl_2$, 100 nM TrioC was incubated with 1 μM RhoA and 1 μM BODIPY-FL-GTP$\gamma$S (ThermoFisher G22183) and the reaction was monitored for 20 minutes, and fluorescence polarization was measured over time using an excitation wavelength of 488 nm and an emission wavelength of 532 nm.

## Results

### TRIPα Purification and Characterization

The TRIPα peptide is the only reported inhibitor of TrioC and was discovered in a yeast two-hybrid screen. TRIPα is a 42 amino acid peptide that was originally shown to bind to TrioC when inserted as a loop of a protein called thioredoxin. The TRIPα sequence itself was later shown to inhibit TrioC independently of thioredoxin and was mutagenized to improve potency to $IC_{50} = 5$ μM.[163] This peptide specifically inhibits the DH domain of

TrioC, and inhibits TrioC and KalirinC, but not p63RhoGEF implying that residues conserved between the Trio and Kalirin DH domains are important for inhibition by TRIPα. The ultimate goal of this work was to determine the molecular basis of TRIPα inhibition in order to determine the basis for GEF inhibition and improve this peptide using rational design or screening. Codon-optimized TRIPα was cloned into an E. coli expression vector and expressed and purified as an N-terminal maltose binding protein (MBP) fusion, with the final sequence as N-MBP-TEV-TRIPα-C. The construct was purified using Ni-NTA resin and subjected to TEV protease cleavage. The resulting reaction was concentrated using a 30 kD cutoff centrifugal concentrator to retain MBP in the top of the filter and allow the peptide to flow through. Peptide was concentrated afterwards using a 3 kD cutoff filter and found to be relatively pure by SDS-PAGE with

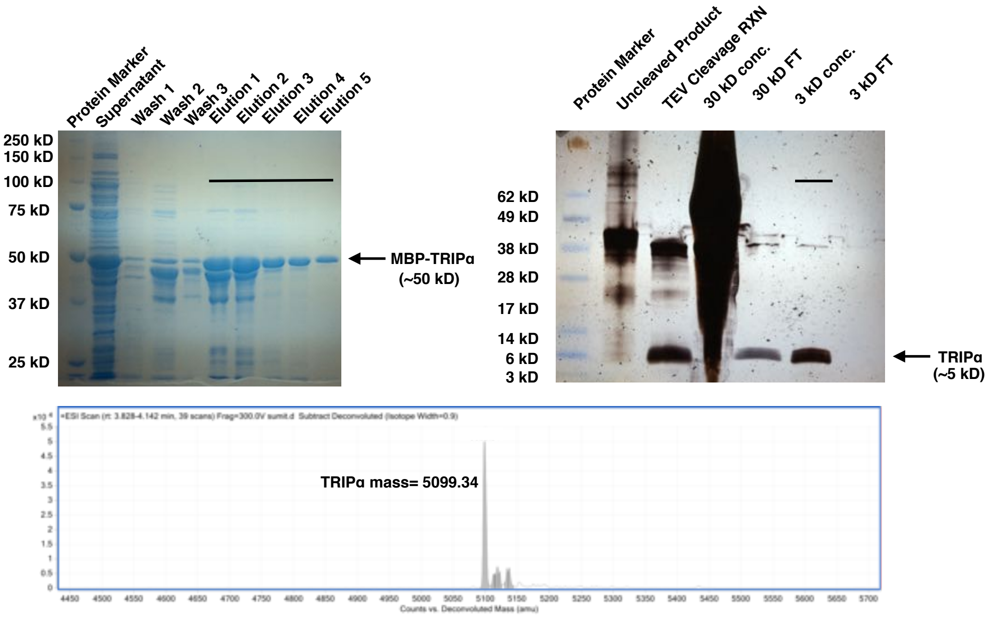


Figure A.2 TRIPα purification scheme shows homogenous peptide confrmed by silver staining and mass spectrometry. Top left shows a Coomassie stained 4-15% gradient gel of an Ni-NTA purification of MBP-TRIPα. Top right shows a silver stained 4-15% gradient gel showing finishing steps of the TRIPα purification using centrifugal concentrators. Bottom panel shows a mass spectrometry scan of purified TRIPα. Black bars indicate samples pooled for the next step of the preparation, or samples pooled for final protein aliquots. Figure A.1 Described RhoGEF inhibitors from the literature. TRIPα is shown as the primary peptide sequence with hydrophobic residues highlighted in blue, and cysteines highlighted in red. Regions deemed indispensable for activity of the peptide are underlined. Shown below in stick models are described GEF inhibitors listed next to their molecular targets and their inhibitory values against those targets. Michael acceptor regions are highlighted in red.

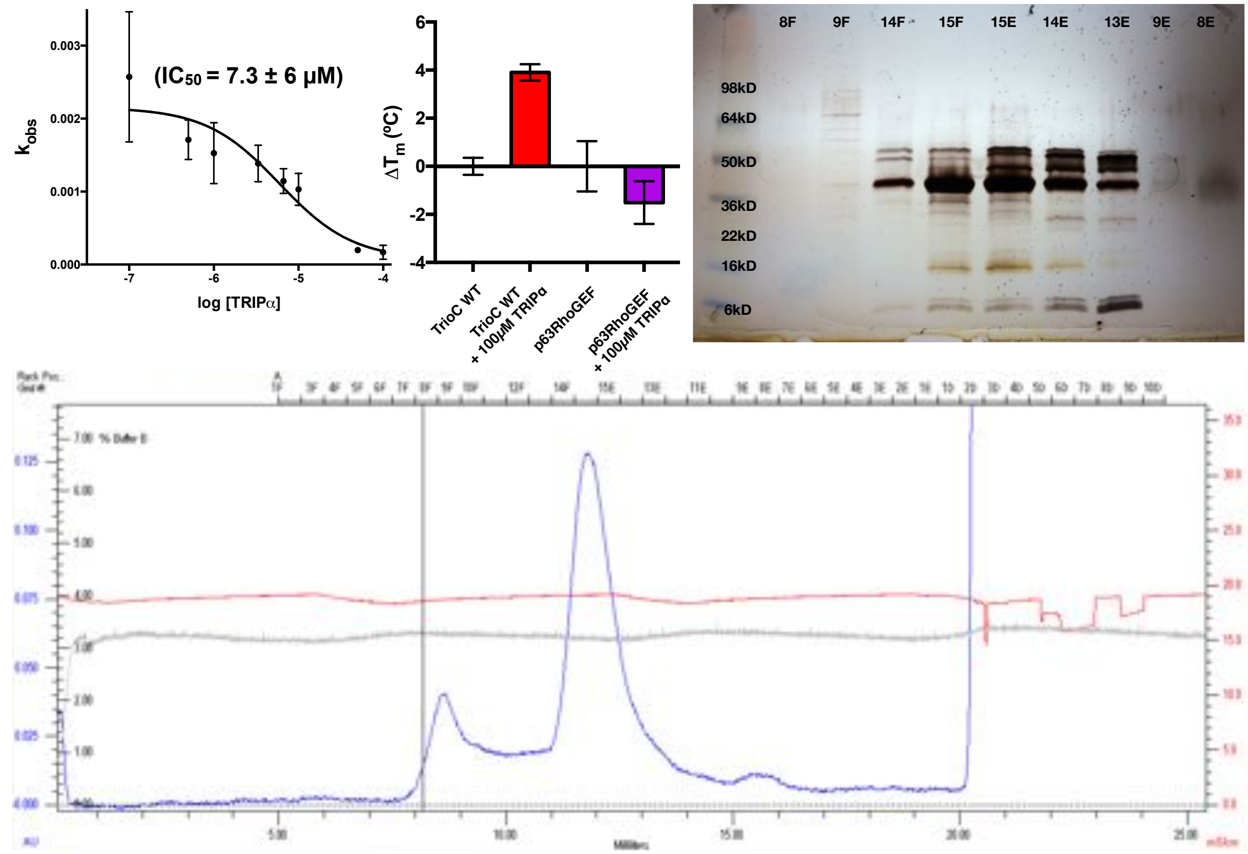


Figure A.3 Characterization of TRIPα peptide purified from E. coli shows it active, capable of thermostabilizing TrioC, and capable of complex formation with TrioC. Top left, titration of TRIPα against TrioC in the GEF activity assay with calculated $IC_{50}$ shown. Data shown is the average of N=2 experiments in duplicate. Error shown is standard deviation. Top middle, DSF assay data plotted as average $\Delta T_m$ values. Data shown is the average of N=3 experiments in triplicate. Error shown as standard deviation. Top right, silver stained 4-15 % gradient gel showing complex formation between TRIPα (5 kD) and TrioC (37 kD). Bottom, corresponding chromatogram showing $A_{280}$ trace (blue line) over the single Superdex 75 column used for TRIPα-TrioC complex formation.Figure 0.1 TRIPα purification scheme. Top left shows a Coomassie stained 4-15% gradient gel of an Ni-NTA purification of MBP-TRIPα. Top right shows a silver stained 4-15% gradient gel showing finishing steps of the TRIPα purification using centrifugal concentrators. Bottom panel shows a mass spectrometry scan of purified TRIPα. Black bars indicate samples pooled for the next step of the preparation, or samples pooled for final protein aliquots.

silver staining and $ESI^+$ mass spectrometry (Figure A.2). This peptide inhibited TrioC when titrated into the GEF activity assay ($IC_{50}$ = 7.4 ± 6 μM) and selectively thermostabilized TrioC 4 °C, yet not p63RhoGEF, in the DSF assay at 100-fold molar excess. TRIPα formed a complex with TrioC via SEC (Figure A.3). TRIPα was also used in the screening optimization described below. Attempts to progress in either screening or co-crystallization were hindered by poor reproducibility in the TRIPα purification procedure. Following attempts to purify TRIPα were not successful in separating the peptide from its MBP fusion partner after TEV cleavage using further Ni-NTA purification or anion exchange. A peptide synthesis company was contacted to synthesize this peptide, but they were unable to produce peptide more than 70 % pure and suggested we cancel the order.

### High Throughput Screening Optimization Attempts

Optimization of several assays for high-throughput screening was attempted. A DSF assay, using TRIPα as a positive control, was attempted at a larger scale. The goal of this assay would be to find compounds that thermostabilize and therefore bind to TrioC. Compounds which thermostabilized would later be tested for inhibitory activity against TrioC. A preliminary result of $Z' = 0.53$ was obtained with $S/B = 1.1$ using TRIPα as a positive control and buffer as a negative control. Labeling of TRIPα was also attempted in order to use the peptide in a fluorescence polarization displacement assay similar to one previously used to discover GEF inhibitors.[164] Labeled peptide could be used to discover small molecules which would bind to TrioC in a similar manner. A variant version of the TRIPα peptide was generated with a lysine-tryptophan dipeptide label appended to its C-terminus. This allowed for amine-directed labeling of the lysine and easier quantitation of peptide amount through either $A_{280}$ or Coomassie staining. Labeling of TRIPα via the installed lysine was performed using the NHS-ester of AF-488. Incubation of TrioC with AF488-TRIPα yielded preliminary $Z' = 0.53$ and S/B of 1.7 using 250 nM AF488-TRIPα as a negative control and adding 50-fold molar equivalent of TrioC for a positive control. A large excess of TrioC is required to get these parameters and as such this assay would require a large quantity of TrioC to run. In addition, a fluorescence polarization version of the GEF activity assay employing BODIPY-FL-GTPγS was attempted with preliminary Z′ = 0.43 and S/B = 1.7 using no enzyme as a negative control and adding TrioC for a positive control. These assay statistics are suboptimal even though the background is artificially low by using no enzyme as opposed to using TrioC with TRIPα.

## Discussion

Efforts to develop such an inhibitor using either rational design based on TRIPα or new scaffold discovery using high throughput screening were initiated. The TRIPα produced was active against TrioC in the GEF assay and thermostabilized TrioC by 4 °C, confirming its potential for use in screening assays. The TRIPα peptide proved difficult to work with when recombinantly produced from E. coli, and this molecule was deemed too difficult to produce by a contracting synthesis group. This is likely due to the relatively large size of TRIPα, its hydrophobicity, and its high number of cysteine residues. This is unfortunate as this peptide is selective, inhibiting TrioC and KalirinC, yet not p63RhoGEF which shares 65% sequence identity. TRIPα also does not inhibit RhoA nucleotide exchange by Lbc, p115RhoGEF, and Dbl, or Rac1 nucleotide exchange by TrioN. Determination of the molecular basis of this selectivity would be important in understanding how a selective GEF inhibitor acts, as there are several possible modes of action for a TrioC subfamily inhibitor. TRIPα most likely acts by sterically disrupting the RhoA binding site on the DH domain of TrioC, as it has in vitro activity against Tgat and the TrioC DH domain. However, since TrioC subfamily members have an autoinhibitory PH domain, an inhibitor may also function by stabilizing this autoinhibited conformation and thereby reducing basal GEF activity.

Screening efforts using activity-based assays for TrioC will require further optimization before they can be used to screen large libraries of small molecule compounds. Screening assays using TRIPα are promising (DSF and displacement FP), but will require TRIPα in order to run. The screening statistics for the activity assay currently derived are suboptimal and will only worsen when applied across an entire screening plate. Because the interface between TrioC-subfamily members and RhoA is extensive and relatively flat, a small molecule interrupting this interaction will likely be large and complex in nature, resembling the TRIPα peptide in that regard.

### Conclusions and Future Directions

The TRIPα peptide still remains a promising lead for inhibitor development as it has micromolar potency and selectivity, an important barrier to cross when developing lead compounds. However, the peptide cannot yet be purified reproducibly in sufficient quantity for biophysical characterization. The peptide could be used in the future as an uncleaved fusion product of MBP. Alternative purification methods for the peptide alone could still be attempted, using dual affinity tags or alternative fusion proteins, but the challenges reported by the peptide synthesis company suggest that the TRIPα sequence itself is problematic.

A mutagenesis campaign screening for active variants of TRIPα easier to purify may be fruitful. As stated by the contract peptide synthesis company, cysteine residues seem to be problematic and those should be prioritized to see if they can be removed while maintaining inhibitory activity. In order to improve the peptide's stability, a series of TRIPα derivatives could be generated, mutating cysteines to serines and hydrophobic residues to hydrophilic ones. These derivatives would be purified from E. coli and profiled for inhibitory activity using the GEF assay and the DSF assay.

Two assay formats that depend on TRIPα have promising preliminary statistics yet will require the peptide at large scale in order to optimize and run the assay. Although the DSF high-throughput assay has a low S/B value of 1.1, the $Z'=0.53$ is acceptable and this assay could be used to find small molecules that thermostabilize TrioC. Counter screens using GEF activity assays would be run to ensure that discovered compounds bind specifically and do not induce the formation of soluble protein aggregates, an artificial way to increase $T_m$ in the DSF assay. The displacement FP assay using AF-488 labeled TRIPα also showed promising $Z'=0.53$ and S/B=1.7, yet currently requires a large excess of TrioC to obtain this signal. Thus the assay would likely need to be optimized in order to use less protein, perhaps by enhancing the labeling efficiency of TRIPα peptide or by using a different fluorophore. Future directions in this aim include optimizing expression and purification conditions for TRIPα and optimizing the DSF assay for use as a primary screening technique to find compounds which thermostabilize TrioC. The DSF assay is robust, requires small amounts of protein and peptide, shows selectivity, and is easily adaptable to HTS. A secondary screen would employ the GEF assay described in 2.2.7 to ensure that thermostabilizing compounds inhibit TrioC GEF activity. In parallel, crystallography and HDX-MS efforts will reveal the structural basis of TRIPα function.

## Appendix B: Table of DNA Vectors for TrioFL project

All vectors in this table were cloned using ligation-independent cloning (LIC) with the exception of vectors annotated “source plasmid”. Ligation independent cloning results in an exogenous serine-asparagine-alanine tripeptide sequence at the N-terminus of the cleaved fusion product.

| Position | Name | Plasmid | Insert (see isoform 1 of UniProtKB O75962) | Purification Successful? | Notes |
|---|---|---|---|---|---|
| A1 | SH3_N | pMCSG9 | 1655-1724 | Yes | |
| A2 | SH3_C | pMCSG9 | 2554-2616 | Yes | |
| A3 | Ig | pMCSG9 | 2680-2780 | Yes | |
| A4 | CTS | pMCSG9 | 61-213 | No | Does not behave when cleaved. |
| A5 | CTL | pMCSG9 | 61-242 | No | Does not behave when cleaved. |
| A6 | 2680-3053 | pMCSG9 | 2680-3053 | No | Kinase; Does not behave when cleaved. |
| A7 | 2680-3097 | pMCSG9 | 2680-3097 | No | See A6 |
| A8 | 2788-3053 | pMCSG9 | 2788-3053 | No | See A6 |
| A9 | 2788-3097 | pMCSG9 | 2788-3097 | No | See A6 |
| B1 | 2680-3053 | pMCSG7 | 2680-3053 | No | No Expression |
| B2 | 2680-3097 | pMCSG7 | 2680-3097 | No | No Expression |
| B3 | 2788-3053 | pMCSG7 | 2788-3053 | No | No Expression |
| B4 | 2788-3097 | pMCSG7 | 2788-3097 | No | No Expression |
| B5 | 2680-3053 | pCDNA7 | 2680-3053 | No | No Expression |
| B6 | 2680-3097 | pCDNA7 | 2680-3097 | No | No Expression |
| B7 | 2788-3097 | pCDNA7 | 2788-3097 | No | No Expression |
| B8 | dTrio 5319 | pRmHA-3 | dTrio 1-2263 (UniprotKB Q7KVD1) | No | Source Plasmid |
| B9 | dTrioFLG9 | pMCSG9 | dTrio 1-2263 | No | Smear on gel |
| C1 | dTrioFLG10 | pMCSG10 | dTrio 1-2263 | No | Smear on gel |
| C2 | dTrioFLH9 | pH9pFB | dTrio 1-2263 | No | No Expression |
| C3 | dTrioFLH10 | pH10pFB | dTrio 1-2263 | No | No Expression |
| C4 | dTrioFLAC5T | pAC5T | dTrio 1-2263 | No | D. Mel Exp. |
| C5 | hTrioFLpEGFP | pEGFP | 61-3097 | No | Source plasmid |
| C6 | hTrioFL-NHT | pEGFP | 61-3097 | No | N-10xHis-TEV Secondary purification failed. |
| C7 | hTrioFLpH9 | pH9pFB | 61-3097 | No | No Expression |

| | | | | | |
|---|---|---|---|---|---|
| C8 | hTrioFLpH10 | pH10pFB | 61-3097 | No | No Expression |
| C9 | hTrioΔ2290pH9 | pH9pFB | 61-2290 | No | No Expression |
| D1 | hTrioΔ2290pH10 | pH10pFB | 61-2290 | No | No Expression |
| D2 | hTrioΔ2290pG9 | pMCSG9 | 61-2290 | No | Smear on gel |
| D3 | hTrioΔ2290pC10 | pCDNA10 | 61-2290 | No | No Expression |
| D4 | hTrioFLpG9 | pMCSG9 | 61-3097 | No | Smear on gel |
| D5 | hTrioFLpG10 | pMCSG10 | 61-3097 | No | Smear on gel |
| D6 | hTrioFLpAC5T | pAC5T | 61-3097 | No | D. Mel Exp. |
| D7 | hTrioΔ1594pG9 | pMCSG9 | 61-1594 | Yes | |
| D8 | hTrioΔ1594pC10 | pCDNA10 | 61-1594 | No | No Expression |
| D9 | hTrioΔ1594pAC5T | pAC5T | 61-1594 | No | No Expression |
| E1 | hTrioΔ1718pG9 | pMCSG9 | 61-1718 | Yes | Smeary but ok |
| E2 | 1284-2638G9 | pMCSG9 | 1284-2638 | Yes | |
| E3 | 1284-2638C10 | pCDNA10 | 1284-2638 | No | No Expression |
| E4 | 1284-1594G9 | pMCSG9 | 1284-1594 | Yes | |
| E5 | 1284-1718G9 | pMCSG9 | 1284-1718 | Yes | |
| E6 | 1284-1959G9 | pMCSG9 | 1284-1959 | Yes | |
| E7 | 1284-2147G9 | pMCSG9 | 1284-2147 | No | Smear on gel |
| E8 | 1284-2290G9 | pMCSG9 | 1284-2290 | Yes | |
| E9 | 1284-2616G9 | pMCSG9 | 1284-2616 | Yes | |
| F1 | 1284-2780G9 | pMCSG9 | 1284-2780 | Yes | |
| F2 | 1284-3053G9 | pMCSG9 | 1284-3053 | Yes | Low exp. |

# Bibliography

1. Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* **3,** 639–650 (2002).
2. Milligan, G. & Kostenis, E. Heterotrimeric G-proteins: a short history. *British Journal of Pharmacology* **147,** S46–S55 (2009).
3. Hamm, H. E. & Gilchrist, A. Heterotrimeric G proteins. *Current Opinion in Cell Biology* **8,** 189–196 (1996).
4. Oldham, W. M. & E Hamm, H. Structural basis of function in heterotrimeric G proteins. *Quart. Rev. Biophys.* **39,** 117–50 (2006).
5. Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66,** 639–678 (1997).
6. Phospholipase C-PI Is a GTPase-Activating Protein for GqIII, Its Physiologic. 1–8 (2003).
7. Kimple, A. J., Bosch, D. E., Giguere, P. M. & Siderovski, D. P. Regulators of G-Protein Signaling and Their G  Substrates: Promises and Challenges in Their Use as Drug Discovery Targets. *Pharmacological Reviews* **63,** 728–749 (2011).
8. De Vries, L., Bin Zheng, Fischer, T., Elenko, E. & Farquhar, M. G. The Regulator of G Protein Signaling Family. *Annu. Rev. Pharmacol. Toxicol.* **40,** 235–271 (2000).
9. O'Hayre, M. *et al.* The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. 1–13 (2013). doi:10.1038/nrc3521
10. Van Raamsdonk, C. D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457,** 599–603 (2009).
11. Van Raamsdonk, C. D. *et al.* Mutations in GNA11 in uveal melanoma. *N. Engl. J. Med.* **363,** 2191–2199 (2010).
12. MD, J. W. H. & PhD, D. L. C. M. A Molecular Revolution in Uveal Melanoma. *Ophthalmology* **121,** 1281–1288 (2014).
13. Luke, J. J. *et al.* Biology of advanced uveal melanoma and next steps for clinical therapeutics. *Pigment Cell Melanoma Res.* **28,** 135–147 (2014).
14. Top. Deep sequencing of uveal melanoma identifies a recurrent mutation in. 1–8 (2016).
15. PPG, A. R. Dysregulated GPCR Signaling and Therapeutic Options in Uveal Melanoma. *Mol Cancer Res* **15,** 1–7 (2017).
16. Augsburger, J. J., Corrêa, Z. M. & Shaikh, A. H. Effectiveness of Treatments for Metastatic Uveal Melanoma. *AJOPHT* **148,** 119–127 (2009).
17. Griewank, K. G. *et al.* Genetic and clinico-pathologic analysis of metastatic uveal melanoma. *Mod Pathol* **27,** 175–183 (2013).
18. Weis, E. *et al.* Management of uveal melanoma: a consensus-based provincial clinical practice guideline. *Curr. Oncol.* **23,** 57–8 (2016).

19. Choudhary, M. M., Triozzi, P. & Singh, A. D. in *Clinical Ophthalmic Oncology* 275–285 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-54255-8_22
20. Woodman, S. E. Metastatic Uveal Melanoma. *The Cancer Journal* **18,** 148–152 (2012).
21. Onken, M. D. *et al.* Oncogenic Mutations in GNAQOccur Early in Uveal Melanoma. *Invest. Ophthalmol. Vis. Sci.* **49,** 5230–12 (2008).
22. Field, M. G. *et al.* Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nature Communications* 1–10 (2018). doi:10.1038/s41467-017-02428-w
23. Moore, A. R. *et al.* Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. *Nature Publishing Group* **48,** 675–680 (2016).
24. Hoiom, V. & Helgadottir, H. The genetics of uveal melanoma: current insights. *TACG* **Volume 9,** 147–155 (2016).
25. Buder, K., Gesierich, A., Gelbrich, G. & Goebeler, M. Systemic treatment of metastatic uveal melanoma: review of literature and future perspectives. *Cancer Med* **2,** 674–686 (2013).
26. Uemura, T. *et al.* Biological properties of a specific Gαq/11 inhibitor, YM-254890, on platelet functions and thrombus formation under high-shear stress. *British Journal of Pharmacology* **148,** 61–69 (2009).
27. Nishimura, A. *et al.* Structural basis for the specific inhibition of heterotrimeric Gq protein by a small molecule. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 13666–13671 (2010).
28. Annala, S. *et al.* Direct targeting of Gαq and Gα11 oncoproteins in cancer cells. *Sci. Signal.* **12,** eaau5948 (2019).
29. Lapadula, D. *et al.* Effects of Oncogenic Gαq and Gα11 Inhibition by FR900359 in Uveal Melanoma. *Mol Cancer Res* **17,** 963–973 (2019).
30. Chen, X. *et al.* Combined PKC and MEK inhibition in uveal melanoma with GNAQ and GNA11 mutations. *Oncogene* **33,** 4724–4734 (2013).
31. Wu, X., Zhu, M., Fletcher, J. A., Giobbie-Hurder, A. & Hodi, F. S. The Protein Kinase C Inhibitor Enzastaurin Exhibits Antitumor Activity against Uveal Melanoma. *PLoS ONE* **7,** e29622–10 (2012).
32. Zuidervaart, W. *et al.* Activation of the MAPK pathway is a common event in uveal melanomas although it rarely occurs through mutation of BRAF or RAS. *Br J Cancer* **92,** 2032–2038 (2005).
33. Carvajal, R. D. *et al.* Effect of Selumetinib vs Chemotherapy on Progression-Free Survival in Uveal Melanoma. *JAMA* **311,** 2397–9 (2014).
34. Komatsubara, K. M., Manson, D. K. & Carvajal, R. D. Selumetinib for the treatment of metastatic uveal melanoma: past and future perspectives. *Future Oncology* **12,** 1331–1344 (2016).
35. Carvajal, R. D. *et al.* Selumetinib in Combination With Dacarbazine in Patients With Metastatic Uveal Melanoma: A Phase III, Multicenter, Randomized Trial (SUMIT). *JCO* **36,** 1232–1239 (2018).

36. Yu, F.-X. *et al.* Mutant Gq/11 Promote Uveal Melanoma Tumorigenesis by Activating YAP. *CCELL* **25,** 822–830 (2014).
37. Vaque, J. P. *et al.* A Genome-wide RNAi Screen Reveals a Trio-Regulated Rho GTPase Circuitry Transducing Mitogenic Signals Initiated by G Protein-Coupled Receptors. *Molecular Cell* **49,** 94–108 (2013).
38. Liu, X. *et al.* Tead and AP1 Coordinate Transcription and Motility. *CellReports* **14,** 1169–1180 (2016).
39. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nature Communications* **6,** 1–16 (2015).
40. Pobbati, A. V. & Hong, W. Emerging roles of TEAD transcription factors and its coactivators in cancers. *Cancer Biology & Therapy* **14,** 390–398 (2014).
41. Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat Cell Biol* **17,** 1218–1227 (2015).
42. Feng, X. *et al.* Hippo-Independent Activation of YAP by the GNAQ Uveal Melanoma Oncogene through a Trio-Regulated Rho GTPase Signaling Circuitry. *CCELL* **25,** 831–845 (2014).
43. Feng, X. *et al.* A Platform of Synthetic Lethal Gene Interaction Networks Reveals that the GNAQ Uveal Melanoma Oncogene Controls the Hippo Pathway through FAK. *CCELL* **35,** 457–472.e5 (2019).
44. Schmidt, S. & Debant, A. Function and regulation of the Rho guanine nucleotide exchange factor Trio. *Small GTPases* **5,** e983880 (2014).
45. Aittaleb, M., Boguth, C. A. & Tesmer, J. J. G. Structure and Function of Heterotrimeric G Protein-Regulated Rho Guanine Nucleotide Exchange Factors. *Molecular Pharmacology* **77,** 111–125 (2010).
46. Lutz, S. *et al.* Structure of Galphaq-p63RhoGEF-RhoA complex reveals a pathway for the activation of RhoA by GPCRs. *Science* **318,** 1923–1927 (2007).
47. Liu, M., Bi, F., Zhou, X. & Zheng, Y. Rho GTPase regulation by miRNAs and covalent modifications. *Trends in Cell Biology* **22,** 365–373 (2012).
48. O'Brien, S. P. *et al.* Skeletal muscle deformity and neuronal disorder in Trio exchange factor-deficient mouse embryos. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 12074–12078 (2000).
49. Correction for Cahill et al., Kalirin regulates cortical spine morphogenesis and disease-related behavioral phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16890–16890 (2009).
50. Aittaleb, M., Nishimura, A., Linder, M. E. & Tesmer, J. J. G. Plasma Membrane Association of p63 Rho Guanine Nucleotide Exchange Factor (p63RhoGEF) Is Mediated by Palmitoylation and Is Required for Basal Activity in Cells. *J. Biol. Chem.* **286,** 34448–34456 (2011).
51. Momotani, K. & Somlyo, A. V. p63RhoGEF: A New Switch for Gq-Mediated Activation of Smooth Muscle. *Trends Cardiovasc. Med.* **22,** 122–127 (2012).

52. Momotani, K. *et al.* p63RhoGEF Couples G q/11-Mediated Signaling to Ca2+ Sensitization of Vascular Smooth Muscle Contractility. *Circulation Research* **109,** 993–1002 (2011).
53. Wuertz, C. M. *et al.* p63RhoGEF—a key mediator of angiotensin II-dependent signaling and processes in vascular smooth muscle cells. *FASEB J.* **24,** 4865–4876 (2010).
54. Rabiner, C. A., Mains, R. E. & Eipper, B. A. Kalirin: A Dual Rho Guanine Nucleotide Exchange Factor That Is So Much More Than the Sum of Its Many Parts. *Neuroscientist* **11,** 148–160 (2016).
55. McPherson, C. E., Eipper, B. A. & Mains, R. E. Multiple novel isoforms of Trio are expressed in the developing rat brain. *Gene* **347,** 125–135 (2005).
56. Yoshizuka, N. *et al.* An Alternative Transcript Derived from the TrioLocus Encodes a Guanosine Nucleotide Exchange Factor with Mouse Cell-transforming Potential. *J. Biol. Chem.* **279,** 43998–44004 (2004).
57. van Unen, J. *et al.* The C-terminus of the oncoprotein TGAT is necessary for plasma membrane association and efficient RhoA-mediated signaling. *BMC Cell Biol* **19,** 1–15 (2018).
58. Debant, A. *et al.* The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 5466–5471 (1996).
59. Steven, R. *et al.* UNC-73 activates the Rac GTPase and is required for cell and growth cone migrations in C. elegans. *Cell* **92,** 785–795 (1998).
60. Awasaki, T. *et al.* The Drosophila trio plays an essential role in patterning of axons by regulating their directional extension. *Neuron* **26,** 119–131 (2000).
61. Fort, P. & Blangy, A. The Evolutionary Landscape of Dbl-Like RhoGEF Families: Adapting Eukaryotic Cells to Environmental Signals. *Genome Biology and Evolution* **9,** 1471–1486 (2017).
62. Zheng, Y. Dbl family guanine nucleotide exchange factors. *Trends in Biochemical Sciences* **26,** 724–732 (2001).
63. Whitehead, I. P., Campbell, S., Rossman, K. L. & Der, C. J. Dbl family proteins. *Biochim. Biophys. Acta* **1332,** F1–23 (1997).
64. Hart, M. J. *et al.* Identification of a novel guanine nucleotide exchange factor for the Rho GTPase. *J. Biol. Chem.* **271,** 25452–25458 (1996).
65. Rossman, K. L., Der, C. J. & Sondek, J. GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* **6,** 167–180 (2005).
66. Activated RhoA Binds to the Pleckstrin Homology (PH) Domain of PDZ-RhoGEF, a Potential Site for Autoregulation. 1–13 (2010).
67. Medley, Q. G. *et al.* The Trio Guanine Nucleotide Exchange Factor Is a RhoA Target: BINDING OF RhoA TO THE TRIO IMMUNOGLOBULIN-LIKE DOMAIN. *J. Biol. Chem.* **275,** 36116–36123 (2000).

68. Nimnual, A. S., Yatsula, B. A. & Bar-Sagi, D. Coupling of Ras and Rac guanosine triphosphatases through the Ras exchanger Sos. *Science* **279,** 560–563 (1998).
69. Lambert, J. M. *et al.* Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. *Nat Cell Biol* **4,** 621–625 (2002).
70. Structure and mutagenesis of the Dbl homology domain. 1–10 (1998).
71. Worthylake, D. K., Rossman, K. L. & Sondek, J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408,** 682–688 (2000).
72. Rossman, K. L. *et al.* A crystallographic view of interactions between Dbs and Cdc42: PH domain-assisted guanine nucleotide exchange. *The EMBO Journal* **21,** 1315–1326 (2002).
73. Kristelly, R., Gao, G. & Tesmer, J. J. G. Structural Determinants of RhoA Binding and Nucleotide Exchange in Leukemia-associated Rho Guanine-Nucleotide Exchange Factor. *J. Biol. Chem.* **279,** 47352–47362 (2004).
74. Chhatriwala, M. K., Betts, L., Worthylake, D. K. & Sondek, J. The DH and PH Domains of Trio Coordinately Engage Rho GTPases for their Efficient Activation. *Journal of Molecular Biology* **368,** 1307–1320 (2007).
75. Cherfils, J. & Chardin, P. GEFs: structural basis for their activation of small GTP-binding proteins. *Trends in Biochemical Sciences* **24,** 306–311 (1999).
76. Rossman, K. L. & Sondek, J. Larger than Dbl: new structural insights into RhoA activation. *Trends in Biochemical Sciences* **30,** 163–165 (2005).
77. Wei, Y. *et al.* Crystal structure of RhoA-GDP and its functional implications. *Nat. Struct Biol.* **4,** 699–703 (1997).
78. Crystal Structure of Human RhoA in a Dominantly Active Form Complexed with a GTP Analogue*. 1–12 (1998).
79. Karnoub, A. E. *et al.* Molecular basis for Rac1 recognition by guanine nucleotide exchange factors. *Nat. Struct Biol.* **8,** 1037–1041 (2001).
80. Bellanger, J. M. *et al.* Different regulation of the Trio Dbl-Homology domains by their associated PH domains. *Biology of the Cell* **95,** 625–634 (2012).
81. Skowronek, K. R., Guo, F., Zheng, Y. & Nassar, N. The C-terminal Basic Tail of RhoG Assists the Guanine Nucleotide Exchange Factor Trio in Binding to Phospholipids. *J. Biol. Chem.* **279,** 37895–37907 (2004).
82. Cash, J. N., Davis, E. M. & Tesmer, J. J. G. Structural and Biochemical Characterization of the Catalytic Core of the Metastatic Factor P-Rex1 and Its Regulation by PtdIns(3,4,5)P3. *Structure/Folding and Design* **24,** 730–740 (2016).
83. Snyder, J. T. *et al.* Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. *Nat. Struct Biol.* **9,** 468–475 (2002).
84. Worthylake, D. K., Rossman, K. L. & Sondek, J. Crystal Structure of the DH/PH Fragment of Dbs without Bound GTPase. *Structure* **12,** 1079–1086 (2004).
85. Rossman, K. L. *et al.* Multifunctional Roles for the PH Domain of Dbs in Regulating Rho GTPase Activation. *J. Biol. Chem.* **278,** 18393–18400 (2003).

86. Baumeister, M. A., Rossman, K. L., Sondek, J. & Lemmon, M. A. The Dbs PH domain contributes independently to membrane targeting and regulation of guanine nucleotide-exchange activity. *Biochem. J.* **400,** 563–572 (2006).
87. Rojas, R. J. *et al.* G q Directly Activates p63RhoGEF and Trio via a Conserved Extension of the Dbl Homology-associated Pleckstrin Homology Domain. *J. Biol. Chem.* **282,** 29201–29210 (2007).
88. Shankaranarayanan, A. *et al.* Gαq allosterically activates and relieves autoinhibition of p63RhoGEF. *Cellular Signalling* **22,** 1114–1123 (2010).
89. Saito, K., Tautz, L. & Mustelin, T. The lipid-binding SEC14 domain. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1771,** 719–726 (2007).
90. Aravind, L., Neuwald, A. F. & Ponting, C. P. Sec14p-like domains in NF1 and Dbl-like proteins indicate lipid regulation of Ras and Rho signaling. *Curr. Biol.* **9,** R195–7 (1999).
91. Ueda, S., Kataoka, T. & Satoh, T. Role of the Sec14-like domain of Dbl family exchange factors in the regulation of Rho family GTPases in different subcellular sites. *Cellular Signalling* **16,** 899–906 (2004).
92. Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J. & Gavin, A.-C. The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *JPLR* **61,** 30–39 (2016).
93. Kostenko, E. V., Mahon, G. M., Cheng, L. & Whitehead, I. P. The Sec14 Homology Domain Regulates the Cellular Distribution and Transforming Activity of the Rho-specific Guanine Nucleotide Exchange Factor Dbs. *J. Biol. Chem.* **280,** 2807–2817 (2005).
94. Ma, X.-M. *et al.* Nonenzymatic domains of Kalirin7 contribute to spine morphogenesis through interactions with phosphoinositides and Abl. *Mol. Biol. Cell* **25,** 1458–1471 (2014).
95. Schiller, M. R. *et al.* Autonomous functions for the Sec14p/spectrin-repeat region of Kalirin. *Experimental Cell Research* **314,** 2674–2691 (2008).
96. Schiller, M. R., Blangy, A., Huang, J., Mains, R. E. & Eipper, B. A. Induction of lamellipodia by Kalirin does not require its guanine nucleotide exchange factor activity. *Experimental Cell Research* **307,** 402–417 (2005).
97. Bi, F. *et al.* Autoinhibition Mechanism of Proto-Dbl. *Mol. Cell. Biol.* **21,** 1463–1474 (2001).
98. Association of the proto-oncogene product Dbl with G protein v subunits. 1–5 (1999).
99. Meier, R., Tomizaki, T., Schulze-Briese, C., Baumann, U. & Stocker, A. The Molecular Basis of Vitamin E Retention: Structure of Human α-Tocopherol Transfer Protein. *Journal of Molecular Biology* **331,** 725–734 (2003).
100. Djinovic-Carugo, K., Gautel, M., Ylänne, J. & Young, P. The spectrin repeat: a structural platform for cytoskeletal protein assemblies. *FEBS Lett.* **513,** 119–123 (2002).
101. de Almeida Ribeiro, E., Jr *et al.* The Structure and Regulation of Human Muscle &alpha;-Actinin. *Cell* **159,** 1447–1460 (2014).

102. Ortega, E. *et al.* The Structure of the Plakin Domain of Plectin Reveals an Extended Rod-like Shape. *J. Biol. Chem.* **291,** 18643–18662 (2016).
103. Localization and Structure of the Ankyrin-binding Site on . 1–6 (2009).
104. Kusunoki, H., Minasov, G., MacDonald, R. I. & Mondragón, A. Independent Movement, Dimerization and Stability of Tandem Repeats of Chicken Brain α-Spectrin. *Journal of Molecular Biology* **344,** 495–511 (2004).
105. Son, K., Smith, T. C. & Luna, E. J. Supervillin binds the Rac/Rho-GEF Trio and increases Trio-mediated Rac1 activation. *Cytoskeleton (Hoboken)* **72,** 47–64 (2015).
106. Vishwanatha, K. S., Wang, Y. P., Keutmann, H. T., Mains, R. E. & Eipper, B. A. Structural Organization of the Nine Spectrin Repeats of Kalirin. *Biochemistry* **51,** 5663–5673 (2012).
107. Kurochkina, N. & Guha, U. SH3 domains: modules of protein–protein interactions. *Biophys Rev* **5,** 29–39 (2012).
108. SH3 domains: complexity in moderation. 1–11 (2001).
109. Schiller, M. R. *et al.* Regulation of RhoGEF Activity by Intramolecular and Intermolecular SH3 Domain Interactions. *J. Biol. Chem.* **281,** 18774–18786 (2006).
110. Otey, C. A., Dixon, R., Stack, C. & Goicoechea, S. M. Cytoplasmic Ig-domain proteins: Cytoskeletal regulators with a role in human disease. *Cell Motil. Cytoskeleton* **66,** 618–634 (2009).
111. Tskhovrebova, L. & Trinick, J. Titin: properties and family relationships. *Nat Rev Mol Cell Biol* **4,** 679–689 (2003).
112. Structural basis for activation of the titin kinase domain during myo®brillogenesis. 1–8 (1999).
113. Giant protein kinases: domain interactions and structuralbasisofautoregulation. 1–12 (2005).
114. Structural Basis for the Autoinhibition of Calcium/Calmodulin-Dependent Protein Kinase I. 1–13 (1997).
115. Molecular Basis of the Death-Associated Protein Kinase–Calcium/Calmodulin Regulator Complex. 1–10 (2010).
116. Bandekar, S. J. *et al.* Structure of the C-terminal guanine nucleotide exchange factor module of Trio in an autoinhibited conformation reveals its oncogenic potential. *Sci. Signal.* **12,** eaav2449 (2019).
117. Sahai, E. & Marshall, C. J. RHO–GTPASES AND CANCER. *Nat Rev Cancer* **2,** 133–142 (2002).
118. Vega, F. M. & Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett.* **582,** 2093–2101 (2008).
119. Jansen, S., Gosens, R., Wieland, T. & Schmidt, M. Paving the Rho in cancer metastasis_ Rho GTPases and beyond. *Pharmacology and Therapeutics* **183,** 1–21 (2018).
120. Taylor, V. G., Bommarito, P. A. & Tesmer, J. J. G. Structure of the Regulator of G Protein Signaling 8 (RGS8)-Gα qComplex. *J. Biol. Chem.* **291,** 5138–5145 (2016).

121. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112,** 531–552 (1999).
122. Otwinowski, Z. & Minor, W. *[20] Processing of X-ray diffraction data collected in oscillation mode*. *Methods in Enzymology* **276,** 307–326 (Elsevier, 1997).
123. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst (2007). 40, 658-674 [doi:10.1107/S0021889807021206]* 1–17 (2007). doi:10.1107/S0021889807021206
124. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst (2008). 41, 641-643 [doi:10.1107/S0021889808006985]* 1–3 (2008). doi:10.1107/S0021889808006985
125. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst (2010). D66, 486-501 [doi:10.1107/S0907444910007493]* 1–16 (2010). doi:10.1107/S0907444910007493
126. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst (2012). D68, 352-367 [doi:10.1107/S0907444912001308]* 1–16 (2012). doi:10.1107/S0907444912001308
127. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst (1997). D53, 240-255 [doi:10.1107/S0907444996012255]* 1–16 (1997). doi:10.1107/S0907444996012255
128. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst (2010). D66, 12-21 [doi:10.1107/S0907444909042073]* 1–10 (2009). doi:10.1107/S0907444909042073
129. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ (2014). M1, 213-220 [doi:10.1107/S2052252514009324]* 1–8 (2014). doi:10.1107/S2052252514009324
130. Mezzasalma, T. M. *et al.* Enhancing Recombinant Protein Quality and Yield by Protein Stability Profiling. *Journal of Biomolecular Screening* **12,** 418–428 (2007).
131. Brecht, M. *et al.* The use of surface plasmon resonance (SPR) and fluorescence resonance energy transfer (FRET) to monitor the interaction of the plant G-proteins Ms-Rac1 and Ms-Rac4 with GTP. *Journal of Biotechnology* **112,** 151–164 (2004).
132. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Cryst (2017). 50, 1545-1553 [doi:10.1107/S1600576717011438]* 1–9 (2017). doi:10.1107/S1600576717011438
133. Trewhella, J. *et al.* 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Cryst (2017). D73, 710-728 [doi:10.1107/S2059798317011597]* **73,** 1–19 (2017).

134. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Cryst (2011). D67, 235-242 [doi:10.1107/S0907444910045749]* 1–8 (2011). doi:10.1107/S0907444910045749
135. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D Biological Crystallography* **62,** 439–450 (2006).
136. Painter, J. & Merritt, E. A. TLSMDweb server for the generation of multi-group TLS models. *J Appl Crystallogr* **39,** 109–111 (2006).
137. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* **2,** 401–404 (2012).
138. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst (2003). 36, 860-864 [doi:10.1107/S0021889803000268]* 1–5 (2003). doi:10.1107/S0021889803000268
139. Zheng, J. *et al.* The solution structure of the pleckstrin homology domain of human SOS1. A possible structural role for the sequential association of diffuse B cell lymphoma and pleckstrin homology domains. *J. Biol. Chem.* **272,** 30340–30344 (1997).
140. Yamada, K. *et al.* Tgat, a Rho-specific guanine nucleotide exchange factor, activates NF-κB via physical association with IκB kinase complexes. *Biochemical and Biophysical Research Communications* **355,** 269–274 (2007).
141. Mori, T., Moriuchi, R., Okazaki, E., Yamada, K. & Katamine, S. Tgat oncoprotein functions as a inhibitor of RECK by association of the unique C-terminal region. *Biochemical and Biophysical Research Communications* **355,** 937–943 (2007).
142. Brown, W. C. *et al.* New ligation-independent cloning vectors compatible with a high-throughput platform for parallel construct expression evaluation using baculovirus-infected insect cells. *Protein Expression and Purification* **77,** 34–45 (2011).
143. Beyett, T. S. *et al.* Carboxylic Acid Derivatives of Amlexanox Display Enhanced Potency toward TBK1 and IKK εand Reveal Mechanisms for Selective Inhibition. *Molecular Pharmacology* **94,** 1210–1219 (2018).
144. Beyett, T. S. *et al.* Perturbation of the interactions of calmodulin with GRK5 using a natural product chemical probe. *Proc. Natl. Acad. Sci. U.S.A.* **116,** 15895–15900 (2019).
145. Irem. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. 1–12 (2004).
146. Carragher, B. *et al.* Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens. *Journal of Structural Biology* **132,** 33–45 (2000).
147. cisTEM, user-friendly software for single- particle image processing. 1–24 (2018). doi:10.7554/eLife.35383.001
148. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Meth* **14,** 331–332 (2017).

149. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23,** 3403–3405 (2007).
150. Cash, J. N. *et al.* Cryo-electron microscopy structure and analysis of the P-Rex1-Gβγ signaling scaffold. *Sci Adv* **5,** eaax8855 (2019).
151. Pascual, J. E. A. Solution Structure of the Spectrin Repeat: a Left-handed Antiparallel Triple-helical Coiled-coil. 1–12 (1997).
152. Hertz, N. T. *et al.* Chemical Genetic Approach for Kinase-Substrate Mapping by Covalent Capture of Thiophosphopeptides and Analysis by Mass Spectrometry. *Current Protocols in Chemical Biology* **2,** 15–36 (2010).
153. Evelyn, C. R. *et al.* High-Throughput Screening for Small-Molecule Inhibitors of LARG-Stimulated RhoA Nucleotide Binding via a Novel Fluorescence Polarization Assay. *Journal of Biomolecular Screening* **14,** 161–172 (2009).
154. Gray, J. L., Delft, von, F. & Brennan, P. Targeting the Small GTPase Superfamily through their Regulatory Proteins. *Angew. Chem. Int. Ed.* 1–26 (2019). doi:10.1002/anie.201900585
155. Crystal Structure of ARF1•Sec7 Complexed with Brefeldin A and Its Implications for the Guanine Nucleotide Exchange Mechanism. 1–9 (2003).
156. Renault, L., Guibert, B. & Cherfils, J. Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* **426,** 525–530 (2003).
157. Schmidt, S., Diriong, S., Méry, J., Fabbrizio, E. & Debant, A. Identification of the first Rho–GEF inhibitor, TRIPα, which targets the RhoA-specific GEF domain of Trio. *FEBS Lett.* **523,** 35–42 (2002).
158. Schmidt, S. & Debant, A. *Aptamer-Derived Peptide Inhibitors of Rho Guanine Nucleotide Exchange Factors*. *Inhibitors of the Ras Superfamily G-proteins, Part A* **33,** 147–168 (Elsevier Inc., 2013).
159. Bouquier, N. *et al.* A Cell Active Chemical GEF Inhibitor Selectively Targets the Trio/RhoG/Rac1 Signaling Pathway. *Chemistry & Biology* **16,** 657–666 (2009).
160. Blangy, A. *et al.* Identification of TRIO-GEFD1 chemical inhibitors using the yeast exchange assay. *Biology of the Cell* **98,** 511–522 (2012).
161. Shang, X. *et al.* Small-molecule inhibitors targeting G-protein-coupled Rho guanine nucleotide exchange factors. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 3155–3160 (2013).
162. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening* **4,** 67–73 (1999).
163. Bouquier, N. *et al.* Aptamer-Derived Peptides as Potent Inhibitors of the Oncogenic RhoGEF Tgat. *Chemistry & Biology* **16,** 391–400 (2009).
164. Hafner, M. *et al.* Displacement of protein-bound aptamers with small molecules screened by fluorescence polarization. *Nat Protoc* **3,** 579–587 (2008).