Characterizing Novel Cyclophilin A-RNA Interactions

Maria T. Carilli

Undergraduate Honors Thesis

Department of Physics, University of Colorado Boulder

2021

Thesis Advisor: Deborah S. Wuttke, Department of Biochemistry

Graduate Mentor: Nickolaus C. Lammer

Table of Contents

Acknowledgements

I would first like to express my deepest gratitude to my advisor, Deborah Wuttke, who made this thesis possible. From my first meeting with her to discuss working in her lab to our most recent meeting, she has provided invaluable scientific insight and direction. Her guidance steadied the course of my undergraduate research experience with excellent structure as well as the freedom to explore exciting science. For consistent encouragement and insight on this thesis as well as for advice about my future academic plans, I cannot thank her enough. In addition to scientific support, she has always been extremely understanding and genuinely invested in the wellbeing of all her lab members, especially in recent, stressful times. I am so grateful to have her as a mentor!

I also can't overexpress my thanks to my graduate mentor, Nick Lammer. He has helped every step of the way and this thesis could not have happened without him. From answering all kinds of questions to helping out when my crazy schedule conflicted with lab work to having conversations about science and life, he has been an incredible person to work with. I couldn't be more thankful: Nick, you are awesome!

Many, many thanks to my other lab mates, too— to Conner Olson, Tom Weiser, and Aly Barbour. They were all always available to answer questions, give awesome advice, and help out, and I am so grateful to them. Thanks also to Neil Lloyd, a previous grad student in the Wuttke lab, for starting this exciting project. What a wonderful and supportive lab!

Abstract

RNA-binding proteins are a diverse and essential class of proteins with critical roles in key processes such as gene regulation, mRNA processing, and cell metabolism[1,2]. While hundreds of RNA binding proteins (RBPs) have functions relating to RNA executed by well characterized RNA binding domains, a recent set of mRNA pulldown studies has uncovered a rich set of unexpected RBPs with no known binding domains and functions that do not obviously require RNA[2-4]. Among these putative RNA binders are proteins belonging to the cyclophilin family, proteins that isomerize the amino acid proline from the lower energy *trans* conformation to the higher energy *cis* conformation[5].

Cyclophilin A (CypA), a highly conserved cyclophilin comprising a single isomerase domain[6], was among the identified novel RBPs[2-4]. CypA is ubiquitous in the cell, as it abundant in the nucleus and cytoplasm as well as being secreted[6,7]. It plays an essential role in protein folding and trafficking, cell signaling, and immune system response[6,7] and has recently been shown to interact with key transcription factors to mediate their transcriptional activity[8,9]. CypA is also biomedically relevant as the target of the immunosuppressive drug Cyclosporin A and is overexpressed in various cancers[7]. In addition to the mRNA studies, CypA and homologues have been implicated as RNA binders in several studies[10-12] though its isomerase activity does not seem to require RNA. It is thus an exciting model to study novel RNA-protein interactions. Here, I take a two-pronged approach to investigate CypA's RNA interactome: *in vitro* binding with previously implicated RNAs and *in vivo* RNA immunoprecipitation with subsequent bioinformatic analysis to identify relevant endogenous RNA classes that CypA may bind. The *in vitro* results suggest that CypA binds RNA with some specificity, as certain RNAs

interacted with CypA while others did not. The *in vivo* results differentiate a set of statistically significant enriched genes from a set of depleted genes, suggesting that CypA interacts sets of transcribed genes in the cell whose functional groups relate to known CypA activity.

Introduction and Background

RNA-binding proteins are essential in diverse cellular processes, ranging from gene expression and regulation to immune system response, governing such activities as mRNA splicing, translational regulation, and RNA stability, among others[1-4]. Previous research has identified prevalent canonical RNA binding domains such as the OB fold[13], RNA recognition motifs (RRMs)[14], Zinc-finger domains[15], and RGG and RG domains[16]. To understand the breadth of protein-RNA recognition, several unbiased studies have been done to characterize the mRNA-proteome of human cells to gain further insight into this important class of proteins[2-4]. In these studies, human cells were cross-linked with UV light to covalently link interacting molecules. All of the mRNA in the cells was then pulled down using a polyT tag that recognized the polyA tails of the mRNA and the attached proteins were identified using mass spectrometry[2-4] (**Figure 1)**. Interestingly, over 50% of the proteins identified using this unbiased strategy had no previously characterized RNA-binding domain[2-4]. Many of the identified proteins were metabolic enzymes whose functions do not apparently require RNA[2-4]. This led to exciting questions about RNA-protein interactions: how is RNA binding these unexpected proteins, what types of RNA molecules bind, and how does RNA interaction

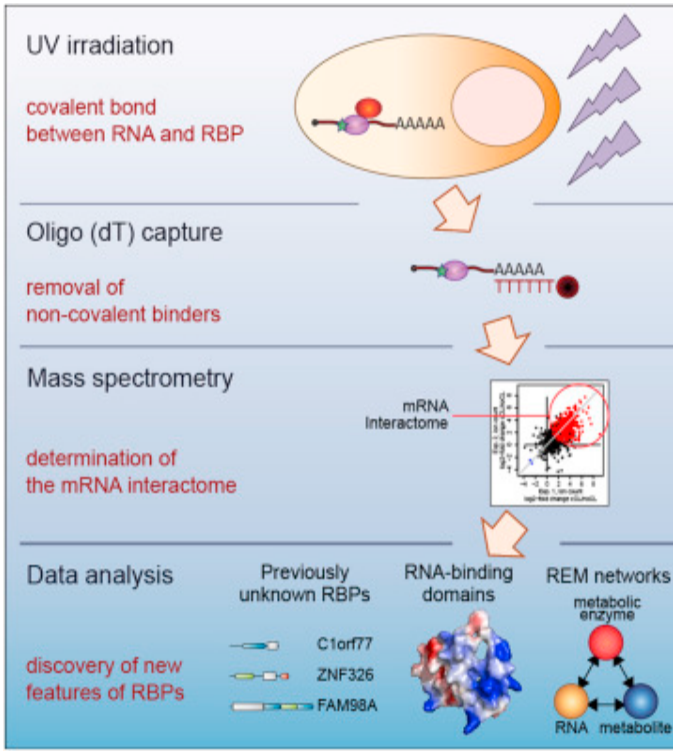


Figure 1. Identification of mRNA Binding Proteins. Study by Castello et al showing workflow. UV crosslinking, pulldown with polyA probe, identification of proteins via mass spectrometry,

contribute to, facilitate, or regulate these proteins' functions? Our lab chose the cyclophilin family of proteins to address these questions, as several were identified among the mRNA associated enzymes[2-4].

The cyclophilin family is an essential group of proteins structurally conserved in diverse organisms across all domains of life[5,6]. There are 17 human cyclophilins with various (some yet unknown) functions[5], but all include a cyclophilin-like domain (CLD) of 109 amino acids that isomerizes the amino acid proline from the *trans* to the *cis* conformation[5,6] (**Figure 2[7]**).

Due to its partial double-bond nature, the peptide bond of proline can exist in two conformations: the *trans* state, with amino acid side chains 180 degrees opposite, and the *cis* state, with adjacent side chains. Although both conformations exist in the cell, the *trans* conformation is slightly more energetically favorable. Cyclophilins stabilize the transition state for this isomerization and are thus critical in protein folding and assembly[5.6]. Cyclophilins also play a role in cell signaling, such as CypA in the regulation of a kinase involved in T-cell activation[7].

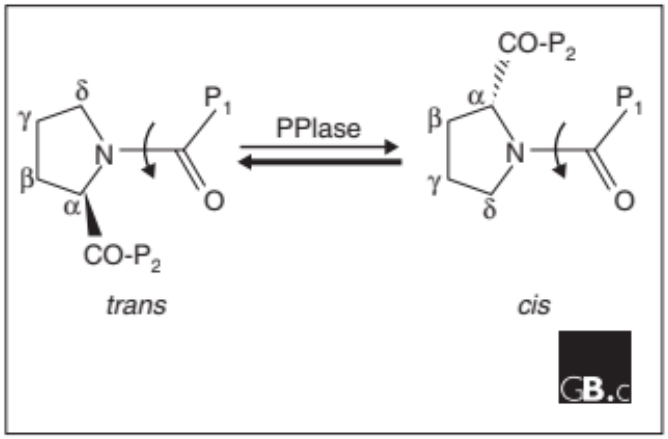


Figure 2. Cyclophilins isomerase the amino acid from the *cis* to the *trans* conformation[7].

Some cyclophilins, such as CypE, are known to have roles in mRNA processing and splicing and contain canonical RRMs (RNA recognition motifs) in addition to a CLD[5], so it is not surprising that they were implicated in the transcriptome-bound proteome studies. However, other cyclophilins which contain only the CLD, specifically CypA and CypB, were also

implicated[2-4]. Experiments conducted previously in our lab show that inclusion of the CLD of CypE increases its RNA binding affinity compared to its RRM alone[10]. RNA was also shown to increase CypE and CypA's isomerase activity[10]. The cyclophilin-like domain, then, is an exciting model to explore as a non-canonical RNA binder.

Cyp A was chosen as a particularly interesting target as an essential, archetypal cyclophilin. It is a small (18kDa, 174 amino acids) protein comprising a single CLD.[5-7] It is particularly ubiquitous— CypA is present in all mammalian tissues, found in the cytosol and nucleus, and is secreted[7]. In fact, it makes up about 0.1% of all cytosolic proteins[6]. Due to its high concentration in the cell, targeting it in *in vivo* studies is convenient: induced expression is not necessary— one can readily work with the endogenous quantities. CypA has been shown to mediate the activity of key transcription factors, including YY1[8], NF-κB[9], and the transcriptional circadian rhythm protein BMAL1.[18] Isomerization by CypA and other proline isomerases of a key proline bond in BMAL1 causes a conformation change that leads to BMAL1 interaction with different activators and repressors.[18] CypA is also involved in cell growth and is overexpressed in lung, breast, and liver cancer, among others[7]. Besides indicative of cancer, CypA is biomedically relevant as the target for the immuno-suppressive drug cyclosporin A.[5-7] Cyclosporin A binds to CypA to form a ternary complex with calcineurin that blocks calcineurin's immune system activation[5-7].

Previous RNA-CypA Interactions

As mentioned above, CypA's rate of isomerization has been shown to increase with the addition of a random RNA oligos in experimental assays[10], thus implying functional significance

of the CypA-RNA interaction. CypA and its yeast homologue Cpr1, which shares 65% sequence identity with CypA, have also been shown to interact with the genomic RNA of the tomato bushy stunt virus (TBSV) to inhibit viral replication[11], and fungal *Pi*CypA has also been shown to directly bind RNA[12]. CypA is thus an evidentially implicated, medically relevant, and cellularly essential protein whose RNA binding activities may provide an unappreciated level of cellular regulation.

Here, I describe my work characterizing CypA-RNA interactions using biochemical and bioinformatic approaches to identify transcribed genes and RNA sequences CypA binds. I took a two-pronged approach: cross-linking and immunoprecipitation *in vivo,* followed by RNA sequencing and bioinformatic analysis, and Fluorescent Anisotropy (FA) binding assays with previously implicated RNA *in vitro*. These two complementary investigations provide good evidence that CypA binds RNA both *in vivo* and *in vitro*, opening exciting new avenues for future research.

Materials and Methods

In vitro

Protein Expression: Nickolaus Lammer, a graduate student in the Wuttke lab, expressed protein in transformed *Escherichia coli* (*E. coli)* cells prior to my joining the lab. CypA was expressed with a His-SUMO tag at its N-terminus in a pET28b plasmid. The plasmid was transformed into *E. Coli* cells, which grew at 37◦C (with 50 g/ml kanamycin) in 2-X-Y-T media, then shocked on ice for 40 minutes. About 1 mL of 1 M of Isopropyl -D-thiogalactopyranoside was added per liter of culture to induce protein expression. Cultures were grown overnight at 18 °C. Cells were centrifuged at 5000 RCF to harvest and pellets stored at −20 °C.

Protein Purification: For the purification, cell pellets were resuspended in lysis buffer (50 mM Tris pH 8.5, 1000 mM NaCl, 10 mM imidazole pH 8.3, 10% glycerol, 0.1% Triton 100X) with an **EDTA**-free protease inhibitor. Lysis of cells was performed with a Misonix Sonicator 3000 with a half inch tip (15 seconds on, 15 seconds off, 10 to 12 times at 110 W) and then centrifuged for 30 minutes at 15,000 RCF. Lysate was poured over nickel-NTA beads and rocked at 4°C for 1 hour, then loaded onto a flow column. The column was washed three times with 15 mL of lysis buffer, then the protein was eluted twice with 15 mL of lysis buffer with 350 mM imidazole added (flow through collected each time). The protease Ulp1 was added to collected elution to cleave the His-tag[E]. This solution was dialyzed overnight with SpectrumSpectra/Por Dialysis tubing with a MW cutoff of 6-8 kDa (Thermofisher) in a 4 L container of dialysis buffer (50 mM Tris pH 8.3, 135 mM KCl, 15 mM NaCl, 10 % glycerol). Cleaved, dialyzed solution was poured over the nickel column to remove uncleaved His-tagged proteins and the flow through containing CypA collected. The flow through was concentrated at

4 °C on a 5K MWCO concentrator to a volume of about 2 mL. It was then run through the G75 gel filtration column and fractions containing CypA were collected and pooled. Pooled fractions were again concentrated to about ~800 uM, aliquoted into 15 20 uL volumes, flash frozen in liquid nitrogen and stored at -70 °C. Protein yield was ~4.32 mg/L

*RNA transcription:*A double stranded 633 nt DNA gBlock was ordered from IDT (Integrated DNA Technologies, **Figure 3**) containing the sequence from the genome of the Tomato Bushy Stunt Virus. In addition, primers designed to segment the DNA into 5 sections with added T7 transcriptase recruiting sequences were ordered.

GGAAATTCTCCAGGATTTCTCGACCTAGTTCGTTTATCTGGTGACTTGCGCTACCGTTG
CTTTGCGTAGAGAATTTCTCTCCATAATTATTATCTTTAGTTGTGGGGTTTGAAGGTTG
GGTCTACCTTTCGGGGGGATAAATTGTAACTTCCAACAAACAAGCGACATGTCTAGAA
GAAACGGGAAGCTCGCTCGCACTCCACAACCTACCAAAGGAGCCTTTGGACGTCTTTC
CCCGTTCAGGAAAGCGGTTTGTGAGAAGGTTGGGGTAGCCCACCGACTTGGGTATGA
TGGGTTTCTGTCATACTACAGCGGTGCGAAACTCCGTACTTACACACGAGCCGTGGAG
AGTCTGCATATCACACCTGTCTCCGAGAGGGATAGTCACTTGACTACCTTCGTAAAAG
CAGAGAACTGCAGAGCGAGTAAGACAGACTCTTCAGTCTGAGTTTGTGGAGATGAGT
GTAAATCTGGCATAGCATACAGGTTACTCTTGTTGGGTATTCCTGTTTACGAAAGTTA
GGTGTCACTTGTGGAAGCGGACCCAGACACGGTTGATCTCACCCTTCGGGGGGGCTA
TAGAGATCGCTGGAAGCACTACCGGACAACCGGAACATTGCAGAAATGCAGCCC

Figure 3. Full Length DI-72 DNA gblock. DNA sequence corresponding to the RNA sequence of the viral DI-72 RNA

First, a PCR (polymerase chain reaction) was run to amplify DNA segments before transcription. The reactions were done in volumes of 50 uL (1x concentrated Taq buffer from 10x concentrated stock (15 mM $MgCl_2$, 500 mM KCl, 100 mM Tris-HCl pH 8.3), 200 uM dNTPs, 0.5 uM forward primer, 0.5 uM reverse primer, ~250 ng template DNA, 1 unit/50 uL Phusion DNA Polymerase, nuclease free water up to 50 uL). DNA was melted at 98 °C for 15 seconds, annealed at 52 °C

for 30 seconds (primers annealed to template), and elongated at 72 °C for 30 seconds. This sequence was repeated 25 times.

After amplification, segmented DNA was transcribed in 200 uL volumes (1X Transcription Buffer as used by Milligan et al.[F], 24 mM MgCl2, 4 mM dATP, dGTP, dUTP, dCTP, 40 uL template from PCR, 97.6 uL nuclease free water, 10 mM DTT, 2 uL PPIase, 4 uL T7 polymerase). Reactions were left in the water bath for 2 hours. RNA purification was done by running a urea gel (7 M Urea, 8% Polyacrylamide) and cutting out RNA bands after visualization via UV shadowing.

Primer	Melting Temperature	Sequences
FWD1	57 °C	aaattctccaggatttctcgacc
REV1-T7	57-58 °C	TAATACGACTCACTATAGGcatgtcgcttgtttgttggaag
FWD2	59 °C	tagaagaaacgggaagctcgc
REV2-T7	59 °C	TAATACGACTCACTATAGGatcatacccaagtcggtggg
FWD3	58 °C	gggtttctgtcatactacagcg
REV3-T7	59 °C	TAATACGACTCACTATAGGttctctgcttttacgaaggtagtcaag
FWD4	58 °C	agcgagtaagacagactcttcag
REV4-T7	57 °C	TAATACGACTCACTATAGGacccaacaagagtaacctgtatg
FWD5	57 °C	attcctgtttacgaaagttaggtgtc
REV5-T7	58 °C	TAATACGACTCACTATAGGctgcatttctgcaatgttccg

Figure 4. Sequences of DNA primers used to transcribe Minus Strand RNA sequences from DI-72 DNA gBlock. The lowercase letters are sections of the larger DI-72 sequence and the capital letters are transcription enzyme (T7) recruitment sequences. Shown are primer names, melting temperature, and sequence.

RNA 3'-end labeling with fluorescein-5-thiosemicarbazide (FTSC): To prepare transcribed RNA for fluorescence anisotropy binding assays, they were labelled at the 3' end with the fluorescent probe FTSC. For the oxidation reaction, 350 pmol RNA were added to 20 mM Sodium Periodate ($NaIO_4$) in a volume of 50uL (nuclease free water up to 50uL) and incubated in the dark at room temperature for 20 minutes. 7.5 uL of 2 M KCl (0.25 M) were added and samples rested on ice for 10 minutes. Reactions were spun for 10 minutes at 16.1k RCF and supernatant transferred to a new tube. 1 uL glycogen and 140 uL of 100% ethanol were added and samples frozen overnight. Samples were then pelleted, washed with 70% ethanol, and resuspended in 50uL labeling solution (100 mM Sodium acetate pH 5.2, 1.5 mM FTSC). They were thoroughly mixed and incubated for 1 hour at 37 degrees Celsius. Samples were spun down, supernatant transferred to a new tube, 125uL of 100% ethanol added, and samples then frozen overnight. Next, they were washed 4 times in 70% ethanol (pelleted via centrifugation, resuspended in 70% ethanol) and resuspended in 30uL of RNA folding buffer. They were then passed through a G-25 spin column and the flow through was collected in an opaque tube. Verification of proper length and labelling effectiveness was done by running a denaturing PAGE gel and a dilution experiment in which intensity of diluted, labelled RNA was measured by a fluorimeter.

Fluorescence Anisotropy: A Fluorescence Anisotropy (FA) binding experiment was conducted with previously expressed and purified CypA and labelled RNA. Experiments were conducted with a constant concentration of RNA (3-6 nM) and 20 titrations of CypA (starting concentration of 66.5 nM, halved 20 times) in 20 uL binding buffer (50mM NaCl, 20 mM Tris, 10 % glycerol).

CypA and RNA incubated in the 384-well plate for 1 hour at room temperature before total anisotropy was measured. Anisotropy curves were fit to a single-site binding equation, $A = B + \frac{[RNA.CypA]}{[CypA]+K_d}$, where A is anisotropy, B is a baseline background, [CypA] is the concentration of CypA, [RNA.CypA] is the concentration of RNA bound to CypA. This simplified equation assumes that the ligand concentration is in sufficiently in excess (in this case, this will hold if the K_d for the CypA is sufficiently greater than RNA concentration). This assumption is valid given our conditions: reported K_d values were about 100-fold higher than RNA concentrations.

In vivo

Crosslinking: Due to COVID training restrictions that precluded me from access to the Biochemistry cell culture facility, a graduate student in the Wuttke lab (Nickolaus Lammer) facilitated the first step of this project by growing and cross-linking human HeLa cells using UV light or formaldehyde. The HeLa cells were grown in DMEM (Dulbecco's Modified Eagle Medium) containing 10% FBS and 1% penicillin-streptomycin solution. Crosslinking was done at 10-12 million cells. To induce crosslinking with UV light, 400 mJ/cm^2 of 254 nm UV light was applied to plated HeLa cells. For formaldehyde samples, cells incubated with 0.1% formaldehyde in 1 X PBS for 10 minutes. Crosslinking was quenched by adding 795 uL of 2.5 M glycine allowing plates to sit for 5 minutes. Both UV and formaldehyde crosslinked cells were then scraped and centrifuged (5 minutes at 1000 RCF), resuspended in 1 mL of PBS, centrifuged again at 4000 RCF for 75 seconds, and flash frozen (after supernatant removal) in liquid nitrogen. Frozen pellets were stored at -70 °C.

CypA-RNA Immunoprecipitation: Cross-linked cells were lysed with a biorupter in ~500uL lysis buffer (50 mM tris, pH 8, 150 mM NaCl, 0.1% SDS, 1% triton X-100, 5 mM **EDTA**, 0.5% sodium deoxycholate, 0.5 mM DTT, 1x Halt protease inhibitor cocktail (PIC) from Thermofisher, 100 units/mL of RNAseout). Cell lysate was collected after centrifugation and diluted with an equal volume of binding/washing buffer (25 mM Tris, pH 7.5, 150 mM KCl, 0.5% NP-40, 5 mM **EDTA**, 0.5 mM DTT, 1x PIC, and 100 U/mL RNAseOUT). This mixture incubated with 33 uL of protein G Dynabeads (Invitrogen) equilibrated with washing buffer (without PIC and RNAse out) for 45 minutes to remove nonspecific bead binders. After magnetically removing beads, a 50uL aliquot was taken as an input and frozen at -20 °C. Then, the supernatant was incubated with 4.4 ug of mouse derived IgG anti-CypA antibody for 2 hours at 4°C. To pull down the CypA antibody, 50 uL of Dynabeads (Invitrogen) equilibrated in binding/washing buffer was added and incubated for 1 hour at 4 °C. The beads were pulled to the side and supernatant removed. Beads were washed twice with binding/washing buffer and frozen at -20 °C.

RNA Isolation: First, 45 uL of RNAse or nuclease free H_2O and 33 uL of 3x reverse-crosslinking buffer (3x PBS (phosphate buffered saline), 6% N-lauroyl sarcosine, 30 mM **EDTA**, 15 mM DTT) and 2 uL RNAseOUT was added to the beads. 36 uL of reverse-crosslinking buffer and 2 uL of RNAseOUT was added to the input samples. Next, 20 uL of proteinase K was added to each input and bead samples and heated at 42 °C for 1 hour and for another hour at 55 °C. To separate RNA from other soluble molecules, 1 mL of trizol and 1.080 mL trizol were added to the beads and inputs, respectively, and vortexed. 200 uL of chloroform were added to beads and 216 uL chloroform to input followed by about 20 seconds of vortex. After centrifuging for

15 minutes at 16k RCF (relative centrifugal force), the aqueous layer was removed and 500 uL of isopropanol (to precipitate RNA) was added per 1 mL of trizol used. After addition of 1 uL of glycoblue and placing samples on ice for 10 minutes, they were centrifuged at 16 RCF for 15 minutes and the supernatant was removed. Samples air dried for about 5 minutes, then 85 uL of nuclease free H_2O was added to each sample. DNA was then digested by adding 10 uL of 10x DNAse 1 reaction buffer and 1 uL of DNAse 1 to each sample, and they were allowed to sit at room temp for 15 minutes. RNA was then purified using Qiagen RNeasy MinElute Cleanup Kit (74204) quick start protocol. RNA samples were frozen at -20°C.

Library Preparation: Libraries were prepared for sequencing with KAPA RNA HyperPrep Kit with RiboErase (HMR) for Illumina Platforms. This protocol consisted of removing rRNA via hybridization to complementary DNA oligonucleotides followed by addition of RNase H to remove rRNA-DNA duplexes and DNase to remove excess DNA oligos. RNA was then fragmentated to desired library size (~300 nt) using heat and magnesium. Due to its catalytic activities, RNA is able capable of self-cleavage, which is facilitated by Mg^{2+} ions stabilizing RNA structures conducive to cleavage and accelerated by heat[C]. After RNA fragmentation for 6 min at 85 °C, the first strand of cDNA was synthesized with random priming, followed by synthesis of the second strand, which incorporated dUTPs to mark the strand and a dAMP to the 3’ end. UMI (unique molecular identifiers) and Illumina adapters were then ligated via 3' dTMPs. Adapter sequences record condition (Formaldehyde or UV input or pulldown) and allow ligation to the flow cell, while UMIs allow accurate counts of each unique RNA molecule. Libraries were then amplified (only the strands without dUTP was amplified so that the original RNA molecules

would be sequenced). Adapter-ligated libraries were sequenced with Illumina NovaSeq Paired End (150 cycles) sequencing. Six samples in total were sequenced (an input and pulldown for two formaldehyde conditions and an input and pulldown for one UV condition). For the first formaldehyde experiment, 22,917,431 reads were obtained for the input and 28,187,490 for the pulldown. For the second, 30,177,674 were obtained for the input and 76,393,313 for the pulldown. For the UV experiment, 31,017,183 reads were obtained for the input and 31,339,559 for the pulldown.

Data Analysis Pipeline: The data pipeline for sequencing data consisted of trimming adapters and low-quality reads with Trim Galore (Barbara Bioinformatics), removing duplicate reads using UMIs with a python script written by Taeyoung Hwang, a postdoc in the Rinn lab, aligning reads to the human genome (GRCh38) using the STAR aligner, and counting the number of reads (RNA molecules) per gene or genomic feature with the Rsubread function featureCounts. The pipeline was primarily developed by a Taeyoung Hwang, a postdoc in the Rinn Lab. Further analysis (calculation of fold change IP/input and input) was done through a combination of bioinformatic software (DESeq2[21]) and python scripts I wrote. Clustering of enriched gene data was accomplished using DAVID[22] tools. Code is available on request.

Results and Discussion

Fluorescence Anisotropy to Assess CypA DI-72 RNA Binding Affinity

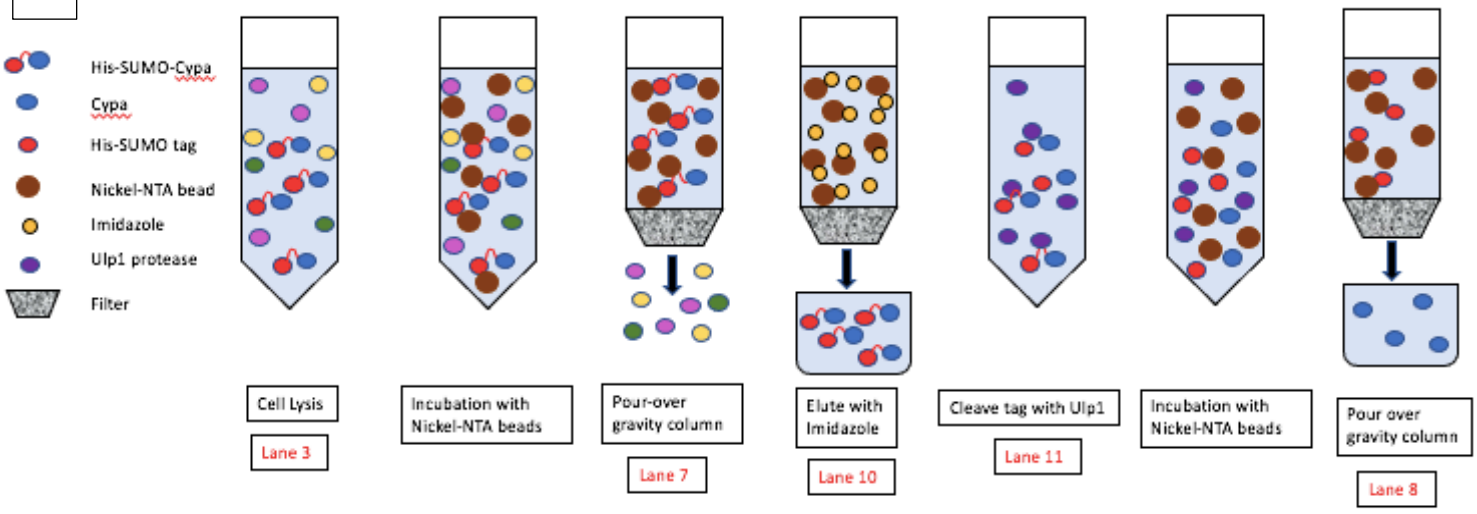
CypA Expression and Purification

Before determining what RNAs CypA interacts with *in vivo,* I sought to ascertain if it binds RNA *in vitro* by expanding on previous studies that suggested RNA-CypA binding[11,12]. Prior to conducting these *in vitro* binding assays, it was necessary to express and purify human CypA. CypA was expressed in *E. coli* cells with a His-SUMO tag at its N-terminus, which allowed for capture of the protein. The His-SUMO tag contained six histidine amino acids and a SUMO protein sequence. The histidines have a high affinity for binding metal ions[23], enabling the use of a commercially available nickel column to separate the tagged protein from other cellular components. The SUMO portion of the tag allows for removal of the tag from CypA after separation as it is recognized and cleaved by the protease Ulp1[B].

CypA purification consisted of lysing cells, pouring the supernatant (soluble cell parts) over a nickel column to which the histidine tag binds, elution of the protein and tag by washing with a high imidazole buffer (imidazole pushed the His-tag from the nickel column by competing for binding to charged metal beads), and cleavage of the tag from CypA with Ulp1 (**Figure 5A)**. Clean-up of the tag was accomplished by another flow over the nickel column, and the resulting flow-through concentrated and further purified by gel filtration.

Figure 5B shows a 15% PAGE gel with samples taken at every step in the protein purification process. PAGE gels denature proteins and are thus used to separate them by size (larger proteins migrate more slowly and appear higher on the gel, while smaller proteins migrate more quickly and appear lower on the gel). Lanes 3 and 7 show the supernatant and

A



B

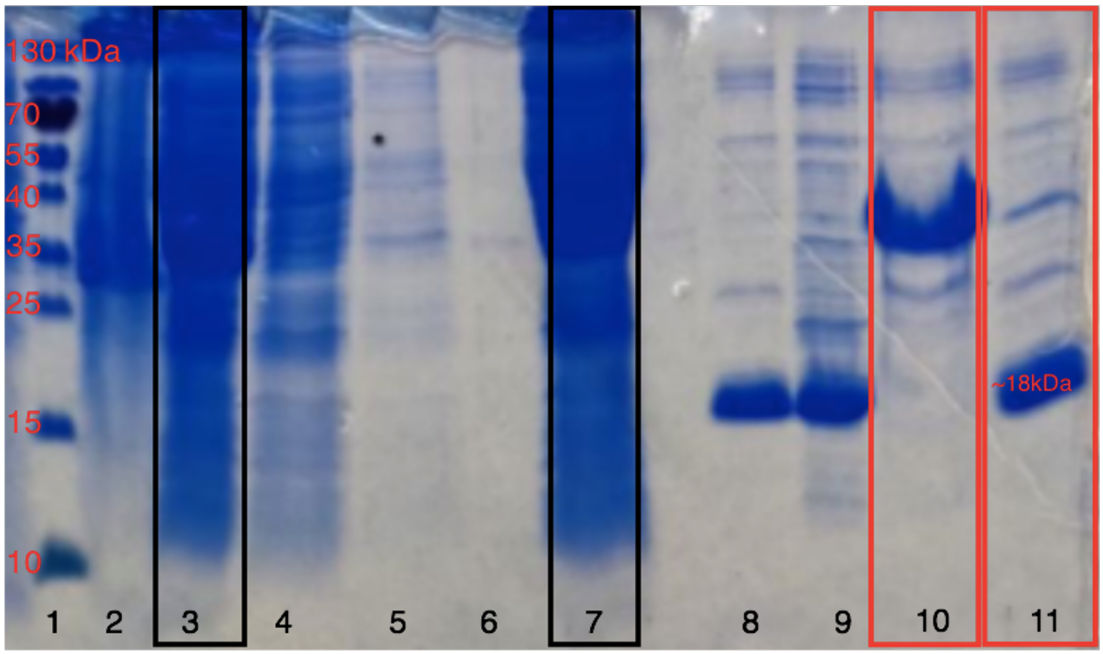


Figure 5. a) Protein Purification Process. Shows major steps of CypA purification: lyse cells, incubate with Nickel-NTA beads, pour over gravity column, elute with imidazole, cleave His-SUMO tag with Ulp1, incubate again with Nickel-NTA beads, pour over gravity column to remove his-SUMO, collect purified CypA. **b)** Denaturing PAGE gel with samples taken at every step in the protein purification process. Lane 1: Ladder (with molecular weight labelled in kDa), Lane 2: Pellet, Lane 3: Lysis supernatant, Lane 4: Wash 1, Lane 5: Wash 2, Lane 6: Wash 3, Lane 7: First flow-through over nickel column, Lane 8: Second flow-through over nickel column, Lane 9: Nickel beads after second flow through, Lane 10: Elution, Lane 11: Dialysis with Ulp1

lanes contain all soluble cell proteins. While lane 7 should not include as high a concentration of His-SUMO-CypA (which is expected to stick to the column), the effect of this one protein construct's removal is indistinguishable from the signal that results from running all soluble cell components. Important to note are lanes 10 and 11. Lane 10 shows the elution (after the His-CypA has been competed off by the nickel column by imidazole). A thick band in the upper third of the lane can be seen. This is His-SUMO-CypA and indicates efficient protein expression and elution. Lane 11 shows a sample taken after dialysis, in which the His-SUMO tag was cleaved from CypA. In this lane, there is thick band that ran further than the tagged CypA, indicating successful cleavage from 32 kDa to 18 kDa. This band is slightly higher than the 15 kDa ladder marker, which is further reassuring as CypA's molecular weight is 18 kDa.

A second flow over the nickel column removed cleaved His-tags but did not purify the sample of unwanted higher molecular weight species (see Lane 8, Figure 3). These were removed by 120 mL size exclusion column, which separated the solution into 2 mL fractions of different molecular weight. **Figure 6** shows fractions collected at and surrounding CypA's molecular weight (18kDa). The band for CypA are clean and, again, run to slightly above the 15 kDa ladder mark, indicating a successful purification process.

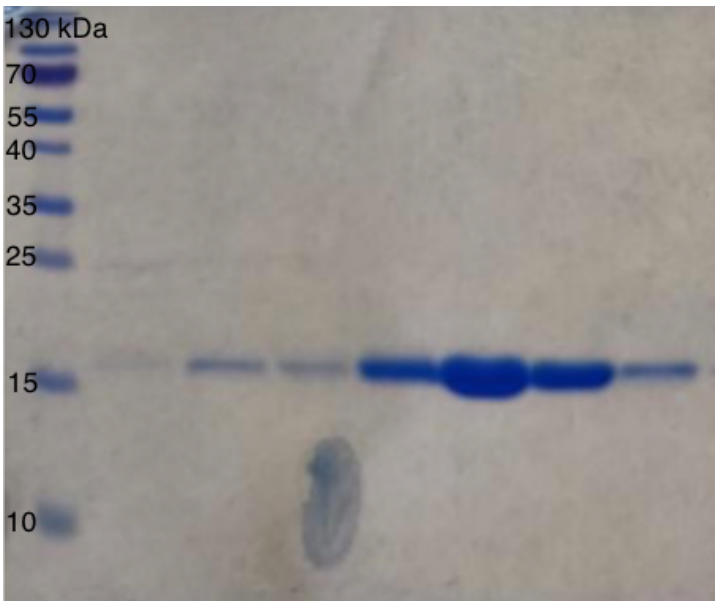


Figure 6. Purified CypA. Denaturing PAGE gel showing (left) ladder with labelled molecular weights and purified CypA fractions.

Selection and Transcription of RNA

With a highly purified supply of CypA, I next obtained a suitable RNA molecule for *in vitro* investigation. Several previous publications provided RNA sequences that bound CypA or CypA homologues in plants and yeast[11,12]. A particular relevant study in 2013 by Kovalev et al. showed that CypA inhibits the replication of the Tomato Bushy Stunt Virus capsid *in vitro,* plausibly by binding to the virus' genomic RNA[11].

Kovalev et al. tested CypA binding with a 633 nucleotide RNA sequence (referred to as DI-72) comprising several non-contiguous segments of the viral genome: the 5' and 3' ends, a template recognition and recruitment element from one of the protein coding sequences, and a replication enhancer[11]. Both the plus strand (coding) and the minus strand (non-coding) were shown to bind CypA in a well-shift EMSA (Electromobility Shift Assay)[11]. The minus strand bound more tightly than the plus strand, with an estimated dissociation constant (K_d) in the micromolar range[11]. K_d measures the thermodynamic tendency of interacting molecules to fall apart and is calculated by dividing free concentrations of molecules by bound concentration at equilibrium: $\frac{[A][B]}{[AB]}$ where [A] and [B] are free concentration of molecules and [AB] is bound concentration. A lower K_d thus corresponds to tighter binding and higher affinity between two interactors. A well-shift experiment shows unbound RNA that runs into the gel lane and bound RNA that does not migrate from the well, thereby making it difficult to estimate an accurate K_d (intermediate concentrations between bound and unbound states are necessary). This estimated CypA-DI-72 K_d, then, was an exciting lead but warranted further investigation.

I thus obtained from the authors the full-length DI-72 sequence and ordered it as double-stranded DNA from which I could transcribe desired RNA segments. As mentioned

previously, the full-length molecule is 633 nt and I sought to hone-in on a smaller RNA sequence that CypA might bind. By designing specific DNA primers, I transcribed from the longer sequence five sections corresponding to the non-contiguous segments out of which they assembled the DI-72 molecule. I amplified the DNA, then transcribed five different sections (minus strand) with the corresponding primers. The five resulting oligos (referred to as Minus 1-5) were between 82 and 167 nucleotides long.

To prepare for a Fluorescence Anisotropy assay, the five RNA strands were labelled at their 3' end with FTSC (fluorescein-5-thiosemicarbazide), a fluorescent probe that absorbs light at 495 nm. The ligation of the fluorescent probe FTSC is a two-step process: first, the introduction of sodium periodate, which opens the carbon ring of ribose at the 3' end of the RNA molecule to form two reactive aldehyde groups and second, incubation with FTSC, which covalently attaches to the opened ring. **Figure 7**[24] illustrates the FTSC labelling process.

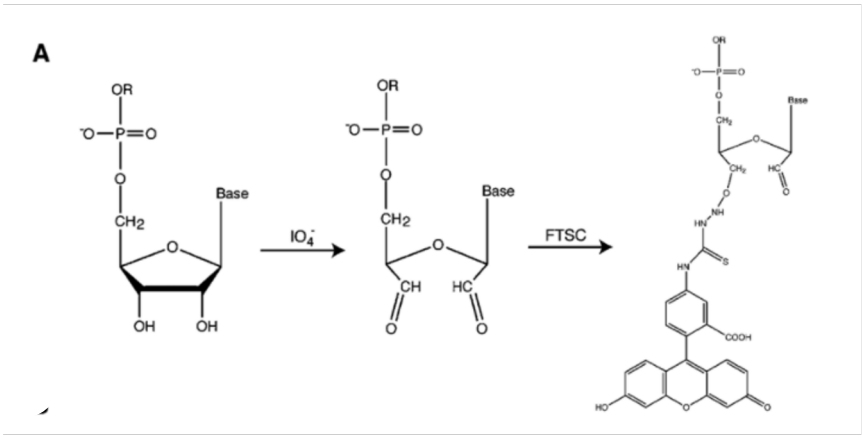


Figure 7. RNA 3' Labelling with FTSC. Illustrates the two steps of FTSC labelling: Opening of the ribose carbon ring with sodium periodate (IO_4) and covalent attachment of FTSC to a resulting aldehyde group. As published by Pagano et al.[24]

To check labelling efficiency, I ran the five RNA molecules on a Urea gel (denatures RNA) at the same concentration and excited the fluorophores with a gel imager using a fluorescent channel. **Figure 8** shows that the RNA molecules were effectively labelled as dark bands appear at the expected relative running distances. The Minus 1 strand was the longest of the five at 167 nucleotides, while Minus 4 strand was the shortest at 82 nucleotides. This was accurately reflected in the running distances. Strands 4 and 5 were not labelled with the same efficiency as the others (weak bands midway through the lane), making it necessary to use higher concentrations of these two strands in FA trials.

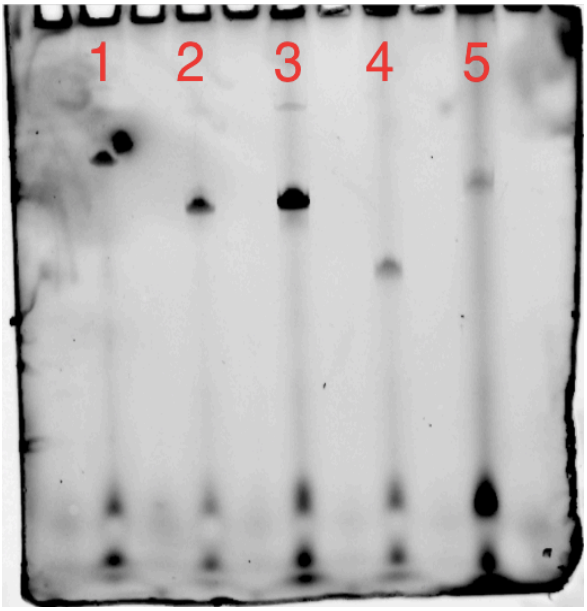


Figure 8. Urea gel with labelled RNA. Lane 1: Minus 1, 167 nt, Lane 2: Minus 2, 121 nt, Lane 3: Minus strand 3, 122 nt long, Lane 4: Minus 4, 82 nt, Lane 5: Minus 5, 131 nt

Binding measured by Fluorescence Anisotropy: I then tested CypA's affinity for the five minus sense strands in several Fluorescence Anisotropy (FP) assays. Fluorescence Anisotropy uses polarized light to excite the fluorophores attached to labelled ligands (in this case, the FTSCT on the RNA). At concentrations of protein below the K_d, most of the ligand is free and, after excitation, tumbles and emits light in different orientations. With higher protein concentration above the K_d, the ligand is bound in a larger complex so tumbles more slowly and emits a greater fraction of light parallel to the excitation light than in other orientations (see **Figure 9**). FP compares the intensity of perpendicular (free, tumbling) light to parallel (bound) light to calculate the fraction of ligand bound to the protein. Specifically, anisotropy is calculated as $\frac{I_{par}-I_{per}}{I_{par}+2*I_{per}}$, where I_{par} is intensity of parallel light and I_{per} is intensity of perpendicular light.

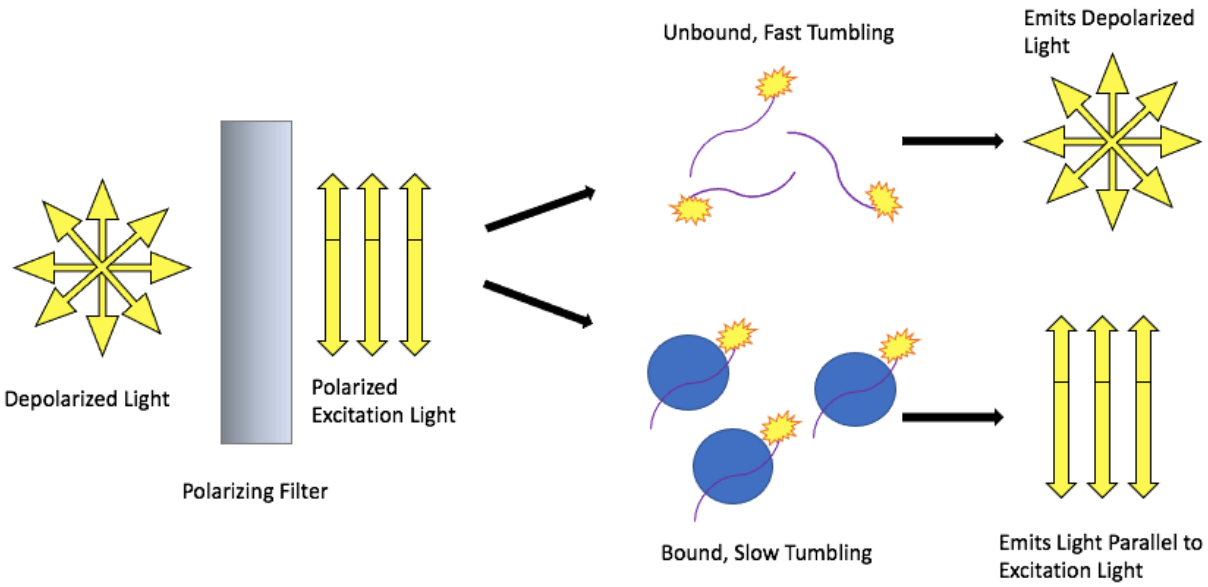
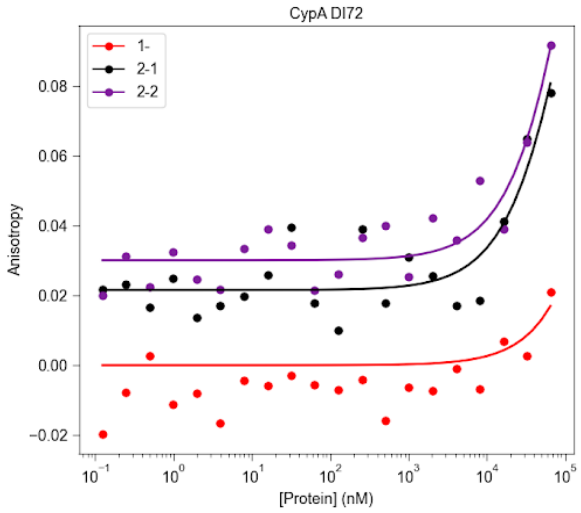


Figure 9. Fluorescence Anisotropy. Fluorescently labelled ligand is excited by polarized light and emits depolarized light when unbound and quickly tumbling and polarized light with bound and slowly tumbling.

The first two FA trials were run on all five RNA strands and a negative control (the biological RNA Gas5). These trials suggested an increased anisotropy as CypA concentration increased for the DI-72 derived RNA, but no shift in anisotropy for the negative control Gas5. These results also indicated preferential binding and implied binding specificity: CypA caused a shift in measured intensities for previously implicated RNA molecules, with different shifts for different sequences, but no shift for an arbitrarily selected negative control. However, the overall intensity of anisotropy for these trials decreased with CypA concentration (likely due to RNA crashing out of solution) and thus a low signal to noise ratio. These results were thus considered qualitative and the issue corrected in a final trial.

I conducted a final FA experiment with the two DI-72 RNAs that showed the most consistent anisotropic shifts between previous, qualitative trials: Minus 1 and Minus 2. **Figure 10** shows that while the Minus 1 strand shows an upward trend at the highest limit of CypA concentration, the Minus 2 strand undergoes a definite shift as CypA concentration is increased. Both the upper curves were generated with the same Minus 2 strand RNA at the same concentration and thus show similar offsets of background intensity. The bottom curve, showing binding to the Minus 1 strand, was



	1-		2-2
Kd (nM)	29450663.17	160981.08	208994.48
S	7.7061	0.206	0.257
O	0.0	0.0216	0.0301
R^2	-0.0645	0.7628	0.8187

Figure 10. CypA binds DI-72 derived Minus 2 RNA. CypA binds two replicates of Minus 2 RNA strand, as indicated by upward trend in anisotropy. Signal is weaker for Minus 1 RNA, which begins to increase in anisotropy at the upper limit of CypA concentration.

generated at a slightly lower concentration, and correspondingly has a lower offset of background intensity.

The trial tested two Minus 2 strand RNAs, which reassuringly produce anisotropy curves of similar trends. The K_d reported (about 185 uM averaged between Minus 2 strand curves) is a lower limit estimation, as the anisotropy curves do not plateau at the upper end of CypA concentration. This lower-level estimation suggests binding and is consistent with that reported by Kovalev et al.[11] (in the micromolar range). These initial FP assays are consistent with the literature and provide good evidence that CypA binds RNA *in vitro*. They also suggest some specificity, as CypA bound certain strands (Minus 2 being the best characterized) more tightly than others and did not bind the negative control Gas5.

In Vivo RNA Immunoprecipitation

Formaldehyde and UV RNA Immunoprecipitation

The first goal of my thesis was to verify, with a well-established equilibrium technique, that CypA is capable of RNA binding *in vitro.* The above results indicate such capabilities, as well as suggesting binding specificity. While ascertaining that CypA binds the DI-72 RNA is important and verified the RNA binding activity of CypA, a further goal of my thesis and the next pertinent problem was to identify what RNA molecules CypA binds in human cells, thus providing information as to the biological relevance of CypA-RNA interaction.

To determine the identity of RNA molecules that interact with CypA in human cells, I performed two types of RNA immunoprecipitations in HeLa cells: formaldehyde (f-RIP) and UV (UV-RIP). A strength of this approach is that CypA is expressed endogenously at high levels, precluding the use of strategies that require overexpression of the target for detection. This ensured that I was observing complexes present at natural conditions. These procedures used two different methods to crosslink (covalently link interacting molecules) cells before extracting the target protein. Formaldehyde (FA) crosslinking covalently links amine groups with high efficiency, allowing extensive linkage of molecules that are proximal but may not be directly interacting[25]. UV cross-linking, which uses 254 nm light to irradiate cells, is more specific and cross-links directly interacting molecules ("zero distance linkage"). However, UV cross-linking efficiency is low: certain studies estimating only up to about 5% efficiency[26]. Both formaldehyde and UV crosslinking immunoprecipitations were performed for deeper insight into CypA's RNA interactome, as formaldehyde linkage is broad and more efficient and UV linkage is specific but less efficient.

The basic steps of an RNA immunoprecipitation are illustrated in **Figure 11.** Following crosslinking, lysed cells were incubated with a mouse derived, Immunoglobulin G (IgG) antibody specific to CypA (anti-CypA, ab58144). The antibody was previously verified by Western Blot. IgG antibodies are made up of four polypeptide chains: two heavy chains and two light chains that together form a flexible Y shape[31]. While the antigen binding region of the antibody is highly variable (to bind diverse antigens), the region of the two heavy chains at the bottom of the Y shape is highly conserved and can be targeted to precipitate antibody-antigen complexes[31]. After incubation with anti-CypA, magnetic beads specific to IgG antibodies were used to immunoprecipitated the antibody bound CypA that was covalently attached to RNA. A proteinase was then added to digest CypA, freeing RNA molecules, and RNA was purified from the samples (referred to as "pulldown" samples). An "input" sample of RNA was also taken and purified before addition of the antibody but after an initial incubation with beads. This

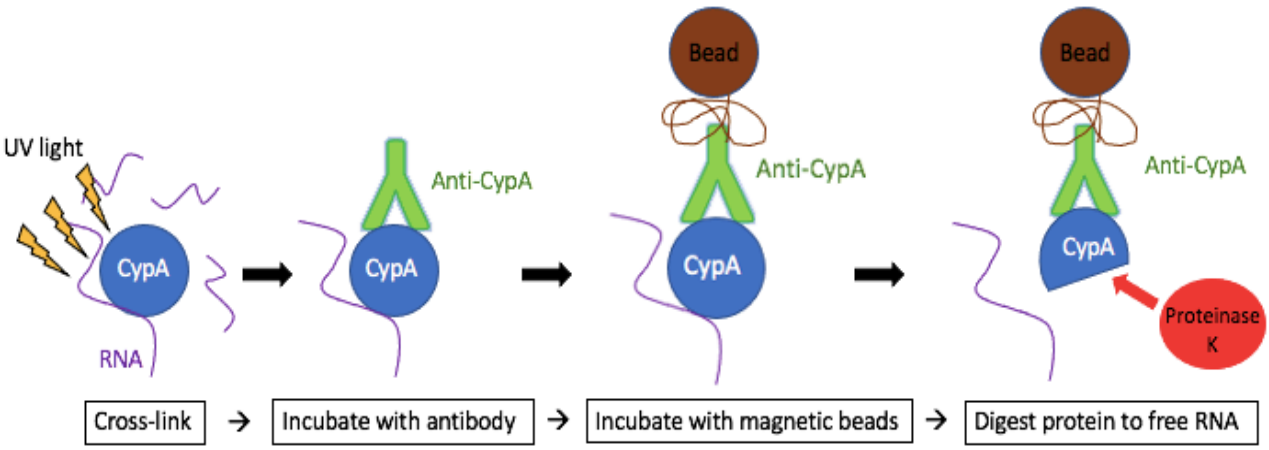


Figure 11. RNA UV Immunoprecipitation. Cells are crosslinked by irradiation with UV light, incubated with anti-CypA, incubated with magnetic beads that bind anti-CypA, pulled down, and RNA is detached by digestion of CypA and other proteins by Proteinase K.

initial incubation removed non-specific molecules that bound to magnetic beads and may thus have otherwise been pulled down with relevant RNA molecules. The result is a pool of RNA molecules that were pulled out due to a covalent interaction with CypA.

While some RNA-protein interaction purification techniques allow for sequencing of the binding site, the described approach frees the entire interacting RNA molecule. Thus, after library preparation and sequencing, only information about the full length RNA molecule is available, not the binding site.

RNA Sequencing

RNA libraries for fRIP and UV-RIP experiments were prepared for sequencing by fragmentation to desirable read size (~300 nt) using heat and magnesium ions (Mg^{2+}). RNAs have catalytic activity and thus the capacity for self-cleavage[C]. Heat and metal ions accelerate the process of self-degradation and Mg^{2+} in particular has been shown to form complexes with RNA molecules to stabilize certain structures which are conducive to RNA catalytic cleavage[C]. Library preparation continued after self-fragmentation of RNA with reverse transcription (converted molecules to complementary DNA), ligation of double stranded DNA Illumina adapters, and PCR (polymerase chain reaction) amplification. Illumina adapters contain sequences necessary for 1) adhesion to an Illumina sequencing flow cell by complementary base pairing and 2) “barcoding” or labelling molecules to identify what sample they came from. This library prep also included the addition of UMIs (Unique Molecular Identifiers), which are unique sequences of nucleotides given to each molecule. These are useful because of the error inherent in amplifying libraries for sequencing—certain molecules may be preferentially

amplified by PCR and thus make up a larger portion of sequencing reads. By counting UMIs rather than sequences, it is possible to accurately determine how many times an identical RNA sequence occurred in each sample. An illustration below (**Figure 12**) clarifies the use of UMIs.

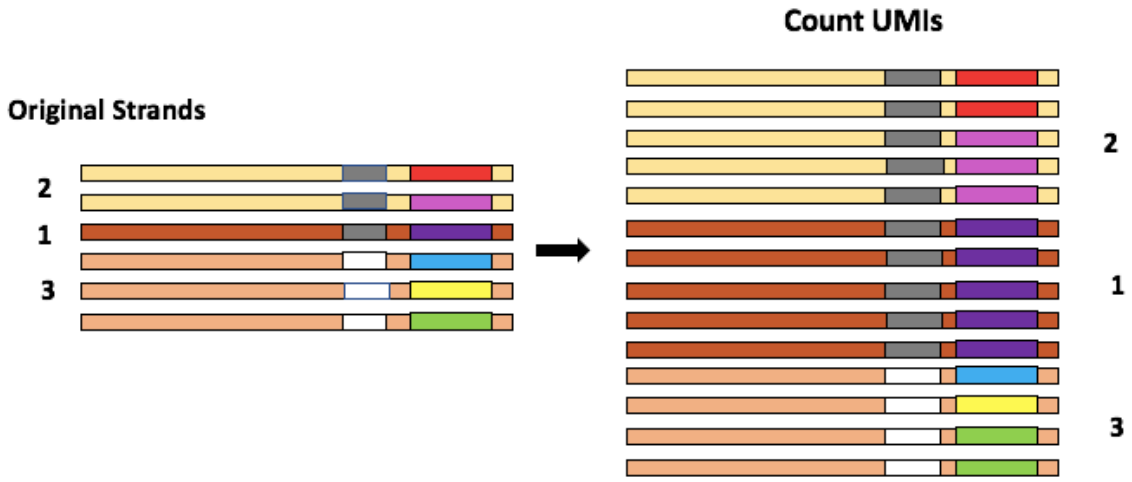


Figure 12. After amplification, UMIs allow accurate counts. Each different color represents a different RNA sequence. All unique RNA molecules (same sequence or not) will be tagged with a unique sequence of nucleic acids. After duplication, the number of UMI sequences is the number of original strands of any sequence the sample contained. Each RNA molecule also contains an adapter (grey or white) that indicates what sample it was from.

Six samples of RNA were sequenced: an input and pulldown from two FA experiments and one UV experiment. Amount of reads per sample ranged from 22,917,431 to 76,393,313 and 86% of reads across samples had a quality score > Q30. Quality scores are calculated based on the number of unidentifiable nucleotides in sequences: a quality score of Q30 represent an average error of 1 in 1,000 nucleotides. Both UV and FA inputs and pulldowns (pre- and post-immunoprecipitation) were processed by removing low-quality and duplicate reads (using UMIs) and aligning unique reads to the human genome (version GRCh38). Unique reads per genomic feature were then counted.

Determination of Statistically Significant Enrichment and Depletion

I then determined which genes were enriched in pulldown samples as compared to input samples using DESeq2[21]. DESeq2 normalizes counts per gene based on total read counts across all genes for a particular sample and averages the normalized counts per gene across sample replicates, which accounts for library size differences. It then calculates a logarithmically scaled fold change: $\log_2$(average normalized pulldown/ averaged normalized input)[21]. For each gene count, DESeq also calculates p-values for enrichment using the Wald test with the null hypothesis that there was no differential expression across sample groups (pulldown and input)[21]. P values were adjusted using the Benjamini and Hochberg method to scale p-values and reduce the number of false positives given a certain p-value cutoff[21]. As it uses averaged read counts, DESeq2 requires replicate experiments. I was able to conduct two formaldehyde immunoprecipitations but only one for UV, so consider just the FA enriched genes in the following analysis.

I considered as significant genes with an adjusted p-value of less than 0.05. **Figure 13** shows a plot with all genes comparing their significance to their log2 fold change. The negative of the $\log_{10}$ of adjusted p-values is plotted on the y-axis such that the higher the gene, the more significant its fold change. The outward and upward sweep of this plot from the origin reveals a general trend for enriched genes: the greater magnitude of the fold change, the more statistically significant the calculation of fold change. This increases confidence in the likelihood

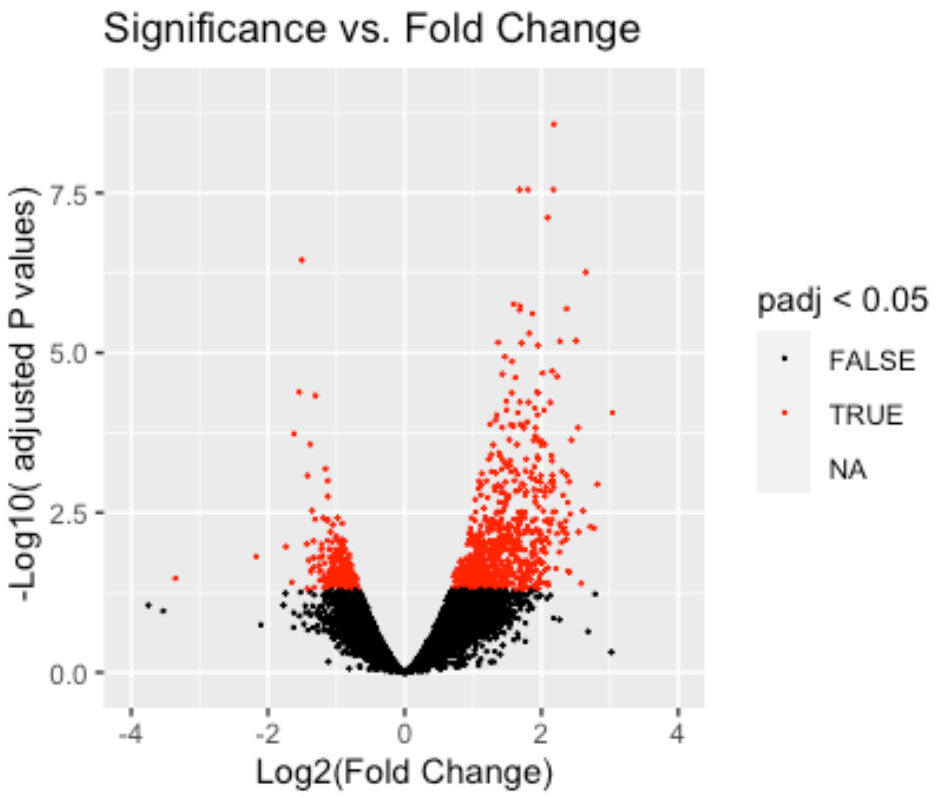


Figure 13. Significance vs. Fold Change of all Genes. Log2 of Fold Change is plotted on the x-axis with -log10 of adjusted p-value on the y-axis for easy visualization of statistically significant genes. Genes with adjusted p-values of less than 0.05 are plotted in red. Genes to the right of the origin (positive log2 fold change) are enriched, while genes to the left are depleted in pulldown vs. input.

that transcribed genes with a higher fold change are indeed interacting with CypA. Figure 8, then, indicates that some transcribed genes are significantly enriched (positive log2 fold change) and others significantly depleted (negative log2 fold change) with respect to CypA interaction, as expected.

Clustering of Enriched Genes

There were 709 enriched genes observed to have a positive log2 fold change and an adjusted p-value of less than 0.05. While this large number provides little information as an unannotated list of genes, patterns which emerge after grouping these genes by biological function annotation reveal specific, enriched pathways. These pathways may suggest a functional role for the RNA-binding activity of CypA.

The DAVID (Database for Annotation, Visualization and Integrated Discovery) clustering tool was used to identify several biological functional groups in which enriched genes occurred[22]. DAVID calculates similarity between genes in a submitted list by comparing gene annotation terms (such as biological function and pathway) obtained from its extensive annotation database[22]. It then clusters genes based on similarity distances[22]. DAVID clustered the 709 enriched genes into groups based on biological function and calculated the significance for each group (based on how many genes belonging to a functional group occurred in the given gene set compared to how many genes would occur in that group by chance). **Figure 14** shows the top clustering results with bar height corresponding to the number of genes that occurred in each group and color to adjusted p-value. Note that this p-value is associated with the significance of the functional group, not the enrichment of the genes within it.

CypA interacts with genes involved in transcription, transcriptional regulation, and kinase activity with the highest significance. There is overlap between top functional groups (zinc fingers and plexins are metal binders, zinc fingers are transcription factors, ATP-binders can be kinases). The diversity of cellular location in functional groups, with cell-to-cell adhesion (extra-cellular) as well as transcriptional regulation (nucleus) is unsurprising due to the ubiquity of CypA: it is found in the nucleus, cytoplasm, and is secreted[7]. In the nucleus, CypA is known to mediate translation through interactions with major transcription factors, specifically YY1[8] and NF-κB.[9] It has also plays a role in kinase signaling by proline-directed regulation of such molecules as the tyrosine kinase Itk[19]. That its most significant set of interactions is with genes coding for proteins involved in transcription, transcriptional regulation, and kinase activity is therefore consistent with our knowledge of the functional roles of CypA.

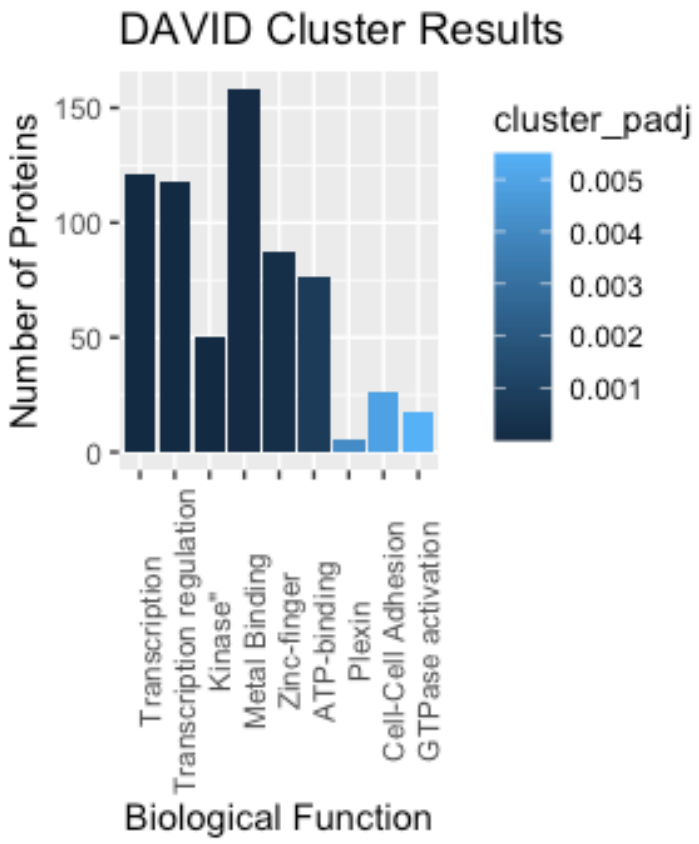


Figure 14. DAVID Clustering Results. Top significant biological functional clusters for the enriched DESeq2 genes were identified with DAVID. Bar heights are number of genes belonging to each cluster, colored according to adjusted p-value for that cluster.

One important caveat should be considered for this set of formaldehyde-linked interactions: formaldehyde links broadly and some of the enriched genes may be due to indirect interactions, or RNA molecules that a protein partner of CypA interacts with. NF-κB, for example, regulates immune system response genes[22] and Rab GTPases are known to influence inflammation and immunity[27]. Among the set of GTPases, then, could be indirect interactors of CypA. YY1, another known CypA mediated protein[8], forms homodimers that are "stabilized by low specificity RNA binding."[28] The diverse set of enriched formaldehyde genes may thus contain YY1 and NF-κB RNA interactions, as well as other indirect or secondary interactions. Thus, in the future it will be critical to cross-reference the formaldehyde crosslinking data with

the UV crosslinking data to determine a refined set of RNA targets that are due to direct interactions.

Consideration of Robustly Enriched Genes

The above analysis does not take into consideration the UV-crosslinked sample, as only one replicate was available. To take advantage of the rich insight that UV-crosslinking would provide, I ran my own analysis to calculate enrichment. I first normalized read counts for each gene by dividing gene counts by "counts per million," or total number of counts for each gene's sample divided by one million. I then calculated fold change (pulldown/input) for each gene (treating FA samples and UV samples separately). I considered as enriched genes with a fold

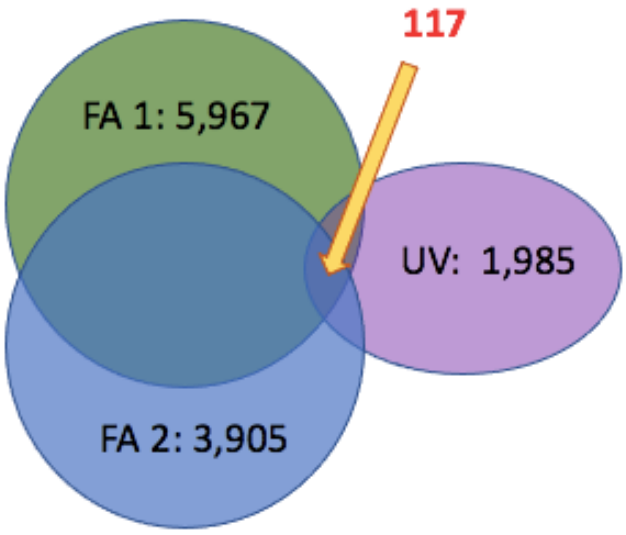


Figure 15. Venn-diagram of Enriched Gene Overlap between experiments. Three RNA immunoprecipitation experiments had different number of enriched genes, with 117 overlap between all three (enrichment calculated as fold change or pull down/input > 2).

change greater than 2. There were 5,967 such enriched genes for one FA trial, 3,905 in another, and 1,985 in the UV trial (**Figure 15**). 117 genes appeared with a fold change of greater than 2 in all three samples (FA1, FA2, UV).

Comparison of these results with enrichment data from DESeq2 analysis yielded 13 genes that were consistently enriched across methods of analysis and samples, referred to as robustly enriched genes (REGs). Other notable genes that emerged were seven from DESeq2 analysis with a strikingly low adjusted p-value ($p\ adj < 10^{-6}$), referred to as strikingly significant genes (SSGs). Next to determine was how the functionality of these two exceptional gene sets relate to known CypA activity. **Figure 16** shows the groups plotted in red (SSGs, extremely low p-adj in FA samples) and green (genes enriched in FA and UV samples) over other DESeq2 enriched genes (positive fold change, adjusted p-value < 0.05).

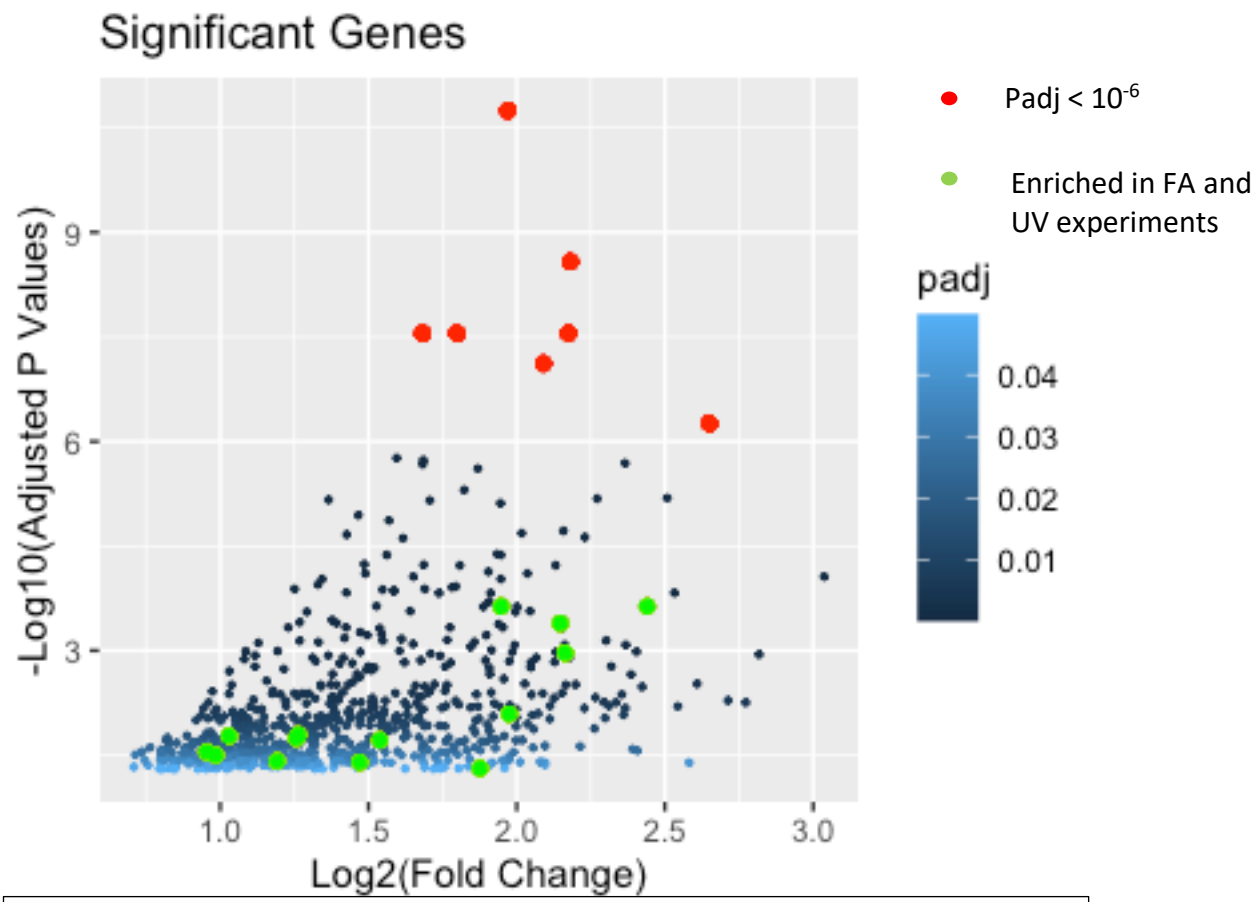


Figure 16. FA enriched genes with overlaid prominent significant genes. 709 significantly enriched genes from DESeq2 analysis (positive log2 fold change and p adj < 0.05) are plotted, colored by adjusted p value. Highlighted are genes that occurred in all methods of enrichment calculation (REGs, green) or had extremely low adj p-values (SSGs, red).

Of the 20 genes belonging to either of the two enriched gene sets, there were 15 with biological roles relating to key CypA activity, including proto-oncogenes and transcription factors. Figure 17 (with information from the UniProt database[29]) summarizes the functions of the SSGs and REGs. Overexpression of CypA has been shown to stimulate cancer cell growth in various studies by enabling cancer proliferation, blocking apoptosis, and regulating cell cycle progression[7]. Four of the SSGs and five of the REGs have been implicated in cancer or abnormal

cell growth, including *SKI,* which codes for a nuclear proto-oncogene[29], *MEGF6,* which induces metastasis[29], and RXRA, a nuclear receptor that has been implicated in renal cancer[29]. CypA also plays a critical role in the transcriptional activity of several key transcription factors. It has been shown to interact specifically with the previously mentioned transcription factor YY1, thereby altering YY1's activity[8]. Also mentioned, it has been shown to interact with a subunit of the general transcription factor NF-κB[9], which regulates genes involved in immune system response[22]. Three SSGs and four of the REGs are involved in transcription and gene regulation.

Figure 17. SSGs, REGs and their functions. Functional Information obtained from UniProt[29]

Gene Name	Enrichment Set	Function	Implicated in Cancer/Abnormal Cell Growth?	Implicated in Transcription/Gene Regulation?
SKI	SSG	Nuclear proto-oncogene	**X**	
MEGF6	SSG	epidermal growth factor, induces metastasis	**X**	
RXRA	SSG	Nuclear receptor, Host-virus interaction, Transcription, implicated in renal cancer	**X**	**X**
CAPN15	SSG	Zinc-finger repeats, transcription factor		**X**
ZNF469	SSG	Zinc-finger, may be transcription factor		**X**
MIRLET7BHG	SSG	Long non-coding RNA, related pathways include metastatic brain tumor	**X**	
SETD1A	REG	Histone Lysine Methyltransferase component, marks TSS of active genes		**X**

PHRF1	REG	Contains PHD and ring-finger, chromatin mediated gene regulation		**X**
GAS8	REG	Binds tyrosine-protein kinase receptors, which are implicated in cell growth and migration, target in cancer therapy	**X**	
CBX4	REG	SUMO-protein ligase, indirectly regulates p53/TP53 transcriptional activation		**X**
MUC1	REG	Cell-surface glycoprotein, overexpressed in many cancers	**X**	
GTF2F1	REG	General transcription initiation factor, binds to RNA polymerase II		**X**
RABL6	REG	GTPase, may play a role in cell growth, oncogene	**X**	
GADD45B	REG	Expression increased with growth arrest and DNA damage	**X**	
BRD3	REG	Contains bromodomain, recognizes acetylated lysine residues of histones (chromatin mediated gene regulation)	**X**	
SF3A2	REG	pre-mRNA splicing, component of splicing factor SF3A		
AP3D1	REG	facilitates vesicles budding from the golgi membrane, and may be directly involved in		

		trafficking to lysosomes		
BAIAP2	SSG	Adaptor protein, links membrane G-proteins to cytosolic		
AC011498.7	REG	unknown		
C13orf46	REG	unknown		

The involvement of SSGs and REGs in cancer and cell growth and/or transcription and gene regulation is striking as CypA plays a major role in these two areas: it is plausible that CypA indeed regulates these genes by interacting with their RNA transcripts.

Potential Models for CypA-RNA Interaction

The presented results are particularly exciting in showing that genes enriched in pulldown samples are involved in CypA-regulated pathways. As CypA has been shown to mediate the transcriptional activities of YY1[8] and NF-κB8[9] as well regulate the tyrosine kinase Itk,[19] it may similarly interact with other transcription factors and kinases. An important distinction must be made, however, between the mRNA transcripts that encode these proteins and the proteins themselves: the immunoprecipitation experiments determined enrichment for RNA, not protein, levels. An intriguing mechanism by which CypA could interact with both the mRNA and protein product of a certain gene is localization by direct RNA-interaction (Figure **18**). CypA could potentially bind the mRNA of the protein it regulates, thus localizing to the area of translation, and be proximally available to edit the side-chain orientation upon (or during) translation. This would accelerate synthesis of fully folded proteins.

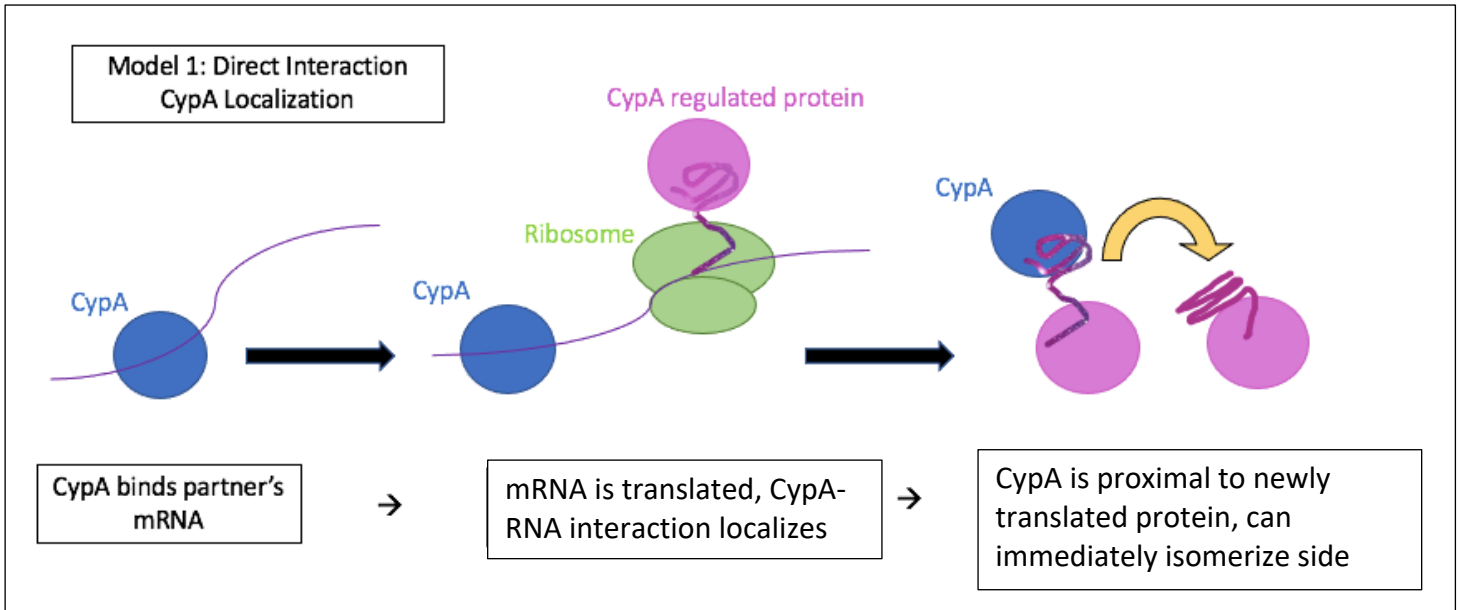


Figure 18. Direct interaction with mRNA allows CypA modification of protein target upon translation. Potential model for CypA-RNA interactions showing that CypA binding to mRNA of a protein it regulates localizes CypA so that it can act quickly to isomerase protein upon translation.

RNA could also act as a bridge between CypA and a protein that it regulates, increasing stability of the complex and efficiency of isomerization. This would be consistent with previous experiments in the lab that showed random RNAs added to an activity assay increased CypA's rate of isomerization[10]: in bringing proteins together and stabilizing their positions, RNA could increase the rate at which CypA finds and acts on the proteins it regulates (see **Figure 19).**

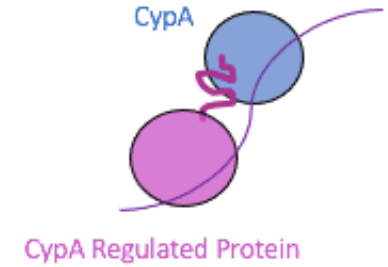


Figure 19. RNA bridges CypA and target protein. RNA could directly interact with CypA and a protein as a stabilizing bridge.

Perhaps, though, it is not direct interaction but rather indirect mediation of protein-RNA binding that implicates CypA in certain gene pathways. A study has shown that interaction between a catalytically active CypA and NS5A, a protein necessary for hepatitis C viral replication, stimulates NS5A's RNA binding[17]. Perhaps, then, CypA's isomerase activity changes the shape of certain proteins to increase their affinity for or enable RNA binding. In this case, the above FA pulldown experiments would include many indirect interactions. **Figure 20** illustrates the potentiality of indirect CypA-RNA interactions.

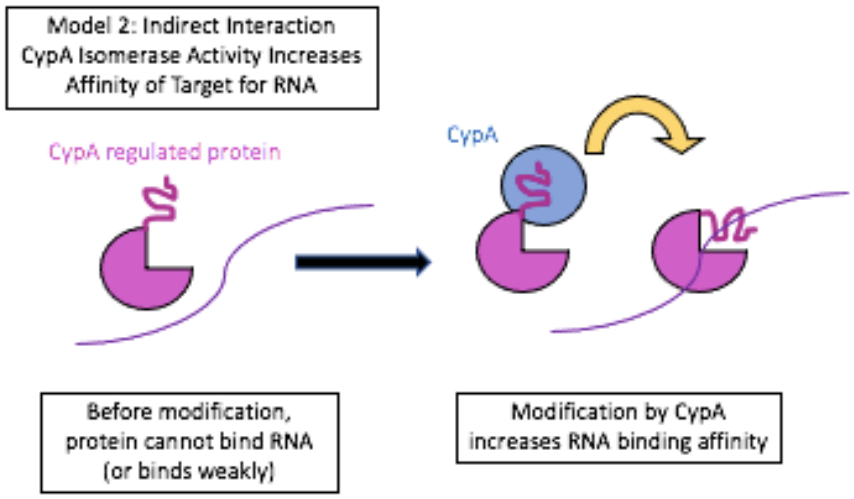


Figure 20. CypA increases target protein's affinity for RNA. By isomerizing target protein, CypA enables target to bind RNA or increases target's RNA binding affinity.

Future Directions

The presented work merely scratches the surface of CypA-RNA interactions, leaving many interesting questions to be addressed. A first step would be to identify certain RNA sequences and/or structures that contribute to the observed specificity of CypA binding. To do this, a similar immunoprecipitation could be performed with the additional step of RNA digestion before sequencing and protein digestion: if RNAse (protein that digests RNA) is added to the sample before RNA is freed from CypA binding, the RNA sequestered by protein binding will protected from digestion, leading to sequencing of merely the RNA binding site. Then, these sequences could be tested *in vitro* to verify CypA affinity. Further, *in vitro* activity assays could be done with an RNA sequence that binds with reasonable affinity to ascertain RNA's effect on CypA isomerase activity.

To investigate if isomerase activity of CypA is required *in vivo* interactions (indirect or direct), an RNA immunoprecipitation could be done with catalytically dead CypA. The CRISPR-Cas9 system could be used to introduce a mutation removing catalytic abilities into the CypA gene of HeLA cells, pulldown performed, and results compared with those reported above. As CypA is an extremely important cellular protein, however, the viability of the cells may be impacted by removing its isomerase activity. An alternative method would transfect a vector with a catalytical dead CypA mutant and an attached his-SUMO tag into HeLa cells. Then, a his-SUMO specific probe rather an antibody could be used for pulldown. Comparing RNAs pull-downed with a catalytically dead CypA to those pull-downed with a catalytically active CypA would provide information on the RNA binding that is related to or depends on CypA's

isomerase activity. This would provide evidence for a previously unknown level of RNA-protein regulation.

Conclusion

RNA-protein interactions have numerous and varied critical biological roles, many yet undiscovered.[1,2,3,4] Here, the novel CypA-RNA interaction have been broadly explored, providing *in vitro* and *in vivo* evidence for CypA-RNA binding. Fluorescence Anisotropy assays with segments of viral DI-72 RNA derived from the genome of the TBSV suggest CypA-RNA binding specificity *in vitro,* as the negative control RNA did not bind CypA while several Minus Strand DI-72 transcripts showed anisotropic shifts indicative of binding. Specific binding, perhaps relating to the sequence or structure of the RNA molecule, is exciting as it could be evidence that RNA plays a functional role in CypA activity. Several RNA immunoprecipitations using both UV and formaldehyde cross-linking methods produced a set of statistically enriched genes in human cells*,* the most significant being transcription related genes. A select group of enriched genes which appeared in all formaldehyde and UV RNA immunoprecipitation samples or had extremely low adjusted p-values are especially interesting as their proposed functions correspond to key CypA activities: transcriptional regulation and cancer progression. While these initial studies show that CypA binds different RNAs with different affinity, further work remains to identify specific RNA characteristics (sequences or structures) that contribute to CypA's preferential binding and determine reliable K_ds for these interactions. Next, activity assays and RNA immunoprecipitations with catalytically dead CypA will increase our understanding of the role RNA plays in CypA activity. This initial exploratory study adds to the body of research implicating CypA as an unexpected yet important RNA binder and opens exciting avenues for future research on CypA-RNA interactions and other novel RNA binding proteins.

Bibliography

(1) Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature reviews. Molecular cell biology*, *19*(5), 327–341. https://doi.org/10.1038/nrm.2017.130

(2) Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell 46, 674–690.

(3) Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., and Hentze, 149 M. W. (2012) Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 149, 1393–1406.

(4) Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M. W. (2016) Comprehensive Identification of RNA Binding Domains in Human Cells. Mol. Cell 63, 696–710.

(5) Davis, T.L., Walker, J. R., Campagna-Slater, V., P. J., Paramanathan, R., Bernstein, G., MacKenzie, F., Tempel, W., Ouyang, H., Lee, W. H., Eisenmesser, E. Z., and Dhe-Paganon, S. (2010) Structural and Biochemical Characterization of the Human 153 Cyclophilin Family of Peptidyl-Prolyl Isomerases. PLoS Biol. (Petsko, G. A., Ed.) 8, e1000439.

(6) Wang, P., & Heitman, J. (2005). The cyclophilins. *Genome biology*, *6(7),* 226.

(7) Nigro, P., Pompilio, G., & Capogrossi, M. C. (2013). Cyclophilin A: a key player for human disease. *Cell death & disease*, *4*(10), e888. https://doi.org/10.1038/cddis.2013.410

(8) Yang, W. M., Inouye, C. J., & Seto, E. (1995). Cyclophilin A and FKBP12 interact with YY1 and alter its transcriptional activity. *The Journal of biological chemistry*, *270*(25), 15187–15193. https://doi.org/10.1074/jbc.270.25.15187

(9) Sun, S., Guo, M., Zhang, J. B., Ha, A., Yokoyama, K. K., & Chiu, R. H. (2014). Cyclophilin A (CypA) interacts with NF-κB subunit, p65/RelA, and contributes to NF-κB activation signaling. *PloS one*, *9*(8), e96211. https://doi.org/10.1371/journal.pone.0096211

(10) Lloyd, N. Discrimination of ssRNA by Pot1 and Identification of a Novel CypE Aptamer Through an Optimized RNA SELEX Protocol. Citable URL: https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/4q77fr444

(11) Kovalev, N., and Nagy, P. D. (2013) Cyclophilin A binds to the Viral RNA and Replication Proteins, Resulting in Inhibition of Tombusviral Replicase Assembly. *J. Virol. 87,* 13330-13342.

(12) Trivedi, D. K., Bhatt, H., Pal, R. K., Tuteja, R., Garg, B., Johri, A. K., Bhavesh, N. S., and Tuteja, N. (2013) Structure of RNA-interacting Cyclophilin A-like protein from Piriformospora indica that provides salinity-stress tolerance in plants. Sci. Rep. 3.

(13) Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic Acid Recognition by OB-Fold Proteins. *Annu. Rev. Biophys. Biomol. Struct. 32*, 115–133.

(14) Maris, C., Dominguez, C., and Allain, F. H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression: The RRM domain, a plastic RNA-binding platform. *FEBS J. 272*, 2118–2131.

(15) Brown, R. S. (2005) Zinc finger proteins: getting a grip on RNA. *Curr. Opin. Struct. Biol. 15*, 94–9

(16) Birney, E., Kumar, S., & Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic acids research*, *21*(25), 5803–5816. https://doi.org/10.1093/nar/21.25.5803

(17) Janeway CA Jr, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. The structure of a typical antibody molecule. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27144/

(18) Gustafson, C. L., Parsley, N. C., Asimgil, H., Lee, H.-W., Ahlbach, C., Michael, A. K., Xu, H., Williams, O. L., Davis, T. L., Liu, A. C., and Partch, C. L. (2017) A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Mol. Cell 66*, 447-457.e7.

(19) Colgan, J., Asmal, M., Neagu, M., Yu, B., Schneidkraut, J., Lee, Y., Sokolskaja, E., Andreotti, A., and Luban, J. (2004) Cyclophilin A Regulates TCR Signal Strength in CD4+ T Cells via a Proline-Directed Conformational Switch in Itk. *Immunity 21*, 189– 201.

(20) Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic acids research*, *15*(21), 8783–8798. https://doi.org/10.1093/nar/15.21.8783

(21) Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, **15**, 550. doi: 10.1186/s13059-014-0550-8.

(22) Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. ***Nature Protoc. 2009;4(1):44-57.***

(23) Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods in enzymology*, *326*, 245–254. https://doi.org/10.1016/s0076-6879(00)26058-8

(24) Pagano JM, Clingman CC, Ryder SP. Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. RNA. 2011 Jan;17(1):14-20. doi: 10.1261/rna.2428111. Epub 2010 Nov 22. PMID: 21098142; PMCID: PMC3004055. FTSC

(25) E. A., Frey, B. L., Smith, L. M., and Auble, D. T. (2015) Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes. *Journal of Biological Chemistry* 290, No. 44, 25406-26411.

(26) Urdaneta, E. C., and Beckmann, B. M. (2019) Fast and unbiased purification of RNA-protein complexes after UV cross-linking. *Methods* 178, 72-82.

(27) Prashar, A., Schnettger, L., Bernard, E. M., & Gutierrez, M. G. (2017). Rab GTPases in Immunity and Inflammation. *Frontiers in cellular and infection microbiology*, *7*, 435.

(28) Verheul, T., van Hijfte, L., Perenthaler, E., & Barakat, T. S. (2020). The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Frontiers in cell and developmental biology*, *8*, 592164. https://doi.org/10.3389/fcell.2020.592164

(29) The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, https://doi.org/10.1093/nar/gkaa1100

(30) Foster, T. L., Gallay, P., Stonehouse, N. J., & Harris, M. (2011). Cyclophilin A interacts with domain II of hepatitis C virus NS5A and stimulates RNA binding in an isomerase-dependent manner. *Journal of virology*, *85*(14), 7460–7464. https://doi.org/10.1128/JVI.00393-11