TICKLED ZINC:
CHARACTERIZING A-TO-I RNA EDITS
IN NERVOUS TISSUE OF OCTOPUS RUBESCENS
IN RESPONSE TO OCEAN ACIDIFICATION

by
Meg A. Mindlin

A THESIS

submitted to

WALLA WALLA UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

May 2025

This thesis for the Master of Science degree

has been approved by the Department of Biological Sciences

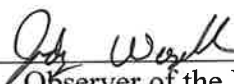and the Office of Graduate Studies

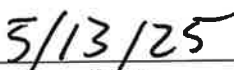Walla Walla University

_____
Major Professor

_____
Committee Member

_____
Committee Member

_____
Committee Member

_____
Dean of Graduate Studies

_____
Observer of the Process - Graduate Representative

_____
Candidate

_____5/13/25_____
Date

ABSTRACT

Global climate change, both rising temperatures and increased acidification in the ocean, has mostly negative effects on octopuses. However, octopuses show resilience in the face of increased acidification, notably their ability to return their resting metabolic rate to normal after prolonged exposure to increased pCO_2. To investigate potential molecular mechanisms of this resilience, a transcriptome wide analysis of octopuses exposed to elevated pCO_2 was performed. A-to-I RNA editing was found to change the bound sequences of Cysteine$_2$-Histidine$_2$ (C2H2) zinc fingers, transcription factors that up or down regulate the proteins they are bound to. In high pCO_2 conditions, transcripts containing C2H2 zinc finger domains were found to be significantly more edited than transcripts lacking C2H2 zinc finger domains. Structural protein prediction software found that one of the changed bound sequences of C2H2 zinc fingers was for the nuclear core complex, which is found to play a vital role in regulating environmental stress. This evidence suggests a regulatory pathway in which the NPC is targeted by a zinc fingers protein whose mRNA is modified by acidification-responsive RNA editing.

TABLE OF CONTENTS

List of Figures

List of Tables

INTRODUCTION

Background and Significance

Ocean Acidification

Human activities, primarily emissions of greenhouse gasses, have undeniably caused global warming, with global surface temperatures reaching a 1.1°C increase from 1850–1900 in 2011–2020[1]. Global surface temperatures have increased faster since 1970 than in any other 50 year period in the last 2,000 years[1]. In 2019, atmospheric CO_2 concentrations reached 410 ppm (Figure 1)[1]. These increases in atmospheric CO_2 far exceed the natural multi-millennial changes between glacial and interglacial periods over the last 800,000 years[1]. Increases in atmospheric CO_2 have direct impacts on the world's oceans as almost 30% of atmospheric CO_2 is absorbed by the ocean[2]. Ocean CO_2 uptake causes pH reductions and alterations in fundamental chemical balances that together are commonly referred to as ocean acidification[2]. This happens because when CO_2 is dissolved in water, it becomes hydrated to a carbonic acid, H_2CO_3[2]. This carbonic acid immediately dissociates to bicarbonate HCO_3^- and a H^+ proton, which reduces the pH[2]. This is seen today as a drop in average surface pH from 8.2 to 8.0 across the surface of the ocean, which represents a 35% rise in H^+ concentrations[1]. If we employ low emission and high mitigation practices, the global average-area surface CO_2 is predicted to peak in 2050 at 442 µatm and decrease to 380 µatm in 2100[3]. However if CO_2 emissions continue to rise unabated over the entire 21st century, the global average-area surface CO_2 is projected to increase steadily to as high as 1,051 µatm in 2100[3]. This correlates to a ~4% increase in H^+ concentrations from low emissions/high mitigation to a ~150% increase in H^+ concentrations from continued

emission/no mitigation[3]. This would result in a decrease of global average-area surface

ocean pH by
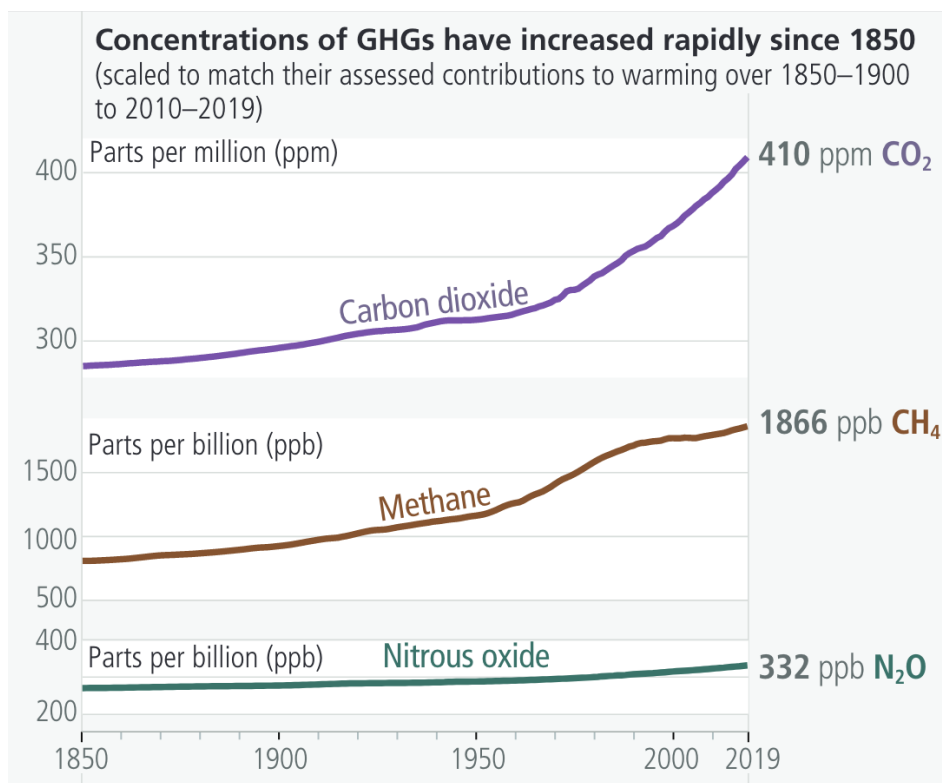
Figure 1: Changes in CO_2, CH_4, and N_2O emissions from 1850 to 2019. Taken from Calvin et al., "IPCC, 2023."

0.01 to 8.06 under low emissions/high mitigation projections, or by 0.39 to 7.68 under

continued emissions/no mitigation projections[3]. This pH change is unprecedented in the

last 800,000 years and will undoubtedly impact our biosphere, from individual organisms

to whole ecosystems[1,3].

Effects of Ocean Acidification on Marine Life

　　With such a large change in ocean chemistry, it seems likely that marine life will

be affected. Early developmental stages are considered the most vulnerable to ocean

acidification (OA) due to the physiological and biochemical requirements that sustain

growth rate during a time of increased morphological and physiological change[4]. Early

stages of development are directly responsible for determining an organism's survival

and overall success, which dictates the response of species population dynamics over

time[4]. In coleoid cephalopods (octopus, cuttlefish, and squid, but excluding nautilus), this

is no different[5].

　　During the embryonic stage the egg capsule acts as a physical barrier to their

outside environment. Despite this, there is an increase of the embryos metabolic

demand due to cellular growth, organogenesis, and muscular activity occurring during

embryonic development[4]. Increased temperature *and* acidification leads to an even

higher metabolic rate with increasing O_2 consumption[4]. To account for this, mollusk eggs

will increase their surface area, which reduces the egg's thickness to maximize

exchanges with the external environment, leading to premature hatching and decreased

survival[4]. This effect was observed specifically in ocean warming (OW) conditions and

increased with the combination of warming and acidification, but not seen solely in OA

conditions[4]. A meta-analysis on the impact of climate change on cephalopods showed also a decrease in paralarvae, juvenile, and adult sizes as well under OA conditions[4].

When it comes to overall survival of cephalopods due to exposure to OA conditions, OA was found to not impact survival in a meta-analysis, while OW does[4]. This highlights the resilience of cephalopods to OA, possibly due to their strong ion and acid-base regulators[4,5]. This is seen specifically in *Sepia officinalis*,when exposed to elevated pCO_z, their blood HCO_3^- content was rapidly increased through active ion-transport processes that partially compensated for the hypercapnia induced respiratory acidosis[6]. A minor decrease in *S. officinalis* intracellular pH from exposure to elevated pCO_2, along with a stable intracellular phosphagen level, indicated an efficient ability to regulate their pH[6]. This compensation occurs through active ion exchange mechanisms in the gills, which elevate blood bicarbonate (HCO_3^-) concentrations to buffer the drop in extracellular pH, thereby mitigating the effects of respiratory acidosis. This enables cephalopods to regulate their internal acidity without disturbing metabolic equilibria or compromising aerobic capacities[6].

Resilience to OA is also seen in *Octopus rubescens*, where it was shown that their resting metabolic rate rapidly acclimates to elevated pCO_2 conditions[7]. However, their hypoxia tolerance was impaired even after a 5 week acclimation period[7]. This shows that *O. rubescens* experiences short term stress in elevated pCO_2 environments, but is able to acclimate, similar to what is observed in *S. officianlis* pH regulation[5,7]. However, a possible mechanism for this acclimation was not identified. Despite this ability to acclimate, the combined effects of OA and hypoxia may present a physiological challenge in octopuses[7].

In contrast to either study, a bathyl species of octopus, *Muusoctopus leioderma*, showed that elevated pCO_2 conditions had no change in resting metabolic rate, critical

partial pressure, noroxygen supply capacity[7]. This ability to be unaffected by elevated

pCO_2 is likely due to physiological adaptations that are directly linked to its phylogeny

and life history[8]. This resilience to OA in terms of metabolic response is only seen in

cuttlefish and octopus, as previous studies in multiple squid species show metabolic

depression in high pCO_2 conditions, directly contrasting the results seen stated

above[4,9,10].

      In addition to reduced metabolic rates, several other physiological impacts have

been observed in squid in response to elevated pCO_2. *Doryteuthis pealeii* raised under

OA conditions demonstrated significant developmental changes that included increased

hatching time, shorter mantle length, and significantly reduced surface area, increased

deformity, and increased porosity of aragonite statoliths[10]. The change in aragonite

statoliths has an impact on squid behavior as aragonite statoliths are critical for balance,

orientation, and movement detection, and could alter squid paralarvae survival in the

wild[10].

      Cuttlefish demonstrate a similar theme, with *S. officinalis* experiencing a

decrease in growth and increase in calcification in their early life stages[11]. Like

discussed earlier, embryonic growth was reduced, leading to much smaller cuttlefish, as

well as hatching being delayed under elevated pCO_2[4,11]. Specific to calcification, the

proportion of their mass that was contributed by their cuttlebone, was increased, leading

to cuttlebones that are significantly more dense in elevated pCO_2 conditions[11]. Denser

cuttlebones are a problem for the survival of cuttlefish, because cuttlebones act as a

buoyancy control[11].

      Immune response is also affected in cephalopods, specifically octopuses, under

OA conditions. In *O. rubescens*, elevated pCO_2 conditions saw an increase in the

number of circulating hemocytes, responsible for inducing a cellular immune response,

indicating stress, as well as an increase in total phagocytosis[12]. The presence of chronic

stress response hints at long-term physiologic consequences for cephalopods, but

needs further research to solidify that hypothesis[12].

These various effects from OA on cephalopods are of great concern because a

decrease in cephalopod populations, would directly and indirectly impact fisheries and

marine ecosystems. This is because cephalopods are ecologically important as both

predator and prey, having a structural role in marine ecosystems as a link between

trophic levels, due to their high growth rate and their voracious prey consumption[13].

Cephalopods are also important as food for human consumption, contributing to the 260

million jobs employed by fisheries and 4 million metric tonnes caught by fisheries each

year, which is 5% of the total harvest of all species from marine waters[13–16].

RNA Editing

The central dogma of biology maintains that genetic information passes from

DNA to RNA to proteins[16]. However, proteome complexity relies on post-transcriptional

processes[17]. There are some tools, such as alternative splicing, that allow organisms to

modify their RNA and increase the transcriptome and proteome diversity[16]. A-to-I RNA

editing is one of these processes. A-to-I RNA editing is characterized by the deamination

of adenosine to inosine that can alter genetic information beyond the genomic

sequence[18]. Unlike alternative splicing, which can shuffle large sections of RNA, A-to-I

RNA editing can target single bases to fine-tune protein function[16]. Biochemically,

inosine base pairs to cytidine, making it a biological mimic for guanosine[19]. During

translation, inosine is processed as guanosine, which in turn can alter codons[17].

Adenosine deaminases acting on RNA (ADAR) changes adenosine to
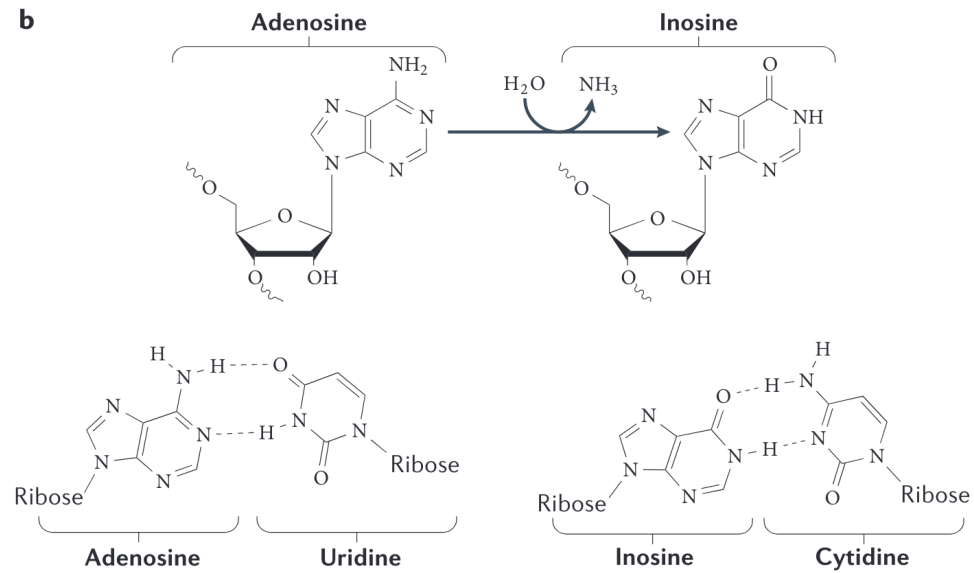
Figure 2: ADARs catalyze a hydrolytic deamination reaction that converts adenosine to inosine (top). Whereas adenosine base pairs with uridine, -inosine behaves like a guanosine,- as -it -base-pairs with cytidine a Watson–Crick -bonding configuration (bottom). Taken from Nishikura 2016.

inosine by hydrolytically replacing the amino group at C6 of the purine ring with a

carbonyl (Figure 2)[20]. This can result in the incorporation of amino acids that are not

directly encoded in the genome[21].

While A-to-I editing is the most common form of RNA editing in metazoans, the

changes rarely result in protein diversification. The vast majority of RNA editing sites are

found in noncoding sequences, such as 5' and 3' untranslated regions (UTR's) and

intronic retrotransposon elements, like non-coding repeat sequences known as *Alu*

elements and long interspersed elements (LINEs)[18]. Despite this, RNA editing of a

glutamate-gated channel mRNA in mammalian brains is important for synaptic

transmission[22]. L-glutamate is a major excitatory neurotransmitter in the vertebrate

nervous system that opens cation channels that mediate fast excitatory synaptic

responses and participates in the establishment and maintenance of synaptic plasticity

underlying learning and memory[22]. Glutamate receptor channels in mouse neurons

contain a codon that has been changed from glutamine to arginine[22]. This is made

possible by an intronic sequence that is imperfectly complementary to the exon

containing the adenosine. Following transcription, the intron base pairs to the exon,

creating a dsRNA and allowing the double stranded RNA binding domain (dsRBD) of the

mouse ADAR to bind to this locus and deaminate the adenosine[22]. When a recombinant

mouse genome that lacks this intron is created, complementarity is removed as well,

making editing by ADAR impossible[22]. As a result, mice will exhibit epilepsy,

hyperexcitability, and do not survive past 3 weeks[22].

Despite this one example, only ~3% of human mRNAs and 1%-4% of *Drosophila*

mRNAs have a recoding site, a site where a nucleotide base is changed, resulting in a

new amino acid to be coded[17]. Looking further into metazoan lineages we find that there
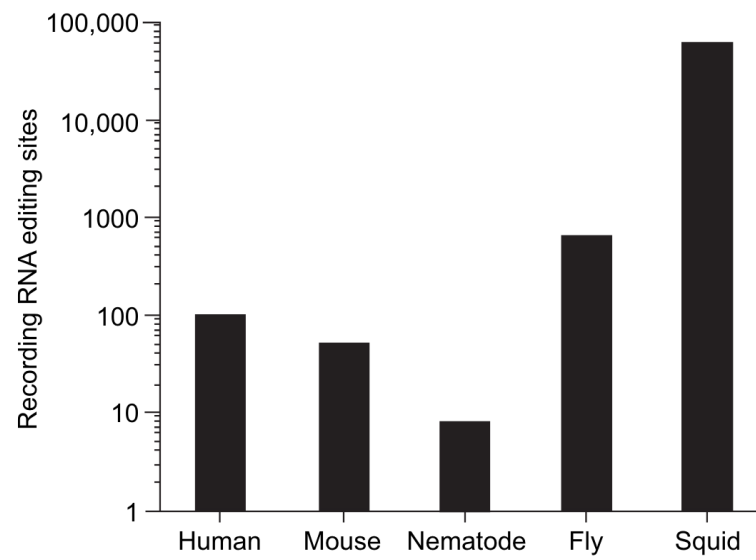
Figure 3: "The extent of recoding RNA editing across taxa. Nematode refers to *Caenorhabditis elegans*, fly refers to *Drosophila melanogaster*, and squid refers to *Doryteuthis pealeii*". Taken from Rosenthal, "The Emerging Role of RNA Editing in Plasticity." 2015.
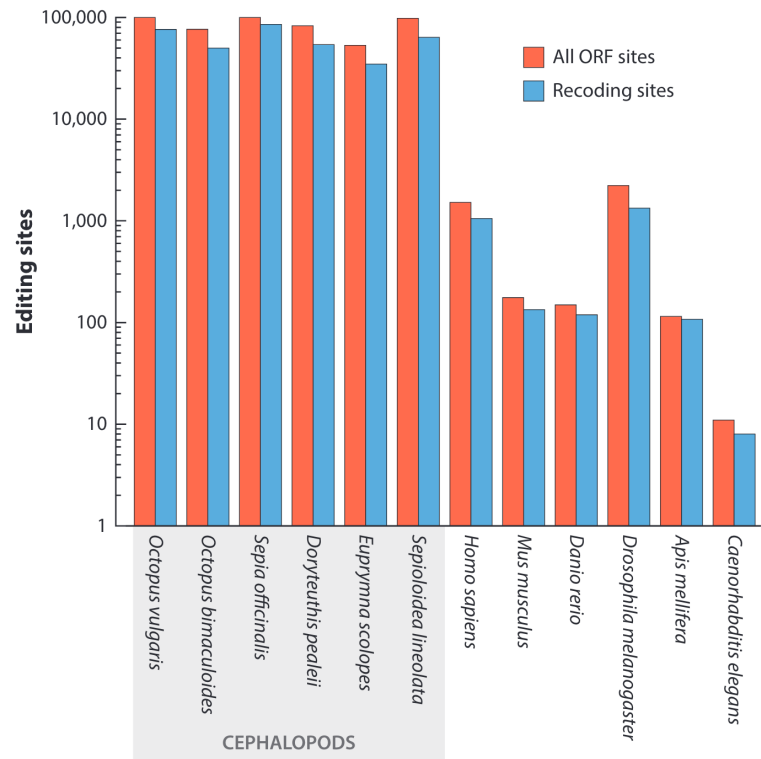
Figure 4: "Editing across taxa. Numbers of all editing sites within the coding region (including synonymous editing) and recoding sites from multiple studies are plotted". Taken from Rosenthal and Eisenberg, "Extensive Recoding of the Neural Proteome in Cephalopods by RNA Editing." 2023.

are 152 recoding sites identified in diamondback moths, ~100 sites in humans, 57 in

leaf-cutter ants, ~50 in mice, and 8 sites in *C. elegans,* with only 34 of these sites being

conserved across mammals (Figure 3)[19]. However, in cephalopods there are

approximately 100,000 known recoding sites, with the highest estimate at ~600,000 total

editing sites, and affect the majority of the encoded proteins (Figure 4)[23]. Recoding

levels in cephalopods vary considerably across tissues, while most prominent in neural

tissues. Recoding accounts for 11-13% of the global RNA editing activity measured in

cephalopods compared to the < 1% in mammals[23]. Additionally, recoding sites are

shared between cephalopods, with ~5,000 recoding sites conserved since the

divergence of the coleoid (squid, cuttlefish, and octopus) lineage[23].

RNA editing in cephalopods was first discovered through the initial investigations

into the market squid giant axon and its sodium and potassium conductances and their

role in the generation of action potentials[20]. The cloning and functional expression of a

squid Kv2 channel (sqKv2), a delayed rectifier K^+ channel, showed a multiplicity of

variants that were the result of RNA editing[21]. Half of the sites were targeted to the T1

domain of the channel, which is important for subunit assembly[21]. The other sites

occurred in the transmembrane spans and the effects on the K^+ channel were

elaborate[21]. Ranging from edited codons affecting the channel's gating kinematics to

several T1 sites regulating functional expression of the gates as well[21]. One particular

edit, R87G, a phylogenetically conserved position, reduced the expression of the

channel 50-fold by regulating the channel's ability to form tetramers[21].

Since that initial investigation in 1997, multiple RNA editing sites have been

identified and described in cephalopods. After the delayed rectifier K^+ channel, the next

site identified in squid regulated a Na^+/K^+ pump[24]. Na^+/K^+ ATPase is a ubiquitous

membrane protein that uses the free energy of ATP hydrolysis to establish and maintain

Na^+ and K^+ gradients across cell membranes[24]. Without the Na^+/K^+ pump, cells lack the driving force required for excitability and the transport of Na^+ in and out of the membrane[24]. As ion gradients across plasma membranes are required to generate electrical signals. The job of maintaining these gradients falls on Na^+/K^+ ATPase[24]. The regulation of this pump is done so by RNA editing, modulating the Na^+/K^+ pump's turnover rate and sodium release by changing 4 codons in ATPase from A to G, converting lysine to glycine[24]. From these two studies, it was therefore hypothesized that A-to-I RNA editing is important for regulating rapid electrical signaling[24].

Up until this point, the data pointed to cephalopods using RNA editing for synaptic transmission. It was when scientists sequenced the site of a delayed rectifier K^+ channel of an octopus from the tropics and compared it to the delayed rectifier K^+ channel of an Antarctic species that we gained another idea of how cephalopods use RNA Editing. The assumption was that two species of octopuses from two different temperature climates would have evolved structural changes to compensate for their thermal environment[25]. Instead, they found that the genes encoding the delayed rectifier K^+ channel differed at only four positions between the species, and displayed similar behavior when expressed[25]. The differences were found in the transcribed mRNAs that were extensively edited and created functional diversity[25]. The one site they described recoded an isoleucine to a valine in the channel's pore, which greatly accelerated the gating kinetics of the pore by destabilizing the open state. This site was extensively edited in cold water octopus species but unedited in the warm water octopus, drawing the conclusion that RNA editing can help the octopus respond to the physical environment[25].

Further evidence for RNA editing helping temperature acclimation came about only recently. It was found that RNA editing responded to cold-induced temperature

change within hours and reached a steady state within approximately 4 days[26]. This was the first evidence to show RNA editing levels change in response to short-term environmental change. These findings were also observed in wild-caught specimens, which exhibited editing changes of comparable magnitude to what was observed in lab settings[26]. The same experiment was conducted in both *O. bimaculoides* and *O.bimaculatus*, yielding similar results, suggesting that temperature-dependent RNA editing at these sites is evolutionary conserved across these two species[26].

Over 21,000 editing sites were found to change editing level in response to colder temperature water[26]. Of those, the change in function induced by the edits in two protein changes were further characterized: kinesin-1, a motor protein that drives axonal transport and the synaptotagmin, a key protein involved in synaptic transmission[26].

In the case of kinesin-1, the site identified recoded a lysine to an arginine[26]. The lysine at this site is universally conserved across 162 species within four phyla, which suggests purifying selection, as the site lies in kinesin's motor domain that faces the microtubule[26]. When faced with a 10°C change in temperature, the site undergoes a 30% shift in percent editing[26]. Using TIRF microscopy, the edited versions of the kinesin-1 protein showed a lower velocity than the wild-type at both warm and cold temperatures[26]. The edited version also displayed a temperature-invariant velocity that was comparable at both 21°C and 11°C, as well as a shorter run length than the wild-type version at both temperatures[26]. The edited kinesin-1 also had a greater tendency to be stationary at both temperatures[26]. When the recoding site of lysine to arginine in kinesin-1 was edited in rats, kinesin-1 resulted in a similar decrease in velocity and run lengths[26].
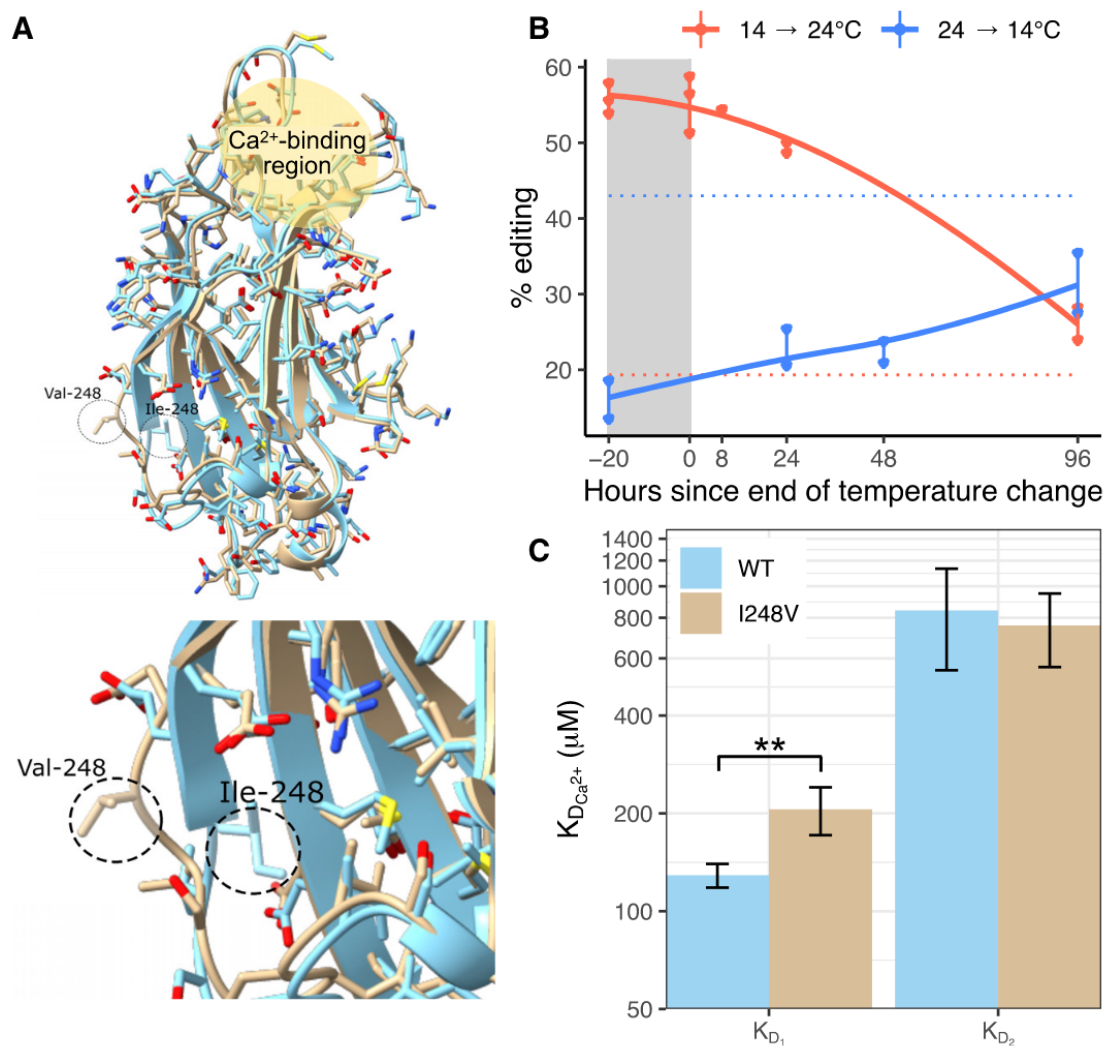
Figure 5: "A cold-induced editing site (I248V) on the C2A domain of synaptotagmin-1 changes protein conformation to alter Ca2+binding affinity". Taken from Birk et al. 2023

In the synaptotagmin edit, the recoding site was changed from an isoleucine to a valine in synaptotagmin-1 (Figure 5). This edit is highly temperature sensitive, increasing by 24% in the cold[26]. In 60 other molluscan species, >60% have either an isoleucine or a valine at this position, with the remaining species mostly having other non-polar residues, such as phenylalanine and cysteine at this position[25]. Synaptotagmin-1 is located at the interface of neurotransmitter-containing presynaptic vesicles and the presynaptic membrane[26]. When the intracellular Ca^{2+} concentration rises during presynaptic excitement, Ca^{2+} ions will bind to synaptotagmin and induce a conformational change that will promote the initialization of vesicles docking to the presynaptic membrane[26]. Synaptotagmin itself is composed of a N-terminal transmembrane domain that is embedded in the synaptic vesicle and two calcium binding domains, C2A and C2B[26]. Each of these C2 domains is capable of binding at least 2 Ca^{2+} ions and additionally phospholipid membranes[26]. The C2A domain is composed of 8 β-sheets that have neighboring high-affinity and low-affinity Ca^{2+} ion-binding sites[26]. Additionally, phospholipid binding, which is promoted by the bound Ca^{2+} ions, occurs at the same location as the ion binding, both at one end of the domain[26].

In rats, this edit is on the opposite side of the C2A domain from Ca^{2+} ion binding sites[26]. The unedited synaptotagmin-1 C2A domain structure is in direct contact with the hydrophobic core of the domain and has a relatively low solvent accessibility surface score[26]. In contrast, the edited synaptotagmin-1 C2A domain structure is completely exposed to solvent with a much higher solvent accessibility surface score due to the elimination of a single methyl group[26]. The β-oop containing this recoding site is relatively rigid, despite being exposed to solvent, compared to the unedited coding site where the loop is relatively flexible[26]. Additionally, the unedited β-loop is flipped inward allowing it to interact with the hydrophobic core, whereas the edited β-loop is flipped

outward[26]. When investigated further, the edited site lowered the binding site affinity of

the first bound Ca^{2+} ion by nearly 60% while the binding affinity for the 2nd Ca^{2+} ion

remained unchanged[26]. These findings demonstrate that the removal of a single methyl

group on the C2A domain of synaptotagmin-1 changes the protein's conformation

sufficiently to alter Ca^{2+} binding dynamics in response to temperature[26]. Additionally,

crystal structure generated by x-ray crystallography of rat C2A domain matched very

closely to the octopus C2A domain crystal structure[32].

If temperature can correlate to changes in RNA editing, can acidification do the

same? The resilience cephalopods show in the face of OA conditions, such as their

metabolic rate, begins a series of questions into the mechanism of this resilience. While

one has been explored in cuttlefish with their strong ion and acid-base regulators, other

molecular and cellular mechanisms had not been explored in octopus until the thesis of

Jaydee Serewit[30] and subsequent thesis of Ricky Wright[31].

Zinc Fingers

Previous work in the lab indicates that octopuses exposed to elevated pCO_2

show suppression in editing levels (Wright, 2024). However, there appeared to be a

disproportionate number of zinc finger proteins (ZFPs) being edited under elevated pCO_2

conditions. In the octopus genome 5.1% of genes are ZFPs, but under OA conditions,

ZFPs were 16.5% of all the acidification-responsive editing sites. In the 51 most

acidification-responsive editing sites, 51.0% of the containing genes are ZFPs. ZFPs are

the largest transcription factor family and contain at least one zinc finger domain, a

finger-like DNA-binding structural motif, and play a significant role in multiple biological

processes[27]. ZFPs primarily function as transcription factors (TFs) in tumorigenesis and

tumor progression in humans[27]. They also play a vital role in multiple complicated

biological processes from metabolism to stem cell differentiation and maintenance[27].

TFs can regulate transcription of genes by recognizing or binding to DNA sequences

directly[27]. Zinc fingers can be categorized into eight different categories depending on

their main-chain conformation and their secondary structure around their zinc-binding

sites[27]. These include Cys_2His_2 (C2H2) zinc finger domains[27], the most prevalent motif

with over 5,000 C2H2-like domains found encoded in the human genome[27].

The C2H2 super family of zinc-finger TFs were found to be massively expanded

upon in the octopus genome[28]. Previously it was thought this family was to be uniquely

enlarged in vertebrates only, with octopus ancestor, the bivalve, lacking them[28]. The

distinguishing features of C2H2 zinc-finger proteins is their strong and specific binding to

a long and unique DNA recognition target sequence and their rapid expansion within

various animal taxa during evolution[29]. The C_2H_2 domain is characterized by a β-hairpin

(an antiparallel β-sheet that consists of two β-strands), followed by α-helix that forms a

left-handed $\beta\beta\alpha$structure (Figure 6)[29,30]. They are called zinc-fingers because the

structure is stabilized by the coordination of a zinc atom with two conserved cysteine

residues at the α-helix terminus[29]. The sequence pattern for C2H2 zinc fingers is X_2-

Cys-$X_{2,4}$-Cys-X_{12}-His-$X_{3,4,5}$-His, containing conservative hydrophobic residues wrapped

in hydrophobic cores, except for the two histidines and cystines[27]. Both the histidine and

cysteine pairs are conserved, as well as the hydrophobic core forming the α-helix[29].

Other amino acid residues in C2H2 domains are highly variable [29]. Individual zinc finger

motifs have been suggested to bind to an adjacent three to five-nucleotide subsequence

and the C2H2-zinc finger domain can be specified to a range of 3 base pair targets[27].

C2H2 zinc fingers can be separated into 3 groups: 1. Proteins with one, two, or

several randomly distributed C2H2 domains; 2. Proteins that have three C2H2 domains
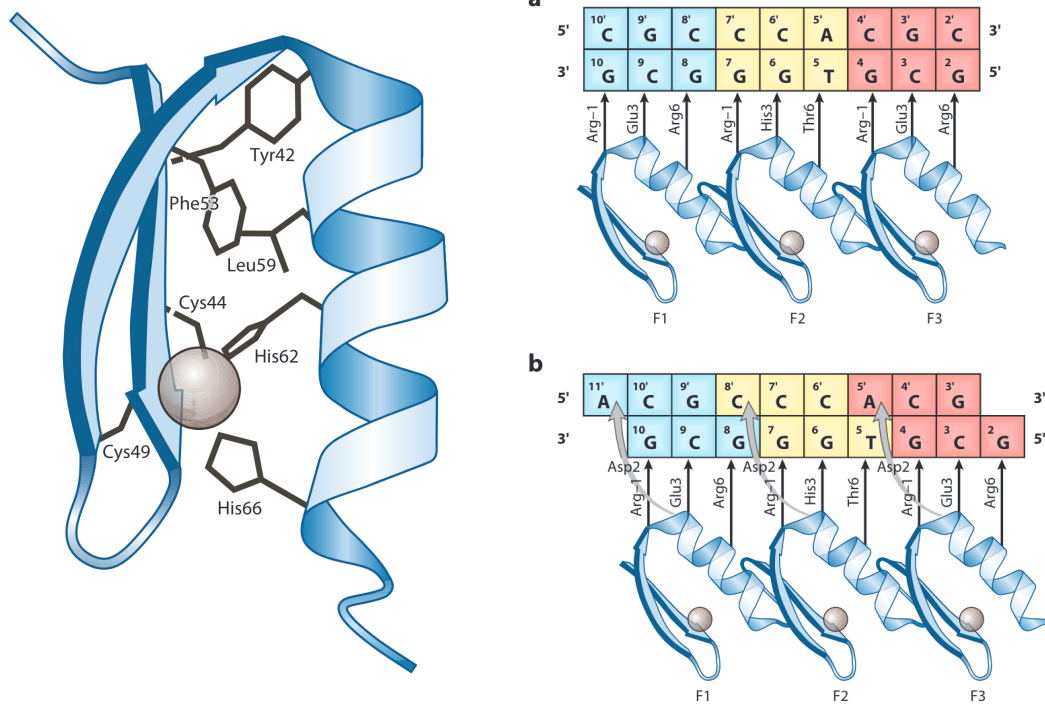
Figure 6: "The structure of a zinc finger from a two dimensional NMR study of a two-finger peptide in solution as well as the first module of modular recognition of DNA by 3 ZFPs (a) vs. the refined model (b)".Taken from Klug, "The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation."

organized into a c-terminus cluster; and 3. Proteins that have more than 3 C2H2

domains that form one or more clusters[29]. Group 2, the most well studied group, includes

conserved TFs that have 3 C2H2 domains, with a large portion of them playing a key

role in the regulation of gene expression in all higher eukaryotes[29]. C2H2-type ZFPs

generally contain anatomical domains such as BTB (Broad-Complex, Tramtrack, and

Bric-a-brac), Kruppel-Associated Box (KRAB) domain, and SCAN (SRE-ZBP, CTfin51,

AW1 and Number 18 cDNA) domain[27]. These domains play a role in regulating immune

response,

cell differentiation, and embryonic development at the transcription and translation level

through specifically binding to the target molecule DNA, RNA, DNA-RNA sequence, and

binding itself to other ZFPs[27].

One of the best known proteins in this family is the highly conserved CTCF

(CCCTC-binding factor) (Figure 7)[29]. CTCF is a transcription factor that plays a key role

in the establishment of chromosomal architecture in vertebrates[29]. But CTCF also plays

a role in many other processes as well, such as embryonic development, the X

chromosome-activation in females, regulation of gene cluster recombination during the

maturation of immunoglobulin genes, and the regulation of alternative splicing[29]. CTCF

contains a cluster of C2H2 zinc finger domains, some of which are highly specific for

binding of the protein to DNA[29]. Proteins that contain C2H2 zinc fingers emerged early

during evolution and are found in many eukaryotes, with many of them being structurally

similar to CTCF[29]. CTCF binding sites are often located at the boundaries of

chromosomal regions, which all have different epigenetic statuses and transcriptional

activity[29]. They are also found at the boundaries of topologically associated domains

(TADs) that spatially separate chromosomes into regions where interactions among

regulatory elements occur[29].
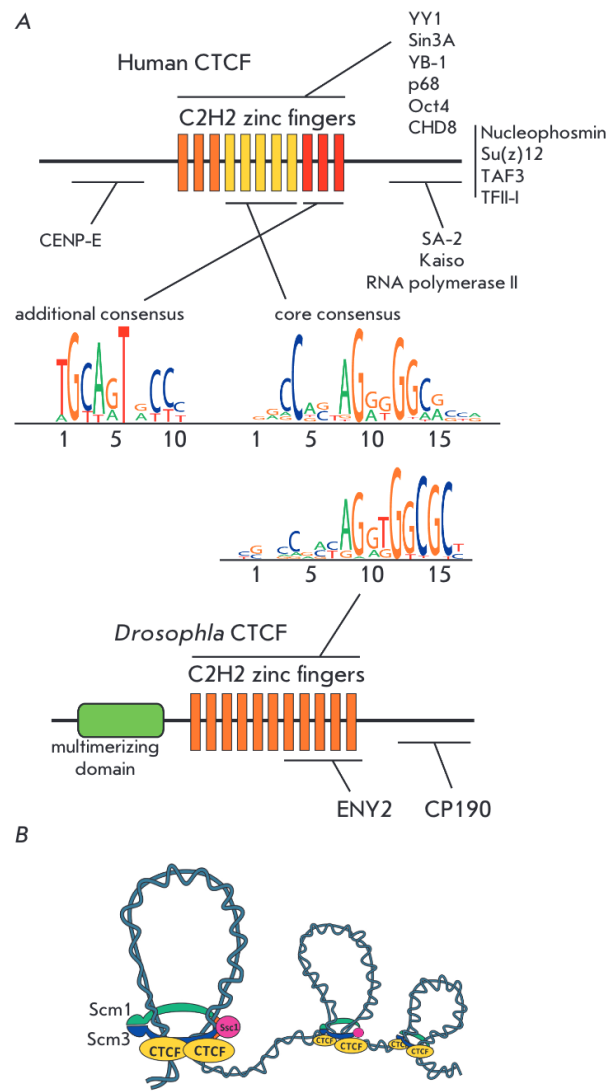
Figure 7: "Comparison of the structures and properties of the Drosophila and human CTCF proteins. (A) The domain structures of the Drosophila and human CTCF proteins. The domains involved in the site-specific DNA recognition and the protein-protein interactions are represented by thin horizontal lines. Drosophila and human [46] CTCFs have similar consensus recognition sites. (B) The mechanism of the long-distance genomic interactions mediated by CTCF and cohesins." Taken from Fedotova et al., 2017

Another notable C2H2 ZFPs domain is KRAB (Figure 8), found only in tetrapods and involved in the repression of transcription[29]. This is a versatile and well-studied mechanism of repression that is characterized by the recruitment of the KRAB-associated protein 1 (KAP 1)[29]. The KAP 1 expression peak is at the early embryonic stages and the transcriptional repression by KRAB C2H2 proteins is critical for early embryonic development[29]. However the majority of KRAB C2H2 proteins are genus and species specific and the evolutionary analysis of the conservation of KRAB C2H2 proteins has shown that these gene families have formed independently in each class of vertebrates[29].

C2H2 ZFPs in invertebrates have one well studied example. In *Anopheles gambiae* (mosquito), *Drosophila melanogaster* (fruit fly), and *Apis mellifera* (honey bee) the zinc finger-associate domain (ZAD) C2H2 proteins (Figure 9) are expressed during oogenesis and early development[29]. This is most well studied in *D. melanogaster*, where ZAD plays a key functional role in development. The Motif 1 Binding Protein (M1BP) is expressed in all tissues at high concentrations and at all stages of development in *D. melanogaster*[29]. M1BP plays a key role in the organization of the architecture of more than 2,000 drosophila promoters with characteristic motifs of (T/C)GG(T/C)CACACTG[29]. Three ZAD C2H2 proteins (Pita, ZIPIC, and Zw5) exhibit properties of insulator/architectural proteins by blocking the interaction between an enhancer and a promoter and maintain long-distance interactions[29]. From the data we have it is clear that ZAD C2H2 TFs play an important role in the organization of the structure and functional activity of promoters, the recruitment of protein complexes, and the formation of the chromosomal architecture[29].
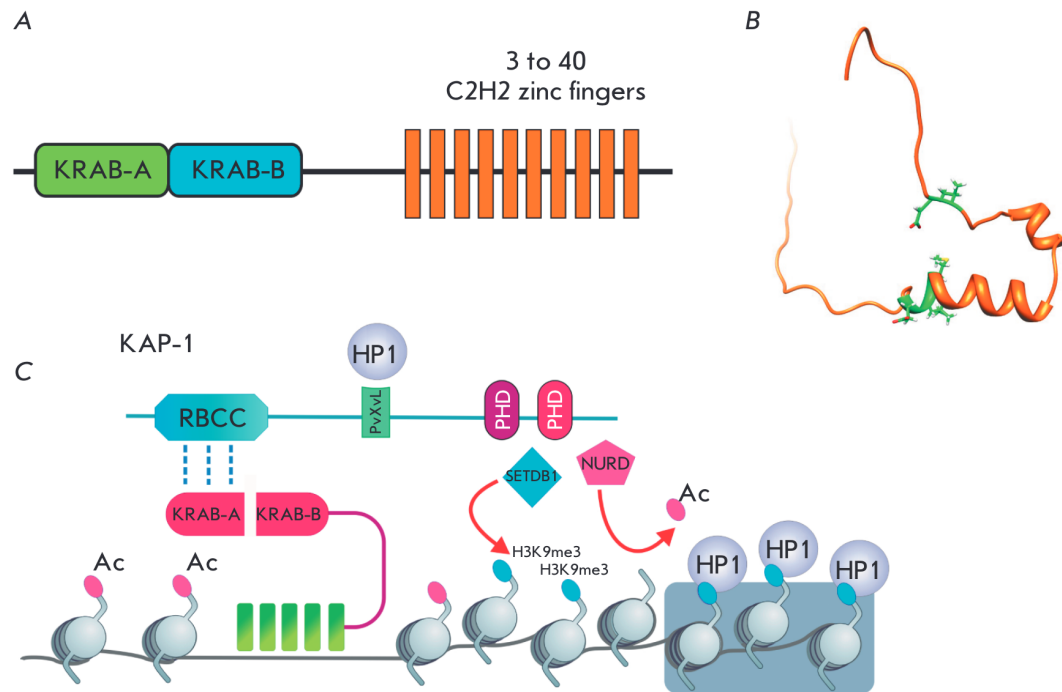
Figure 8:"The structure and properties of the KRAB domain. (A) A typical domain structure of the KRAB C2H2 proteins. (B) The NMR structure of KRAB A: 5 mammalian conserved aa are shown in green (DV in positions 6 and 7, and MLE in positions 36–38); they are essential for the KAP-1 recruitment [PDB 1V65]. (C) The mechanism of KAP 1 recruitment and the subsequent formation of the repressive complex." Taken from Fedotova et al., 2017
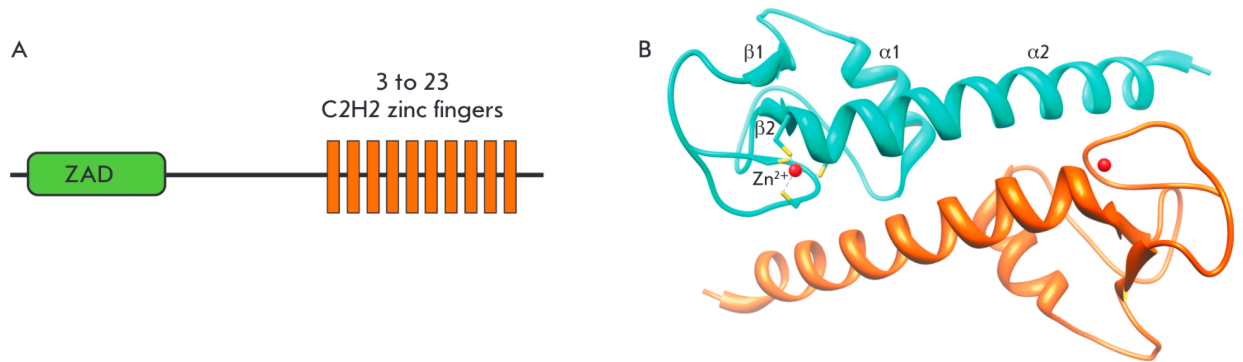
Figure 9: "The structure and properties of the SCAN domain. (A) A typical domain structure of the SCAN C2H2 and SCAN KRAB C2H2 proteins. (B) The crystal structure of a SCAN domain dimer from the Zfp206 protein [110]." Taken from Fedotova et al.,2017

RNA editing in Octopus rubescens optic lobe

Previous work in the Onthank Lab has investigated changes to RNA in the nervous system of *Octopus rubescens* in response to elevated environmental CO_2. Six octopuses were exposed to normal pCO_2 conditions (750 µatm) and 6 octopuses to elevated pCO_2 conditions (1500 µatm)[31]. Octopuses were acclimated to the aquaria for 1 week and then exposed to elevated or control CO_2 levels for 7 days. After 7 days the octopuses were euthanized and optic lobe tissue was harvested. Six samples, three from the elevated CO_2 treatment and three from the control CO_2 treatment, had both gDNA-seq and mRNA-seq performed, the other 6 were saved as biological replicates for verification. The transcriptome was assembled *de novo* from the mRNA reads using trinity. Then, Bowtie2 was used to align both gDNA and mRNA to the transcriptome. Editing sites were detected by identifying loci where gDNA reads disagreed with the transcriptome at a particular locus, or where some number of the mRNA reads disagreed with gDNA reads. In both cases, all gDNA reads had to be uniform at a specific locus to be categorized as a potential editing site to avoid mistaking heterozygotic sites as editing sites. To compare editing sites between normal and high pCO_2 conditions a permutation t-test was done using custom R script[31]. A Benjamin-Hochberg p-value correction for multiple comparisons was used to adjust p-values.

Approximately 12,000 editing sites showed different editing levels between octopuses exposed to high and low CO_2 environments (hereafter referred to as acidification-responsive edits), and overall there appeared to be editing suppression in octopuses subjected to elevated pCO_2. Further, a set of edits was identified that would be a high priority for future investigation. These high priority edits were defined as A-to-G nonsynonymous edits with a mean editing level difference of at least 20% between treatments and were evidenced by at least 10 mRNA reads in each octopus and at least

10 gDNA reads from all octopuses combined, and had a permutation t-test adjusted p-value of less than 0.005, yielding 51 such editing sites. ZFPs appeared to be over-represented in proteins containing editing sites that were differentially edited between high and control CO_2 treatments. ZFPs are 5.3% of the total number of genes in the *Octopus bimaculoides* genome, but were 16.3% of the genes containing differentially edited sites in Ricky's data and 51% of genes containing high priority edits.

Hypothesis

I hypothesize that the ZFPs that are differentially edited in elevated pCO_2 conditions are TFs that are up or down regulating protein production. I hypothesize that the proteins coded for those genes are helping the octopus acclimate to elevated pCO_2.
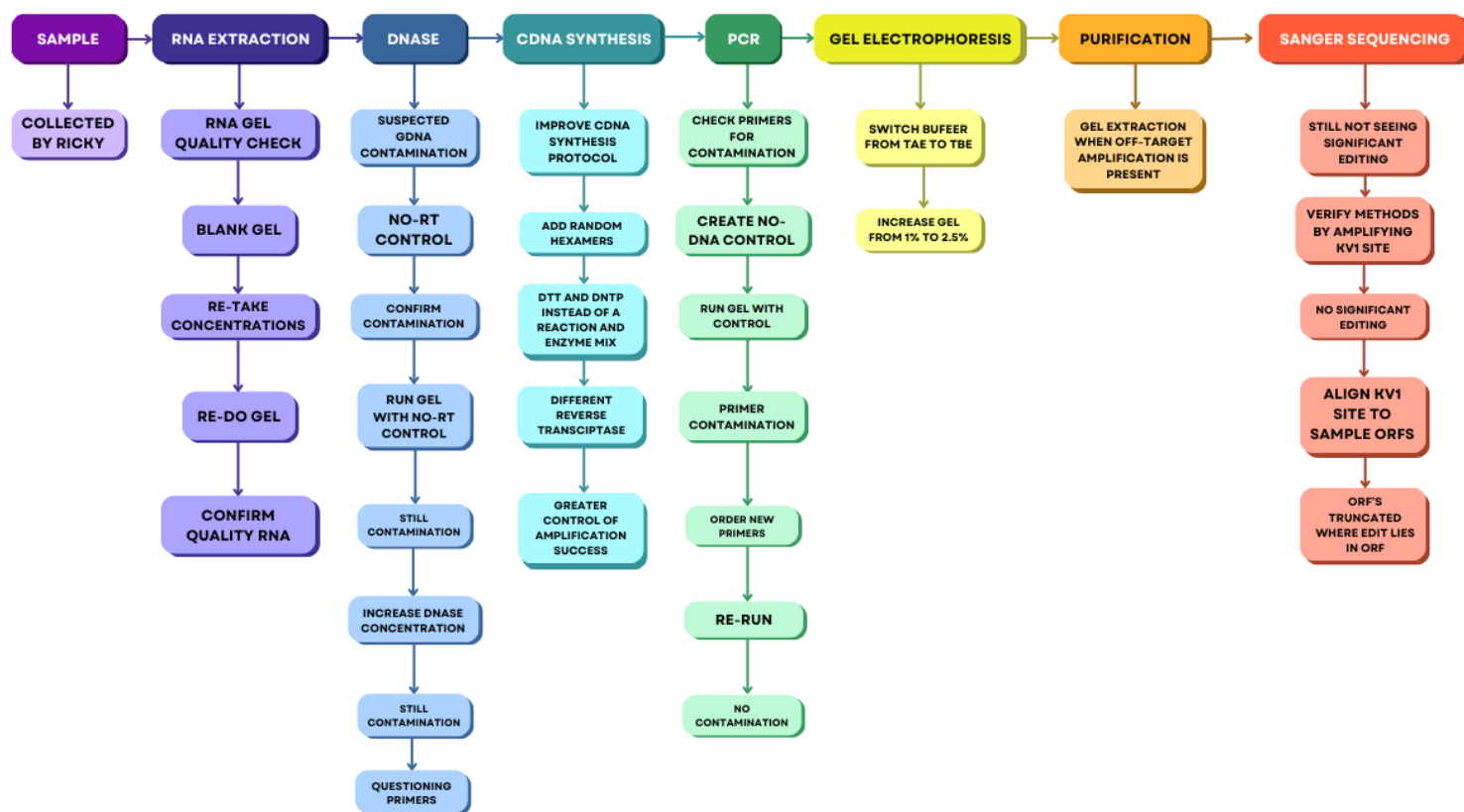
Figure 10: flow chart of laboratory methods and each troubleshooting step taken.

METHODS

Verification of RNA Editing Sites

Verifying the RNA editing sites found to be significantly differentially edited between elevated and control CO_2 treatments in previous Onthank lab member's research is important to confirm that these sites are genuinely occurring and are not false positives[32]. Previous attempts at verification with these samples have not been successful. This is most likely due to gDNA contamination (Joshua Rosenthal, per comms). gDNA contamination could cause the apparent editing rates of specific sites to be heavily skewed toward the unedited state, giving results that differed from his results from sequence analysis[32].

Control for gDNA contamination

Initial detection of contamination

To confirm gDNA contamination, cDNA synthesis was re-run using NEB Protocol #E6560. PCR amplification was run using previously developed primers for sites of interest and NEB protocol #M04096 with a No-Reverse Transcriptase (No-RT) control[32]. In a no-RT control water takes the place of the reverse transcriptase enzyme, and is run alongside each sample.

PCR products from no-RT control samples were run on a 2% agarose gel at 100V in 1x Tris-Acetic acid-EDTA (TAE) buffer, and showed strong bands in No-RT control wells, confirming the presence of gDNA contamination in the template.

First DNAse Treatment Attempt

To eliminate gDNA contamination, samples were treated with DNAse following NEB protocol #M0303, using 10 µg RNA, 10µL DNAse I Reaction Buffer from NEB, 1 µL DNAse I from NEB, and sufficient nuclease-free water to result in a total volume of 100 µL. However, amplification was still observed in the No-RT controls, as evidenced by bright bands on the resulting agarose gels, indicating contamination was not removed.

RNA Integrity Check

Next, to ensure the RNA samples were not degraded, a RNA gel was run using a modified protocol from Molecular Cloning: A Laboratory Manual (Sambrook 2001, Michael Morgan, pers comm). For each sample, the original concentration measurements (taken prior to the first DNAse treatment) were used to calculate the volume required to obtain 20 µg of RNA. Nuclease-free water was then added to bring the total volume to 20 µL. Next 2 µL of 10x MOPS electrophoresis buffer was added, then 4 µL of formaldehyde, 10 µL of formamide, and 1 µL of ethidium bromide. Samples were then incubated for 60 minutes at 55℃, chilled in ice water for 10 minutes, and then centrifuged for 5 seconds. 2 µL of 10x formaldehyde gel-loading buffer was added to each sample and returned to ice. A 2% agarose gel was run at 150V for 35 minutes in 1x MOPS electrophoresis buffer. The gel was visualized on a UV transilluminator.

No bands were observed on the first gel, indicating that there was significantly less RNA in the samples than expected. RNA concentrations of the samples were measured using a ThermoScientific Nanodrop Lite Spectrophotometer (nanodrop) and found they were lower than values obtained prior to the first round of DNAse. These inaccurate RNA concentrations lead to an insufficient load of RNA into the gel. A second

RNA gel was run, with corrected loading volumes and the resulting bands were bright, clear, and un-smeared, indicating the RNA was of good quality.

Modified DNAse Protocol

Next, a DNAse protocol was performed to remove gDNA contamination that other investigators have had success with (Michael Morgan, pers comm). The MessageClean protocol from GenHunter uses a higher concentration of DNAse with a phenol/chloroform extraction cleanup instead of heat inactivation. For DNAse digestion, concentrations of all samples were measured again using a nanodrop. Based on concentration, 10-50 μg of each sample were aliquoted and volume was brought to 50 μL with DEPC treated water. To each sample, 5.7 μL of 10X Reaction Buffer and 2 μL of DNAse 1 were added. Samples were mixed by pipetting up and down and incubated at 37°C for 30 minutes.

For the phenol/chloroform extraction, 40μL of a 3:1 phenol-chloroform solution was added to the DNAse digestion, vortexed for 30 seconds, and then incubated on ice for 10 minutes. Samples were then spun in a centrifuge at 4°C for 5 minutes at max speed. After centrifuging, the upper phase was collected and saved.

The collected upper phase was added to 5 μL of 3M sodium acetate (NaOAc) and 200 μL of 100% ethanol (EtOH), mixed well, and sat for at least 1 hour at -80°C. Afterwards, the RNA were pelted by spinning the samples for 10 minutes at 4°C. Removing the supernatant, the RNA pellet was then rinsed with 0.5 mL of 70% EtOH without being disturbed. The RNA pellet was then spun again for 5 minutes and EtOH was removed and spun one more time to remove any residual liquid. The RNA pellet was then re-dissolved in 10-20 μL of DEPC treated water, with the volume of DEPC water dependent on the size of the pellet, using more water for a larger pellet. Finally,

RNA concentration was measured using a nanodrop. DNAse-treated RNA was then stored in -80°C. DNAse procedure was done only on 3 octopus samples to preserve the other samples until sanger sequencing verification methods can be confirmed.

Verification of Decontamination

To confirm samples no longer had contamination, cDNA synthesis, PCR, gel and electrophoresis was repeated with the 3 samples that were DNAsed (octopus 4, 6, and 10).

cDNA synthesis originally involved NEB protocol, #E6560. This protocol was from the NEB First Strand Synthesis kit, but the kit was not used and instead individual reagents were purchased. To facilitate a more comprehensive representation of the original RNA, random hexamers were added to cDNA synthesis. Additionally, instead of using the NEB ProtoScript II Reaction and Enzyme Mix, 5X First strand buffer, 0.1 M DTT and a 10 mM dNTP mix were used, preceded by Invitrogen's SuperScript III Reverse Transcriptase. This resulted in more reliable amplification, higher yields of cDNA, and allowed flexibility that was crucial for a non-model organism (Michael Morgan, pers comm).

For cDNA synthesis, 1 µg of DNAse-treated RNA was added to a final volume of 9 µL using DEPC water. Then 2 µL of Oligo dT12-18 primers and 1 µL of random hexamers were added to the RNA, and pipetted up and down to mix. The mixture was then heated to 70°C for 10 minutes, transferred immediately to ice afterwards to chill. Next, 4 µL of 5X first strand buffer, 2 µL of 0.1 M DTT, and 1 µL of 10mM dNTP were added in order as listed and mixed by pipetting up and down, then centrifuged. After spinning, the mixture was then incubated at 37°C for 2 minutes. Next, 1 µL of

Superscript II RT was added for a total volume of 20 μL. Another No-RT control was made alongside newly DNAsed samples to confirm if decontamination was successful.

The cDNA synthesis mixture was then pipetted up and down to mix, incubated at 37℃ for 1 hour, and then gradually raised to 50℃ in 1 minute increments. To terminate the reaction, mixture was immediately placed on ice. cDNA (RT rxn) was diluted 1:100 with nuclease-free water to use in PCR reactions and then stored at -80℃.

For PCR, originally primers were used at their original 100 mM concentration. This was too high a concentration and resulted in high primer dimer formation and inefficient use of reagents. Going forward, primers were diluted to 10mM. Additionally, a modified PCR protocol other investigators had success with was implemented (Michael Morgan, per comms). The reaction mixture consisted of 1 μL of 10 mM forward primer, 1 μL of 10 mM reverse primer, 2.5 μL of 1:100 RT Reaction, 10 μL of 2x TAQ Master Mix from NEB, and 5.5 μL of nuclease-free water, for a final volume of 20 μL instead of 50 μL (Michael Morgan, pers comm). Both 1 μL and 2.5 μL of cDNA synthesis 1:100 was tested, with ultimately 2.5 μL of 1:100 producing a brighter and more distinct band. Reaction was mixed by pipetting up and down and briefly spun. PCR parameters were denature for 30 seconds at 95℃, run 40 cycles of 15 seconds at 95℃, followed by 30 seconds at annealing temperature designated by each specific primer pair (Table 1), and another 30 seconds at 68℃. This differed from the original PCR parameters in the number of cycles run, 30 instead of 40, and reducing the duration of each cycle step from 30 seconds to 15 and 60 seconds to 30, respectively. This allowed the same yield of amplified sequence in a shorter run time.

Final extension was at 68℃ for 5 minutes and held at 4℃. PCR products were then run on a 2.5% agarose with a 50 bp ladder. Instead of running the gel in TAE buffer, Tris-Borate-EDTA (TBE) buffer was used because TBE provides high resolution

of small DNA fragments. Resulting gels were visualized on a UV transilluminator. Bands remained visible in no-RT control lanes, indicating persistent contamination.

Primer Contamination Detection

While No-RT control samples were still amplifying, they were showing fainter bands than previously. Confident in the effectiveness of the updated DNAse procedure, other possible sources of contamination were investigated. To check that previously developed primers were not contaminated, a No-DNA control was made, in which water replaced the DNA template in the PCR reaction.

The resulting gel showed bands in No-DNA control lanes, indicating primer contamination. New primers were ordered, PCR and gel electrophoresis were repeated, and the resulting gels no longer showed contamination in lanes in the No-RT control, nor the No-DNA control, resolving the observed contamination.

Sanger Sequencing Verification

Editing Sites Verified

After decontamination was confirmed, two out of 7 editing sites had successful amplification. A site previously identified, TRINITY_DN90_c0_g1_i16:425-2539, and zinc finger 4969, with editing site TRINITY_DN4969_c1_g1_i4:c748-224 (Ricky Wright, 2024). TRINITY_DN90_c0_g1_i16:425-2539 used primers PP1, while TRINITY_DN4969_c1_g1_i4:c748-224 used primers ZFP_4969 (Table 1). Both successful amplifications were purified using the Monarch PCR & DNA Cleanup Kit from NEB. Concentrations were taken with a nanodrop to verify samples met the minimum threshold of 20-40 ng/rxn required by Lone Star Labs (Houston, TX).

Forward Primer	Sequence	Tm (C)	Reverse Primer	Sequence	Tm (C)	Annealing Temp (C)	Date Ordered	Work?
NEW_ZF271_F	ACAGGAG AGAAGCA ATATCC	53	NEW_ZF271_R	CACTGATAC GGTTTCTCT CC	55	47	07/12/24	No
PP1_F	ACACGATC TGTCAGAA AACACA	58	PP1_R	GGTAAATG AGTGAGGC AACTAG	56	49	07/22/24	Yes
ZFP_4969_F	GTAAGGCA GCTACACT CACTAC	55	ZFP_4969_R	GTATGAAC GCGCTTGTG TTAG	60	62	08/01/24	Yes
ZFP_43482_F	GAACACAA GCGCATTC ATACAG	61	ZFP_43482_R	CGCTTGCAT TCAGTCAAC TTAC	60	62	08/01/24	No
Kv1-Forward	AGAAGAG GGATTTAT CAAAGAA G	47	Kv1-Reverse	TTATTTTTG ATATGATTA AACCC	41	45	11/07/24	No
Kv1_F	AGATGAGG GATTTAAC AATCGGG	62	Kv1_R	TTATTTTTG ATATGATTA AACCC	52	47	01/22/25	No
JOSH_Kv1_F	CGTTTCGC TTCCTGTC CTGTG	65	JOSH_Kv1_R	CGGTTGTCA TGGTAACG ACGG	66	62	02/07/25	Yes

Table 1: Table of primers used in PCR amplification.

Sanger sequencing of these samples was performed by Lone Star Labs (Houston, TX)). Samples were sequenced in both the forward and reverse directions to increase the likelihood to obtain at least one high quality chromatogram.

Sequencing Analysis

Sequence files (.ab1 format) were converted to chromatograms using R. The editing site was located, and a cursory visual inspection of the chromatogram was performed to determine if editing was present at the site, which would be evidenced by a "G" peak in addition to, or in replacement of the expected "A" peak at that position. Using this method, editing was not detected at each site.

To further verify the sites of interest, chromatograms were analyzed using MultiEditR v1.1.0, an R package specifically designed to detect and quantify RNA editing from sanger sequencing data[38]. This algorithm, which was adapted from EditR (originally developed for analyzing CRISPR-Cas9 DNA base editing[38,39]), uses an ab1 file along with an ab1 or fasta file of a control sample[38]. Ab1 files were inputted into MultiEditR, with gDNA of octopus 6 used as control. No significant editing at the sites of interest was detected. To address this issue, two approaches were pursued: verifying the bioinformatics pipeline, and verifying the post-bioinformatics verification methods.

Bioinformatics Pipeline Verification

To verify initial edits site detection, raw genomic and transcriptomic reads were obtained from optic lobe tissue of *Doryteuthis pealeii*, for which editing sites had been previously identified[16]. The reliability of the bioinformatics pipeline for identifying editing

sites in *O. rubescens*[41] was assessed by applying it to the published *D. pealeii* dataset*,* with the expectation that it would reproduce the transcriptome-wide editing patterns reported in the original study. This manner of verification had previously been performed on the *O. rubescens* gill transcriptome dataset[40], but not on the optic lobe data on which this study is based[41].

Post-bioinformatics Method Verification

To verify the post-bioinformatics methods, an attempt was made to replicate editing in Kv1, an ion-gated channel previously verified in *O. rubescens*[25]. The Kv1 site was amplified from octopus samples. Primers provided by Rosenthal successfully amplified Kv1, but off-target amplification could not be resolved even after raising annealing temperatures. Therefore, sequences were gel extracted and purified using Thermo Scientific's GeneJET Gel Extraction kit and the DNA sequence determined. No significant editing was found in the resulting chromatograms using MultiEditR.

Transcriptome Check

To confirm this site is significantly edited in the samples, the published *O. rubescens* Kv1 sequence (accession # JQ246413[25]) was blasted against the transcriptome. The only matching transcript was truncated at position 498, leaving out the edited site of interest at position 882. The truncation may have occurred during transcriptome assembly or open reading frame (ORF) identification (Figure 11).
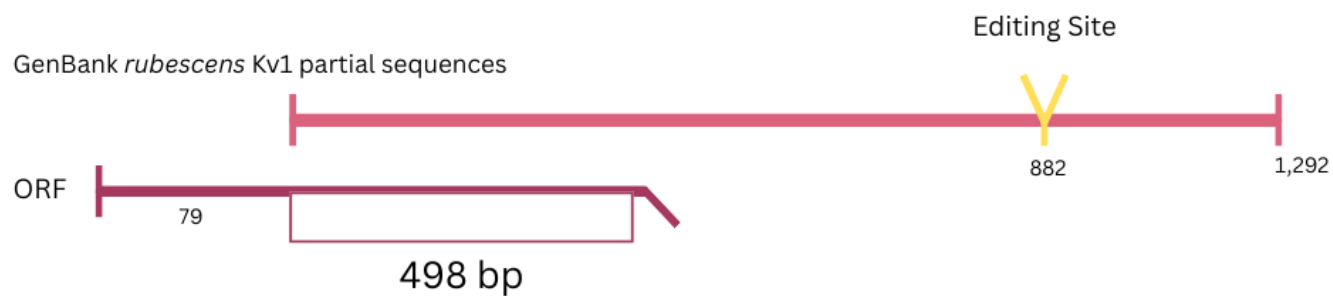
Figure 11: Why Kv1 site was unable to be verified in the transcriptome. Kv1 partial

sequence aligned to an ORF that was truncated before the editing site occurred.

Zinc Finger Protein Prediction

Initial Prediction Pipeline

Previous analyses suggested that zinc finger domains containing proteins were overrepresented among transcripts with editing sites responsive to elevated CO_2[32]. However, zinc finger identification in that analysis relied on annotations generated through homology to sequences in the SWISS-Prot database. This homology-based annotation can introduce errors due to incomplete sequence information, functional divergence of homologs, or inaccuracies in database entries. To address these limitations, a direct prediction approach was employed to identify C2H2 zinc finger domains based on their conserved sequence and structural features using the software PWMpredict.

Improved Zinc Finger Prediction Pipeline

First, the whole transcriptome open reading frames (ORF's) were inputted into R and translated to amino acid sequences (Figure 12). Then, PWM predictor was employed to predict C_2H_2 zinc finger domains in a transcriptome, based on the amino acid sequence[33]. From there the predicted zinc finger domains were cross-referenced with editing sites using R to investigate patterns of RNA editing and presence of zinc finger domains within the transcriptome. The distribution of CO_2-responsive edits between zinc finger containing transcripts and non-zinc finger containing transcripts was compared using a two-sample permutation test.
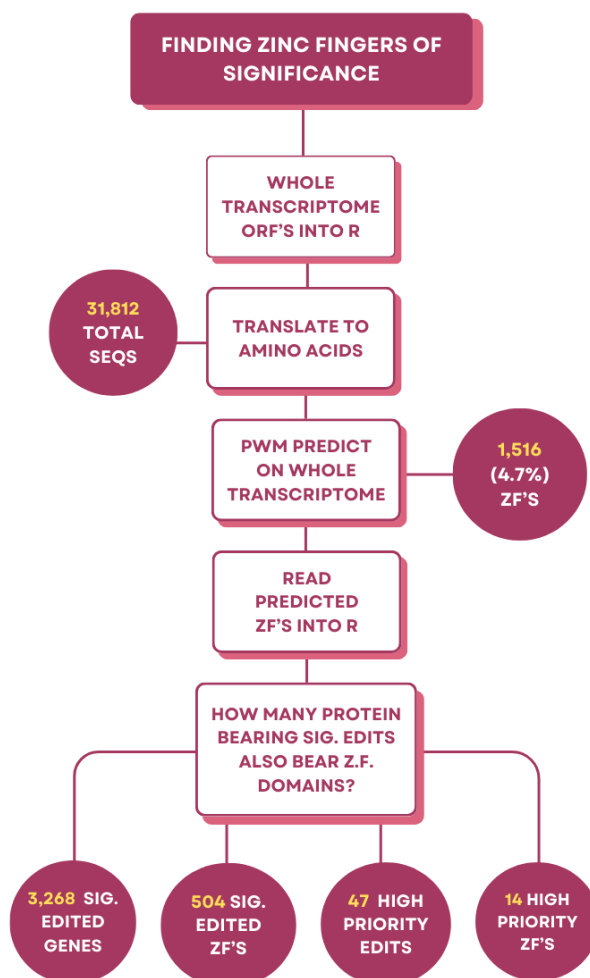
Figure 12: Flow chart of how zinc fingers of significance were found

Determination of Zinc Finger Targets

Edited vs Unedited Sequence Analysis

Using a custom Python script, the edited and unedited versions of the sequences for all nonsynonymous A to G CO_2-responsive edits were generated. These sequences were then translated into their amino acid sequences and put through PWM predict again to predict the nucleotide sequences to which each unedited and edited zinc finger domain binds. Using another custom python script, the PWM output was converted into fastq files, and any bases that had a 50% or less probability were converted to N. A third script identified cases where the predicted zinc finger binding sequences were different between edited and unedited versions (Figure 13).

Functional Prediction of Zinc Finger Targets

Target Transcript Identification

To identify transcripts that may be targets of zinc fingers of interest, the zinc finger binding sites that changed between the unedited and edited versions were then aligned to the *O. rubescens* transcriptome and genome using STAR align[34]. From this a table was created of transcripts that aligned to predicted zinc finger binding targets.

Function Prediction

Most of the predicted zinc finger target transcripts did not have meaningful annotations, nor did BLAST search of these predicted targets reveal similar, well characterized homologs. Therefore, the I-TASSER web server was used to be able to predict the function of these bound proteins of interest. I-TASSER is a web-server based application that models the structures and functions of multi-domain proteins[35].

After I-TASSER predicts the protein structure, the web-server predicts the function using COFACTOR, which deduces protein function based on structural similarity, rather than amino acid sequence comparison[36]. Amino acid sequences corresponding to the predicted zinc finger binding targets, that were identified by aligning bind site sequences to the transcriptome, were submitted to I-TASSER.

Data Availability

All data generated in this thesis are archived on Zenodo.
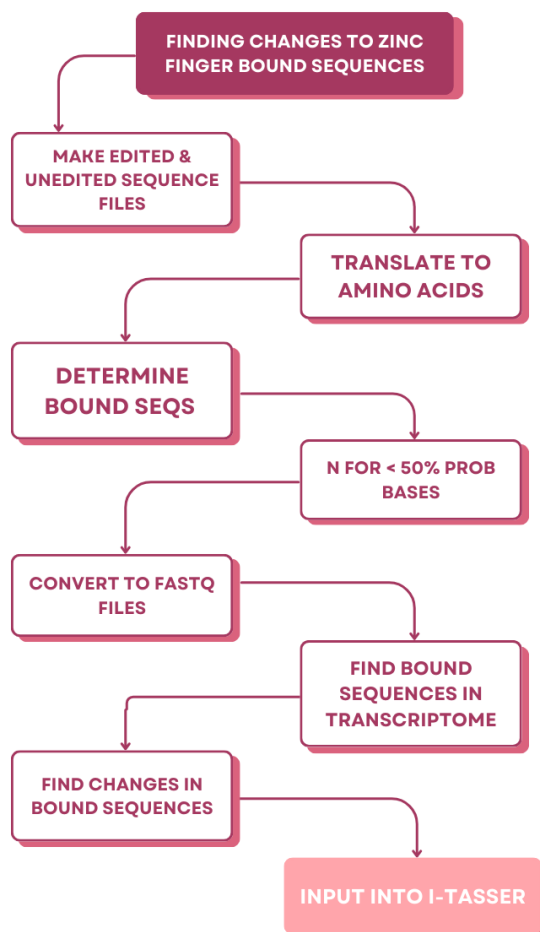
Figure 13: Flow chart of how changes in bound sequences of zinc fingers were found.
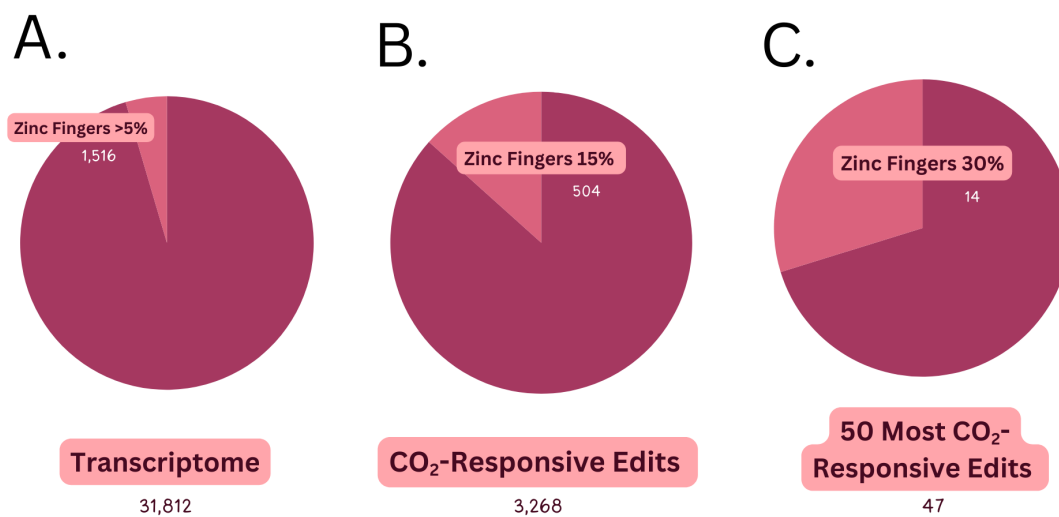
Figure 14: The proportion of zinc fingers throughout the transcriptome. A. Zinc fingers take up less than 5% of the entire transcriptome. B. Zinc fingers take up 15% of all the edits that are significantly differentially edited between treatments. C. 14 out of 47 most significantly differentially edited sites are zinc fingers.

Zinc Finger ID	Unedited Bound Sequence ID	Edited Bound Sequence ID	Unedited Bound Sequence ID	Edited Bound Sequence ID
TRINITY_DN210_c1_g2_i6:c2214-1	TRINITY_DN268_c0_g1_i1	TRINITY_DN268_c0_g1_i1	ctg_p_l_000067_0	ctg_p_l_000067_0
TRINITY_DN26251_c1_g1_i4:17-634	did not align	TRINITY_DN227_c2_g1_i10	did not align	ctg_p_l_000045_0
TRINITY_DN43482_c0_g1_i1:61-525	TRINITY_DN10497_c7_g1_i1	TRINITY_DN49772_c0_g1_i1	ctg_p_l_000083_0	ctg_p_l_001707_0
TRINITY_DN4969_c1_g1_i4:c748-224	TRINITY_DN22626_c0_g2_i1	TRINITY_DN4257_c1_g1_i26	ctg_p_l_000033_0	ctg_p_l_000042_0
TRINITY_DN46147_c0_g2_i1:c1242-112	did not align	did not align	ctg_p_l_000019_0	ctg_p_l_000118_0
TRINITY_DN2112_c0_g1_i1:c749-3	did not align	did not align	ctg_p_l_000029_0	did not align

Transcriptome Alignment   **Genome Alignment**

Table 2: The 6 zinc fingers that had their bound sequence change and aligned to either the transcriptome or genome. Transcriptome sequence or genomic scaffold that each zinc finger aligned to is also listed. The other 22 zinc fingers that were found to have their bound sequence change with editing did not successfully align.

RESULTS

To test whether zinc finger proteins were disproportionately affected by RNA editing in response to ocean acidification, C2H2 domains were predicted from translated transcript sequences and cross-referenced with transcripts containing significantly edited sites. From this analysis of the *Octopus rubescens* transcriptome, it was found that 1,516 transcripts contained zinc-finger domains (ZFPs) making up less than 5% of all transcripts (Figure 14A). Among editing sites that are responsive to ocean acidification (defined as those showing a significant difference in the proportion of edited reads between CO_2 treatments) 504, or 15% were located in transcripts encoding ZFPs (Figure 14B). However, analysis of the transcripts containing the top 50 most responsive sites (which was 47 transcripts) revealed that 30% of them are ZFPs (Figure 14C).

The number of acidification-responsive edits differed between non-zinc finger and zinc finger containing transcripts (p-value < 2.2×10^{-16}). Transcripts that do not contain zinc fingers have a 1 in 3 chance of being edited, with a large majority of them having no edits at all, while zinc finger containing transcripts on average have 1 edit (Figure 15). Additionally, 194 transcripts were identified as having more than 10 significant edits throughout the whole transcriptome.

Zinc Finger Characterizations

Zinc finger–bound sequences were predicted using PWM predictor and compared between edited and unedited forms using custom Python scripts. The function of the target proteins was then predicted using structural modeling with I-TASSER. Twenty-eight ZFPs in the transcriptome were found to have their bound sequence change with editing, and all of those were subsequently aligned to the transcriptome and genome to identify potential target regions (Table 2). Out of the twenty-eight, only six

sequences aligned to either the transcriptome or genome. Of the 4 that aligned to the transcriptome, bound sequences were sent to I-TASSER for further investigation. All mapped targets correspond to a currently uncharacterized protein-coding sequence.

For the majority of predicted protein structures there were no high confidence scores in either structural protein modeling (C-score < -1.4), structural analogs (template modeling, or TM-scores < 0.5) nor in predicted molecular function, biological process, or cellular component. Except for the edited bound sequence of Zinc Finger 4969, whose structural analog, a IR subunit from the nuclear pore complex protein of *Xenopus laevis* (PDB accession # 7wkk), received a TM-score of 0.932, with the highest score obtainable being 1.0 (Table 3).

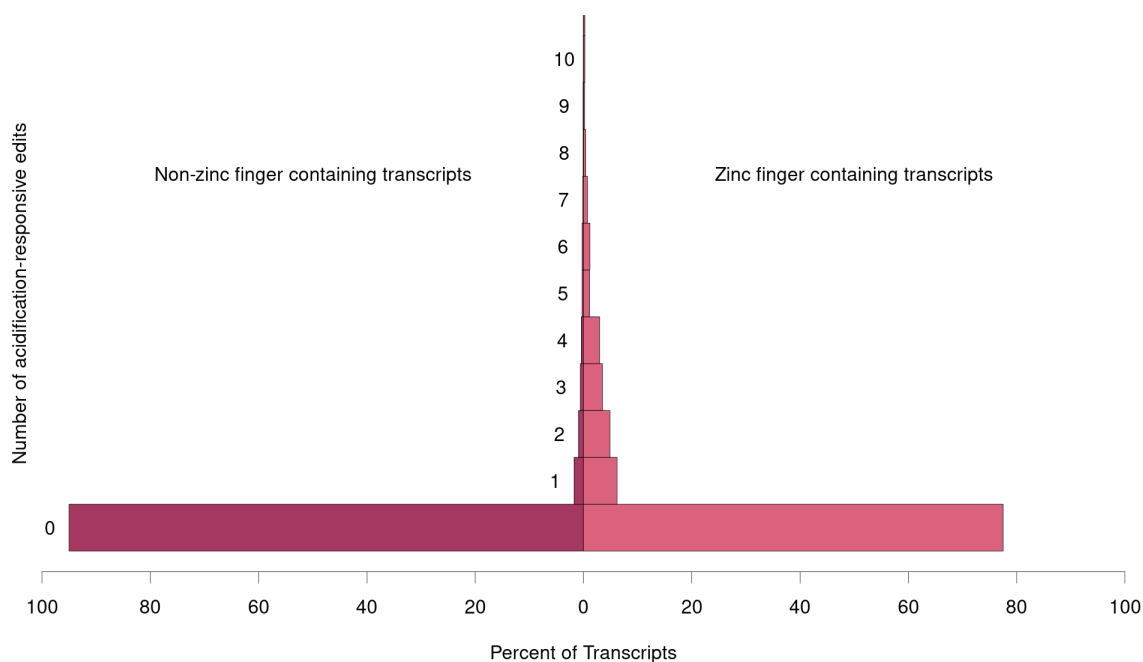Figure 15: Distribution of OA-responsive edits in zinc finger domain containing (right) and non-zinc finger domain containing (left) transcripts plotted as percent of total transcripts in each category. Zinc finger domain containing transcripts harbor significantly more OA-responsive edits than non-zinc finger domain containing transcripts (Two sample permutation test, Z=-11.459, p < 2.2 x 10-16)

Zinc Finger	Edited or Unedited	C-Score	PDB Accession #	Analog Name	TM-Score	Molecular Functions	GO-Score	Biological Processes	GO-Score	Cellular Component	GO-Score
43482	Unedited	-4.03	7sa6A	Factor H binding protein	0.592	molybdenum ion binding	0.13	electron transport chain	0.12	periplasmic space	0.13
43482	Edited	-2.29	5xb7A	GH42-alpha-L-arabinopyranosidase	0.469	ion binding	0.47	primary metabolic process	0.47	cytoplasm	0.32
46147	Unedited	-2.22	1dgjA2	Aldehyde oxidoreductase	0.434	metal cluster and ion binding	0.45	carboxylic acid biosynthetic process	0.44	cell periphery	0.44
46147	Edited	-3.98	6gyhA	Family A G protein-coupled receptor-like protein	0.519	oxidoreductase activity	0.07	oxidation-reduction process	0.13	membrane integrity	0.13
4969	Unedited	-4.2	4mftA	ChpT protein	0.491	acetyltransferase activity	0.47	N-terminal protein amino acid acetylation	0.07	intracellular role	0.37
4969	Edited	-1.28	7wkkB	IR subunit of NPC	0.932	protein transport activity	0.5	virus-host interaction	0.39	organelle envelope	0.59
210	Unedited	-1.41	3ja4a	RNA-directed RNA polymerase	0.826	RNA-directed RNA polymerase	0.58	Viral genome replication	0.58	Virion component	0.45
210	Edited	-1.41	3ja4a	RNA-directed RNA polymerase	0.826	RNA-directed RNA polymerase	0.58	Viral genome replication	0.58	Virion component	0.45
26251	Unedited - Did not align	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
26251	Edited	-2.28	4d8mA1	Bacillus thuringiensis	0.658	Transmembrane transporter activity	0.58	Protein localization to nucleus	0.58	Pore complex	0.58

Table 3: Table of each individual zinc finger (ZF) that was further investigated along with their edited and unedited bound sequences. Analogous structure name, PDB Accession #, molecular function, biological process, and cellular components were identified via I-TASSER. C-score is for the bound sequence structure prediction. Only the top hit was included for each category. For a full list of all results, see appendix.

DISCUSSION

Building on previous work conducted in this laboratory, this study refines zinc finger protein identification in the octopus nervous system transcriptome and confirms that the apparent overrepresentation of editing in zinc finger genes is not an artifact of annotation. It also represents an initial effort to characterize the potential functional consequences of editing in these zinc finger proteins by identifying changes in their predicted binding sequences and structural targets. This study also adds support to the hypothesis that RNA editing is involved in cephalopod environmental acclimation.

The identification of C2H2 zinc fingers' role in protein diversification in ocean acidification is unsurprising in the context of the C2H2 zinc finger family being greatly expanded in the octopus genome[28]. I found the *O. rubescens* transcriptome to contain 1,516 C2H2 zinc finger containing transcripts, quite similar to the the *Octopus bimaculoides* genome containing 1,790 C2H2 zinc finger-containing genes, and much greater than the 764 found in the human genome. Based on the evidence presented that zinc fingers harbor significantly more acidification-responsive editing sites than other proteins (Figures 15 & 16), and that, for at least some, the nucleotide sequences they bind to are changed with editing (Table 3). These findings suggest that changing gene regulation through modifying zinc fingers may be a major mechanism employed by octopuses to adapt to low environmental pH.

C2H2 zinc fingers are the largest family of transcription factors in eukaryotes and are commonly used to mediate responses to abiotic stress, including low environmental pH[37]. In plants, C2H2 zinc fingers have been called the "master regulators of abiotic stress responses"[38]. This is due to the role C2H2 zinc fingers play in salt tolerance, osmotic stress regulation, cold resistance, drought resistance, oxidative stress, high-light stress, and hormonal stress resistance[38]. When looking at how plants deal with low soil

pH, C2H2 zinc fingers make another appearance. Specifically, the STOP1 family of C2H2 zinc fingers has been shown to have a central factor in modulating the response to pH (proton) stress in numerous plants.

C2H2 zinc fingers are involved in responses to abiotic stress across diverse animal taxa as well. The diamondback moth, *Plutella xylostella,* upregulates the expression of 4 zinc fingers (*PxZFP320, PxZNF568, PxZNF93, and PxZNF266)* to mitigate oxidative damage induced by elevated temperatures[39,40]. *Apis cerana*, a honeybee from China, upregulates zinc finger AcZFP271 during exposure to oxidative stress including extreme temperature[41]. In mice cell models, zinc finger Znf179 was found to induce the expression of Sp1, a redox-regulated transcriptional activator, after undergoing oxidative stress[42]. A zinc finger protein that is conserved from insects to mammals, known as metal-responsive transcription factor-1 (MTF-1), modulates the expression of multiple genes in response to heavy metal stress[43]. From insect gene knockouts, to mammalian cell cultures, multiple lines of evidence support the hypothesis that C2H2 zinc finger proteins act as modulators of abiotic stress tolerance by gene regulation.

To date, this study provides the first evidence for stress-triggered RNA editing changes in a C2H2 ZFP mRNA. However, this observation raises the question of how could RNA editing plausibly alter the function of a zinc finger protein?

One route could be the modification of interactions with coactivators and corepressors. For example, the KRAB domain is a potent transcriptional regression module that helps ZF proteins interact with KAP-1, which will recruit other transcriptional factors that repress the genes that KRAB ZFPs bind to[44]. An edit to the KRAB domain could change how zinc fingers are able to repress the genes they regulate. RNA editing could also alter post-translational modification. For example, zinc finger protein GATA-1

is a transcription factor that is acetylated at lysine residues near one of the zinc finger domains, which modulate GATA-1's interactions with other proteins[45]. If RNA editing recoded one or more of these lysine codons to a codon for another amino acid, it would disrupt GATA-1 acetylation and alter the protein's function. Additionally, changes in codon usage can influence translation efficiencies, with certain codons, even synonymous codons, being translated more or less efficiently depending on tRNA availability and ribosome dynamics[46]. RNA editing that results in codon substitutions, even synonymous changes, could therefore impact zinc finger translation and subsequent abundance. Finally, RNA editing that alters the amino acid sequence of the zinc finger domain itself, may modify its nucleotide-binding behavior. Because each C2H2 zinc finger domain recognizes a specific 3-nucleotide sequence, such changes could weaken or strengthen binding affinity, or shift the binding specificity entirely. This is supported by the identification of 28 zinc finger transcripts whose predicted binding sequences were modified by RNA editing. This last mechanism, modification of nucleotide binding specificity, likely represents at least one method by which RNA editing modulates zinc finger function in octopuses.

Of the potential mechanisms by which RNA editing could modulate zinc finger function, This study further investigated how editing might alter the DNA-binding specificity of these proteins by identifying their potential targets of the edited and unedited version in the octopus transcriptome and genome. Identification of zinc finger targets proved difficult largely due to the poor characterization of the octopus genome. Of the 33,638 predicted genes in the *O. bimaculoides* genome (genbank ASM119413v2), 13,494 are unannotated, and 6,431 of the annotated genes are labeled as uncharacterized, meaning 59% of the genome has no functional annotation. This made assigning functionality to the zinc finger targets particularly challenging.

Nevertheless, one zinc finger target was characterized in greater detail using structural analysis. The edited version of zinc finger 4969's predicted binding site aligned to transcript DN4257 in the *O. rubescens* transcriptome. While there were no available annotations for this transcript, the structural prediction of the amino acid translation of this transcript was an exceptionally strong match (TM-score=0.94) for the *Xenopus laevis* (the African clawed frog) inner ring of the nuclear pore complex (NPC), which mediates nucleocytoplasmic shuttling[47].

The NPC is well-known as a key regulator in molecular traffic between the cytoplasm and the nucleus[48]. However, the NPC is also emerging as an important regulator of gene expression, as it is found to be a central unit for a network of proteins and ribonucleoproteins positioned along the nuclear basket structure and out into the nuclear periphery, interconnecting other NPC's that ensure efficient control of gene expression[48]. Nucleoporins (Nup) is a family of proteins that make up the NPC, and play key roles in responses to environmental stress. For instance, Nup54 was found to be one of the top regulation-genes involved in adapting endurance to high-pH stress in Pacific white shrimp, *Litopenaeus vannamei* [49].

This is not the only case in which the expression of nucleoporins changes as a result of environmental stressors. In the two toned pygmy squid, *Idiosepius pygmaeus,* transcriptomic analysis following exposure to elevated pCO_2 led to the identification of several central nervous system (CNS)-specific hub genes (genes that are correlated with both treatments, in this case CO_2, and the outcome of interest, in this case a set of behavioral assays), of which 3 nucleoporin proteins were identified, Nup160, Nup155, and Nup205[50]. As hub genes, expression of these genes was not only impacted by elevated CO2, but also correlated with the behavioral measures of the study such as average speed moved and total distance moved during the experiments. Both Nup155

and Nup 205 are primary components of the inner ring of the nuclear core complex in vertebrates[48].

Together, this evidence suggests a regulatory pathway in which the NPC is targeted by a zinc fingers protein whose mRNA is modified by acidification-responsive RNA editing. In cephalopods, environmental acidification appears to alter RNA editing patterns, including increased editing of ZF4969. The edited form of the ZF4969 shows enhanced predicted binding to the mRNA of a NPC inner ring nucleoporin, possibly increasing its expression, as has previously been shown in cephalopods under elevated pCO_2. The upregulation of this nucleoporin may, in turn, contribute to the behavior changes that have been documented in cephalopods exposed to ocean acidification. Currently, this proposed mechanism is based on correlative and predictive evidence, and future experiments will be needed to validate it. However, if confirmed, this would represent a novel mechanism by which RNA editing can modulate responses to environmental stress by altering the function of zinc finger proteins.

Verification efforts were unsuccessful both in the ZFP targets and the positive control site, Kv1, likely due to multiple factors. MultiEditR did not detect significant editing at the site and visual inspection did not suggest editing was present. Attempts to confirm editing in the transcriptome were unsuccessful, which is likely attributable to a truncated Kv1 sequence in the *O. rubescens* transcriptome used in this study (Figure 11). However it is likely the site is edited, considering Garrett & Rosenthal previously verified editing in Kv1 in *O. rubescens* [25] and the site was found to be edited in gill tissue (Kirt Onthank, per comms)[25].

One possible explanation for verification failure is mislabeling and misidentification of samples obtained from previous grad students; if tissues were incorrectly identified and optic lobe tissue not used, the absence of editing could be

explained.Another possible failure point in verification lies in the post-sequencing analysis. I could not find any sites for which MultiEditR was able to confirm editing in any chromatograms in this study, which is unexpected given the prevalence RNA editing is employed by octopuses. Ultimately, the challenges with verification appear to arise from problems beyond the specific ones I addressed in this study.

Additionally, this study generated many avenues for future work to proceed upon. As this represents the first identification of C2H2 zinc finger modification by RNA editing in response to environmental conditions, many routes can be taken to further describe these protein changes and characterization. The wide array of functions of zinc finger proteins makes this avenue of research particularly promising. The next logical step would be to recreate the bioinformatics pipeline to use the genome instead of the transcriptome, which would give a broader idea of the amount of zinc finger proteins being employed. One of the limitations of using the transcriptome and only ORF's is that there are likely more edits in the UTR regions of zinc fingers. Some could be bound to regions that are not only in the ORF. I am unsure if using the genome would fix problems with verification, but repeating my steps after the genome has been used would be helpful in further amending the verification problem.

In conclusion, these findings expand our understanding of how RNA editing may contribute to environmental acclimation in cephalopods. By uncovering a potential mechanism in which RNA editing alters nucleotide-binding specificity of C2H2 zinc finger proteins in response to environmental acidification, this study provides a foundation for future work exploring how transcriptomic changes can mediate the cephalopod response to ocean acidification. While key aspects of this mechanism remain to be experimentally validated, the evidence presented here highlights how RNA editing and zinc finger transcriptional factors may work in concert to coordinate responses to environmental

stress. Continued investigation into the targets and downstream effects of edited ZFPs will be essential for clarifying their role in the broader context of cephalopod resilience and adaptation.

ACKNOWLEDGEMENTS

LITERATURE CITED

1. Calvin, K. *et al. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland.* https://www.ipcc.ch/report/ar6/syr/ (2023) doi:10.59327/IPCC/AR6-9789291691647.

2. Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. Ocean acidification: the other CO_2 problem. *Annu. Rev. Mar. Sci.* **1**, 169–192 (2009).

3. Jiang, L. *et al.* Global Surface Ocean Acidification Indicators From 1750 to 2100. *J. Adv. Model. Earth Syst.* **15**, e2022MS003563 (2023).

4. Borges, F. O., Sampaio, E., Santos, C. P. & Rosa, R. Climate-Change Impacts on Cephalopods: A Meta-Analysis. *Integr. Comp. Biol.* **63**, 1240–1265 (2023).

5. Boletzky, S. von. Biology of early life stages in cephalopod molluscs. *Adv. Mar. Biol.* **44**, 144–204 (2003).

6. Gutowska, M. A. *et al.* Acid–base regulatory ability of the cephalopod (Sepia officinalis) in response to environmental hypercapnia. *J. Comp. Physiol. B* **180**, 323–335 (2010).

7. Onthank, K. L., Trueblood, L. A., Schrock-Duff, T. & Kore, L. G. Impact of short- and long-term exposure to elevated seawater P_{CO_2} on metabolic rate and hypoxia tolerance in *Octopus rubescens*. *Physiol. Biochem. Zool.* **94**, 1–11 (2021).

8. Trueblood, L. A. *et al.* Bathyal octopus, *Muusoctopus leioderma*, living in a world of acid: First recordings of routine metabolic rate and critical oxygen partial pressures of a deep water species under elevated pCO2. *Front. Physiol.* **13**, 1039401 (2022).

9. Rosa, R. & Seibel, B. A. Synergistic effects of climate-related variables suggest future physiological impairment in a top oceanic predator. *Proc. Natl. Acad. Sci.* **105**,

20776–20780 (2008).

10. Kaplan, M. B., Mooney, T. A., McCorkle, D. C. & Cohen, A. L. Adverse Effects of Ocean Acidification on Early Development of Squid (Doryteuthis pealeii). *PLoS ONE* **8**, e63714 (2013).

11. Sigwart, J. D. *et al.* Elevated pCO2 drives lower growth and yet increased calcification in the early life history of the cuttlefish Sepia officinalis (Mollusca: Cephalopoda). *ICES J. Mar. Sci.* **73**, 970–980 (2016).

12. Culler-Juarez, M. E. & Onthank, K. L. Elevated immune response in *Octopus rubescens* under ocean acidification and warming conditions. *Mar. Biol.* **168**, 137 (2021).

13. De La Chesnais, T., Fulton, E. A., Tracey, S. R. & Pecl, G. T. The ecological role of cephalopods and their representation in ecosystem models. *Rev. Fish Biol. Fish.* **29**, 313–334 (2019).

14. Teh, L. C. L. & Sumaila, U. R. Contribution of marine fisheries to worldwide employment. *Fish Fish.* **14**, 77–88 (2013).

15. González, Á. F. & Pierce, G. J. Advances in the study of cephalopod fisheries and ecosystems. *Fish. Res.* **242**, 105975 (2021).

16. Alon, S. *et al.* The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *eLife* **4**, e05198 (2015).

17. Liscovitch-Brauer, N. *et al.* Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* **169**, 191-202.e11 (2017).

18. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).

19. Rosenthal, J. J. C. The emerging role of RNA editing in plasticity. *J. Exp. Biol.* **218**, 1812–1821 (2015).

20. Patton, D. E., Silva, T. & Bezanilla, F. RNA Editing Generates a Diverse Array of Transcripts Encoding Squid Kv2 K+ Channels with Altered Functional Properties. *Neuron* **19**, 711–722 (1997).

21. Rosenthal, J. J. C. & Bezanilla, F. Extensive Editing of mRNAs for the Squid Delayed Rectifier K+ Channel Regulates Subunit Tetramerization. *Neuron* **34**, 743–757 (2002).

22. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11–19 (1991).

23. Rosenthal, J. J. C. & Eisenberg, E. Extensive Recoding of the Neural Proteome in Cephalopods by RNA Editing. *Annu. Rev. Anim. Biosci.* **11**, 57–75 (2023).

24. Colina, C., Palavicini, J. P., Srikumar, D., Holmgren, M. & Rosenthal, J. J. C. Regulation of Na+/K+ ATPase Transport Velocity by RNA Editing. *PLOS Biol.* **8**, e1000540 (2010).

25. Garrett, S. & Rosenthal, J. J. C. RNA Editing Underlies Temperature Adaptation in K+ Channels from Polar Octopuses. *Science* **335**, 848–851 (2012).

26. Birk, M. A. *et al.* Temperature-dependent RNA editing in octopus extensively recodes the neural proteome. *Cell* **186**, 2544-2555.e13 (2023).

27. Li, X. *et al.* Structures and biological functions of zinc finger proteins and their roles in hepatocellular carcinoma. *Biomark. Res.* **10**, 2 (2022).

28. Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–224 (2015).

29. Fedotova, A. A., Bonchuk, A. N., Mogila, V. A. & Georgiev, P. G. C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae* **9**, 47–58 (2017).

30. Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation

and Genome Manipulation. *Annu. Rev. Biochem.* **79**, 213–231 (2010).

31. Sereewit, J. A-to-I RNA editing in *Octopus rubescens* in response to ocean acidification. (Walla Walla University, 2022).

32. Wright, R. O. CHARACTERIZING A-TO-I RNA EDITS IN THE NERVOUS SYSTEM OF OCTOPUS RUBESCENS IN RESPONSE TO OCEAN ACIDIFICATION.

33. Persikov, A. V. & Singh, M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108 (2014).

34. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

35. Zhou, X. *et al.* I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat. Protoc.* **17**, 2326–2353 (2022).

36. Zhang, C., Freddolino, P. L. & Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Res.* **45**, W291–W299 (2017).

37. Kundu, A. *et al.* GhSTOP1, a C2H2 type zinc finger transcription factor is essential for aluminum and proton stress tolerance and lateral root initiation in cotton. *Plant Biol.* **21**, 35–44 (2019).

38. Han, G. *et al.* C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Front. Plant Sci.* **11**, 115 (2020).

39. Miao, X. *et al.* A zinc finger protein shapes the temperature adaptability of a cosmopolitan pest. *Open Biol.* **15**, 240346 (2025).

40. Li, T. *et al.* Zinc finger proteins facilitate adaptation of a global insect pest to climate change. *BMC Biol.* **22**, 303 (2024).

41. Guo, H. *et al.* Identification of an Apis cerana zinc finger protein 41 gene and its involvement in the oxidative stress response. *Arch. Insect Biochem. Physiol.* **108**,

e21830 (2021).

42. Chuang, J.-Y. *et al.* Specificity protein 1-zinc finger protein 179 pathway is involved in the attenuation of oxidative stress following brain injury. *Redox Biol.* **11**, 135–143 (2017).

43. Egli, D. *et al.* Knockout of 'metal-responsive transcription factor' MTF-1 in Drosophila by homologous recombination reveals its central role in heavy metal homeostasis. *EMBO J.* **22**, 100–108 (2003).

44. Lupo, A. *et al.* KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Curr. Genomics* **14**, 268–278 (2013).

45. Jen, J. & Wang, Y.-C. Zinc finger proteins in cancer progression. *J. Biomed. Sci.* **23**, 53 (2016).

46. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.* **107**, 3645–3650 (2010).

47. Huang, G. *et al.* Cryo-EM structure of the inner ring from the Xenopus laevis nuclear pore complex. *Cell Res.* **32**, 451–460 (2022).

48. Strambio-De-Castillia, C., Niepel, M. & Rout, M. P. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 490–501 (2010).

49. Huang, W. *et al.* Analysis of the transcriptome data in *Litopenaeus vannamei* reveals the immune basis and predicts the hub regulation-genes in response to high-pH stress. *PLOS ONE* **13**, e0207771 (2018).

50. Thomas, J. T. *et al.* Transcriptomic responses in the nervous system and correlated behavioural changes of a cephalopod exposed to ocean acidification. *BMC Genomics* **25**, 635 (2024).

APPENDIX

Zinc Finger	Edited or Unedited	C-Score	PDB Accession #	Analog Name	TM-Score	Molecular Functions	GO-Score	Biological Processes	GO-Score	Cellular Component	GO-Score
43482	Unedited	-4.03	7sa6A	Factor H binding protein	0.592	molybdenum ion binding	0.13	electron transport chain	0.12	periplasmic space	0.13
			4ayiD	Complement factor H	0.588	electron transfer activity	0.13	transport	0.13		
			2kdyA	Outer membrane lipoproteins	0.478	nitrate reductase activity	0.13	nitrate assimilation	0.13		
			3pqsA	Transferrin-binding protein	0.454	4 iron/4sulfure cluster binding	0.13	cell adhesion	0.07		
			6om5A	haemophore	0.451	DNA-directed RNA polymerase activity	0.07	carbohydrate metabolic process	0.07		
						DNA binding	0.07	DNA-dependent transcription	0.07		
						alpha-galactosidase activity	0.07				
43482	Edited	-2.29	5xb7A	GH41-alpha-L-arabinopyranosidase	0.469	ion binding	0.47	primary metabolic process	0.47	cytoplasm	0.32
			1eh9A3	Glycosyltrehalose trehalohydrolase	0.455			multi-organism process	0.39		
			2zs6B2	Hemagglutinin components HA3	0.438						
			6vbu92	Bardet-Biedl syndrome 18 protein	0.435						
			3zmrA1	cellulase	0.433						
46147	Unedited	-2.22	1dgjA2	Aldehyde oxidoreductase	0.434	metal cluster and ion binding	0.45	carboxylic acid biosynthetic process	0.44	cell periphery	0.44
			6mrfA	Methionine aminopeptidase	0.434			cellular amine metabolic process	0.44		
			1sijA	Aldehyde oxidoreductase	0.432						
			5u47A2	Penicillin binding protein 2x	0.429						
			3figB	2-isopropylmalate synthase	0.426						
46147	Edited	-3.98	6gyhA	Family A G protein-coupled receptor-like protein	0.519	oxidoreductase activity	0.07	oxidation-reduction process	0.13	membrane integrity	0.13
			8adnJ	Proteasome inhibitor 31-like	0.518	heme binding	0.07	G-protein coupled receptor protein signaling pathway	0.07	proteasome core complex	0.07
			4fbzA	deltarhodopsin	0.51	electron carrier activity	0.07	defense response to bacterium	0.07	cytoplasm	0.07
			1e0pA	bacteriorhodopsin	0.507	monooxygenase activity	0.07	protein-chromophore linkage	0.07	plasma membrane	0.07
			8jh0A	xanthorhodopsin	0.506	4 iron 4 sulfur cluster binding	0.07	cell wall macromolecule catabolic process	0.07		
						lysozyme activity	0.07	phototransduction	0.07		
						ion channel activity	0.07	proteolysis involved in cellular protein catabolic process	0.07		
						photoreceptor activity	0.07	peptidoglycan catabolic process	0.07		
						protein binding	0.07	cytolysis	0.07		
4969	Unedited	-4.2	4mftA	ChpT protein	0.491	acetyltransferase activity	0.47	N-terminal protein amino acid acetylation	0.07	intracellular role	0.37
			1nmtA	N-Myristoyl Transferase	0.49	N-acetyltransferase activity	0.37	induction of apoptosis by intracellular signals	0.07		
			7txaA	Class II Fructose-1,6-Bisphophatase	0.488			activiation of pro-apoptotic gene products	0.07		
			7ojuA	Chaetomium thermophilum Naa50 GNAT-domain	0.487			N-terminal protein myristoylation	0.07		
			3rojA	D-fructose 1,6-bisphosphatase	0.483			glycerol metabolic process	0.07		
								in utero embryonic development	0.07		
								protein lipoylation	0.07		
4969	Edited	-1.28	7wkkB	IR subunit of NPC	0.932	protein transport activity	0.5	virus-host interaction	0.39	organelle envelope	0.59
			7mvxA	Nucleoporin Nup188	0.762	protein binding	0.5	mitosis	0.39	nucleolus	0.43
			7wo9A	Nucleoporin Nup188	0.731	RNA binding	0.43	protein export from nucleus	0.39	cytosol	0.34
			5ijoJ	Nuclear pore complex protein Nup155	0.657			gene silencing	0.37	ribonucleoprotein complex	0.31
			6lk8A	MGc83295 protein	0.564			RNA metabolic process	0.34	kinetochore	0.31
								mRNA transport	0.31	cajal body	0.31
										annulate lamellae	0.31

Zinc Finger	Edited or Unedited	C-Score	PDB Accession #	Analog Name	TM-Score	Molecular Functions	GO-Score	Biological Processes	GO-Score	Cellular Component	GO-Score
210	Unedited & Edited	-1.41	3ja4a	RNA-directed RNA polymerase	0.826	RNA-directed RNA polymerase	0.58	Viral genome replication	0.58	Virion component	0.45
			1n1hA	Polymerase lambda3	0.731	Nucleotide binding	0.53	Transcription, DNA-dependent	0.40		
			5zvs2	RNA polymerase	0.717	RNA binding	0.53				
			6pnsA	RNA-dependent RNA	0.650	Ion binding	0.37				

				polymerase							
			7xr3Z	Mud crab reovirus	0.617						
26251	Unedited - Did not align	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
26251	Edited	-2.28	4d8mA1	Bacillus thuringiensis	0.658	Transmembrane transporter activity	0.58	Protein localization to nucleus	0.58	Pore complex	0.58
			4v3iA	bacterial type VI secretion system component TssL	0.658	Transition metal ion binding	0.45	Nuclear import	0.58	Endomembrane system	0.58
			7p3xA2	AP-3 complex	0.657	Protein binding	0.40	Protein import	0.58	Organelle envelope	0.58
			50qlE	90S pre-ribosome	0.655	Endopeptidase activity	0.32	Protein targeting	0.50		
			7wb4l	NR subunit of NPC	0.653	Signal sequence binding	0.31				

Table 4: Full table of each individual zinc finger (ZF) that was further investigated along with their edited and unedited bound sequences. Analogous structure name, PDB Accession #, molecular function, biological process, and cellular components were identified via I-TASSER. C-score is for the bound sequence structure prediction.