CHARACTERIZATION OF KPAP, A NOVEL POKEWEED ANTIVIRAL PROTEIN ISOFORM

FERNAND DE LA SELLE

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2022

ABSTRACT

Phytolacca americana is a perennial plant that harbors the pokeweed antiviral protein (PAP) genes which belong to the N-glycosidase family. These proteins are called ribosome inactivating proteins (RIPs) which remove an adenine base from the sarcin/ricin loop (SRL) of the 28S rRNA in eukaryotes. This depurination results in protein translation inhibition leading to a concentration-dependent cytotoxic consequence. Many RIPs in the plant kingdom are being discovered with new genomic sequencing technologies but, many of them remain uncharacterized. Recently, our lab discovered a new RIP called Novel PAP/KPAP in the sequenced pokeweed genome. KPAP transcript levels did not behave the same way as other isoforms under phytohormone stimulated biotic stress, however, they showed a slight decrease in drought. In this study, I provide preliminary groundwork on the KPAP gene model, expression in abiotic stresses, protein structure, depurination activity and potential binding arrangement on the SRL. My results show the presence of predicted upstream open reading frames (uORFs) in the long leader intron, perhaps belonging to retrotransposons. The enzymatic activity recorded from KPAP in an *in vitro* translation assay resembles the activity of PAP-I. Moreover, the exclusive expression of KPAP in leaves compared to other isoforms and its downregulation in drought may indicate a function related to photosynthesis, carbon metabolism, plant growth or leaf expansion. Structurally the predicted protein was docked *in silico* to an adenine base and showed a similar disposition of the active site compared to PAP-I. Lastly, KPAP was docked *in silico* to the *E.coli* SRL and residue interactions are discussed in the context of other RIPs. This early research shows KPAP's similarities and differences compared to the other isoforms, however, much remains to be learned from this isoform and RIPs in general.

TABLE OF CONTENT

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation and gratitude towards my supervisor, Dr. Kathi Hudak, for having me as a student over these past few years. I learned tremendously in her laboratory, and I value the critical thinking, the experimental design, the data analysis and troubleshooting; and the countless hours we spent over this project. From her, I have also learnt a great deal about human values that were key to grow as a person. Her encouragements, positive attitude and patience has given me the courage to persevere through difficult times and for that I am truly grateful.

In addition to my supervisor, I would like to tank my advisor, Dr. White, for providing feedback and advice on my project which have been really useful in goal development and troubleshooting. Moreover, Dr. Donaldson has been of great help with the structural aspect of my project by providing avenues of possibilities for structural investigations. Structural biology is a field that I would like to pursue in the future thanks to Dr. Donaldson.

Thank you to my previous and present lab members: Kira Neller, Alex Klenov, Camille Diaz, Jennifer Chivers and Tanya Prashar. They have been mentors and friends who have helped me grow as a scientist and as a person. Additionally, my spice tolerance has increased with your help.

Thank you to my friends and colleagues Mohamed Salem (Peng lab), Heyam Hayder (Peng lab), David Miller (Benchimol lab), Nick Bragnolo (Audette lab), Marjan Moallem (Rosonina lab), Esra Eshaghi (Peng lab), Shahir Morcos (Campos lab), Neushaw Moradi (Hood lab) for troubleshooting and discussing various parts of my project. A special thanks to Naveed Samari-Kermani (Audette lab) who has driven through many feet of snow to help me with protein purification. Truly unforgettable times.

I would like to also extend my thanks to my family for their strong support which has been the backbone of this degree. Thank you for the encouragements and mostly thank you for your will to understand the science which allowed me to think about ideas differently.

STATEMENT OF CONTRIBUTION

The presented thesis was written by Fernand de La Selle (FDLS). Dr. Katalin Hudak and Dr. Andrew White assisted with editorial tasks. All experiments and analyses were performed by FDLS. Exceptions are listed here. The second iteration of the sequencing of the pokeweed genome was assembled by Kyra Dougherty (KD), and therefore the KPAP gene model based on the genome sequencing was facilitated by KD. The MashUp reverse transcriptase used in reverse transcription was made and provided by Alex Klenov (AK). Figures were generated by FDLS unless otherwise stated.

LIST OF FIGURES

LIST OF TABLES

LIST OF COMMON ABBREVIATIONS

12-OPDA	12-oxo-phytodienoic acid
aa	Amino acid
ABA	Abscisic acid
bp	Base pair
BMV	Brome mosaic virus
BDX	Cell wall protein BIIXI
CBF	C-repeat binding factor
cDNA	Complementary DNA
CDS	Coding sequence
CRE	Cis regulatory sequences
DEAE	Diethylaminoethyl
dH_2O	Distilled water
DTT	Dithiothreitol
dNTPs	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EF-1G	Elongation factor 1 gama
GAC	GTPase associated center
gDNA	Genomic DNA
GTPase	Guanosine triphosphate hydrolase
HOAc	Acetic acid
JA	Jasmonic acid
JAZ1	JA-containing zim domain 1
Kb	Kilobase
KOAc	Potassium acetate
LB	Luria broth
MOPS	3-(N-morpholino) propanesulfonic acid
nt	Nucleotide
ORF	Open reading frame
OsbHLH148	Oryza sativa basic helix-loop-helix 148
OsDREB1	Dehydration-responsive-element-binding protein 1
OSRIP	Oryza sativa ribosome inactivating protein
PAP	Pokeweed antiviral protein
PAPx	Active site mutant of PAP
PCR	Polymerase chain reaction
PCI	Phenol, chloroform, isoamy-alcohol
PDB	Protein databank
PEG	Polyethylene glycol
PTC	Peptidyl transferase center
PVP	Polyvinylpyrrolidone
RIP	Ribosome inactivating protein
RT-qPCR	Reverse transcription quantitative PCR
RT PCR	Reverse transcription PCR
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SRL	Sarcin/ricin loop
TF	Transcription factor
trGTPase	Translational GTPase
TSS	Transcription start site
uORF	Upstream open reading frame
UTR	Untranslated region

1 INTRODUCTION

1.1 Pokeweed Antiviral Protein (PAP) Overview

1.1.1 *Phytolacca americana* and PAP

Phytolacca americana also called pokeweed is a perennial plant that belongs to the order Caryophyllales and Phytolaccaceae family. It has mainly been studied in the context of the pokeweed antiviral protein (PAP), an N-glycosidase belonging to the type I ribosome inactivating protein (RIP) class of ribotoxins. These proteins are involved in a process called depurination which catalytically removes a purine base from the sarcin/ricin loop (SRL) of the large ribosomal RNA (rRNA; Marchant & Hartley 1995). This damage inhibits the elongation step of translation and therefore reduces global protein translation (Irvin 1983). In the same capacity PAP was shown to be antiviral for its ability to limit viral protein synthesis (Aron & Irvin 1980). This activity was also depicted in stages of viral life cycles in plant and animal RNA viruses (He et al. 2008; Mansouri et al. 2009; Karran & Hudak 2008; Krivdova & Hudak 2015). The antiviral activity of PAP was reported to inhibit the proliferation of HIV-1 (Rajamohan et al. 1999a, Rajamohan et al. 1999b, Uckun et al. 2003), poliovirus (Ussery et al. 1977), herpes simplex virus (Aron & Irvin 1980), influenza virus (Tomlinson et al. 1974), cucumber mosaic virus (Tomlinson et al. 1974), tobacco mosaic virus (Chen et al. 1993), brome mosaic virus (Picard et al. 2005), turnip mosaic virus (Domashevskiy et al. 2012), lymphotic choriomenigitis virus (LCMV; Uckun et al. 2005), Japanese encephalitis virus (Ishag et al. 2013); and potato virus X and Y (Lodge et al. 1993). Beyond its antiviral effects pokeweed is also known to be a hyperaccumulator of heavy metals like cadmium and zinc showing a potential role in phytoremediation (Fu et al. 2011; McBride et al. 2019; Zhao et al. 2011). While pokeweed is not a model organism, the genome has been sequenced twice (Neller et al. 2019; Dougherty & Hudak 2022 *unpublished*) and transcriptome experiments under various stress conditions have also been done (Neller et al. 2016; Neller et al. 2019; Chen et al. 2017).

1.1.2 Family of PAP genes

Typically plant species synthesizing RIPs have many isoforms like the saporins (Tartarini et al. 2010) or the rice OSRIPs (Jiang et al. 2008). In Phytolacca americana, various PAP genes

have been reported. PAP-I was first isolated from spring leaves (Irvin 1975) and has extensively been studied. PAP-II has been isolated from early summer leaves (Barbieri et al. 1982) and the PAP-α cDNA clone was obtained from leaves, roots, and seeds (Kataoka et al. 1992). PAP-S1 and S2 have been identified in seeds (Honjo et al. 2002) and novel PAP has been identified in the pokeweed transcriptome (Neller et al. 2016). In this paper KPAP and novel PAP will be used interchangeably ('K' for Kira, the former PhD student, who discovered KPAP). These PAP isoforms were confirmed in the recent genome analysis (Neller et al. 2019), however, ambiguous isoforms such as PAP-III (Kurinov & Uckun 2003) was identified to be PAP-II; and PAP-C (Barbieri et al. 1989) and PAP-R (Bolognesi et al. 1990) were identified to be PAP-I. The small biochemical differences observed in these ambiguous isoforms were attributed to post-translational modifications and allelic diversity for PAP-III and PAP-C/ PAP-R respectively (Neller et al. 2019).

1.2 Structure and enzymatic activity of RIPs

1.2.1 Structure differences among RIPs

The PAP isoforms are considered type I RIPs which highlights the single catalytic subunit with a molecular weight around 29kDa (Irvin 1975). There is also type II which consists of a catalytic and a lectin domain (which allows to penetrate cell membranes) and a type III which consists of a single catalytic subunit with an extended N-terminus pro-peptide that gets cleaved upon activation (**Figure 1**; Domashevskiy & Goss 2015). Type I RIPs are amaranthin, bouganin, saporins (R1-4; S5,6; L1-3), PAPs, lychnin, momorcharin, gelonin, trichosanthin, ME1, Be27 to name a few, while type II are ricin, abrin, modeccin, ebulin (as reviewed by Wang et al. 2016). Lastly type III RIPs are more uncommon than their counterparts and exist in corn and barely (Domashevskiy & Goss 2015). More recently, however, type III RIPs have been described as containing a RIP domain plus an unknown functional C-terminal domain (Zhu et al. 2018). It seems likely that the classical way of naming RIPs will need to be updated since new RIPs are being discovered with unusual characteristics and a diversity of mechanisms of action (Wong et al. 2020).

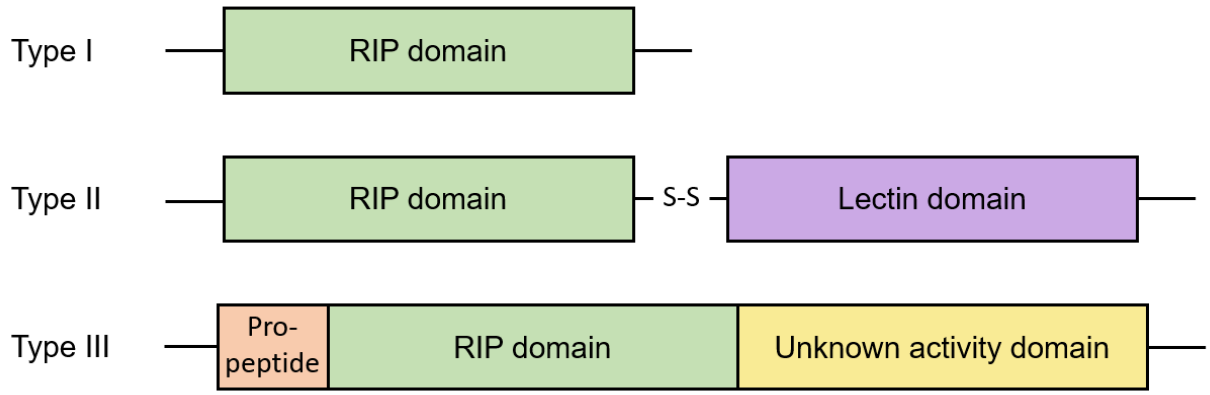


Figure 1: Protein domains in type I, II and III RIPs. Type I consists of a single catalytic RIP domain. Type II consists of that same RIP domain as type I and added is a lectin domain held by a disulfide bond. The lectin domain allows type II RIPs to enter cells through the plasma membrane. Type III contains a pro-peptide that needs to be cleaved before activating the RIP domain. It is generally accompanied by a C-terminal domain with unknown function.

1.2.2 Active site and enzymatic activity

The RIP catalytic domain contains 62 well conserved amino acids including Y72, Y123, E176, and R179 in PAP-I (Di Maro et al. 2014; Rajamohan et al. 2001). These residues form a pocket where an adenine or guanine has been co-crystalized with PAP-I (Protein Databank (PDB): 1qci and 1d6a respectively; Kurinov et al. 1999a; Kurinov et al. 1999b). The role of R179 is to protonate the N3 of the substrate adenine to destabilize the N9-C1' bond linking the base to the ribose sugar. The intermediate state is the formation of an oxocarbenium transition state which is resolved by E176 through the nucleophilic reaction of a water molecule (Chen et al. 2000). The amino acid role in the depurination mechanism is depicted in bouganin with minor differences compared to PAP-I (see **Figure 2**; Fermani et al. 2009). Mutating E176 or R179 significantly reduces the enzyme activity 1000-fold (Rajamohan et al. 2000). The purpose of Y72 and Y123 is to properly orient the nitrogenous base through π-stacking to the residues (Rajamohan et al. 2001). Mutational analysis identified Y82 and Y123 as essential for cytotoxicity in yeast (Hudak et al. 2004). Additional residues promote the purine-specific binding. The carbonyl of the backbone of V73 and the hydroxyl of S121 interact with the adenine promoting specificity with this substrate. Other ligands like guanine have been observed to be depurinated by PAP-I, however 100-fold slower (Kurinov et al. 1999b).

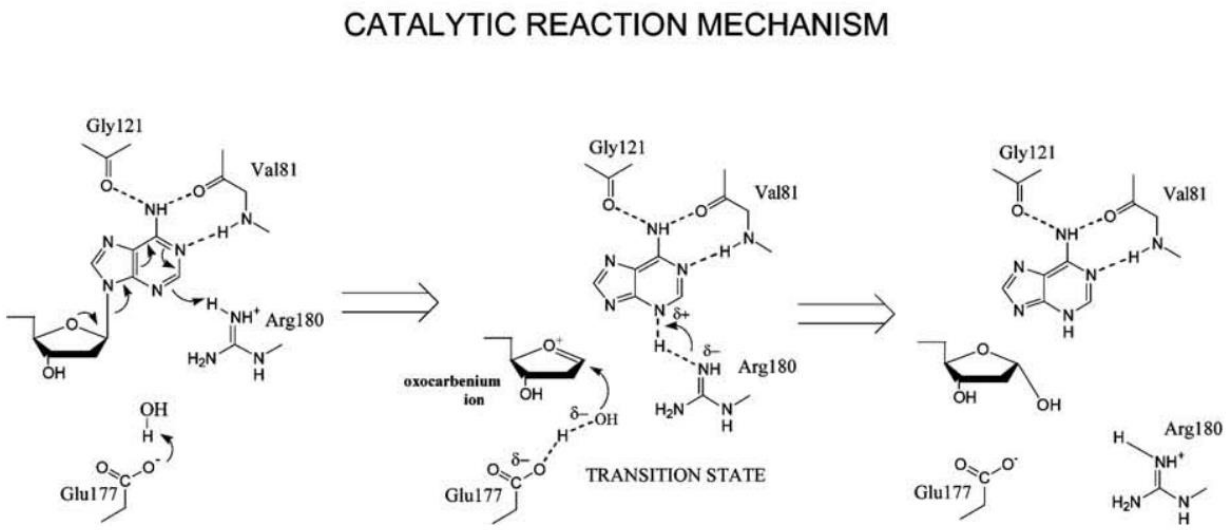


Figure 2: The catalytic mechanism of depurination from bouganin obtained from Fermani et al. (2009). The reaction starts by protonating the N3 of the adenine base by Arg180 which triggers an electron cascade to stabilize the charge thereby unlinking the base to the ribose sugar and transferring the charge to the oxygen on the ribose. This intermediate transition state is called a ribooxocarbenium ion, which is subsequently stabilized by the negative portion of a neighboring hydroxyl obtained from a water molecule interacting with E177. The result is an abasic site where the rRNA backbone is now separated from the base.

1.3 The sarcin/ricin loop (SRL)

1.3.1 What is the SRL

Classically, RIPs act on the SRL of 28S rRNA as the primary RNA substrate (Rajamohan et al. 2001; Marchant & Hartley 1995) which inhibits protein translation (Irvin 1975; Poyet et al. 1997a). It has also been reported that RIPs can act on various RNA species such as rRNA, viral RNA, mRNA; and DNA (Aron & Irvin 1980; Hudak et al. 2002; Barbieri et al. 1997). The structure of the SRL resembles a distorted hairpin however, the components are remarkably conserved (**Figure 3**; Szewczak & Moore 1995). It consists of the stem, a flexible region, the G-bulged cross strand stack, and the GAGA tetraloop (Havrila et al. 2013). Compared to the stem which follows the classical Watson-Crick base pairs, the remainder of the structure is mainly stabilized through π-stacking interactions. This atypical loop structure organization is recognized by the local translation machinery and by RIPs (Rajamohan et al. 2001; Marchant & Hartley 1995). The GAGA tetraloop is held together by π-stacking (**Figure 3**). G2661 stabilizes A2660 (base numbering in *E.coli*; G3028 and A3027 in yeast) in this way at the top of the loop which means A2660 is fully exposed without forming any hydrogen bonds making it accessible to

surrounding factors (Heus et al. 1991; Moazed et al. 1988). A2660 is the target adenine for most RIPs and because it is not hydrogen bonded it has a greater flexibility of rotation (Marchant & Hartley 1995). This structure is important for recognition by RIPs, however, mutation of the tetraloop from GAGA to GAAA does not prevent PAP-I from depurinating the RNA. In addition, the bulged G is not necessary for the recognition of the GAGA tetraloop by PAP-I, as it was able to depurinate the GAGA tetraloop of the 16S rRNA of *E.coli* consisting entirely of Watson-Crick base-pairs (Marchant & Hartley 1995).

1.3.2 The importance of the SRL

The ribosome is the central element of protein translation, the place for nascent peptide synthesis. While the small subunit of the ribosome is involved in genetic information decoding from mRNA, the large subunit is the site for the catalytic peptidyl transferase center (PTC) where aminoacyl-tRNAs deliver amino acids synthesized into a polypeptide chain (Liljas & Ehrenberg 2012). Moreover, the large ribosomal subunit harbors the GTPase associated center (GAC) essential for the continuous motion of the ribosome during translation. The important actors at the GAC are proteins that hydrolyze GTP called translational GTPases (trGTPase) and they provide the unidirectional motion of the translational machinery at the expense of GTP hydrolysis (Maracci & Rodnina 2016). The GAC is responsible for recruiting trGTPases and initiating GTP hydrolysis in the center of the GAC consisting of the SRL and the ribosome stalk composed of ribosomal proteins exclusively (Wahl & Moller 2002).

1.3.3 RIPs accessing the SRL

RIPs access the SRL in different ways. Reports suggest that PAP-I accesses the SRL on the ribosome by binding to ribosomal protein L3 (Hudak 1999; Rajamohan et al. 2001). However, in other RIPs the landing site is different. Ricin A-chain (RTA) was shown to bind the yeast ribosome P-stalk pentamer through protein P2 (Li et al. 2013; Shi et al. 2016). Additionally, trichosanthin (Chan et al. 2007), the active form of type III maize RIP (MOD; Yang et al. 2010), and the type II shiga toxin bacterial RIP (McCluskey et al. 2008, 2012) have shown to interact with protein P2 using the same binding site as RTA. A model has been proposed suggesting the main docking part is the ribosomal P-proteins, however, the L3 protein seems to be the landing platform for PAP-I (as reviewed in Grela et al. 2019). Zeng and colleagues (2003) have engaged in a discussion involving the structural model of the 50S

ribosome and protein L3 and deemed it very unlikely for PAP-I to interact with both the SRL and L3 at the same time due to more than 35Å separation. It was hypothesized that a multistep process should be considered involving L3 as the landing platform followed by SRL depurination.

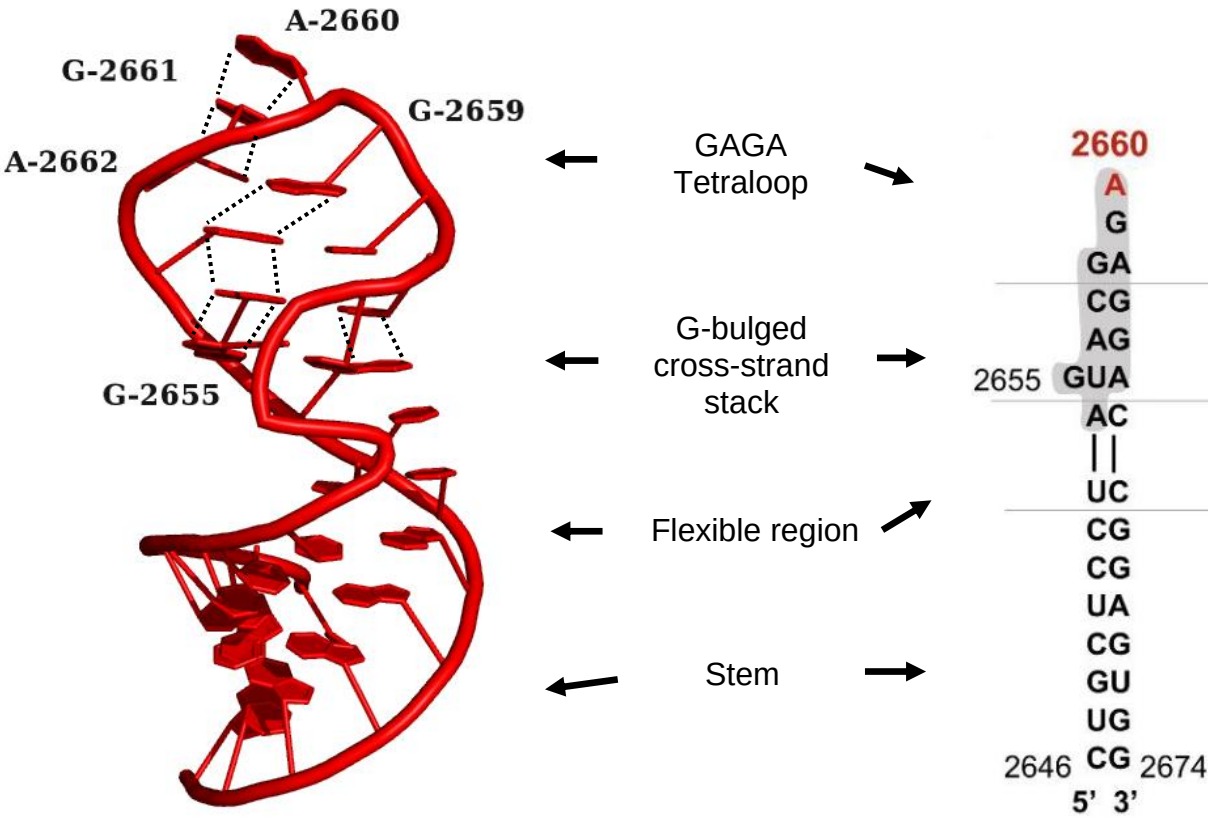


Figure 3: The *E.coli* α-sarcin/ricin loop (SRL) nucleic acid structure and sequence depicting the four components: the GAGA tetraloop, the G-bulged cross-strand stack (G2655), the flexible region and the stem. The structure image generated using PyMol but was inspired from Grela et al. (2019) and the SRL sequence diagram was taken from Grela et al. 2019. The dash lines represent the π-stacking interaction between bases instead of the normal Watston-Crick base pairing

1.4 Plant regulation of abiotic stress response

1.4.1 Abscisic acid (ABA)

ABA is a small molecule and an important phytohormone in many physiological processes. In seeds, it is important for protein synthesis, dessication, dormancy and inhibition of phase transitioning between germination, vegetative or reproductive stages (Leung & Giraudat 1998; Rock 2000). ABA is well-known for its abiotic stress response in salt, cold, and drought

where ABA levels can increase up to 10-fold and quickly decrease within a few hours (Kushiro et al. 2004). Plants synthesizing ABA tends to induce stomatal closure and halt growth through reduced canopy expansion (Ye et al. 2012). This stress-related response allows the plant to survive periods of drought. Additionally, the stomatal closure plays a role against invading pathogens that can use these pores as entry sites to the leaf. By closing them, ABA indirectly prevents biotic stress in the plant (Melotto et al. 2006; Ton et al. 2009).

1.4.2 Drought

Drought is one of the most prevalent conditions affecting the growth and development of crops and plants. For this reason, plants have evolved to adapt to drought stress through morphological, physiological and biochemical modulations (Bohnert et al. 1995). The plant naturally transpires through the stomata, an opening that allows the exchange of CO_2 and H_2O (Schroeder et al. 2001). The strategy that plants use is based on the regulation of the opening and closing of the stomata to improve drought resistance (Umezawa et al. 2010; Yadav et al. 2019). While cytokinins stimulate stomatal opening, drought, cold, high CO_2, darkness and phytohormone like abscisic acid (ABA) and ethylene trigger stomatal closure (Acharya & Assmann 2009). Moreover, elevated levels of jasmonic acid precursor 12-oxo-phytodienoic acid (12-OPDA) are known to decrease stomatal aperture thereby improving drought tolerance (Daszkowska-Golec & Szarejko. 2013). In rice, a basic helix-loop-helix (bHLH) protein (OsbHLH148) derepresses the dehydration-responsive-element-binding protein 1 (OsDREB1) transcription factor which allows the expression of drought resistance genes (Seo et al. 2011; as reviewed by Wang et al. 2021).

1.4.3 Cold

Cold stress adversely affects plant growth and can be organized into two categories: low temperature stress and freezing. Under low temperatures within the range of 0-15°C, plant cells undergo noticeable dysfunctions, whereas freezing generates ice crystals which cause mechanical damage to cells and their organelles (Ding et al. 2019). Plants have evolved mechanisms to alleviate the damage. The induction of cold tolerance genes can be triggered by abscisic acid (ABA). Upon cold stress ABA accumulates and triggers a signaling cascade. ABA is sensed by receptors which relay the signal to SNF1-related protein kinase 2 (SNRK2; Umezawa et al. 2010). The phosphorylation activates the ABRE-binding bZIP transcription

factor (ABF) which then binds ABA response element (ABRE; **Figure 4A**; Kim et al. 2013) used to express the NAC-type transcription factor responsible to express genes involved in cold tolerance. (Nakashima et al. 2012).

Cold stress has also been investigated in the JA signaling pathway. The C-repeat binding factor/dehydration-responsive-element-binding protein 1 (CBF/DREB1) has been investigated for its role in cold damage mitigation in Arabidopsis and rice using the ICE-CBF transcriptional regulation pathway (Kim et al. 2015). In Arabidopsis, ICE1/2 is a bHLH transcription factor that regulates the expression of CBF using the CANNTG cis regulatory element of CBF promoters (**Figure 4B**; Kim et al. 2015). Under regular conditions, genes involved in the ICE-CBF pathway will be repressed by complexed ICE1/2 and JAZ1/4 transcription factors. In rice, cold stress mainly regulates signaling genes like COI1 (Du et al. 2013). The JAZ-COI1 co-receptor sequesters JAZ in the presence of JA-Ile allowing the activation of repressed genes and also targeting JAZ for degradation through polyubiquitination (Yang et al. 2019). This results in the production of cold-related genes to promote survival and development during cold stress.

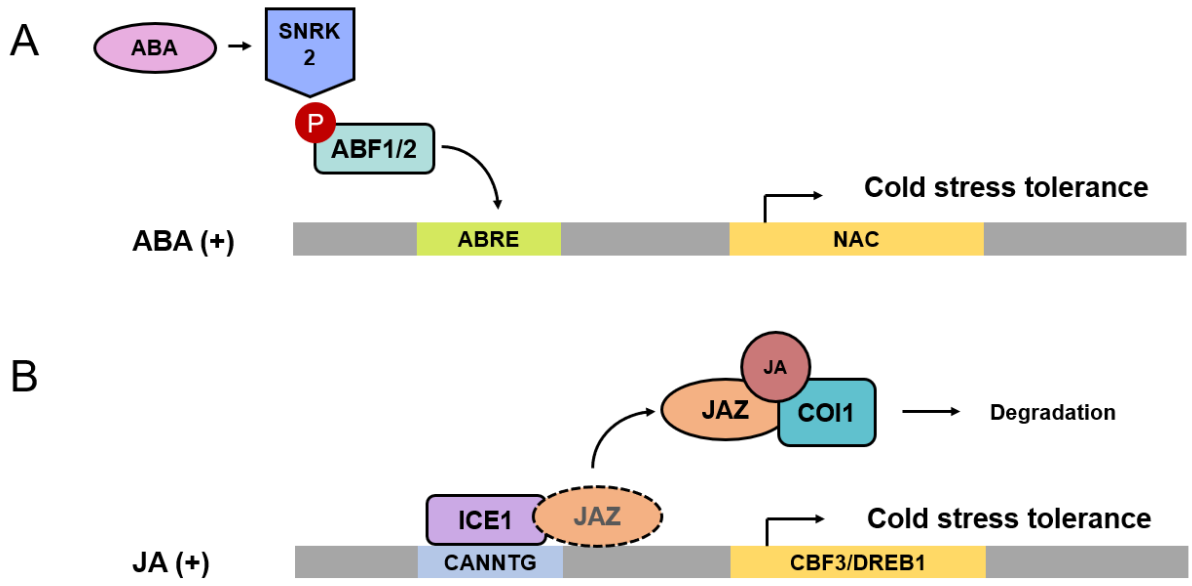


Figure 4: Cold response pathways in plants in ABA and JA. (A) Under ABA response to cold SNRK2 is activated which will phosphorylate ABF1/2 and promote binding to ABRE motifs. In turn, NAC-type transcription factors (TFs) will be expressed and will initiate the expression of genes involved in cold stress tolerance. (B) In another pathway, ICE1 TF will bind the CANNTG recognition element along with JAZ. Upon JA response to cold, JAZ will be sequestered by COI1 for degradation and derepress the transcription of the CBF3/DREB1 involved in upregulating genes that will provide tolerance to cold stress.

1.4.4 RIPs in response to stress

Plants respond to stress primarily through phytohormones that relay signals throughout the plant to initiate a survival-dependent response. Phytohormones include jasmonic acid (JA), salicylic acid (SA), ethylene (ET) and abscisic acid (ABA). Abiotic stress such as salinity, drought, cold, or mechanical and biotic stress such as viruses, insects, or fungi are common stresses that plants have to face on a daily basis. RIPs are common in many plants and have diverse roles when it comes to stress. For example, OSRIP1 (from rice) is upregulated in drought, salinity and ABA (Jiang et al. 2008), while PAP-I and PAP-S1 (pokeweed) are upregulated by JA and SA (Neller et al. 2019). Curcin-L (*Jatropha curcas*) is upregulated by ABA, SA, drought, temperature and UV (Qin et al. 2009) and saporin L3 and S6 (*Saponaria officinalis*) are upregulated in wounding and ABA (Tartarini et al. 2010; as reviewed in Mishra et al. 2022). The diversity of responses that RIPs exhibit relates to the diversity of stresses that the plant encounters and RIPs are part of the line of defense necessary for survival.

1.5 Research objectives

As recently discovered by Neller et al. (2019), Novel PAP or KPAP, is one of the least known isoforms and the transcript shows a different behavior compared to other PAP isoforms in stress treatments. Moreover, the presence of an unknown long sequence in the 5'UTR and the discontinuation between the promoter and the coding sequence (CDS) encountered by previous students made it an interesting gene to study. Traditionally RIPs have been studied in the defense-related context of biotic stress including herbivores, pathogens, fungus, and viruses which are regulated mainly through the JA signaling pathway. Since KPAP was insensitive to JA, SA, and wounding stresses (Neller et al. 2019), it furthered the perspective of an exception among the other PAP isoforms. Therefore, the goal of this study was to characterize the gene model of KPAP, to investigate the mRNA expression under various abiotic stress, and to investigate structurally and functionally the activity of the protein. With the new sequencing of the pokeweed genome, the PCR walking experiments helped validate the gene model. Also, the investigation of abiotic stress under PEG, cold and ABA would complement the expression profile already produced for biotic stresses (Neller et al. 2019) which would be emphasized by the identification of cis-regulatory elements (CREs) related to abiotic stress. Lastly the investigation of the *in-silico* interaction of the SRL with KPAP has led to identify important

residues involved in the interaction. This very early work on KPAP improved our understanding of this new isoform, but also highlighted the work remaining to be done.

2 MATERIALS AND METHODS

2.1 Plant cultivation

2.1.1 Seeds

Phytolacca americana seeds were originally obtained from Rutgers University, New Jersey, USA and plants were propagated at York University since 2002.

2.1.2 Pre-treatment

To promote germination, pokeweed seeds were vortexed with sulfuric acid for 2-3min and washed in a sieve with running tap water for 2min. Treated seeds were collected and submerged in distilled water for a few days until light cracking was observed on the surface of >95% the seeds.

2.1.3 Sowing

The soil mixture was composed 1 bag Pro-Mix (Berger BM 1, 107L), 1 bag garden soil (Alltreat Farms 3-way Mix, 15kg), ½ bag cattle manure (Alltreat Farms, 18kg), 2/3 bag play sand (20kg) and 7L of water mixed for 25min. Cracked seeds were planted in pairs to ensure germination rate of >95% in 2.5-inch square pots about 1 inch below the soil surface. Sowed pots were placed in trays with a clear plastic dome.

2.1.4 Growth conditions

All pokeweed plants were kept at 20°C for 14 hours of light and 10 hours of darkness at 24°C under 75% fluorescent and 25% incandescent lighting and 65% chamber fan speed. The growth chamber was Econair controller 6000.

2.2 Abiotic stress treatments and manipulation

2.2.1 Cold time course

A total of 9 four-leaf stage pokeweed plants per time point were selected and arranged to account for the small growth discrepancies. The treatment consisted of three time points (2, 8, 24h) subjected to 5°C and a room temperature control (Zhao et al. 2018). The chamber had the same light/dark cycle as stated previously under growth conditions.

2.2.2 Dehydration time course

Similar to the cold time course, a total of 9 four-leaf stage pokeweed plants per time point were selected and arranged to account for the small growth discrepancies. The treatment consisted of three time points (8, 24, 96h) treated with 20mL of 20% PEG8000 in dH_2O every 24h while the control was treated with 20mL dH_2O every 24h (Xu et al. 2020). The plants were covered with a plastic dome and grown in the growth chambers.

2.2.3 ABA time course

Similar to the cold time course, a total of 9 four-leaf stage pokeweed (3 plants per biological replicate; 3 biological replicate) plants per time point were selected and arranged to account for the small growth discrepancies. The treatment consisted of 4 time points (2, 8, 24, 72h) that were sprayed with 0.5mM abscisic acid (ABA) on the front and back of the two top most leaves (Agarwal et al. 2005). Plants were placed under a plastic dome to prevent cross-talk. Cross-talk refers to the signaling of one plant to others in conditions of stress through the use of small molecules and phytohormones. In this case, plants used in the control were isolated from the treated plants by using a plastic dome. Only the treated leaves were harvested.

2.2.4 Plant tissue harvest

Stress treated plant tissues were harvested by cutting the midvein of the youngest leaves (2-3g) and flash frozen in liquid nitrogen. Three plant leaves were combined in one single biological replicate to accommodate for growth disparities. Three biological replicates were collected per time point for each time course

2.3 DNA isolation (small scale)

Overnight *E.coli* (DH5-α) culture (5mL) in Luria broth (LB) expressing the plasmid of interest in a selective antibiotic were used for plasmid extraction. Culture incubation was performed at 37°C with shaking at 250rpm. The bacterial cells were pelleted the following day at 16,000 x g for 2min. The pellet was resuspended in 200uL solution 1 (100mM Tris pH7.5, 10mM EDTA, 100ug RNase A) followed by the addition of 200uL of solution 2 (200mM NaOH, 1% SDS) for lysis. The solution was mixed by inverting and allowed to incubate at room temperature for 2min. Immediately after 200uL solution 3 (3M KOAc, 5M HOAc) was added to neutralize the reaction and the solution incubated on ice for 10min. Large white precipitate

formed and were centrifuged at 16,000 x g for 5min. The supernatant was added to equal volume isopropanol for precipitation and incubated at room temperature for 30min. Nucleic acid precipitates were collected by centrifugation at 16,000 x g for 20min. The supernatant was discarded, and the pellet air dried for 5min. The plasmid DNA was resuspended in 50uL dH_2O and quantified using a spectrophotometer and its quality assessed by agarose gel electrophoresis.

2.4 DNA isolation (large scale)

E.coli culture (250mL) in LB media were inoculated with 2mL of saturated 5mL *E.coli* culture in LB and appropriate antibiotic grown overnight. The 250mL culture was grown until OD 0.8 at 37°C and shaking 250rpm. Culture was centrifuged at 3,200 x g for 10min at 4°C. The cell pellet was resuspended in 10mL of solution 1 (same as miniprep) followed by 10mL of solution 2 (same as miniprep) and incubated at room temperature for 5min after several inversions. Immediately following incubation 10mL of chilled solution 3 (same as miniprep) was added and allowed to incubate on ice for 20min. Large white precipitates formed and the supernatant was collected through gravity funnel filtering (4 layers of cheese cloth stacked on 2 layers of filter paper). The resulting filtrate was syringe filtered and applied to a DEAE gravity column (Qiagen Cat#: 12162) that was equilibrated with 10mL buffer QBT (50mM MOPS pH7, 750mM NaCl, 15% isopropanol, 0.15% Triton X-100). The column was washed with 60mL buffer QC (50mM MOPS pH7, 1M NaCl, 15% isopropanol) and the plasmid was eluted with 15mL buffer QF (50mM Tris pH8.5, 1.25M NaCl, 15% isopropanol). The DNA was precipitated using 10.5mL (0.7 volumes) of isopropanol and split into microfuge tubes and centrifuged 16,000 x g for 30min at 4°C. The pellet was washed with 1mL 70% EtOH and allowed to air dry for 10min followed by resuspension in 50uL. Aliquots were used for quantification using a spectrophotometer and quality by agarose gel electrophoresis.

2.5 Polymerase chain reaction (PCR)

PCR was used to amplify specifically selected DNA sequences. Depending on the application, the template was either gDNA (250ng), plasmid DNA (10ng) or cDNA (0.5µL). Each reaction used the same master mix: 5µL of 5X Taq buffer, 0.5µL of 10µM forward primer, 0.5µL of 10µM reverse primer, 0.5µL of 10mM dNTPs, 0.5U of Taq enzyme (NEB Cat#: M0495), and template to a final volume of 25µL. The PCR program was started with an initial denaturation step at 95°C for 90sec followed by 38 cycles alternating between annealing (60-

70°C; 30sec) and extension (72°C; 30sec per kilobase). To assess that the correct size fragment was amplified, the PCR product (6μL) was visualized through agarose gel electrophoresis (1-2%) at 100V for 30min.

2.6 Low melt gel extraction

PCR products were isolated through low melt gel extraction upstream of cloning using Gibson assembly. PCR products were separated through agarose gel electrophoresis at 80V for 40min on a low melt agarose gel stained with Midori green Xtra (Nippon Genetics Cat#:MG10) to visualize DNA under blue light. After electrophoresis, a stand held a LED light with a filter to illuminate the agarose gel with blue light. Desired DNA bands observed were excised using a razor blade. The DNA was extracted from the gel slice using the DNA gel extraction kit (NEB #T1020) following the manufacturer's protocol. Quantification was assessed using a spectrophotometer.

2.7 Cloning using TEDA assembly

The gel purified insert and vector were added to a 0.2mL PCR tube in a 1:3 insert to vector molar ratio (insert = 40fmol; vector = 120fmol) with 7.5uL 2X TEDA mix (500mM Tris-HCl pH7.5, 50mM MgCl2 50mM DTT, 0.25g PEG 8000, 10U T5 exonuclease (NEB Cat# M0363)) to a final volume of 15uL. The reaction used the following program: at 50°C for 45min facilitating the annealing of the 3' overhangs generated by the T5 exonuclease, and 70°C for 10min for enzyme inactivation. The products (0.25-2uL) were transformed into competent DH5α cells.

2.8 Transformation

Competent Rosetta 2 cells (BL21 DE3 pLysS) were thawed on ice and 1-10ng of plasmid DNA was incubated with cells on ice for 30min. Heat shock occurred at 42°C for 35 seconds and after cells were placed on ice for 2min. Recovery happened by adding 300uL of LB media to the cells and incubating at 37°C for 1h on a shaker (250rpm). Cells were then pelleted by centrifugation (quick spin: 16,000 x g for 15sec) and the supernatant was reduced to 100uL. The cells were resuspended and streaked on an LB agar plate containing appropriate antibiotics. The plate was incubated at 37°C overnight and colonies were observed and selected for colony PCR.

2.9 Colony PCR

Colony PCR was used to screen bacterial colonies from the transformed cloning plates to identify successful assembly. Colonies were picked from the cloning plates and resuspended in a 0.2mL PCR tube with 40uL dH_2O. Between 5-20 colonies were screened per construct depending on the number colonies obtained. An aliquot of 5uL from each resuspended colonies was plated on selective LB plates and incubated at 37°C. The remaining 35uL was incubated at 95°C for 10min on a thermocycler to induce cell lysis. Samples were snap cooled on ice for 2min and centrifuged at 16,000 x g for 5min to pellet the cell debris. The supernatant (5uL) was to a PCR master mix containing the following: 5uL 5X Taq buffer, 1.25uL 10uM of the forward and reverse primers, 0.5uL 10mM dNTPs, 0.5U of Taq polymerase and dH_2O to a final volume of 25uL. The PCR program initialized with a single denaturation step of 95°C for 1min 30sec, followed by 35 cycles of denaturation (95°C for 30sec), annealing (55-65°C for 30sec) and extension (72°C for 30sec); the program was terminated after a single extension step at 72°C for 5min. The PCR product was separated (6uL) on an agarose gel (1-2%) stained with ethidium bromide at 100V for 30min and subsequently imaged under UV gel imager.

2.10 RT-qPCR/RT-PCR

2.10.1 Pokeweed RNA extraction

Total RNA was isolated from 4-leaf pokeweed plants using the acid PCI method. Frozen tissue (3g) was ground in a pre-chilled mortar with a pestle and liquid nitrogen. The ground powder was mixed with 4mL pre-extraction buffer (20mM citrate buffer pH4, 10mM EDTA, 1% polyvinylpyrrolidone(PVP), 1%SDS, 2% β-mercaptoethanol) and 4mL of acid phenol:chloroform:isoamyl alcohol (PCI)(25:24:1 pH4.5) and vortexed. The mixture was centrifuged at 3270 x g for 15min and the top layer was transferred to another 50mL containing an equal volume of acid PCI and spun again with the same parameters. The upper layer was transferred to an equal volume of chloroform:isoamyl alcohol (CI)(24:1) and centrifuged with the same parameters. The resulting upper layer was transferred and split into microfuge tubes where 1.5 volumes of isopropanol was added. After the samples incubated on ice (30min), the mixtures were centrifuged at 16,000 x g for 10min and the pellets were washed once with 70% ethanol and allowed to air dry for 5min. The pellets were resuspended in 50μL dH_2O and allowed to sit on ice for 10min for complete resuspension. Quantification was assessed using a spectrophotometer. For the treatment of co-purified genomic DNA (gDNA), a DNase treatment

consisted of 1U of DNase I (NEB Cat#: M0303) for every 100ug of RNA in DNase buffer (NEB Cat#: B0303) and incubated at 37°C for 1h. An equal volume of acid PCI was added and the vortexed mixture was centrifuged at 16,000 x g for 5min. The supernatant was precipitated with 1/10 volume of 3M NaOAc pH5.5 and 2.5 volumes of 100% EthOH and incubated at -20°C overnight. The precipitate was centrifuged at 16,000 x g for 30min at 4°C and the pellet was washed once with 70% ethanol. The pellet was allowed to air-dry for 5min and was resuspended in 50uL dH_2O. After 10min incubation on ice, the RNA sample was quantified by a spectrophotometer and its quality was assessed by agarose gel electrophoresis.

2.10.2 cDNA synthesis

Pokeweed total RNA (1.25µg) was combined with 2.5µL 10µM gene specific reverse primer to a final volume of 10uL. Samples were denatured at 65°C for 10min and were subsequently snap-cooled for 2 minutes on ice. The reverse transcriptase master mix (10µL) was added to the denatured product (10uL; total 20uL) and contained the following components: 4µL 5x MashUp buffer (Alekseenko et al. 2021), 1uL 0.1MDTT, 1µL 10mM dNTPs, 0.5µL RNase Inhibitor Murine (NEB Cat#: M0314), 1µL MashUp enzyme (Alekseenko et al. 2021), 2.5µL dH2O. Samples were incubated at 50°C for 60min and were heat inactivated at 70°C for 30 seconds. cDNA was stored at -20C until further use in RT-PCR or RT-qPCR.

2.10.3 Quantitative PCR (qPCR)

cDNA samples (1.5uL) were combined with 2uL forward and 2uL reverse gene specific primers (10uM), 33uL of 2X Sybr Green qPCR Master Mix (Bimake Cat# B21202) to a final volume of 66uL. Each sample was thoroughly mixed and aliquoted in 20uL triplicates in 0.1mL 4-strip PCR tubes (Progene Cat# 87-C100-4-RG). qPCR was performed using a QIAGEN Rotor-Gene Q thermocycler with the following program: hold 95°C for 10min (initial denaturation and hot start polymerase activation); 40 cycles of 95°C for 15 seconds denaturation, 68°C for 45 seconds annealing/extension. Data analysis was performed in Excel using the ΔΔCt method (Livak & Schmittgen, 2001). The selected internal controls were as follows: dehydration assay BDX and EF-1gamma; Cold assay BDX and EF-1alpha; ABA BDX Tubulin; Developmental stages BDX and EF-1gamma. Melting curve analysis was performed to confirm only one product was formed. gDNA contamination was also checked for each primer pair in each RNA sample to assess the contribution of gDNA to the signal.

2.11 Structural bioinformatics

2.11.1 Protein prediction using SWISS-MODEL

Protein prediction was divided into two major parts: the folding prediction and the validation. The protein prediction was performed using the SWISS-MODEL web server (https://swissmodel.expasy.org/interactive). The input required is the peptide sequence. SWISS-MODEL relies on the ProMod3 homology pipeline which extracts structure information from the template structure, and it relies on an in-house backbone-dependent rotamer library for non-conserved side chains. Multiple models were generated and were validated through PROCHECK and Verify3D (for both https://saves.mbi.ucla.edu/ Version 6.0)

2.11.2 N-terminal prediction using SignalP5.0

SignalP5.0 is a predictor of a signal peptide and cleavage site in prokaryotes and eukaryotes based on deep convolutional and recurrent neural network architecture (https://services.healthtech.dtu.dk/service.php?SignalP-5.0). To predict the signal peptide and cleavage site, the web tool requires the full length peptide as input and the kingdom of life. Optional output display selection between long and short output can be chosen; the long output was always selected. The result page displays the first 70 amino acids from the N-terminus and three colored lines can be observed representing the probability of the signal peptide (red), the cleavage site (dotted green), and the sequence associated with no kind of signal peptide (light orange). As a positive control PAP-I was used and the graph was compared to KPAP.

2.11.3 Protein ligand docking: Autodock4

Autodock4 (https://autodocksuite.scripps.edu/autodock4/) is a software used in predicting docking conformations between a receptor and a ligand based on empirical free binding energy and a Lamarckian genetic algorithm. The software works in two phases. The first involves the generation of interaction energies of the receptor from a user input grid box which sets the scene for docking. The second phase is the docking based on the grid information. **Preparing the receptor.** The KPAP SWISS MODEL pdb file was added to the workspace. All hydrogen atoms were removed, and only polar hydrogens were added. Kollman charge (based on electrostatic potential) of the protein was subsequently calculated and verified for equal distribution around the protein and the file was saved as a .pdbqt file (atom file plus partial charges). **Preparing the ligand.** The adenine molecule was extracted from the PAP-I (1qci) crystal structure using

PyMol, polar hydrogens were added, and overall charge was calculated. Next, a torsional tree was calculated which allows the software to understand the flexibility of the molecule (limit is 32 rotatable bonds) and similarly to the receptor molecule, the ligand was saved as a pdbqt file. **Generating the grid.** Autodock4 uses the ligand to generate a map. It makes a map from separate ligand atoms at each point in a 3-dimensional grid around the receptor. In addition, it generates a desolvation energy and electrostatic map. The adenine was selected as the ligand to derive the map and a grid box was manually placed around the presumed active site of KPAP. An output file was generated for the grid dimension and Autogrid4 was run. **Generating the docking.** The search parameters were generated from the genetic algorithm (GA) which is based on a population model selecting the better conformations from a random pool that progressively chose conformations over a set number of generations. Fifty runs of the GA were done with a population size of 300 and 27,000 generations. The conformation with the lowest free binding energy was selected and uploaded to PyMol for visualization.

2.11.4 Haddock2.4

The Haddock2.4 (**H**igh **A**mbiguity **D**riven protein-protein **DOCK**ing) web server (https://wenmr.science.uu.nl/haddock2.4/) is a flexible docking approach for modeling biomolecular complexes. Haddock uses ambiguous interaction restraints (AIRs) for predicting protein interfaces and guiding the docking process. Protein and nucleic acid files were submitted in the .pdb file format to the server. The KPAP SWISS MODEL file was used and E177 and R180 were used as unambiguous interaction residues. The *E.coli* SRL (Protein Databank ID: 480d) file was used as the nucleic acid ligand and A2660 was used as the unambiguous interaction residue intended to interact with E177 and R180. Molecule parameters such as allowing Haddock to use buried residues and adjusting to a value of 7 the minimum percentage of relative solvent accessibility (RSA) to consider a residue accessible (default: 15) for the protein only were used in addition to default settings. Docking parameters used were those set by default. The results are displayed in clusters and the 4 better conformations for each cluster is available as a pdb file. Additional graphical and numerical information is provided for each cluster. Cluster conformations were evaluated based on their similarity to the adenine complexed to PAP-I (Protein Databank ID: 1qci) and adenine docked to KPAP and the closest conformation was selected as the final solution. Analysis of amino acid interactions with the ligand and within the protein was performed using PyMol.

2.11.5 PyMol

PyMol (https://pymol.org/2/) is a python-based molecule visualization software to produce high quality 3-D images of small molecules to multi-protein complexes. From a pdb file, the cartoon version of the protein was selected and chosen amino acids or ligands were converted to sticks for visualization purposes. From a selection of amino acids, RNA backbone, or adenine ligand, hydrogen bonds were identified by using the find function. Dash color was set to black, and labels were turned on. After color-coding the protein against its ligand and finding the right angle, the molecule was rendered using the ray function, and a picture was captured.

2.12 Protein purification

2.12.1 Culture and induction

Transformed Rosetta 2 cells (BL21-DE3 pLysS) were cultured in 5mL LB media at 37°C overnight with chloramphenicol and appropriate plasmid antibiotic selection. An aliquot of the saturated culture was transferred to a large LB culture (50, 250, or 500mL) containing the appropriate antibiotics and grown at 37°C until the optical density (OD600nm) reached 0.8. Induction was triggered by adding IPTG (0.6mM final concentration) and transferring the culture to an 18°C shaking water bath in a cold room for 16h. The culture was centrifuged at 10,000 x g for 20min at 4°C. The resulting whole cell pellet was stored at -20°C or directly used in the lysis step.

2.12.2 Lysis by sonication

The whole cell pellet was resuspended in 10-30mL lysis buffer (50mM Tris pH8.0, 300mM NaCl, 5% glycerol) and sonicated (10sec ON; 30 sec OFF) at 18 microns peak-to-peak for 5-15cycles depending on large culture size (ie. larger cultures required more sonication due to more cells). The lysate was centrifuged at 9,800 x g for 30min at 4°C in Nalgene bottles (250mL). The supernatant was collected and passed through a syringe filter (0.22µm) to remove potential debris that did not pellet in the spin.

2.12.3 Nickel column purification

Using 5mL propylene gravity columns (Thermo Scientific Cat#29922), 500uL of 1:1 nickel sepharose bead slurry (GE Healthcare Cat#175318-01) in 20% ethanol was used and layered on the column bed provided by the column manufacturer. The column was washed once with 5mL of dH_2O, then 5mL lysis buffer and allowed to empty by gravity flow. The supernatant

was applied to the column followed by one 30mL wash of 1mM imidazole in lysis buffer followed by one 30mL wash of 10mM imidazole in lysis buffer. The elution was performed with 3mL of 250mM imidazole in lysis buffer. Concentration of the product was obtained by a Bradford assay and the quality of the purification was assessed by Coomassie-stained SDS-PAGE analysis.

2.12.4 Column concentration

The column concentration allows for buffer exchange and protein concentration. The Amicon Ultra-15 centrifugal filter unit concentrator (UFC901008) has a 10kDa threshold and will retain heavier proteins. It was first washed with 10mL dH_2O and centrifuged at 3270 x g for 15min (until it reached the 1mL mark). The eluted protein from the nickel purification was added to the concentrator and centrifuged in the same manner. Protein storage buffer (20mM Tris pH8.1, 1mM EDTA, 1mM DTT, 100mM NH_4Cl, without glycerol) was then added (10mL) to the concentrator and centrifuged. It was repeated two more times to dilute the imidazole required for the elution. After buffer exchange, the protein was concentrated to 500uL and transferred to a fresh microfuge tube.

2.12.5 Bradford assay

The Bradford assay is used to measure the concentration of a protein sample. The Bradford reagent (0.1g Coomassie brilliant blue G250, 50mL 100% ethanol, 100mL 85% H_3PO_4) was diluted 5-times to working concentration. The standard curve was generated by creating a gradient of bovine serum albumin (BSA) in protein storage buffer (20mM Tris pH8.1, 1mM EDTA, 1mM DTT, 100mM NH_4Cl, without glycerol) with the following amount in 500uL Bradford reagent in a cuvette: 1ug, 500ng, 250ng, 125ng, 62.5ng. The correlation factor (r^2) was greater than 0.99. The protein samples (10uL) were added to 500uL Bradford reagent in a cuvette and measured at OD600. The concentration was extrapolated from the absorption on the standard curve and was recorded.

2.12.6 Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE)

This method relies on protein weight for separation through an acrylamide matrix. A gel casting apparatus was prepared by placing a small plate in front of the bigger plate which is designed to allow a gap between the two plates. Within the gap was poured the separating gel (0.375M Tris pH8.8, 12% acrylamide/bis, 0.1% SDS, 0.05% ammonium persulfate, 0.1%

TEMED) and layered on top was 100% isopropanol. Once the gel solidified, the isopropanol was poured out and the stacking gel (1.25M Tris pH6.8, 4% acrylamide/bis (Stock: 0.8% w/v N'N' – Bis-methylene-acrylamide, 29.2% w/v acrylamide), 0.1%SDS, 0.06% ammonium persulfate, 0.1% TEMED) was added until the top of the small plate. A 10-lane comb was inserted between the plates and allowed to solidify. The plates containing the gel was transferred to an electrophoresis tank (Biorad). Ice cold SDS running buffer (25mM Tris pH 8.3, 192mM Glycine, 0.1% SDS) was added to the inner chamber until it reached the top and to the outer chamber until it submerged the electrode. The samples were mixed with equal volume 2X SDS sample buffer (125mM Tris pH6.8, 20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.05% bromophenol blue) and boiled for 5min at 95°C. The comb was removed from the gel, and the samples were added to the wells. The gel ran at 200V for 40min.

2.12.7 Coomassie staining

Coomassie staining is used for visualizing protein on an SDS-PAGE gel. At the end of the electrophoresis, the gel was removed from the plates and the stacking gel was cut and discarded. The remaining separating gel was washed three times with dH_2O and microwaved in dH_2O for 7 seconds. The dH_2O was discarded and Coomassie brilliant blue stain (0.25%w/v Coomassie brilliant blue R250, 45% methanol, 10% acetic acid) was added and the gel was microwaved 7 seconds (~50°C). The gel was rocked for 1min and the stain was discarded. The gel was washed 3 times with dH_2O. Destain solution (40% methanol, 10% acetic acid) was added with 3 kim wipes rolled in a ball to absorb the dye released by the gel.

2.12.8 Western blot

A western blot is used to determine a protein identity based on interaction of immobilized proteins and antibodies. Protein samples ran on a SDS-PAGE gel and the gel was layered on a nitrocellulose blotting membrane (GE Healthcare Cat#: 10600003) between two layers of filter papers and two sponges hydrated in western transfer buffer (25mM Tris pH8, 192mM glycine, 20% methanol, 2% SDS). The layers were held together in a sandwich and placed in a transfer apparatus (Biorad) filled with western transfer buffer. The transfer happened at 120V for 40min at 4°C. After, the gel was discarded and the membrane was blocked in blocking buffer (1X TBS-T (0.1% Tween-20), 5% BSA) for 2h rocking at room temperature. The blocking buffer was discarded and the polyclonal PAP antibody (gifted by N.Tumer from Rutgers U; 1:5,000) was

added in blocking buffer and rocked at 4°C overnight. The membrane was washed 4 times with 1X TBS-T for 10minutes each with rocking at room temperature. The membrane was then incubated with anti-rabbit secondary antibody (Sigma Aldrich Cat#: A4914; 1:10,000) in blocking buffer and rocked at room temperature for 1.5h. Next, the membrane was washed as described above and 1mL of detection reagent 1 and 2 (Thermo Fischer Cat#: 32109) were added. The blot was imaged using a MicroChemi (Bio-Imaging Systems).

2.13 *In vitro* transcription

DNA construct pcDNA3-Luc was linearized using XbaI in 10X Cutsmart buffer (500mM potassium acetate, 200mM Tris-acetate pH7.9, 100mM magnesium acetate, 100ug/mL BSA) and incubated at 37°C overnight. Following digestion, equal volume of basic phenol: chloroform: isoamyl alcohol (PCI; pH8) was added and sample was vortexed and centrifuged at 16,000 x g for 5 min. About 80% of the top layer was collected in a fresh tube and precipitated with 0.1 volume 3M NaOAc and 2.5 volumes of 100% ethanol. The mixture was vortexed and incubated at -20°C overnight. The solution was centrifuged at 16,000 x g for 30min and the supernatant was discarded. Washing the pellet was performed with 500uL of 70% ethanol, the tube was inverted several times and centrifuged at 16,000 x g for 5 min and the supernatant was discarded. The washing process occurred a total of two times after which the pellet was allowed to air dry for 10 min. The pellet was resuspended in 50uL dH_2O and incubated in ice for 10min. The linear DNA concentration was quantitatively measured using a nanodrop 1000 (Thermo Fischer) and was also analyzed on a 1% agarose gel against a circular plasmid. Good quality linear plasmid DNA (5ug) was added to the in vitro transcription mix containing 5X transcription buffer (200mM Tri-HCl pH7.9, 30mM MgCl2, 5mM DTT, 2mM spermidine), RNase Inhibitor Murine, 10mM NTPs, T7 polymerase and dH2O in a total volume of 100uL. The mix was incubated at 37°C for 2h. Subsequently, 4U of DNase I was added and incubated at 37°C for 30min to digest the linear DNA template. The sample volume was increased to 200uL by adding 100uL of dH2O and 200uL of acidic PCI (pH4.5) was added to the reaction followed by vortexing. The sample was centrifuged at 16,000 x g for 5min and 80% of the top layer was collected for precipitation with 0.5 volume 7.5M NH4OAc and 2.5volumes of 100% ethanol and was stored at -20°C overnight. Washing of the RNA pellet was performed identically to linear DNA pellet and it was resuspended in 18uL dH2O. The concentration quantification was performed with the nanodrop spectrophotometer and analyzed on a 6% urea/acrylamide

2.14 Aniline treatment

The aniline treatment was used to identify the presence of abasic sites in an RNA sample when treated with a RIP. The RIP (500ng) was incubated with 20uL of tobacco ribosomes (previously isolated) with dH_2O to a final volume of 100uL and the mixture was incubated at 30°C for 30min. Next, 100uL of 2X extraction buffer (240mM NaCl, 50mM Tris-HCl pH8.8, 20mM EDTA, 2% SDS) was added to the mixture followed by 200uL of acid PCI (25:24:1). The vortexed mixture was centrifuged at 16,000 x g for 10min and the supernatant was added to 200uL of CI (24:1). The mixture was vortexed and centrifuged at 16,000 x g for 10min. The supernatant was added to 1/10 volume 3M NaOAc (pH5.5) and 2 volumes 100% ethanol and incubated at -20°C overnight to precipitate the RNA away from the RIP and ribosomal proteins. The precipitate was centrifuged at 16,000 x g for 30min at 4°C and the pellet was washed with 70% ethanol. The pellet was allowed to air dry for 5min and was dissolved in 10uL dH_2O. The sample was split into two 5uL aliquots where one was frozen at -40°C while the other one was added 25uL aniline mix (25uL aniline, 30uL 100% acetic acid, 200uL dH_2O) and incubated on ice for 30min. Next, 1/10 volume 3M NaOAc (pH5.5) and 2 volumes 100% ethanol was added and incubated at -20°C overnight. Once centrifuged at 16,000 x g for 30min and one 70% ethanol wash, 10uL of formamide buffer (950uL formamide, 40uL 0.5M EDTA, 10uL dH_2O) was used to resuspend the pellet. The frozen aliquot at -40°C was thawed and 5uL of formamide buffer was added. The two samples were incubated at 65°C for 10min and snap cooled on ice for 2min. The samples were allowed to run on a 7M urea/6% acrylamide gel at 200V until the leading dye reached 2/3 of the cassette. The gel was stained in ethidium bromide in dH_2O for 15min and destained with dH_2O for 15min. The gel was imaged under a UV imager.

2.15 *In vitro* translation

The translation mix was made using a rabbit reticulocyte lysate kit (Promega Cat# L4960). The master mix had the following components: 9.25uL rabbit reticulocyte lysate, 0.25uL complete amino acid mix, 0.25uL RNase Inhibitor Murine (NEB Cat# M0314). From that mix 9.25uL was added to 1.5ug luciferase mRNA and PAP-I/KPAP (0.04nM or 0.4nM; 12pg or 1.2pg respectively) and dH_2O for a total sample volume of 12.5uL. The samples were incubated at 30°C for 90min followed by a 5min incubation at 30°C with 2uL of 10mg/mL of RNase A added to the mix. On a 96-well plate, 50uL of luciferase assay reagent (LAR; Promega Cat# E1500) was thoroughly mixed with 2.5uL of the incubated translation mix and quickly inserted

into a plate reader (Synergy H4 hybrid microplate reader) to measure luminescence. Raw values were recorded and plotted against each other on a bar graph. Values were normalized to the average of the sample containing no protein and values of samples were expressed as percent protein synthesis

3 RESULTS

3.1 The discovery of the KPAP gene model

A gene model allows one to understand the arrangement of the features and provides basic knowledge for transcription and translation of a gene. In **Figure 5**, the well-known PAP-I gene model is displayed as means of comparison for the KPAP gene model which proves to be more complicated. Near the promoter the first 5'UTR is observed while the second portion of the 5'UTR is beside the coding sequence (CDS) and the 3'UTR. This gene model was possible thanks to the second sequencing iteration of the pokeweed genome assembled by our lab (Dougherty & Hudak 2022 *unpublished*). Previously our lab performed PCR experiments on the genomic DNA and cDNA of KPAP with primers on the first 5'UTR (near the promoter) and 3'UTR (data not shown). The data confirmed a product on the cDNA but not the gDNA despite increased extension time. This indicated that the two portions of 5'UTRs were flanking a large piece of gDNA that would be spliced like introns. Additionally, another student performed a genome walking assay (data not shown) confirming 750bp of intron upstream of the second 5'UTR. Genomic data allowed us to see that the length for this intron was 43.75kb. Prior to this recent genome sequencing, limited data were available from the first sequencing iteration. The problem was that the small reads from the first sequencing placed the promoter and first 5'UTR on another piece of assembled gDNA (contig) than the CDS. Therefore, to confirm and connect the two contigs a PCR-sequencing approach was used to walk upstream of the CDS until no product was observed. This was achieved by designing primers upstream of the known CDS to confirm the sequencing data of the contig. At the bottom of **Figure 5**, agarose gels are color labeled with the regions probed confirming the existence of the product. The 5'UTRs were confirmed on both contigs by sequencing PCR products. Certain regions were difficult to probe due to the nature of the sequence (ie. AT rich/repeat regions). One such region was the inter-contig region which was probed with multiple different primer pairs. With the new genomic sequencing, the probed regions were confirmed and the contigs were connected. Overall, the KPAP gene model seems to follow PAP-I gene model where a large intron is flanked by 5'UTRs followed by the CDS and the 3'UTR.

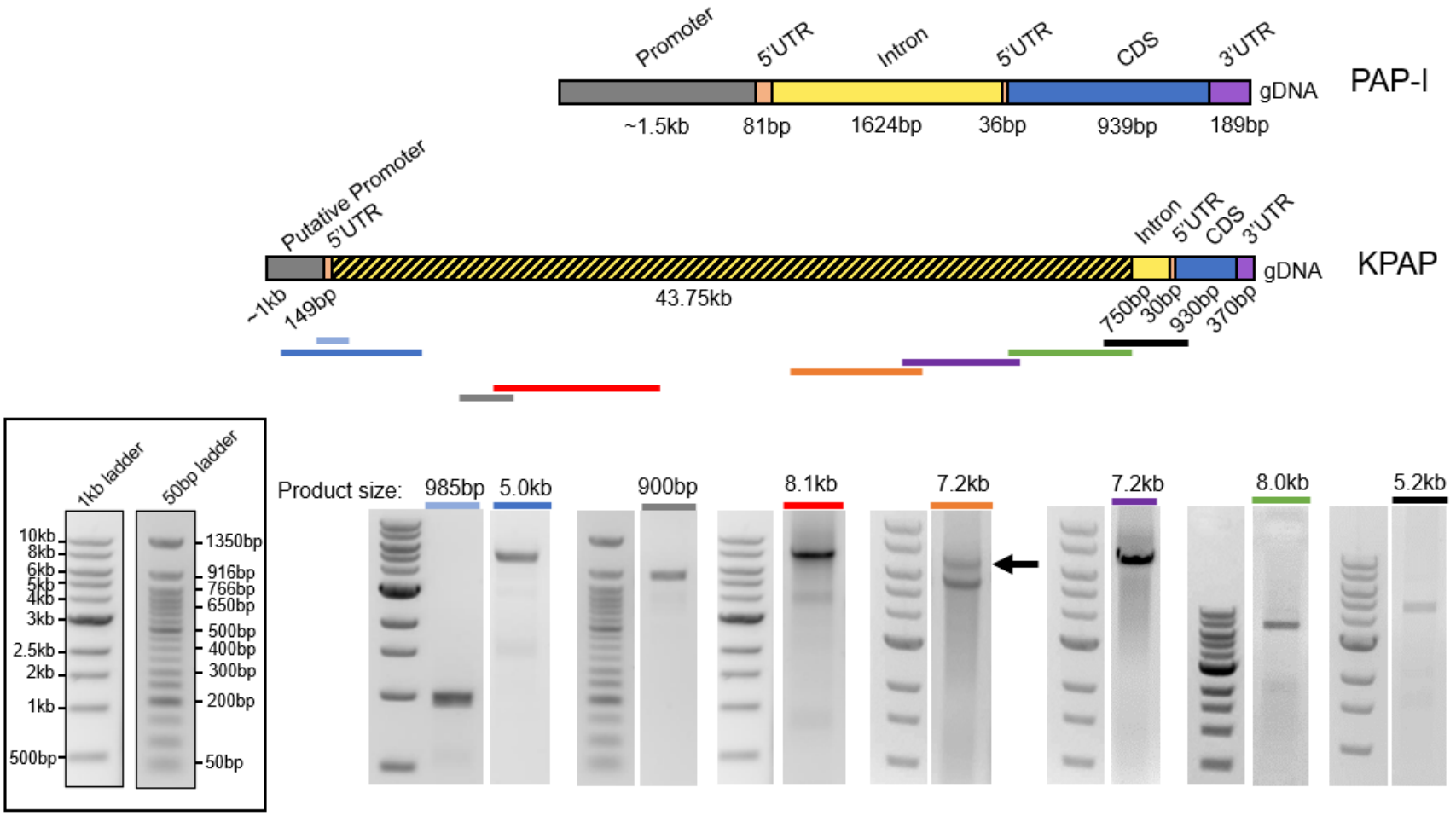


Figure 5: The development of the KPAP gene model by using genomic sequencing data and PCR walking. The well-known gene model of PAP-I at the top with the promoter (grey), the first 5'UTR (pale orange), the intron (pale yellow), the second 5'UTR (pale orange), the coding sequence (CDS; blue), and the 3'UTR (purple). The size of each feature is stated below. The KPAP gene model follows the same color pattern as the PAP-I model with the yellow striped area featuring an unknown area potentially related to the intron. The colored bars below refer to the PCR products shown below. The product sizes for each PCR are displayed at the top and the colored line matches the colored line on the gene model to provide context

3.2 Upstream open reading frames (uORFs) in the long intron

It is uncommon to see such large introns, however, this one seems to harbor coding sequences. Using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) 22 upstream open reading frames (uORFs) were identified having greater than 100 amino acids (see **Table 1**). They were confirmed in Snapgene Viewer. They were observed in all 6 frames and the largest one was nearly as large as KPAP with 252aa. When these uORFs were blasted (Basic Local Alignment Tool (BLAST)) similar hits were obtained including reverse transcriptase, RNaseH, integrase, and gag which are elements of retrotransposons, group II introns and endogenous retroviruses (Schaffer et al. 2003).

3.3 Amino acid alignment of the PAP genes and features

The amino acid alignment of the PAP genes allows us to distinguish specific amino acid conservation among the different isoforms. At the N-terminus, the underlined amino acids refer to the signal peptide that is cleaved in the mature protein as a necessary precursor of transport to the apoplast (**Figure 6**). For PAP-I, this is 22aa for the N-terminus and 29aa for the C-terminus (Lodge et al. 1993). The potential signal sequence of other genes was predicted using SignalP 5.0. Among them KPAP and PAP-I were compared (**Figure 7**). The prediction obtained from the PAP-I sequence was accurate with the literature and predicted the cleavage site with high confidence (100%). The KPAP prediction was less accurate and showed lower confidence (<50%). The cleavage site can be predicted between Val18 and Iso25. SignalP confirms its prediction on the bottom line where Arg23 is labeled as the first amino acid that is not part of the signal peptide. Despite the confidence of this prediction, it seems very likely that KPAP possesses a signal sequence in the vicinity of Arg23. Despite the alignment showing a several amino acid gap between the KPAP and PAP-I signal sequences, they are both the same length (22aa). The identity similarity of the alignment of KPAP compared to PAP-I is 34% as obtained with the basic local alignment search tool (BLAST). In **Figure 6**, the bolded amino acids refer to known highly conserved amino acids in other RIPs such as ricin A-chain, abrin A-chain, trichosanthin, mirabilis antiviral protein and saporin-6 (Funatsu et al. 1991). The light orange refers to the cysteine pair required for the first disulfide bond while the dark orange represents the second disulfide bond. KPAP predicted to have only one disulfide bond was different from the other PAP genes that have two disulfide bonds. The green refers to the active site amino acids which have been well studied (Hudak et al. 2004).

Table 1: Characteristics of upstream open reading frames (uORFs) from the 5'UTR intron

Name	Frame	Nucleotide length (bp)	Peptide length (aa)	Peptide Weight (kDa)	Distance from TSS (bp)	BLAST results (% identity)
uORF 6	3	759	252	29.1	+21,182	RT (57%), RNaseH (55%)
uORF -12	-2	549	182	19.9	+7,031	Extensin (42%), uncharacterized proteins(42%)
uORF -1	-3	534	177	20.7	+40,755	gag (44%)
uORF -13	-3	513	170	18.9	+2,919	Sugar phosphate permease (30%), uncharacterized proteins (29%)
uORF -9	-3	477	158	17.2	+30,144	RT (34%), gag (37%), integrase (44%)
uORF -4	-2	474	157	18.3	+37,820	RT (45%), RNaseH (47%), integrase (46%)
uORF -8	-1	438	145	16.7	+32,143	RT (29%), integrase (31%), gag (28%)
uORF -6	-2	438	145	15.9	+34,724	Cysteine synthase (37%), uncharacterized proteins (28%)
uORF -2	-1	420	139	15.9	+39,247	RNaseH (41%), gag (45%)
uORF -7	-3	414	137	15.2	+34,068	Cyclin-dependant kinase inhibitor 3 (26%)
uORF 9	1	405	134	15.3	+40,872	Protocadherin-like polarity protein stan (31%), protein detoxification (31%)
uORF 1	2	390	129	14.7	+988	RNaseH (40%), RT (40%)
uORF 2	1	387	128	14.5	+1,446	Endonuclease/exonuclease/phosphatase (33%), LINE1 RT (37%), RT (32%), Rnase H (35%)
uORF 8	3	378	125	14.5	+31,748	NAD(P)-binding domain (42%)
uORF -10	-1	372	123	14.3	+28,189	Mitochondrial translation initiation factor 2 (32%)
uORF -3	-1	363	120	13.9	+38,245	RNaseH (60%), uncharacterized proteins (59%), RT (59%)
uORF -5	-2	357	118	13.5	+36,296	RT (58%), gag (63%)
uORF 5	1	354	117	13.1	+20,811	Uncharacterized proteins (37%), gag (37%), RNaseH (35%), integrase (37%)
uORF -11	-3	342	113	12.6	+20,136	UDP-N-acetylglucosamine-1-phosphate (33%), uncharacterized proteins (27%)
uORF 4	3	339	112	13.0	+19,895	Chromo domain-containing protein (33%), uncharacterized protein (37%)
uORF 3	3	339	112	12.3	+6,695	Uncharacterized protein (48%)
uORF 7	1	336	111	13.1	+23,898	Integrase (39%)

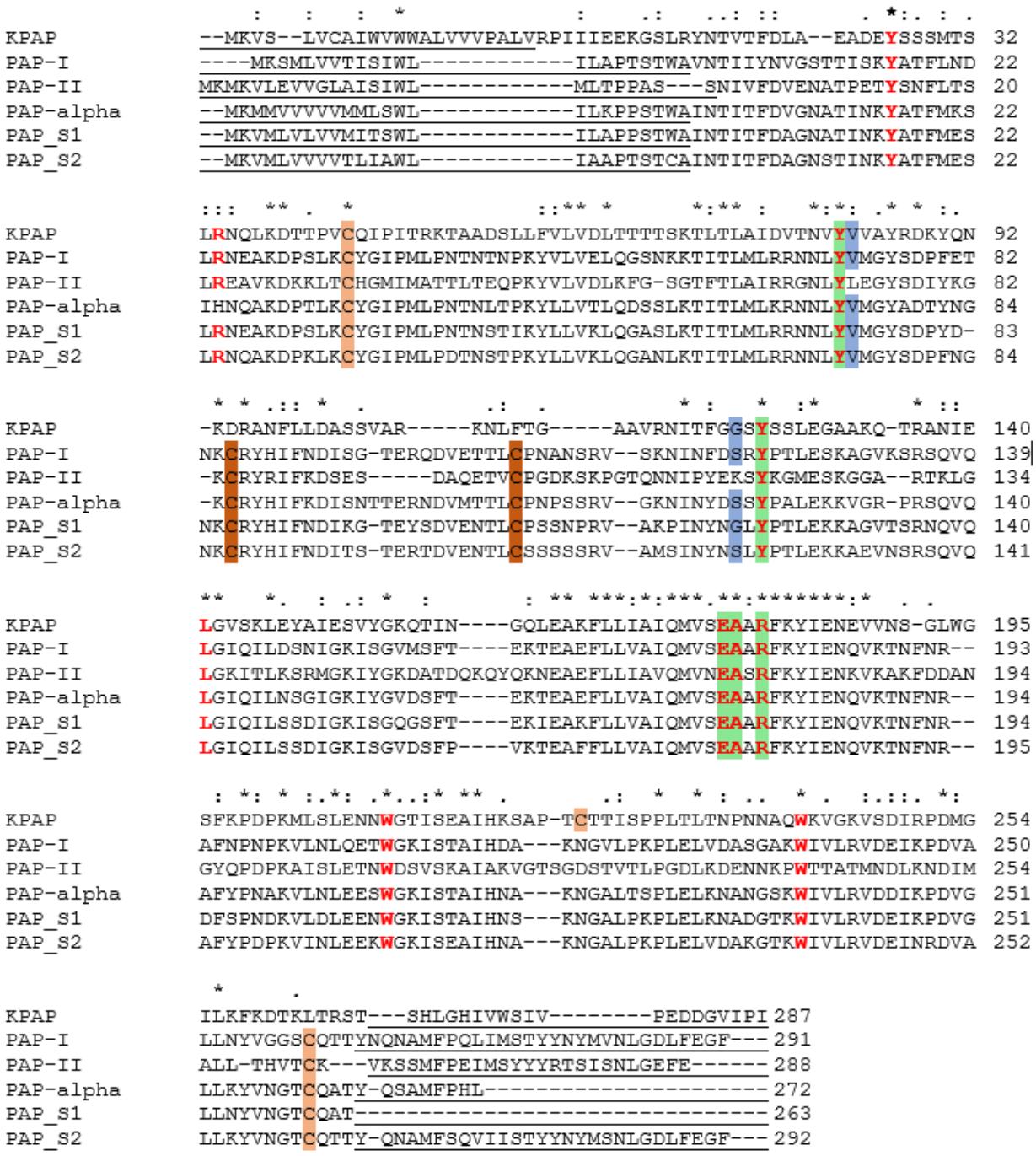
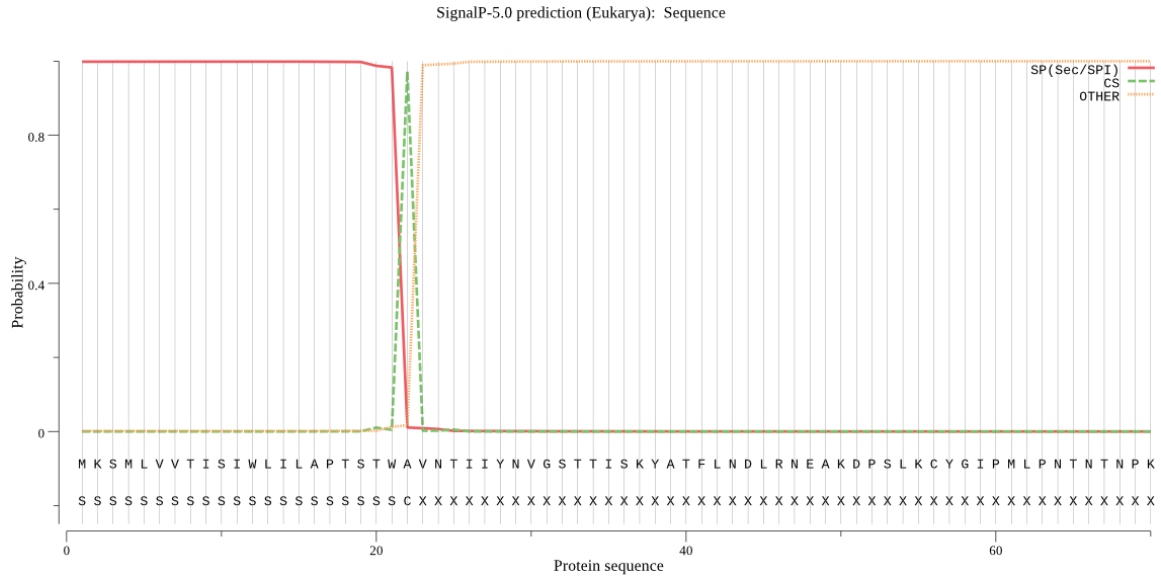


Figure 6: Amino acid alignment of all known PAP isoforms using Clustal Omega. The notation above the alignment refers to the degree to which the residue is conserved across all isoforms: (*) fully conserved, (:) residues share common properties, (.) weakly conserved residues. The underlined residues refer to the signal peptides of the isoforms in the N- and C-termini predicted by SignalP 5.0. The bolded residues refer to known highly conserved amino acids among RIPs. The light orange refers to the first disulfide bond and the dark orange the second, absent in KPAP. The green refers to active site amino acids. The blue refers to additional amino acids involved in the active site

A



B

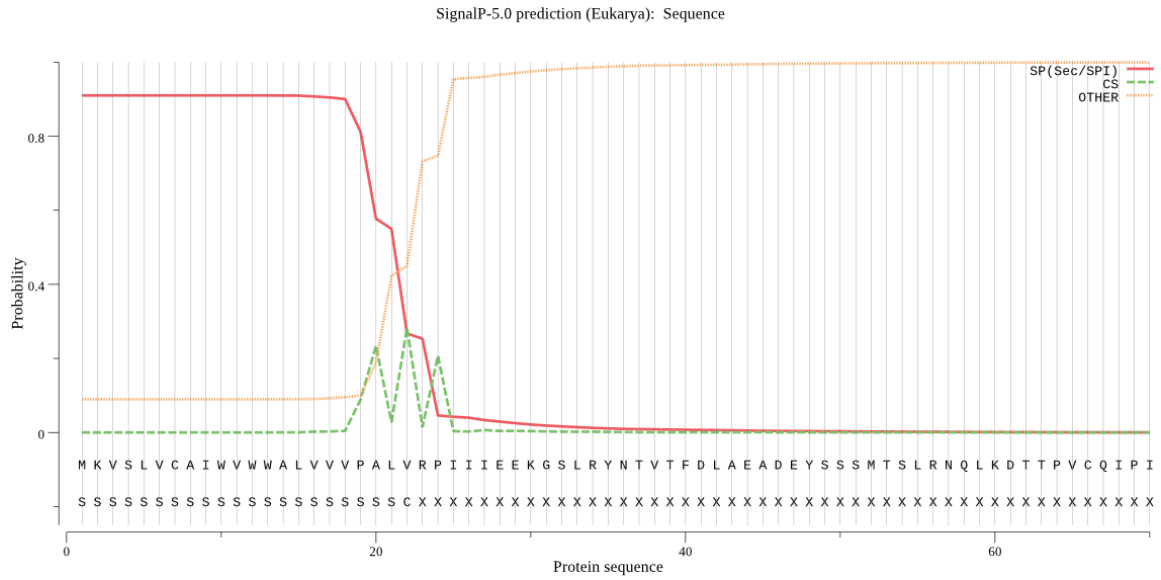


Figure 7: Comparison of predicted PAP-I and KPAP N-terminal signal sequence. The red line represents the probability of the amino acid being a portion of the signal peptide, the green dotted line refers to the cleavage site and the pale orange line represents the other amino acids that are not part of the signal peptide. (A) The graph for PAP-I shows a sharp decrease for the signal peptide at amino acid Ala22 and a sharp green spike reaching 100% probability of cleavage at Ala22. The pale orange indicates the first amino acid Val23. (B) The graph for KPAP shows a stepwise decrease of the signal peptide and 3 green spikes below 40% probability. The pale orange line also increases in a stepwise manner. The bottom line labels amino acids as either part of the signal peptide (S), part of the cleavage (C), or other amino acids unrelated to signal peptide (X).

The blue refers to additional amino acids involved in binding the RNA base in the active site. From the amino acid alignment, it was determined that KPAP differed from PAP-I (34% similar).

3.4 KPAP protein structure prediction overlayed on PAP-I crystal structure

To further elucidate the function of KPAP, a 3D protein structure was predicted and superimposed on the PAP-I (1qci) crystal structure (**Figure 8**). This way structural similarity between the two proteins could imply that they behave the same way. The prediction was performed with SWISS MODEL and the image was viewed with PyMol. The N- and C-termini are labeled indicating the beginning and the end of the protein respectively. When observing the superimposed structures, the secondary structures like α-helices and ß-sheets overlap well with some differences in the smaller helices and sheets. The major differences lie in the unstructured domains in between the secondary structures. Despite amino acid differences, structurally, KPAP and PAP-I show high similarity.

3.4.1 Comparison of SWISS-MODEL and Alphafold algorithms

In order to predict with accuracy the KPAP protein, several algorithms were used and the better model was selected from their validation scores. Here is a description of SWISS-MODEL and Alphafold and how they work. SWISS-MODEL is a web server that streamlines homology modelling workflows through an automated workflow. Upon input of an amino acid sequence, the algorithm will first attempt to find evolutionarily related protein structures that it will compare to a curated template library (Waterhouse et al. 2018). This step is performed mainly through BLAST for closely related templates and HHblits for remote homology. Next the algorithm will rank the templates according to their quality which is performed by Global Model Quality Estimate (GMQE) and Quaternary Structure Quality Estimate (QSQE; Waterhouse et al. 2018). In the model building step, the conserved amino acids from the query to the selected template will be used as the starting point to assign atom coordinates followed by loop modelling for lesser conserved regions. This step is performed by the OpenStructure framework and ProMod3 (Waterhouse et al. 2018). Lastly, the software will attempt to rank the quality of the models by using the QMEAN scoring function which is based on the combination of individual models and other models that use different templates. Overall, the software provides a variety of results based on different templates ranked according to the scoring function. In addition to

SWISS-MODEL, other algorithms were used to model KPAP, including Alphafold, which is gaining popularity in the field due to the use of deep learning neural networks. Alphafold begins by comparing the sequence input to evolutionarily related proteins by using several multiple sequence alignment tools including JackHMMER and HHBlits (Jumper et al. 2021). In addition, a template-based method gathers pairwise features of amino acids between the query and the template library of structures. Next, the algorithm uses an attention-based deep learning architecture which has the effect of refining the atomic details. Attention is the process by which programmers can introduce pieces of information to favor a certain selection of outcomes. Therefore, the algorithm can learn and refine models based on previous empirical and experimental evidence. It begins with the Evoformer block that exchanges information with the multiple sequence alignments and the pairwise features (previously generated) to establish spatial and evolutionary relationships (Jumper et al. 2021). The structure module follows which will initiate the building of the 3D model using rotation and translation parameters for each residue. Initially the parameters remain trivial however, with the iterative refinement which loops the output back into the pipeline, it rapidly improves the accuracy of the model (Jumper et al. 2021). As a final step, AMBER is used in the form of an energy refinement through a gradient descent. The gradient descent allows for energy minimization through a set number of cycles (typically 100-500) to relax the protein energetically. This process has the effect of removing stereochemical violations within peptide bond geometry. When the final structures of KPAP were validated and compared, the validation scores of the SWISS-MODEL structure were superior to the Alphafold prediction. Therefore, the SWISS-MODEL structure was selected and used in the in silico applications downstream.

3.5 Comparison of key amino acids in the active site of PAP-I and KPAP

After generating the predicted 3D structure of KPAP, elucidating the differences in the active site in comparison to PAP-I was investigated. The crystal structure of PAP-I (PDB: 1qci) active site shows Tyr72 and Tyr123 involved in the π-π-stacking of the adenine base (**Figure 9A**). The catalytic Arg179 (1.7Å away from the adenine N3) and the supporting Glu176 are surrounding the base. Additional amino acids such as Val73 and Ser121 also help in stabilizing the base in the active site. Hydrogen bonding is observed at distance 2.7Å to 3.3Å represented by the black dash lines. In contrast to PAP-I, the predicted docking of an adenine base into the active site of the KPAP SWISS MODEL prediction was performed using Autodock4 (**Figure**

9B). The key amino acids are very similar to PAP-I where we can observe Tyr82 and Tyr125 involved in the base stacking, the catalytic Arg180 and Glu177 are also present and the additional Val83 and Gly123 (substituted from Ser in PAP-I). The overlap of these amino acids can be observed when KPAP is superimposed on PAP-I (in PyMol using the superimposition command; **Figure 9C**). The overlap shows that the general orientation of the base and the active site are relatively similar with the minor differences observed in the topmost Tyr (KPAP Y82 and PAP-I Y72), the catalytic Arg (KPAP R180 and PAP-I R179) and the substitution of PAP-I Ser121 for KPAP Gly123. Additional similarities can be observed with KPAP Trp210 and PAP-I Trp208 (**Figure 10**). This amino acid is conserved across the PAP genes and could facilitate the orientation of the arginine coplanar to the tryptophan rings. Presumably this orientation would allow the arginine to make hydrogen bonds to the adenine due to their close proximity. In addition, Arg34 from KPAP is shown to be conserved in the alignment. Its role may be explained by its interaction with other residues (**Figure 11**). In KPAP the Arg34 hydrogen bonds with Glu185, Ala178, and Ala179. This interaction occurs at the junction of three helices where one holds the catalytic Arg180 responsible for the depurination initiation. Moreover, while most of the protein retains its shape through hydrophobic interactions, this residue could act as an anchor to maintain structural integrity of three helices. There is also high orientation similarity with PAP-I suggesting an important structure element that is perhaps conserved in a greater range of RIPs. Arg34 is conserved across PAP genes except for PAP-α that has a histidine. The analysis of the 3D structure of PAP-I and KPAP allows us to identify the differences that exist structurally, which is helpful to understand the activity of the active site. Moreover, structures can also elucidate the role of conserved amino acids not present in the active site.

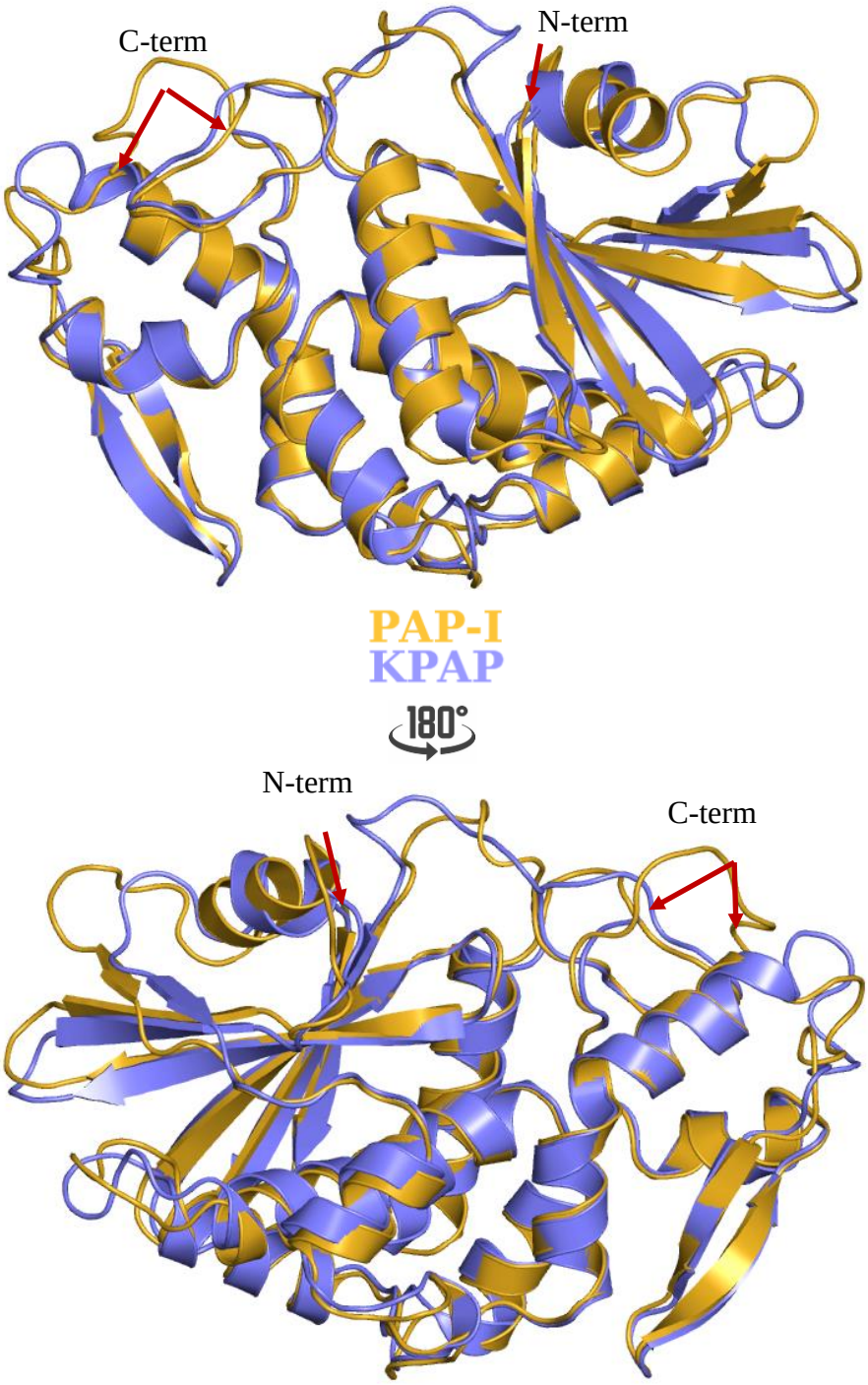


Figure 8: Protein structure overlap of PAP-I crystal structure (1qci; gold) and the SWISS MODEL predicted KPAP (blue). Arrows indicate the N-terminus and C-terminus of the proteins. The two panels compare the protein by rotating it 180 degrees in the direction of the arrow. High similarity is observed in helices and β-sheet, and unstructured domains show the most variability. The Cα RMSD between KPAP and PAP-I (1qci – chain A) is 1.28Å.

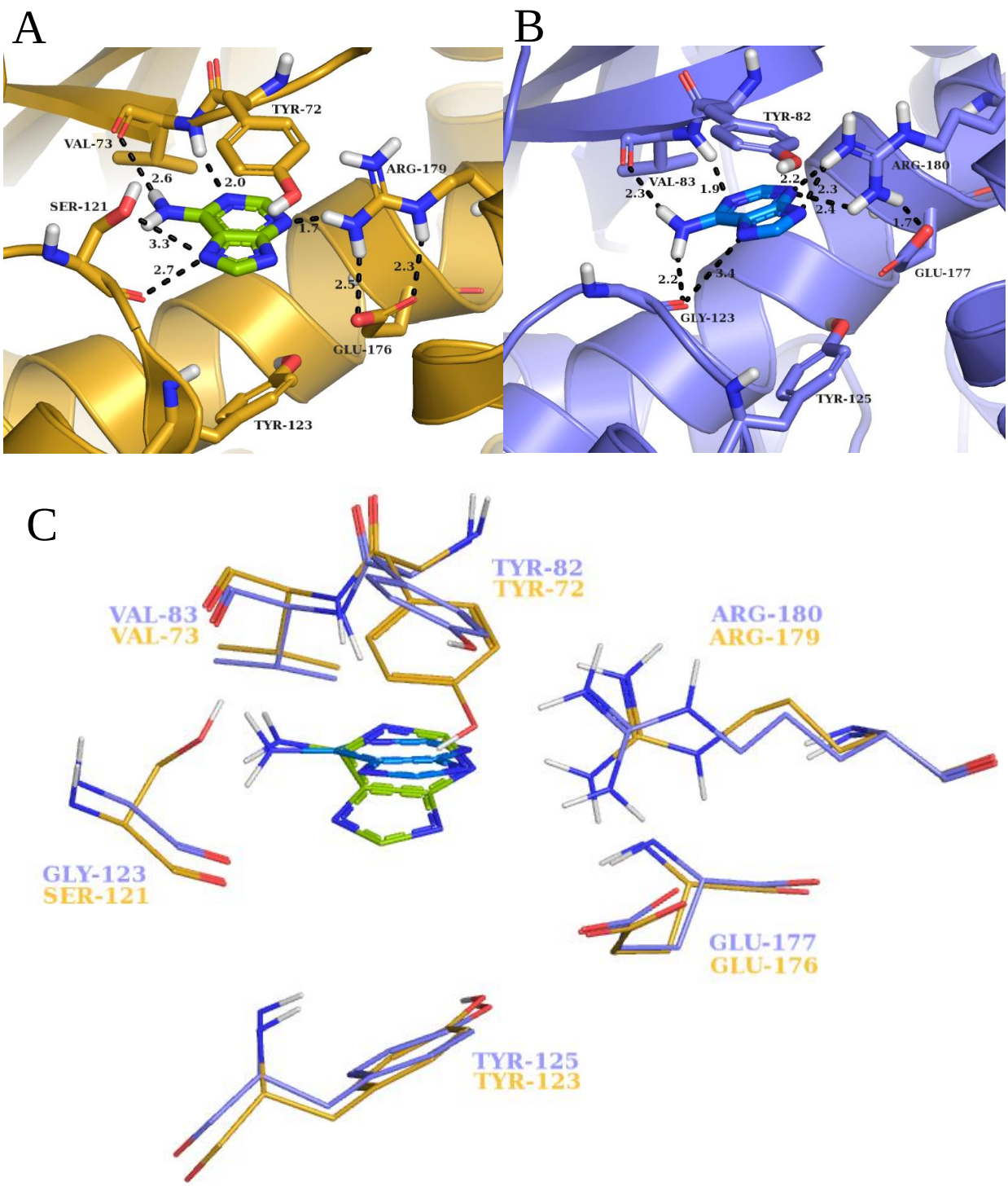


Figure 9: Active site representation of key amino acids involved in the interaction with the adenine ligand. (A) PAP-I (1qci; gold) shows labeled amino acid interaction with the adenine (green). (B) KPAP SWISS MODEL (blue) shows the adenine (sky blue) in the active site generated by Autodock4 . Black dashes represent the hydrogen bonding. The numbers represent the distance in Ångstroms of the closest dash line. (C) Key amino acids from A and B overlapped (PAP-I is gold; KPAP is blue) Figures were made in PyMol

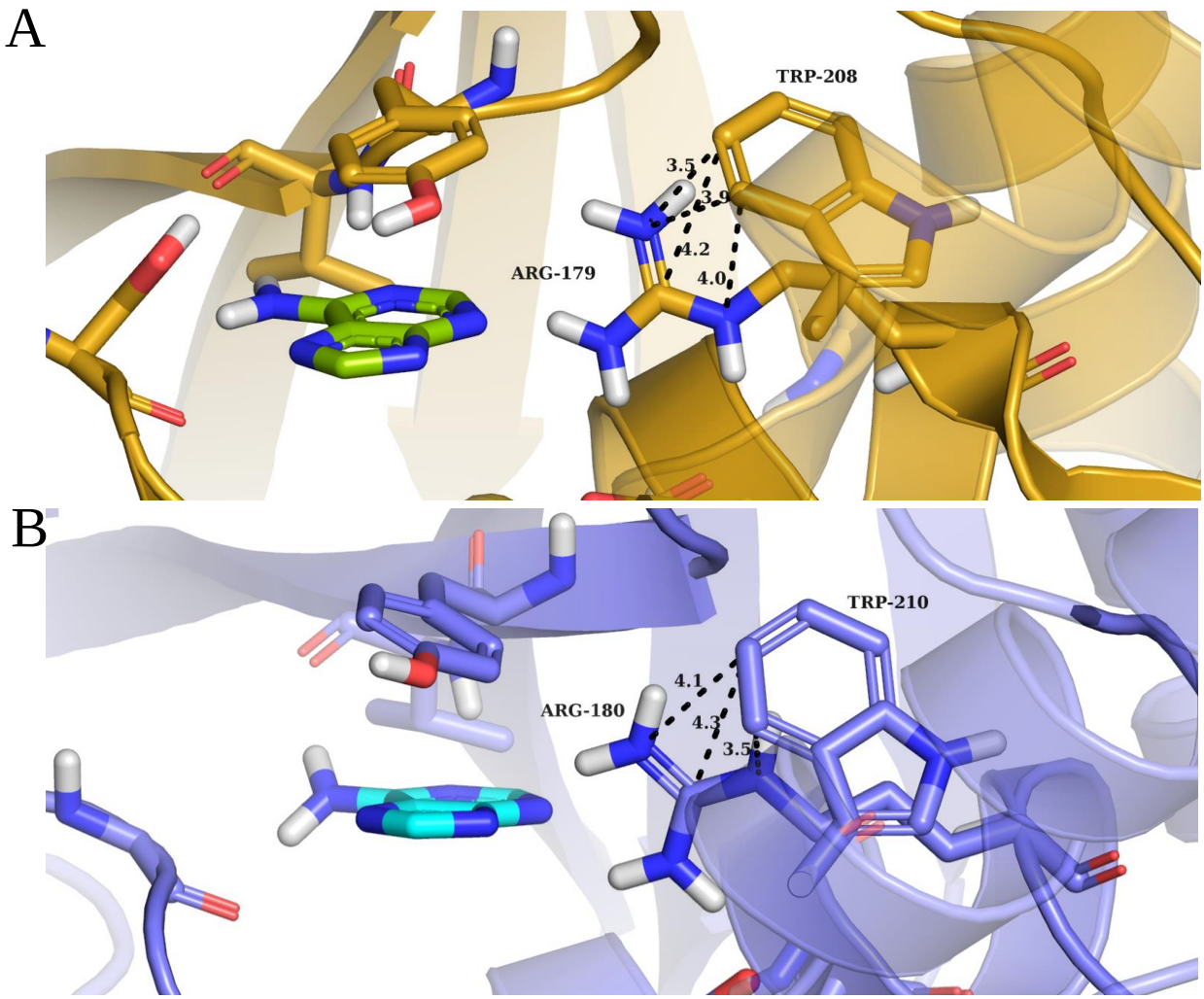


Figure 10: Representation of the active site with a focus on a tryptophan residue beside the catalytic arginine which could serve to orient the arginine to more readily hydrogen bond with its ligand adenine. The dashes show the distance in Ångstroms between atoms. PAP-I is represented in gold and KPAP is represented in blue. Tryptophan is likely coupled to arginine through cation-pi interaction.

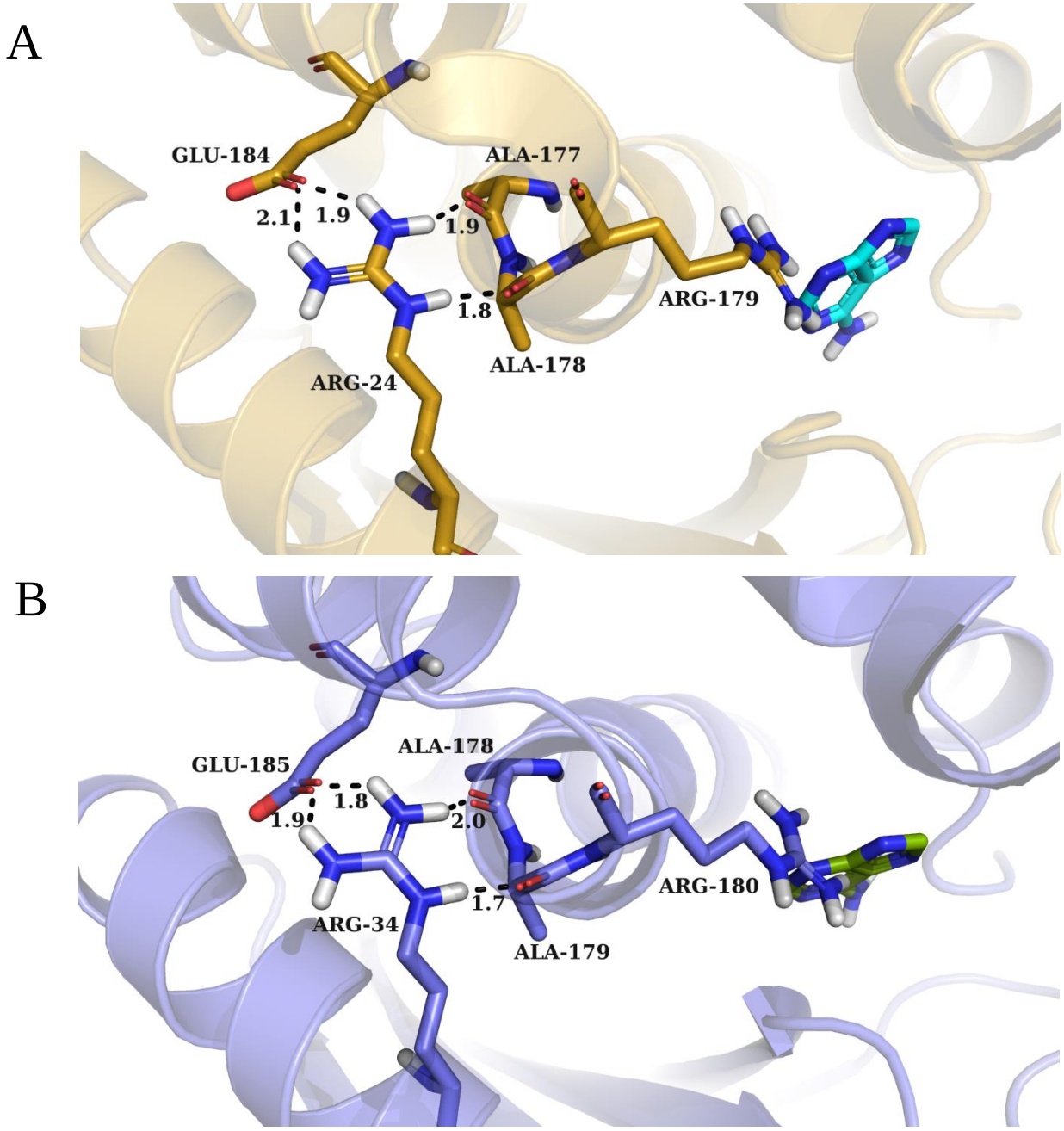


Figure 11: Focus on the conserved Arg24 (PAP-I; gold) and Arg34 (KPAP; blue) and its interactions. The arginine forms hydrogen bonds with two alanines located beside the catalytic Arg179/Arg180 (PAP-I/KPAP) from the active site. Additionally, it forms a hydrogen bond with Glu184/Glu185 (PAP-I/KPAP). The skyblue adenine (A) was taken from PAP-I (PDB: 1qci) and the green adenine (B) was taken from the autodock prediction of the KPAP SWISS MODEL. Black dashes represent the distance between atoms in Ångstroms.

3.6 Protein structure comparison of KPAP and bouganin

As previously stated, the amino acid similarity between PAP-I and KPAP is 34%. The protein that showed the highest similarity with KPAP is the RIP bouganin from *Bougainvillea spectabilis*. They are 58% identical. Understanding which protein is most similar can help in potentially identifying the functions of KPAP which remain unknown. **Figure 12** shows the alignment of KPAP to bouganin obtained using clustal omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with some essential residues labeled. One major difference from the comparison of KPAP to PAP-I is that KPAP and bouganin share one disulfide bond while PAP-I has two. Despite the relatively low similarity identity there are a lot of substituted residues with the same properties throughout the protein which increase the relatedness as function of residues may be retained. The bouganin RIP has been crystalized (PDB: 3ctk). When the KPAP protein is overlapped with bouganin a very high similarity is observed in secondary structure and disordered regions (**Figure 13A**). When looking at the active site, the residues are surprisingly well overlapped and oriented (**Figure 13B**). A potential reason for the high similarity in 3D structure may come from the protein predicting algorithms which most begin with a psi-blast to identify the closest reference model from which the query protein is built. Since bouganin was the only high scoring hit, it is possible that the protein prediction algorithms refer solely on it for the 3D prediction. Other protein prediction software (AlphaFold2, Phyre2, ORION) produced the same bias (data not shown). The other possibility could be that KPAP is relatively easy to model due to the many secondary structures. Further, crystalizing it may not bring other important information aside from confirming the current structure. Overall it seems that KPAP is a hybrid of a *P.americana* and a *B.spectabilis* RIP.

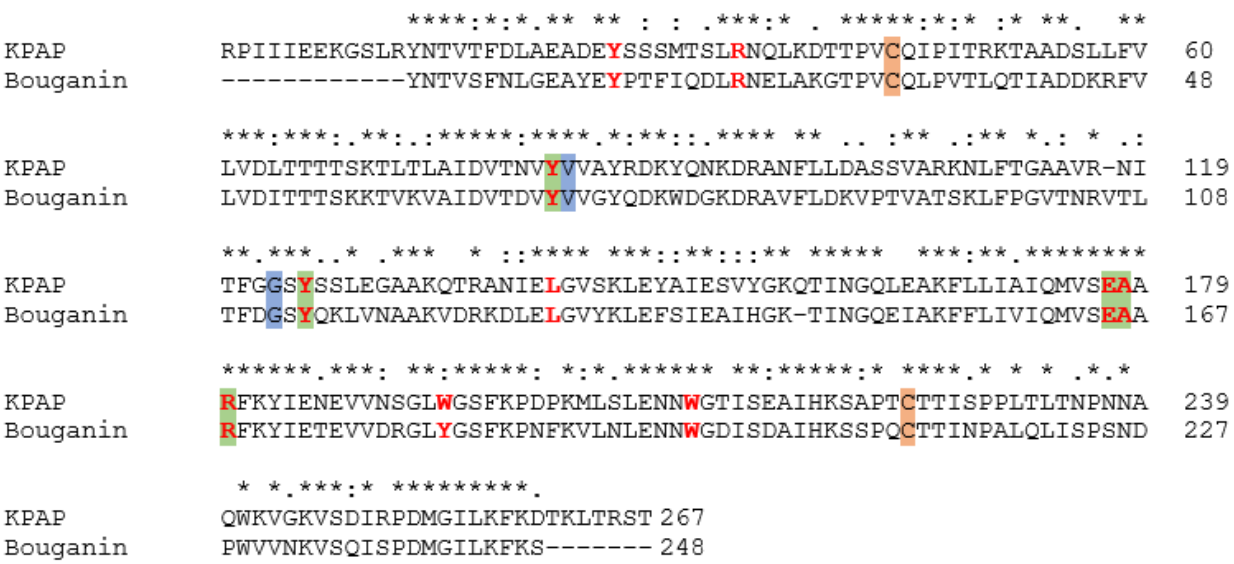


Figure 12: Alignment of KPAP and bouganin using clustal omega. The notation above the alignment refers to the degree to which the amino acids are conserved: (*) fully conserved; (:) residues share common properties; (.) residues are weakly conserved. The underlined residues are the signal peptides of the proteins in the N- and C-terminus. The bolded red amino acids refer to conserved amino acids across other RIPs. The green refers to active site residues. The pale orange refers to the only disulfide bond. The blue highlights are additional residues involve in stabilizing the ligand.

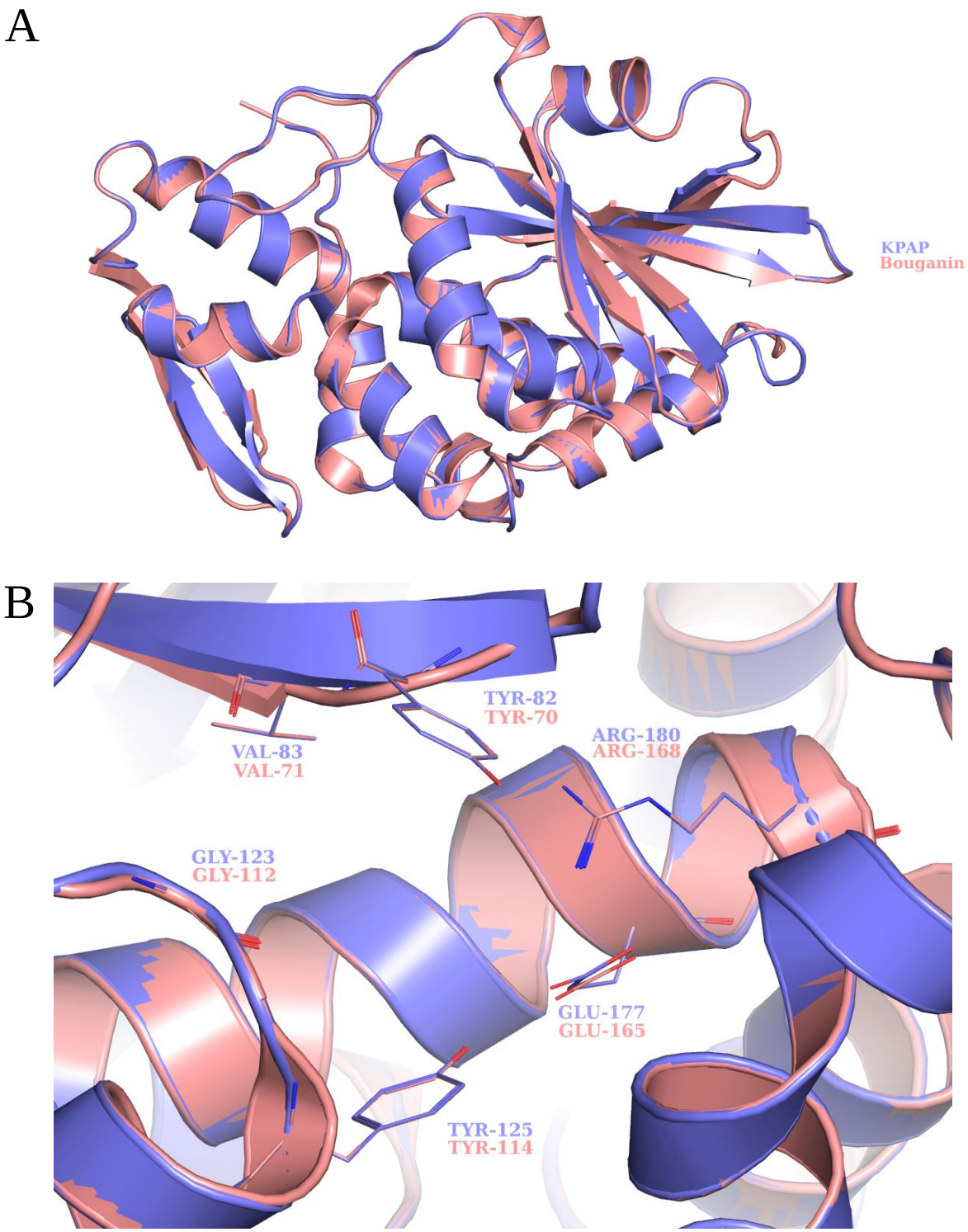


Figure 13: Comparison of KPAP (blue) to bouganin (pink; PDB: 3ctk) through an overlap of the proteins and the active site amino acids. (A) The protein overlap shows very high similarity in secondary structures and disordered regions. (B) The overlap of residues of the active site shows very high similarity in orientation. Figures were made in PyMol. The Cα RMSD between KPAP and bouganin (3ctk) is 0.27Å.

3.7 Simulating KPAP binding to *E.coli* SRL of the 23S rRNA

So far, the adenine has been the model to characterize the active site of KPAP, however, the adenine is the result of depurination. A more biologically relevant ligand would be the sarcin/ricin loop (SRL), the well-known target of PAP-I and the presumed target of KPAP. Using the HADDOCK 2.4 web server, the KPAP protein was docked with the *E.coli* SRL of the 23S rRNA (PDB: 480d). The overall interaction is shown in **Figure 14A** and a zoomed-in view (B) allows to see the base in the active site of KPAP. This simulation puts into perspective the size of the protein compared to the RNA loop and also suggests the orientation of adenine 2660 (A2660) observed at the top of the loop. A detailed look at the protein's active site during this interaction with A2660 reveals the location of the key amino acids involved in orienting the ligand (**Figure 15**). A2660 seems to be stabilized by Tyr82 through π-stacking and Arg180 is in close proximity to the adenine to initiate depurination. Glu177 hydrogen bonds with Tyr125 and the 2'OH of the ribose sugar. Finally, **Figure 15C** summarizes the orientation comparison between the simpler model complexed with adenine and the model complexed to A2660 showing very similar positions for each amino acid suggesting there is very little change involved to bind either ligand.An additional amino acid, Glu207, appears to hydrogen bond the 2'OH of the ribose sugar but does not seem to interact with the adenine alone (**Figure 16**). Perhaps the function of this amino acid is only important when binding larger ligands and maybe this tight binding between Arg180, Glu177 and Glu207 to the ribose functions to place the adenine rings in the proper position. One could also extrapolate that DNA would probably not interact the same way since it lacks the 2'OH group. In addition to the active site, there is another interesting feature of this interaction at Pro229. The proline seems to behave like an anchor at G2655 of the G-bulge, that could promote the correct positioning of A2660 at the entrance of the active site (**Figure 17**).

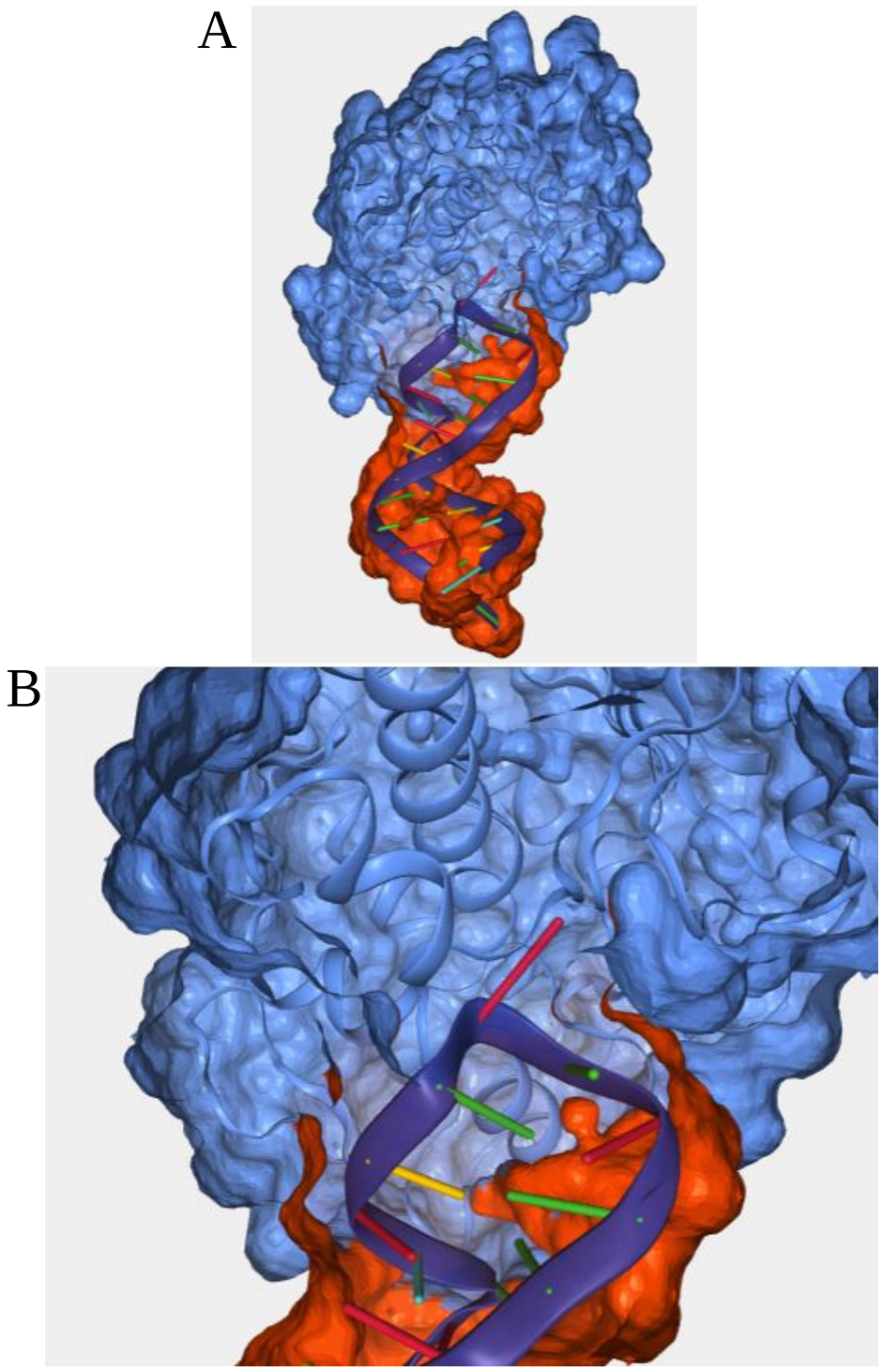


Figure 14: Haddock2.4 generated image of KPAP SWISS MODEL (blue) docked to the sarcin/ricin loop (SRL) from *E.coli* (protein databank ref: 480d) of the 23S rRNA containing 27 bases. (A) The general overview of KPAP binding the SRL is observed. (B) Adenine 2660 is represented by the pink stick observed in the active site. The surrounding orange surface around the RNA is the electron cloud from the bases and the backbone. Similarly, the blue surface around the cartoon of KPAP depicts the electron cloud from the protein atoms.

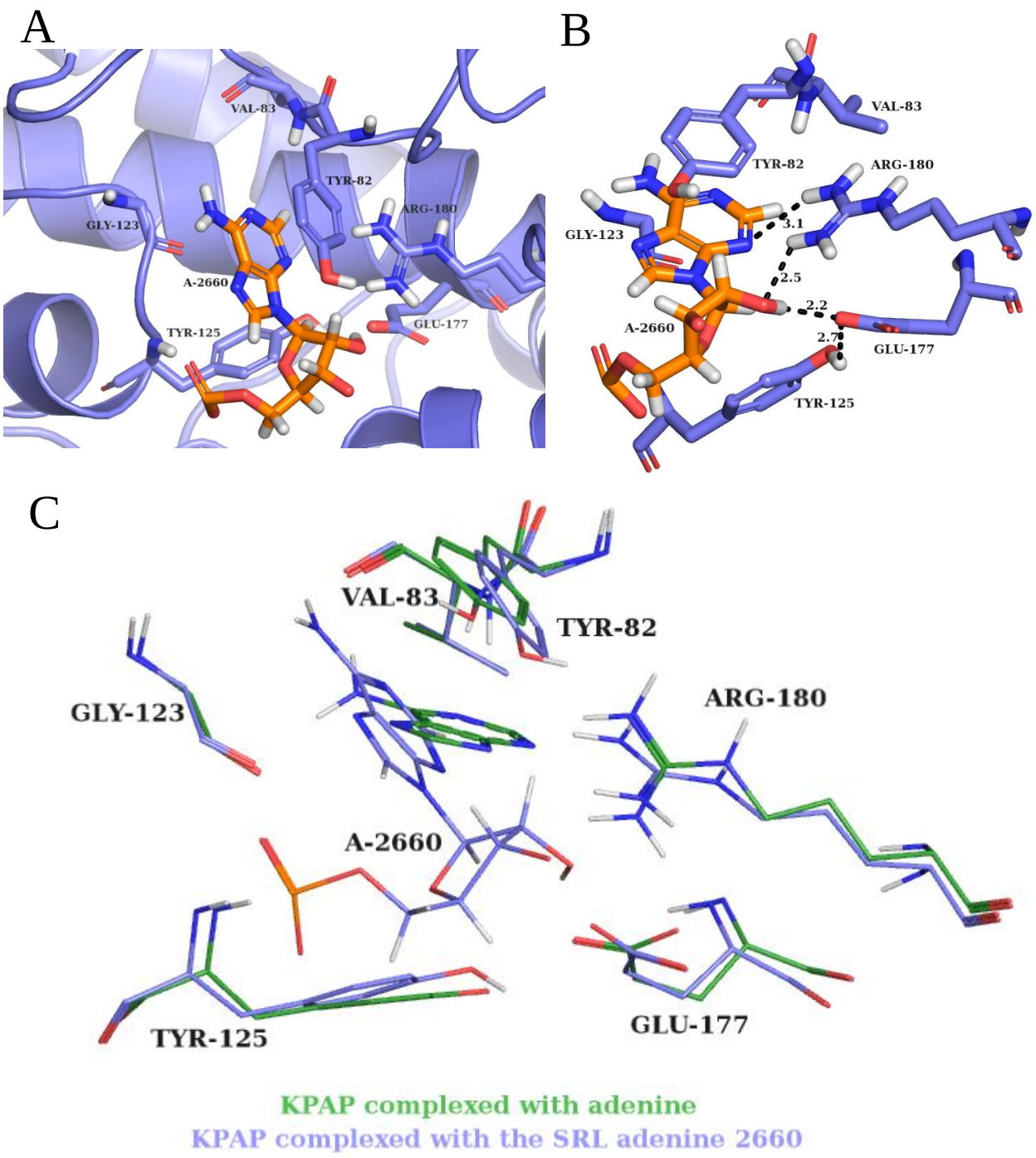
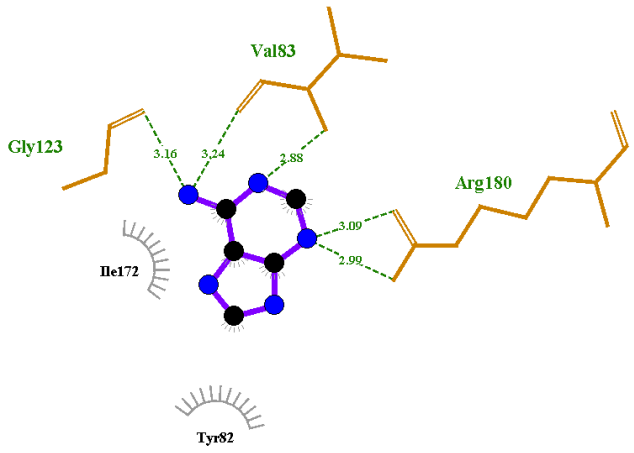
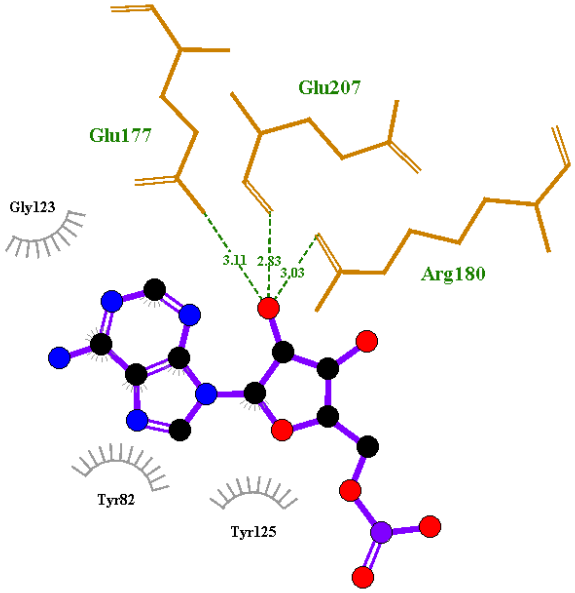


Figure 15: KPAP active site (blue) interaction with A2660 (orange) from the *E.coli* SRL (PDB 480d). (A) The overall placement of the nucleoside in the context of the active site of the protein is observed. (B) The adenine base is most likely interacting with Tyr82 through pi-pi interaction while, the 2'OH of the ribose seems to interact with Glu177 and Arg180 through hydrogen bonding. (C) The overlap comparison of the active site amino acids complexed with the adenine and complexed with SRL adenine 2660 nucleoside. The dashes represent the distance in Ångstroms. Figures generated with PyMol.

A


Val83
Gly123
3.16
3.24
2.88
Arg180
3.09
2.99
Ile172
Tyr82

B


Glu207
Glu177
Gly123
3.11
2.83
3.03
Arg180
Tyr82
Tyr125

C

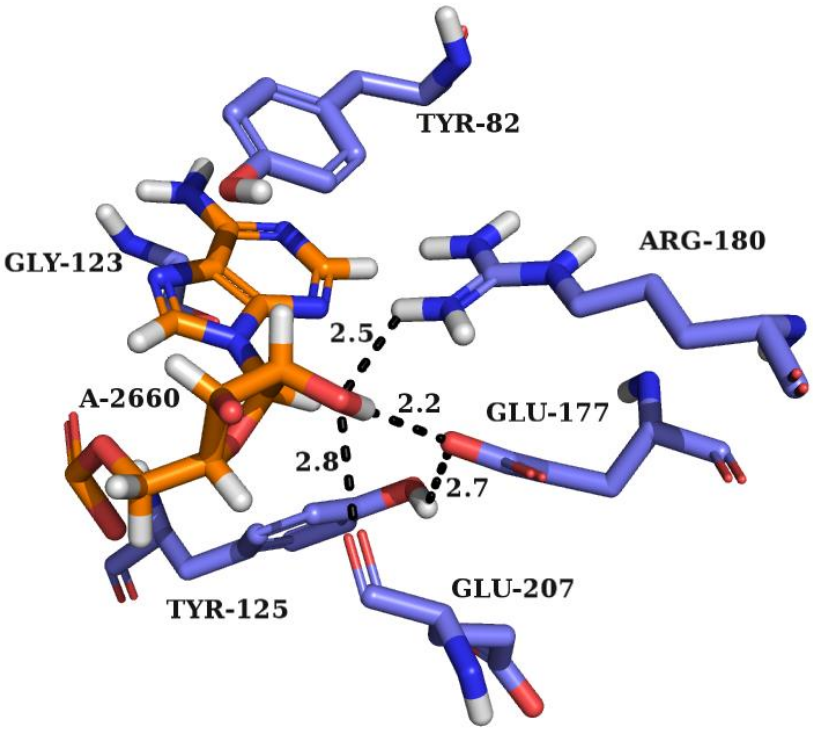

TYR-82
GLY-123
ARG-180
2.5
A-2660
2.2
GLU-177
2.8
2.7
GLU-207
TYR-125

Figure 16: Active site representation of Glu207 in LigPlot+ and PyMol. The simplified 2D representation from LigPlot+ recapitulates the interactions in KPAP docked to an adenine (A) and identifies other amino acids binding the A2660 (B). The LigPlot+ images (A, B) show the KPAP protein amino acids that hydrogen bond in dark yellow with green labels, and the protein's amino acids that interact through hydrophobic interactions are shown in grey with the circular spikes. The ligand is shown in a ball and stick format with purple bonds. The black balls represents carbon atoms, the blue represents the nitrogen, the red represents the oxygen, and the purple represents the phosphorus. In the PyMol picture (C), the KPAP amino acids are labeled blue (like the previous figures) and the A2660 is labeled orange. The black dashes refer to the distance between atoms in Ångstroms. Glu207 appears to hydrogen bond with the 2'OH group from the ribose sugar of the nucleotide in the 2D image (B) with surrounding amino acid Glu177 and Arg180. The same interactions can be observed in PyMol with the addition of Tyr125 which hydrogen bonds with Glu177.

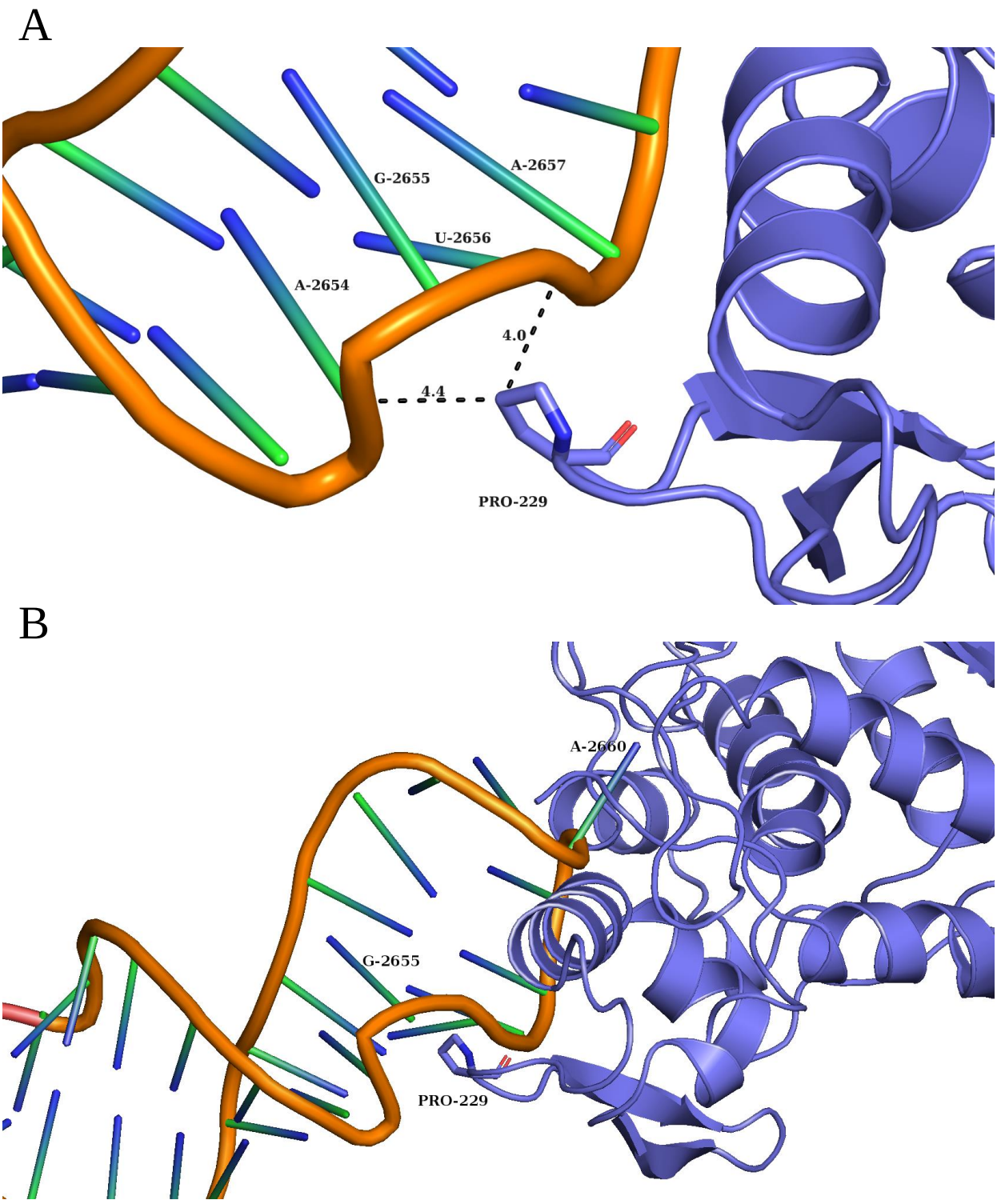


Figure 17: Proline299 from KPAP (blue) surrounded by the RNA backbone (orange) from A2654 to U2656. The first panel (A) shows a close view with distances of the backbone to P229 and the second panel (B) shows the position of P229 in relation to A2660 present in the active site.

3.8 *In vitro* assays comparing PAP-I and KPAP enzymatic activity

The enzymatic characteristics of KPAP were investigated due to the lack of knowledge of the protein. RIPs are generally characterized using *in vitro* assays including an aniline assay and a cell-free translation assay to confirm their ribotoxic activity.

3.8.1 Enzymatic activity on the ribosome

The aniline assay qualitatively shows whether a RIP has depurinated rRNA. The initial depurination of rRNA by the incubated RIP removes a single base leaving the backbone intact. The most well-known target of RIPs on rRNA is the adenine located at the top of the SRL. Incubation with acidic aniline will cleave the phosphodiester bond at the abasic site yielding the Endo fragment (~360nt). **Figure 18A** shows PAP-I and non-functional mutant PAPx (E176V; Wang & Tumer 1999) compared to KPAP and non-functional mutant KPAPx (E177V) with and without aniline treatment. Two fragments are clearly observed around 360nt, one in the PAP-I lane and one in the KPAP lane with aniline treatment and they are not observed in the non-functional mutant lanes treated with aniline. This result indicates that PAP-I and KPAP have enzymatic activity producing the same size fragment. This is a typical characteristic of RIPs showing depurination activity. Potential endonuclease contaminants from protein samples would be observed in the lanes without aniline treatment producing bands that would be absent in the lane containing RNA alone.

3.8.2 Activity on the ribosome inhibits translation

In order to gain more insight on the consequences of the cleaved rRNA fragment observed in the aniline assay, an *in vitro* translation assay was carried out. This assay will measure the inhibitory effect of the RIP activity on translation. A luciferase transcript was added to a commercial rabbit reticulocyte lysate in the presence of different concentrations of RIP (0.4nM and 0.04nM). After the incubation, luciferase assay reagent was added to the lysate and luminescence was measured and normalized to the positive control (no RIP control). **Figure 18B** shows the identity of the proteins added to the translation assay expressed and purified from *E.coli*; and **Figure 18C** shows the translation assay results. The two concentrations selected were taken from the literature (Ready et al. 1983; Poyet & Hoeveler. 1997a). Both PAP-I and KPAP concentrations showed a significant decrease in protein translation (in both concentrations) compared to their respective mutants which can be associated with the active site catalytic

activity. The point of interest is the decreased amount observed in KPAP 0.4nM compared to PAP-I 0.4nM, it is the same. KPAP and PAP-I did not depurinate the transcripts. Prior incubation of the luciferase transcript and the RIPs followed by *in vitro* translation without RIPs showed no significant difference with untreated control in translation (*data not shown*). Therefore, in this assay KPAP inhibits translation to the same extent as PAP-I by depurinating the rRNA. Combined with the aniline assay, I have shown that the incubation of KPAP with intact ribosomes yielded an rRNA fragment comparable to PAP-I indicating similar enzymatic activity and the incubation of KPAP in a cell-free translation system resulted in protein synthesis inhibition also comparable to PAP-I. From these activity assays and previous structural analysis it seems very likely that KPAP is a RIP that targets rRNA with the goal of inhibiting translation. Whether KPAP targets the very same adenine as PAP-I on the SRL remains to be tested, but structural data based on active site similarity suggests that it is likely.

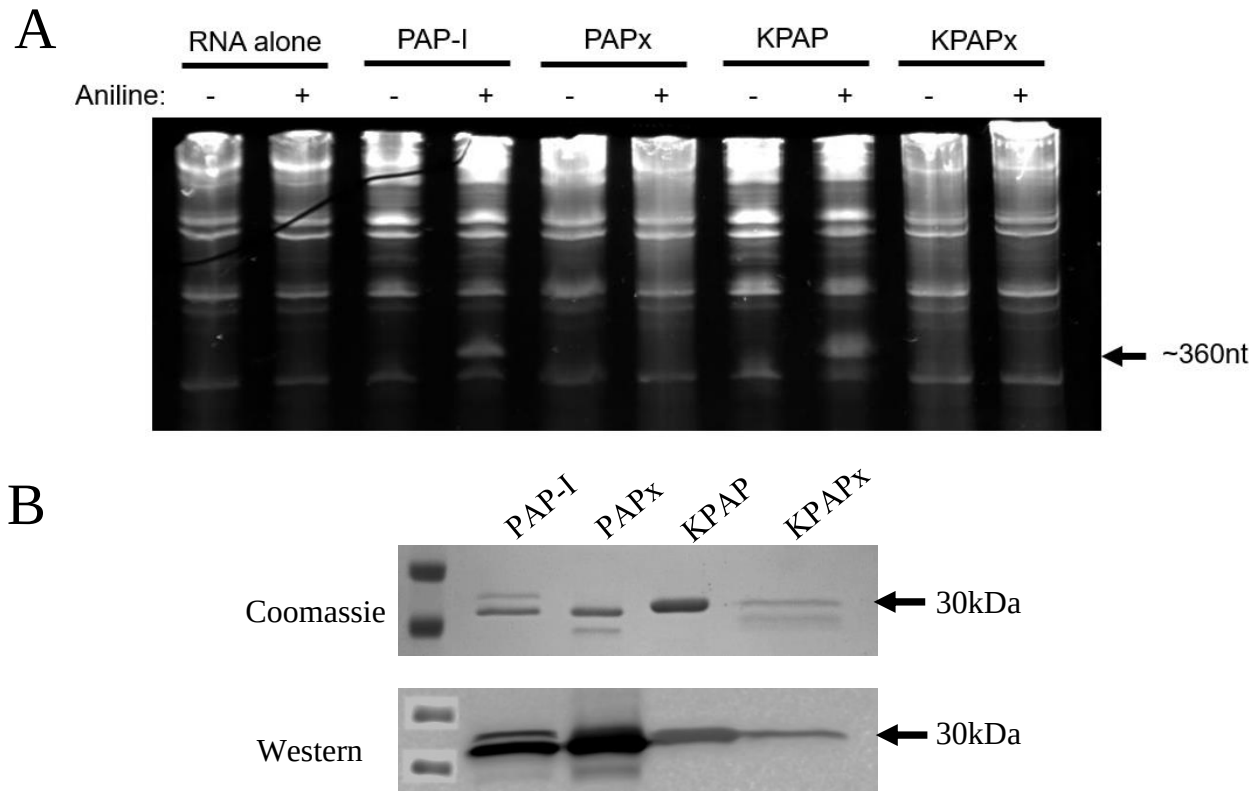

A
RNA alone
PAP-I
PAPx
KPAP
KPAPx
Aniline:
- + - + - + - + - +
~360nt
B
PAP-I
PAPx
KPAP
KPAPx
Coomassie
30kDa
Western
30kDa
C
% Protein Synthesis
100
50
0
****
***
****
*
-RNA
+RNA
PAP-I 0.4nM
PAP-I 0.04nM
PAPx 0.4nM
PAPx 0.04nM
KPAP 0.4nM
KPAP 0.04nM
KPAPx 0.4nM
KPAPx 0.04nM

Figure 18: Enzymatic activity of KPAP compared to PAP-I. (A) The comparison of PAP-I and KPAP with their associated non-functional mutants (PAPx and KPAPx) in an aniline assay on tobacco ribosomes is shown on a 7M urea/6% acrylamide gel. The cleaved rRNA fragment is observed at ~360nt in PAP-I and KPAP treated with aniline while absent in other lanes. (B) The enzymes used in this assay are shown in a Coomassie stained gel and a western blot. (C) The in vitro translation assay showing percent protein synthesis compared to different concentrations of PAP-I or KPAP (or mutants) and their impact on protein synthesis (n=3). Error bars are SD. (*) $p<0.05$, (***) $p<0.001$, (****) $p<0.0001$.

3.9 PAP transcripts in different tissues

The PAP genes remain to be well studied and identifying their presence in various tissues is one step closer to understanding them. An RT-PCR experiment was done on a leaf, stem, root, and flower total RNA and probed for the PAP genes using specific primers. This method allows us to identify the presence of transcripts in the different tissues. Tissues were collected from 4-leaf stage plants except flowers which were collected from blooming mature plants. KPAP mRNA only appears in the leaf fraction (**Figure 19**). PAP-I and PAP-II transcripts are both present in all tested tissues. PAP-α appears to be expressed in the leaf, stem, and root but not in the flower fraction. PAP-S1 transcript appears in the leaf, stem, and flower but not in the root fraction. Overall, KPAP seems to be exclusively expressed in leaves which could probably involve the function of the protein in matters of the leaves (ie. storage or photosynthesis) although this remains to be tested. PAP-α is one of the most obscure isoforms and it is interesting to see that it does not appear in flowers. PAP-S1 is expressed in flowers, and this could potentially be explained by its well-known presence in seeds (hence its name S1- seed 1) and flowers are precursors to seeds. Its absence from roots is also a mystery. This provides new ground on which to investigate the PAP genes.

3.10 Quantitative mRNA analysis of the PAP genes during abiotic stresses

3.10.1 Drought time course

The steady state levels of the PAP transcripts were measured from a dehydration treatment to elucidate gene expression patterns of the PAP isoforms under drought stress. Selected pokeweed plants were watered with 20% PEG-8000 daily and the time course spanned 96h. The plant leaf tissues were harvested, and the extracted RNA was used in RT-qPCR. PAP-I is the most well-known and prevalent PAP gene in pokeweed and its purpose in this assay was to compare its behavior to the other genes. PAP-I displays a significant increase in mRNA levels initially (UT-8h) followed by a reduction to original levels (8-24h) and increase again at 96h (**Figure 20**). In contrast to PAP-I, KPAP shows a different pattern. From the initial time point to 24h and 96h, a significant ~1-fold and ~2-fold decrease is observed respectively. PAP-II also shows a distinct pattern where mRNA levels are relatively stable for the first 24h until 96h where it increases almost 2-fold. In contrast, to PAP-II, PAP-α shows a significant 2-fold increase at 8h followed by a drop to initial levels. Lastly, PAP-S1 does not show any significant changes across the time course. Since all datasets followed a normal distribution, the parametric one-way

ANOVA was used followed by a Tukey multiple comparisons test. Statistically significant comparisons are labeled with a (*) and (**) representing a $p<0.05$ and $p<0.01$ respectively. Looking at these individual pieces of information, the PAP genes behave very differently from each other under drought stress. The genes could potentially be divided into two categories, those that are upregulated (ie. PAP-I, II, α) and those that are downregulated (ie. KPAP) omitting those that are unaffected by the drought stressor (ie. PAP-S1).

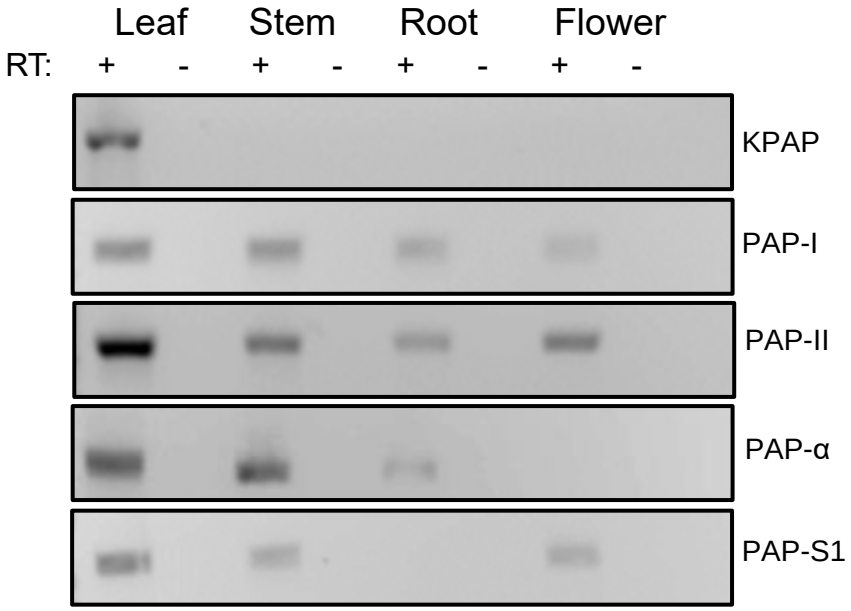


Figure 19: Presence of PAP transcripts in different plant tissues. RT-PCR products imaged on a 1% agarose gel of the PAP genes in leaves, stems, roots, and flowers. KPAP transcripts show presence only in the leaf fraction. PAP-I and PAP-II are present across all tissues. PAP-α is present in the leaf, stem and root fraction. PAP-S1 is present in the leaf, stem, and flower fraction. RT refers to reverse transcription.

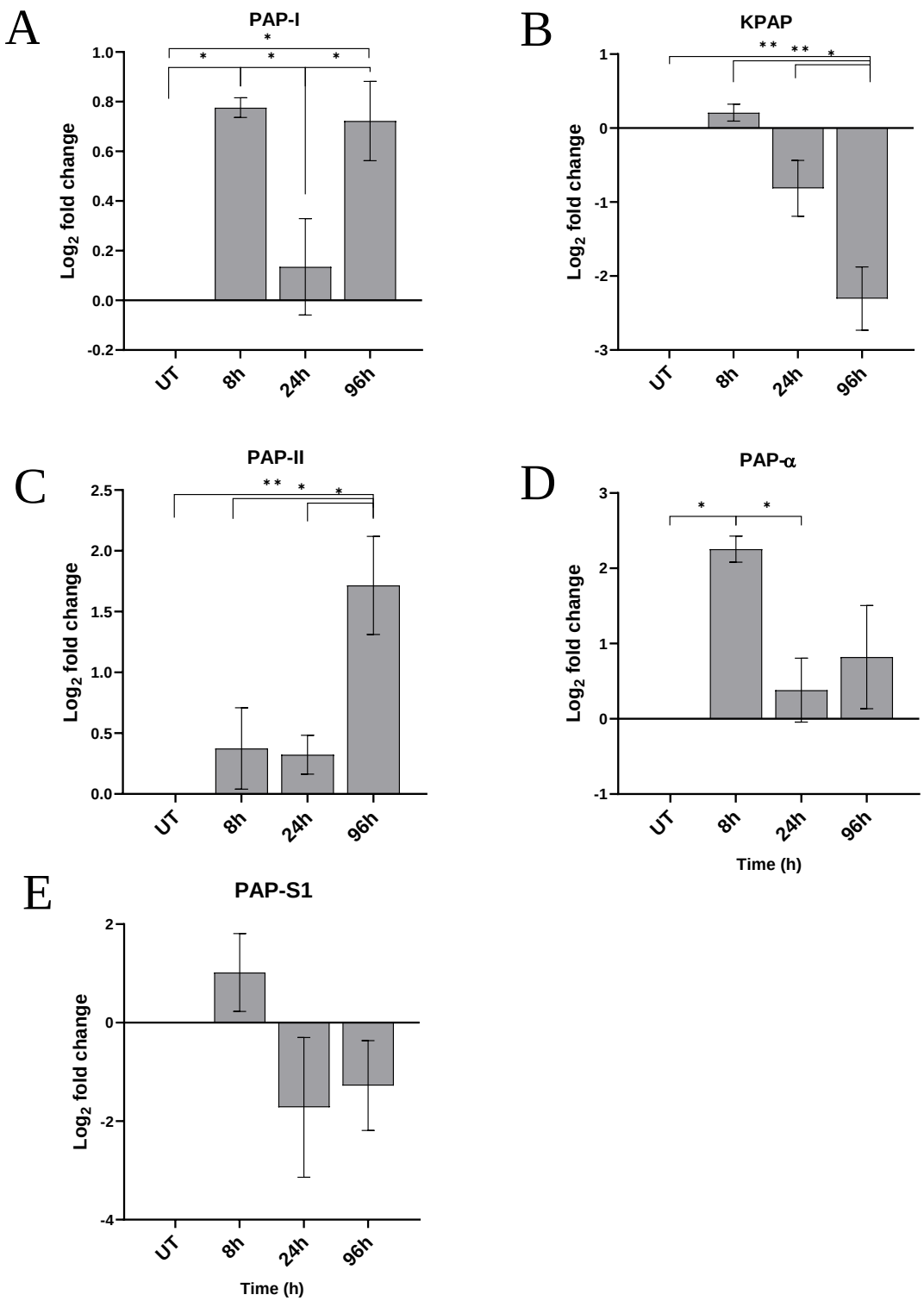


Figure 20: Quantitative analysis of steady-state levels of PAP transcripts in a dehydration treatment (20% PEG) over the course of 96h. Time course of transcript levels shown as $\log_2$ fold change for (A) PAP-I, (B) KPAP, (C) PAP-II, (D) PAP-α, and (E) PAP-S1 (n=3). Tukey's multiple comparison analysis displays (*) when $p<0.05$ and displays (**) when $p<0.01$. The error bars represent the standard error of the mean (SEM). UT: untreated.

3.10.2 Cold time course

Similarly to the drought experiment, steady-state levels of the PAP transcripts were measured from cold (4°C) treated pokeweeds in a time course in order to identify if they were sensitive to cold stress (**Figure 21**). The plant leaf tissues were harvested, and the extracted RNA was used in RT-qPCR. PAP-I mRNA do not significantly change across the time course. KPAP shows a significant change between 8-24h portrayed by a drop of less than 1-fold where 24h is comparable to untreated levels. PAP-II, PAP-α and PAP-S1 show no significant change across the time course. Compared to the drought treatment, the cold treatment does not seem to impact PAP transcript levels with the same intensity and can be suggested that PAP isoforms may not be sensitive to cold stress.

3.10.3 ABA time course

In addition to the previous abiotic stresses, a time course of 0.5mM abscisic acid (ABA) was performed on pokeweed plants to measure the differential expression of the PAP genes. The plant leaf tissues were harvested, and the extracted RNA was used in RT-qPCR. The time course consisted of 5 time points that reached 72h. PAP-I and KPAP (**Figure 22A, B**) mRNA levels did not significantly change during this treatment, however, PAP-II, α, and S1 (**Figure 22C, D, E**) showed an upregulation to various degrees. PAP-II displayed an upregulation from the untreated (UT) and 2h compared to 8h ($p<0.01$ and $p<0.05$ respectively) and additionally the UT and 2h compared to 72h were also significantly upregulated (both $p<0.01$). The 72h time point is only representative of n=2 due to the elimination of an outlier that was 3-$\log_2$ fold change lower than the other two replicates. PAP-α shows two combinations that are upregulated, 2h-24h and 2h-72h (both $p<0.05$). Lastly, the exact same trend is observed in PAP-S1 where 2h-24h and 2h-72h are significantly upregulated (both $p<0.05$). The trend initially starts with a decrease in transcript levels followed by an increase at point 24h and 72h. This initial drop at 2h is also observed in PAP-I and KPAP although it is not significant. Overall, the increase due to ABA reaches almost 2 $\log_2$ fold change in PAP-II and is closer to 1 and 1.5 $\log_2$ fold change in PAP-α and PAP-S1.

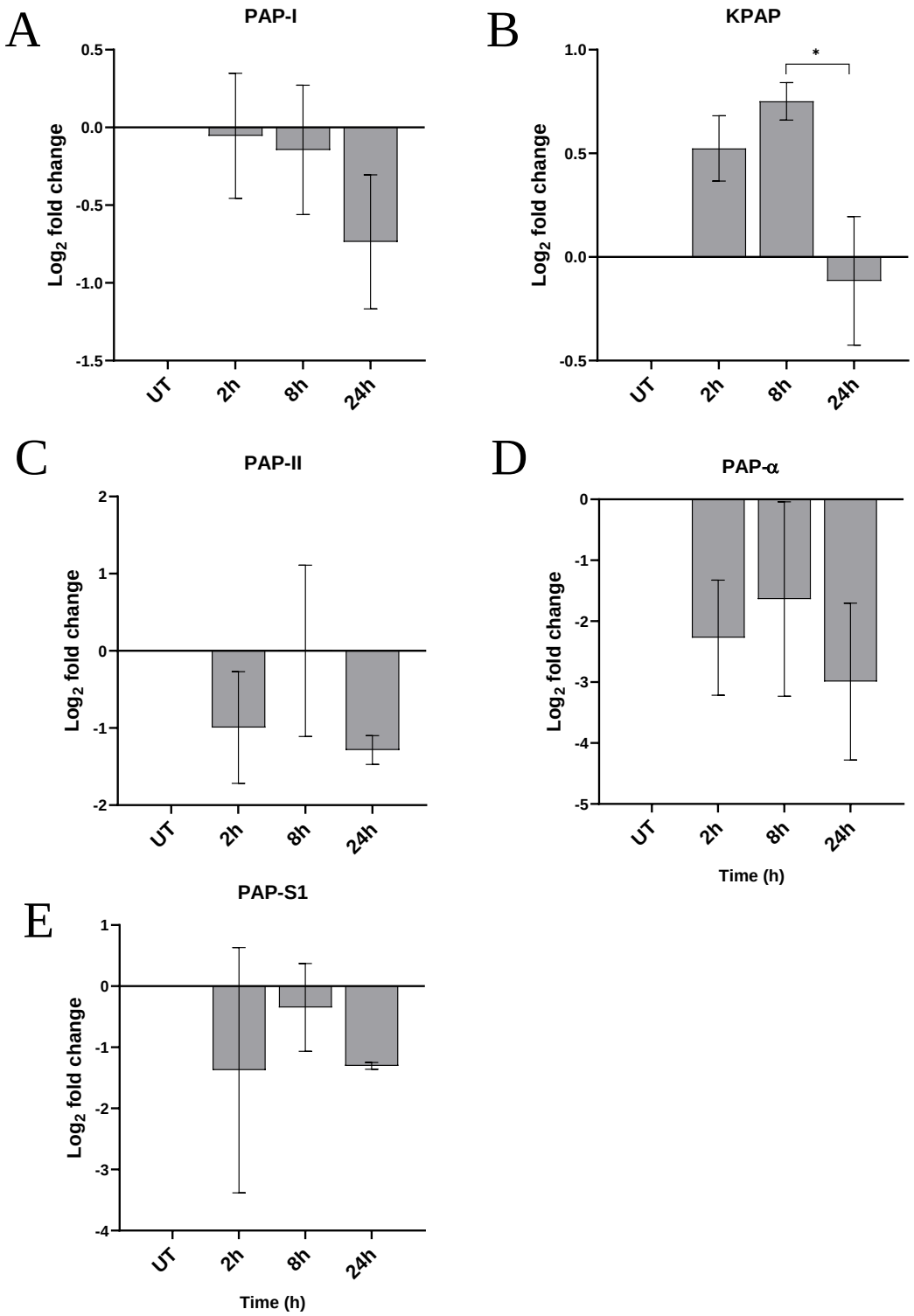


Figure 21: Quantitative analysis of steady-state levels of PAP transcripts in a cold treatment (4°C) over the course of 24h. Time course analyzing the following genes: (A) PAP-I, (B) KPAP, (C) PAP-II, (D) PAP-α, and (E) PAP-S1 (n=3). Tukey's multiple comparison analysis displays (*) when $p<0.05$ only in KPAP. The error bars represent the standard error of the mean (SEM). UT: untreated.

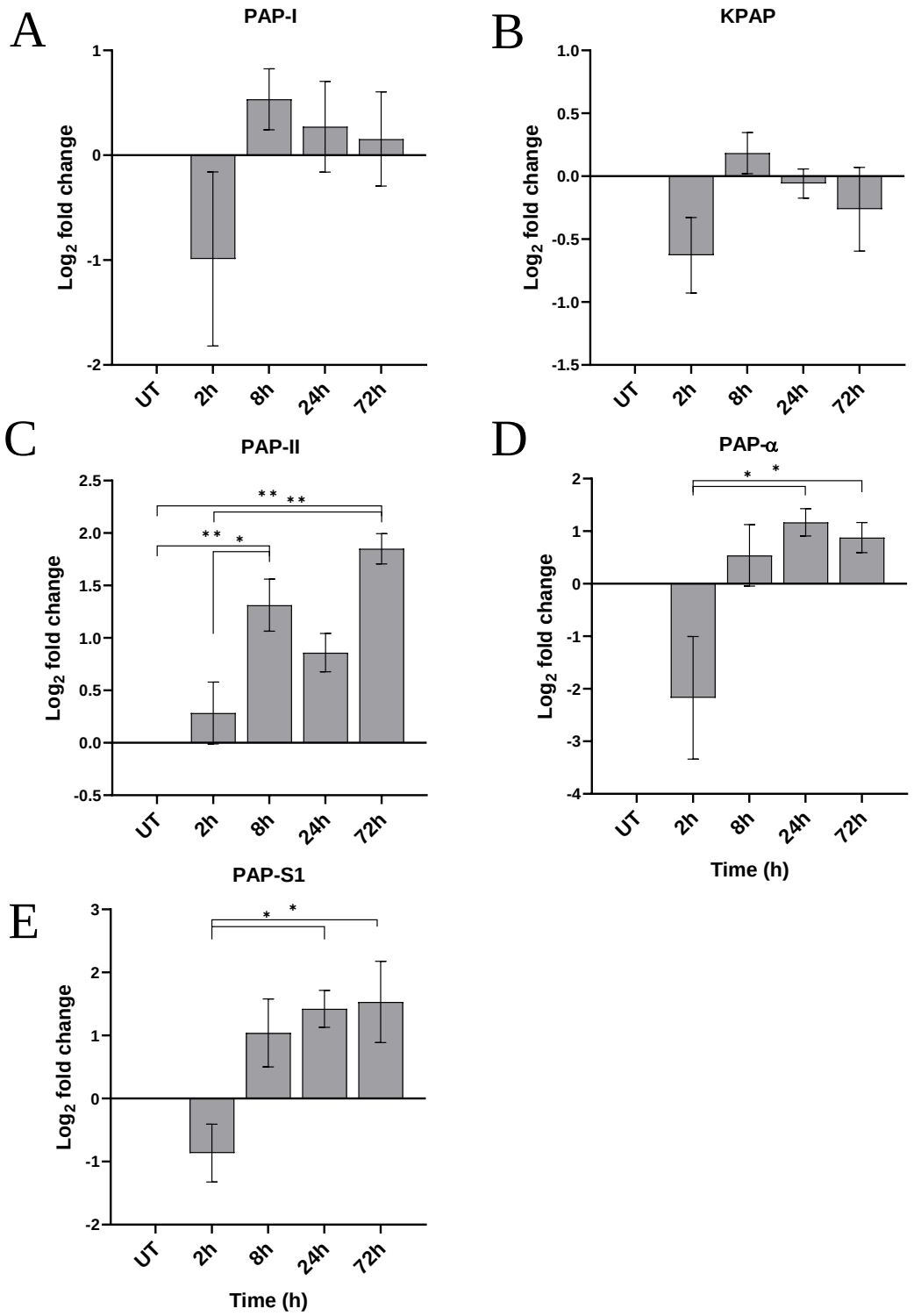


Figure 22: Quantitative analysis of steady-state levels of PAP transcripts in an ABA (0.5mM) time course spanning 72h. The genes involved were (A) PAP-I, (B) KPAP, (C) PAP-II, (D) PAP-α, (E) PAP-S1. Tukey's multiple comparison test identifies (*) as $p<0.05$. Sample size n=3, except PAP-II 72h n=2 (outlier was removed). The error bars represent the standard error of the mean (SEM). UT: untreated.

3.11 PAP transcripts analyzed in developmental stages

The difference in age of the plant was thought to be a factor for differential expression of the PAP genes. A younger plant (4-leaf stage) typically selected for experiments was compared to an older plant (8-leaf stage) with a more mature development. Initial observations of the two plants (**Figure 23A, B**) shows the biggest difference in height and size of the leaves where the 8-leaf stage was ~21cm and has broader, thicker leaves while the 4-leaf stage was ~10cm and has much smaller and thinner leaves. The plant leaf tissues were harvested, and the extracted RNA was used in RT-qPCR. The analysis of steady-state levels of mRNA reveals no significant change in PAP-I (**Figure 23C**), KPAP (**D**), and PAP-II (**E**) whereas PAP-α and PAP-S1 show a significant 3.5 and 4 $\log_2$ fold change respectively ($p<0.01$). This indicates an important change for PAP-α and PAP-S1 and perhaps they play a more important role in an older plant while it seems that PAP-I, KPAP and PAP-II may be more stably expressed across development.

3.12 KPAP promoter and putative abiotic stress cis-regulatory elements (CREs)

To provide additional insight on the potential regulation of KPAP in abiotic stress, a list of cis-regulatory elements was established (**Table 2**). The proximal promoter used was 1.4kb upstream of the transcription start site (TSS). The TSS has been validated through sequencing of the transcript and 5'RACE performed previously by other lab members (data not shown). A TATA box motif was found at -28 consistent with the PAP-I promoter. In addition, a variety of CREs were identified related to abiotic stress: ABRE (ABA), DRE2CORE (drought, ABA), EBOX (cold), GT1CONSENSUS (light inducible), LTRE (cold), MYB (ABA, drought, GA), and more. Approximately 8 CREs overlap in the region -1219: ABREs (x4), DPBFCOREDCDC3 (x1), EBOXBNNAPA (x1), MYB1AT (x1) and ACGTATERD1 (x1). Among these CREs, 7 out of 8 are involved in ABA and drought response. It is possible that the drought response observed in KPAP occurs from that region. GT1CONSENSUS appeared 13 times within the probed region and is related to light induction. Along with the drought responsive elements promoter studies could investigate these elements to confirm their importance in ABA-induced drought and light induction.

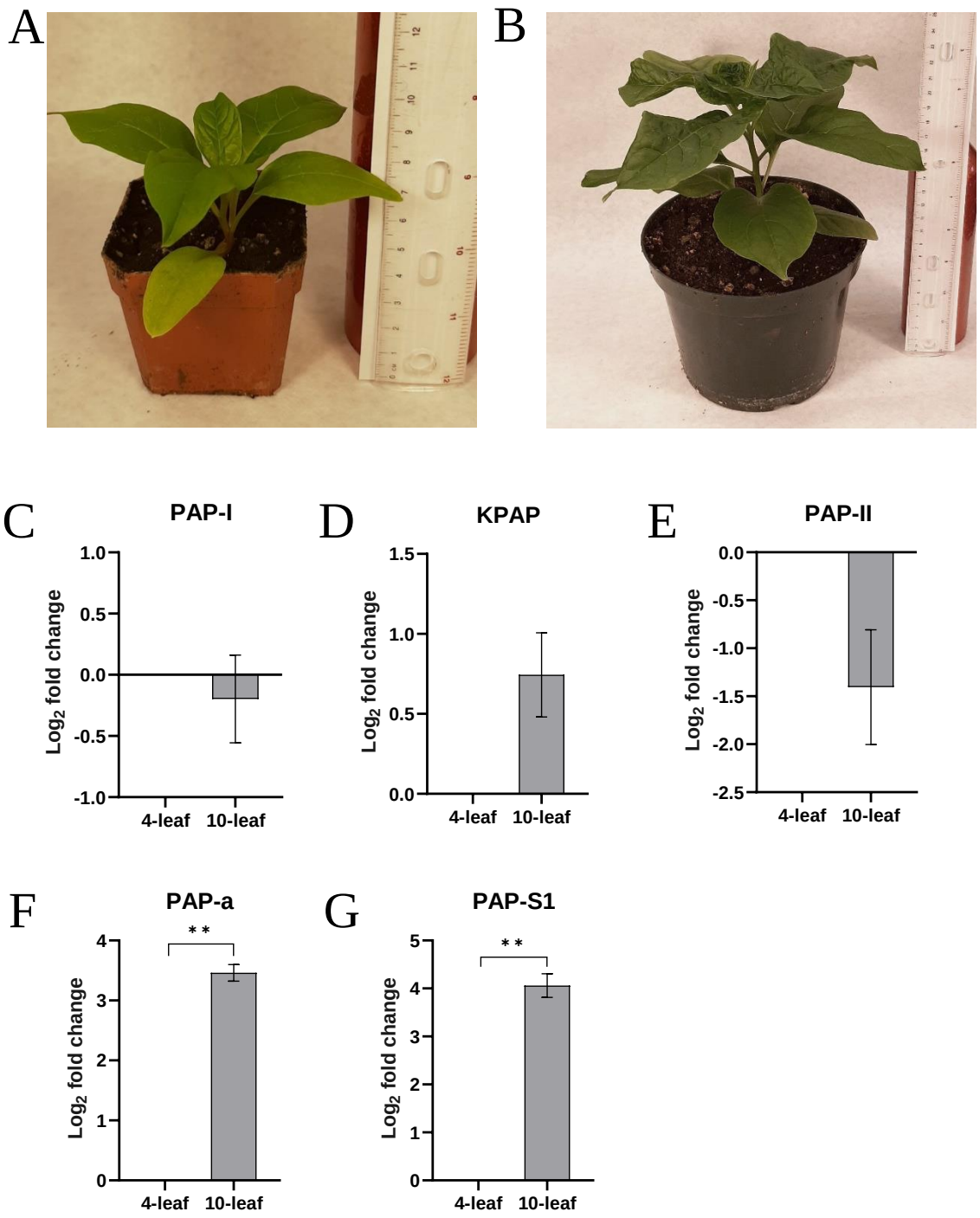


Figure 23: Quantitative analysis of steady state levels of PAP transcripts in a 10-leaf stage pokeweed plant compared to a 4-leaf stage plant. (A) The 4-leaf stage plant shows two sets of young leaves with the cotyledons and has a height of approximately ~10cm with ~4cm above soil. (B) The 10-leaf stage plant shows 4 sets of two mature leaves with a height of approximately ~22cm with ~13cm from the soil. The investigated PAP isoforms were PAP-I (C), KPAP (D), PAP-II (E), PAP-α (F) and PAP-S1 (G). The error bars represent the standard error of the mean (SEM). Sample size n=3. (**) indicates a $p<0.01$.

Table 2: Cis-regulatory elements (CREs) potentially involved in abiotic stress in KPAP promoter

CRE	Motif	Response	Copy number (KPAP promoter)	Reference
ABRELATERD1	ACGTG	ABA; Dehydration	3	Simpson et al. 2003
ABRERATCAL	MACGYGB	ABA; Ca2+	2	Kaplan et al. 2006
ACGTATERD1	ACGT	Dehydration	8	Simpson et al. 2003
ANAERO1CONSENSUS	AAACAAA	Low oxygen	1	Mohanty et al. 2005
CAREOSREP1	CAACTC	GA	1	Sutoh & Yamauchi 2003
DPBFCOREDCDC3	ACACNNG	ABA	2	Kim et al. 1997
DRE2COREZMRAB17	ACCGAC	ABA; Dehydration	1	Kizis & Pages 2002
DRECRTCOREAT	RCCGAC	Cold; Dehydration; High salt	1	Dubouzet et al. 2003
EBOXBNNAPA	CANNTG	Cold	8	Agarwal et al. 2006
GARE1OSREP1	TAACAGA	GA	1	Sutoh & Yamauchi 2003
GAREAT	TAACAAR	GA	1	Ogawa et al. 2003
GT1CONSENSUS	GRWAAW	Light inducible; SA	11	Buchel et al. 1999
GT1GMSCAM4	GAAAAA	Pathogen; high salt	3	Park et al. 2004
LTRE1HVBLT49	CCGAAA	Cold	3	Dunn et al. 1998
LTRECOREATCOR15	CCGAC	Cold; Dehydration	3	Kim et al. 2002
MYB1AT	CANNTG	ABA; Dehydration	2	Abe et al. 2003
MYBCORE	CNGTTR	ABA; Dehydration	2	Abe et al. 2003
MYBGAHV	TAACAAA	GA	1	Gubler et al. 1995
PREATPRODH	ACTCAT	Hypoosmolarity	1	Satoh et al. 2004
PYRIMIDINEBOXOSRAMY1A	CCTTTT	Sugar repression; GA	3	Mena et al. 2002

4 DISCUSSION

4.1 Length of the KPAP intron

KPAP was discovered during the first pokeweed genome assembly (Neller et al. 2019) but the gene model was incomplete. Here, the work presented, elucidates the gene model. The KPAP gene organization follows that of PAP-I consisting of a single catalytic domain that is known to depurinate the α-sarcin/ricin loop (**Figure 5**; Di & Tumer 2015). The KPAP gene also shows an intron flanked by two portions of the 5'UTR and ends with a 3'UTR. Due to the large size of the intron, PCR primers were designed to walk along the sequenced contigs to confirm the sequence. The 5'UTRs were sequenced from cDNA and 750bp upstream of the second 5'UTR had been sequenced by previous students. The second assembly of the pokeweed genome confirmed the long intron. The two contigs from the previous genome connected together with the addition of a 732bp sequence identified with the second sequencing iteration of the genome. The KPAP intron length is now confirmed to be 43,752bp. Chang and colleagues (2017) analyzed the content of *Arabidopsis* genes containing large introns. The largest intron was 57, 631bp (AT5G32690.1-1) and next 4 highest were between 10-15kb. Unfortunately, this gene remains unstudied. According to the *Arabidopsis* database, AT5G32690.1, is considered a pseudogene of AT2G29880 with 58,519bp total length and AT2G29880 is associated to Myb/SANT-like DNA-binding domain protein (https://www.arabidopsis.org). This suggests that it is possible for an intron of this size to exist. Among the PAP isoforms (excluding KPAP) the average length of their intron is ~3.2kb (Neller et al. 2019). Within the KPAP intronic sequence, I have identified several upstream open reading frames (uORFs) coding for relatively large proteins. A threshold was applied to filter only the ones with more than 100 residues, yielding 22 potential proteins. The largest protein was 252aa weighing 29.1kDa, nearly as much as KPAP. Interestingly, a BLAST search revealed that several of these proteins were identified as reverse transcriptases, ribonuclease H, integrase or gag-containing domain. These proteins are common of retrotransposons, group II introns, and endogenous retroviruses (Schafer et al. 2003). I inputted the sequence of AT5G32690.1 into the NCBI ORFfinder identifying all potential peptides starting with ATG and filtered to output only those putative peptides that are larger than 150aa. Results provided 132 ORFs. The top 5 largest peptides are outlined in **Table 3** with their BLAST analysis. Since the annotation of this gene has not been released, the position of the intron

Table 3: uORF annotations from AT5G32690.1 using ORFfinder

Name	Frame	Nucleotide length (bp)	Peptide length (aa)	Peptide Weight (kDa)	BLAST results (top hits)
ORF 1	-3	3729	1242	141.3	RT, integrase
ORF 2	+3	1119	372	42.2	gag
ORF 3	-2	1092	363	41.0	uncharacterized
ORF 4	-1	1023	340	39.0	Putative transposon protein
ORF 5	-2	726	241	27.8	DNaseI, proteoglycan 4-like

remains unknown. Interestingly, 3 out of 5 ORFs show genes like RT, integrase and gag typically belonging to retrotransposons.

Within the retrotransposon family, it seems likely that the KPAP intron sequence contains a long terminal repeats (LTR) retrotransposon rather than a SINE or LINE due to its length and presence of the gag capsid genes (as reviewed by Schulman 2013). At the end of the gag gene, it is not uncommon to observe +/- 1 translational frameshifting (Gao et al. 2003), therefore the uORFs in KPAP can be out of frame and yet still be translated together. Retrotransposons are an important part of plant genomes considering that 80% of the maize genome is composed of them (Feschotte et al. 2002). In wheat, retrotransposons have been shown to alter expression of adjacent genes by either activating or silencing them (Kashkush et al. 2003), however, since they are in high copy number throughout the genome, they have the possibility to affect several different pathways. BARE-1 retrotransposon in barley has a copy number greater than 50,000 (Suoniemi et al. 1996). Even some with small copy number like Tos17 in rice affects 14 functional classes of genes (top 3 classes: cell growth/division, cell structure, disease/defense; Miyao et al. 2003). Whether KPAP is activated or silenced by the retrotransposon-like sequence remains unknown, however, identifying the identity of that retrotransposon may elucidate the potential role it has in KPAP expression.

4.2 KPAP structural characteristics

4.2.1 The active site

The current KPAP residues involved in the active site have been well studied in PAP-I and ricin (Hudak et al. 2004; reviewed in Shi et al. 2016). The well-known catalytic arginine is also observed in KPAP (R180) and is in the vicinity of the essential Y82, E177, G123, Y125 in KPAP (**Figure 9**). The glutamic acid E177 is essential for resolving the ribooxocarbenium

transition state at the end of the depurination reaction (**Figure 2**; Fermani et al. 2009) and Y125 is involved in base stacking through a π-π-interaction with the adenine base (Huang et al. 1995).

Tyr82 which is shown to be involved in π-π-stacking has also been shown in abrin (Y74) to behave like the gate of the active site, where a closed conformation interacts with abrin G111 and the open conformation rotates the tyrosine opening the site (Banisa et al. 2019). This opened and closed behavior is also noted in momorcharin (Huang et al. 1995), in trichosanthin (Gu & Xia 2000), and PAP-I (Kurinov et al. 2008). Considering G123 in KPAP it is likely that Y82 closes the active site using G123. Overall, the active site of KPAP is very similar to other RIPs and the molecular dynamics of their residues could be applied to KPAP.

4.2.2 The secondary active site

The additional glutamic acid (E207) residue found near the primary E177 in KPAP (**Figure 10**) has been studied in trichosanthin (Wong et al. 1994). In trichosanthin, the primary E160 is responsible to resolve the oxocarbenium transition state intermediate that occurs at the end of the depurination reaction. In ricin A-chain (RTA), the mutation E177A, corresponding to trichosanthin E160 found that glutamic acid essential for RTA catalysis (Schlossman et al. 1989). The secondary E189 in trichosanthin has been attributed a backup function in case E160 was mutated. When the mutation E160A is compared to E160D with intact E189, E160A was more cytotoxic. It was hypothesized that E189 can move to replace the role of E160, when the smaller alanine amino acid mutation would allow the rescue of the depurination activity by E189 compared to the bulkier aspartic acid which would prevent the access to the active site. The corresponding alanine mutation in RTA has also been found to make use of the secondary glutamic acid to stabilize the transition-state ribooxocarbenium ion (Kim et al. 1992). The activity of trichosanthin reduced 1800-fold when the mutation E160A and E189A were introduced which indicate that the residues participate directly in the catalysis reaction (Wong et al. 1994). Interestingly the class III maize RIP only contains the primary glutamic acid and lacks the secondary (Mak et al. 2007). Due to the smaller active site it is not possible to accommodate two glutamic acid residues and it is thought that this RIP may represent an intermediate state in the evolution of RIPs. I would like to postulate an additional function for the secondary E207 in KPAP.

4.2.3 RIP interaction with the upstream nucleotide of A2660 from the SRL

An amino acid from other RIPs have been found to interact with the upstream guanine base adjacent to the generally targeted adenine from the SRL through an asparagine. Fermani and colleagues (2009) have identified that this residue is absent in bouganin but other groups have confirmed the interaction in ricin through N78 (Olson & Cuff 1999) and in PAP-I through N70 (Rajamohan et al. 2001). KPAP N80 is 4.6Å away from 2'OH of G2661 (data not shown). At this distance it is unlikely to form a hydrogen bond, however, proteins and ligands rarely remain static and perhaps the dynamic nature of proteins would allow closer proximity to hydrogen bond. For bouganin the lower efficiency in protein synthesis inhibition is thought to be due to the lack of an asparagine residue (Asp68 instead; Fermani et al. 2009). If this hypothesis were to be true for bouganin and KPAP N80 were to hydrogen bond with G2661 2'OH, then perhaps this difference would make KPAP more efficient at inhibiting protein synthesis resembling more PAP-I and ricin A-chain.

Another interesting feature of the docked SRL is the hydrogen bond between G211 and the phosphate backbone of the base G2661 that are 2.4Å apart (*data not shown*). This amino acid has been mutated in RTA (G212E) and showed a 20-fold decrease in depurination activity on yeast ribosomes and 5-fold decrease activity on total RNA (Li et al. 2013). Additionally, G212E mutant dissociated faster from the ribosome than wild type. Considering these findings, G212E mutant was less cytotoxic than the wild type RTA suggesting that the interaction of G212 with the SRL backbone is important for depurination activity for RTA on yeast ribosomes. Similarly with KPAP, G211 interacts with the SRL backbone and would presumably carry a consequent role for the cytotoxicity of KPAP.

4.2.4 Disulfide bonds in RIPs

From the crystal structure of PAP-I (1qci), two disulfide bonds can be observed one at C34-C259 and C85-C106 and across the other PAP isoforms. On KPAP however, only one is observed C44-C224 and it overlaps with PAP-I C34-C259 on the crystal structure. While type II RIPs have domains held together by a single disulfide bond and several other disulfide bond in the lectin domain (Lord et al. 1994), the catalytic domain from type I RIPs also contain disulfide bonds. The two disulfide bonds observed in PAP-I were thought to confer thermal stability (Kung et al. 1990). However, the carboxy terminal of PAP-I (residues 244-259) is homologous

to the consensus sequence of prokaryotic lipoprotein attachment site (Hur et al. 1995; as reviewed in Tumer et al. 2000) and is cleaved by a signal peptidase that cleaves after the cysteine residue (Hayashi & Wu 1990). Despite that these membrane attachment sites are not processed the same way in eukaryotes, it is possible that PAP-I relies on its C-terminus to be secreted. The homologous consensus sequence is different in KPAP and may behave in a different way.

While other RIPs like the low molecular weight RIP luffin P1, has two disulfide bonds (Ng et al. 2011), RIPs like bouganin only have one similarly to KPAP (**Figure 13**). From a general standpoint nascent peptides from the endoplasmic reticulum (ER) have their disulfide bridges formed by protein disulfide isomerase (PDI) very early on once the peptide has been translated which helps in the folding of the protein (Depuydt et al. 2011). Perhaps it is possible that the remaining disulfide bond is necessary for the accurate folding of the protein, while the loss of the other one is compensated by hydrophobic interactions to maintain the proper folding.

Alternatively, a type III RIP in corn has been observed to possess two cysteines residues (C51 and C206) which were too far apart to create a disulfide bond (Mak et al. 2007). One characteristic of this RIP is the relatively small active site compared to other RIPs. Perhaps there is a correlation between having no disulfide bond and a small active site, however, this remains unstudied. It seems that the number of disulfide bonds vary across RIPs and therefore, that variability probably entails that disulfide bonds may not be essential to the function of RIPs since they are not well conserved.

4.3 The functional activity of KPAP compared to other RIPs

The ability of KPAP to depurinate the ribosome and inhibit protein translation is comparable to PAP-I (**Figure 18**). In the aniline assay, KPAP produces the same fragment size as PAP-I on tobacco ribosomes. A similar product of 367nt was observed with PAP-I on the 26S yeast ribosomes (Tumer et al. 1997), indicating similar enzymatic activity between KPAP and PAP-I. In the *in vitro* translation inhibition assay (modified to use luciferase instead of radioactivity; Bai et al. 2009), 0.4nM of KPAP was necessary to reduce protein synthesis at 50% in a rabbit reticulocyte lysate which was similarly observed in PAP-I. This result reproduces the findings of Poyet and colleagues (1997a), where 0.4nM PAP-I was observed to inhibit translation to ~50% also in a rabbit reticulocyte lysate. In a wheat germ lysate, PAP-I reduced protein synthesis to 75% (Bonness et al. 1994). The dataset of the wheat germ experiment was scattered, and the

hypothetical curve of best fit was grossly aligned to the datapoints. However, considering the lesser effectiveness of PAP-I on a wheat germ lysate, we could hypothesize that PAP-I has a lesser affinity on plant ribosomes. This behavior has also been observed in RTA comparing rat and wheat germ ribosomes (Cawley et al. 1979). Instead, just like RTA, PAP-I seems more effective on mammalian ribosomes, and this aligns well with the defense related characteristics of PAP-I, that is, to protect the plant against herbivores. Moreover, KPAP could behave similarly, although that remains to be tested.

4.4 PAP isoforms expression in different tissues

KPAP has been identified only in leaf tissue compared to PAP-I which was identified in all the tested tissues (ie leaf, stem, root, and flower). Additionally, other PAP isoforms were probed for their expression in different tissues. The most interesting result comes from KPAP which is only expressed in leaves displaying tissue-specific expression. A similar occurrence was noted in *Oryza sativa* where 7 RIP genes among 20 tested were expressed at least 2-fold more in young leaf tissue compared to other tissues (OSRIP1, 24, 25, 26, 27, 29, 30; Jiang et al. 2008). Interestingly 15 genes of the 20 were expressed in single tissue, indicating a tissue-specific function for these RIPs. A similar expression is noted in *Saponaria officinalis* where RIPs were categorized based on tissue. SO-L1, 2, 3 are from leaves; SO-R1, 2, 3 are from roots; SO-S5, 6, 8, 9 are from seeds (Ferreras et al. 1993). A review outlines the isolation of RIPs from specific tissue from plants, algae, fungi and bacteria (Girbes et al. 2004), but the field lacks a detailed analysis of the distribution of these in RIPs in various specialized tissues of the organisms. The knowledge of tissue-specific RIPs could help to elucidate the potential function of these proteins. Another way to contribute to the understanding of the function of a RIP comes from observing expression under various stresses.

4.5 Comparison of KPAP expression in drought with previous data

The aim of this study was to develop a time course to analyze the expression changes of the PAP isoforms in drought conditions. The dehydration experiment used PEG-8000 over 8h, 24h and 96h. Neller et al. (2019) investigated a similar treatment (**Figure 20**). Comparing the 24h time points, Neller et al. (2019) and this study show some variation (**Table 4**). While PAP-I, KPAP and PAP-S1 seem to show the least separation, PAP-II and -α show the largest. A possible reason for the discrepancies might be the unknown PEG molecular weight used in Neller, which

PAP-II and -α might be most sensitive to. Moreover, this study used 20% PEG while Neller used 10% PEG. On the other hand, Neller used RNA seq to obtain these data and validated them with qPCR over 5 biological replicates. This study did not use RNA seq and kept only 3 biological replicates which indicates that this study potentially has more variation. To reduce the variation more replicates could have been used to truly narrow the values and validation through RNA seq could strengthen the data.

Table 4: Comparison of differentially expressed PAP genes in drought compared to Neller et al. (2019)

PAP isoforms	Differentially expressed genes in this study at 24h (log_2 fold change)	Differentially expressed genes in Neller et al. 2019 at 24h (log_2 fold change)
PAP-I	0.16	0.3
KPAP	-0.72	-0.72
PAP-II	0.34	-1.22
PAP-α	0.50	-1.20
PAP-S1	-0.34	-0.65

4.6 Upregulated PAP isoforms in PEG treatment

Upregulation of some PAP isoforms in drought have been observed: the significant log_2 fold change of PAP-II at 96h (1.82) and PAP-α at 8h (2.28). In rice, OSRIP1, 2, and 26 were induced by 30% PEG up to 3 log_2 fold change within 3h (Jiang et al. 2008) while OSRIP18 was also found to increase drought tolerance under 30%PEG for 1h (Jiang et al. 2012). In two Malaysian rice cultivar MR220 and MR211, a RIP (Os03t0688000-00) was also upregulated under 6% PEG for 24h (Othman et al. 2019). Another group studying *Bougainvillea glabra*, identified most RIP isoforms were expressed in summer and autumn leaves and they suggested it was due to typical conditions of the seasons such as elevated heat and drought (Lin et al. 2021). It seems likely that the upregulation observed in PAP-II and PAP-α is due to drought sensitivity in the same purpose as other RIPs, however, their role in drought remains to be characterized.

4.7 Downregulated KPAP in PEG treatment

It seems plausible that a gene such as KPAP decreasing in expression under drought (at 96h -2.17 log_2 fold change) stress may be related to functions in the plant that are inhibited. In the large RIP study in rice investigating drought, no RIP was shown to decrease with drought (Jiang et al. 2008). Either they were upregulated or unaffected. In Neller et al. (2019), it is also interesting to see that KPAP is only showing downregulation similarly to the data presented here.

Additionally, the abundance of transcript was also measured and appeared to be the third most expressed isoform. Perhaps is it responsible for a role in the cell that requires constitutive expression exclusively in leaves. Drought stress affects the normal pathways in plants (Basu et al. 2016). Photosynthesis, for instance, is reduced in drought stress through the closing of stomata thereby limiting the intake of CO_2 (Tezara et al. 1999). Metabolism also reduces in periods of drought stress due to the changes in the photosynthetic carbon metabolism (Lawlor & Cornic 2002). Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is essential for the biochemical efficiency of photosynthesis and has been shown to be inhibited during drought stress (Parry et al. 2002). In another pathway, it has been observed that drought, reduces the levels of phytohormone gibberellic acid (GA) which correlated with reduced plant growth (Wang et al. 2008). Additionally, increased levels of ethylene during drought inhibits shoot and leaf expansion (Munne-Bosch & Alegre 2004). Similarly to ethylene, fascilin-like arabinogalactan proteins (FLAs) involved in cell expansion during salt stress were shown to be down-regulated in 10% PEG, indicative of sensitivity to osmotic change (Neller et al. 2019). It is possible that KPAP somehow relates to these pathways since they correlate with KPAP's downregulation. Additionally, CREs found in the KPAP promoter related to light induction further supports this hypothesis (see **Table 2**). However, RIPs have been traditionally reported to promote defense related mechanism. KPAP behaves like a RIP according to its enzymatic activity, but perhaps it plays a secondary role either in photosynthesis, carbon metabolism, plant growth, or leaf expansion.

4.8 Crosstalk between JA and ABA in drought

As indicated in Neller (2019) PAP isoforms are mainly upregulated by jasmonic acid (JA) and have thus far been attributed to defense against biotic stressors, however, a crosstalk exists between JA and abiotic stress. A report indicated that a JA-Ile analog improved tobacco growth under PEG-induced stress (Xu et al. 2020). In addition, Yosefi and colleagues (2020) conducted a similar investigation in strawberry where 0.05mM JA alleviated the symptoms of PEG-induced dehydration. This evidence confirms the idea of a crosstalk between the JA-signaling pathway and abiotic stress like drought. Another group furthered this study by investigating the jasmonate zim domain 5 (JAZ5), a transcriptional repressor of the JA-pathway, in rapeseed (Cao et al. 2022). Results indicated that JAZ5 regulated the drought response in an ABA-dependent manner affecting stomatal density. An overexpressing JAZ5 transgenic plant had a survival rate nearly

halved compared to wild type in drought stress which was rescued by MeJA treatment (Cao et al. 2022). The recent studies describing crosstalk between JA and ABA-signaling pathway may pave the way for understanding the implication of JA-induced PAP isoforms in drought related stress. When looking at the JA data presented in Neller (2019), PAP-I, PAP-II, and PAP-α are upregulated. However, with the ABA data presented here, it becomes apparent that PAP-II has the most potential in the crosstalk since it is also upregulated in ABA. The role of RIPs in drought has not been investigated. However, it was hypothesized that they play a role in protecting the plant against pathogens that entered the leaves through the stomata (Jiang et al. 2008). It was also suggested that jasmonates in abiotic stress regulate the antioxidant defense system (Per et al. 2018). This is sensible since, upon stomatal closure, photosynthesis can no longer convert the high energy UV radiation thereby accumulating reactive oxygen species (ROS; Agurla et al. 2018). Some RIPs have shown superoxide dismutase (SOD) activity (Li et al. 1997; Sharma et al. 2004; Iglesias 2015; Zhu et al. 2016). SODs constitute the first line of defense for the mitigation of ROS produced in moments of abiotic stress by converting $O_2^{\bullet-}$ to H_2O_2 (Saibi & Brini 2018). Whether PAP-II is involved in pathogen defense or has SOD activity when stomata close remains to be studied.

4.9 PAP isoform expression in cold

The PAP isoforms were mostly unresponsive to cold treatment across the time course, only KPAP shows a weak upregulation at 8h followed by a drop of expression at 24h (**Figure 21**). In rice OSRIP1 and 7 were downreguated in cold stress while the remaining 28 RIPs remained unchanged (Jiang et al. 2008). Curcin 2 from *Jatropha curcas* was upregulated by 4°C on a northern blot (Qin et al. 2005). Another group investigated curcin-L promoter in transgenic tobacco and showed to be sensitive to cold temperature through the MYB binding site (YAACKG; Qin et al. 2009). Stomatal closure also occurs in cold stress. There are two factors that trigger stomatal closure: water loss and reduced water intake (Agurla et al. 2018). In cold stress, roots show reduced permeability to absorb water, therefore, to maintain the same water levels as well as the same respiration rate, guard cells need to close the stomata. RIPs have not been extensively tested in cold conditions, yet some, mentioned here, show upregulation (Qin et al. 2005, 2009). It is possible that the upregulation observed in KPAP is due to the variability of the qPCR assay previously described.

4.10 ABA-responsive PAP isoforms

Amongst PAP isoforms, PAP-I and KPAP were unresponsive to ABA treatment (**Figure 22**). PAP-II showed a 2-fold increase at 72h and PAP-α and S-1 showed an initial reduced expression at 2h followed by an increase to 1-fold and 2-fold for PAP-α and PAP-S1 respectively. A RIP from bitter melon (*Momordica charantia*) leaves is reported to be induced by ABA after 24h (Xu et al. 2007). Similarly, a RIP from *Phytolacca insularis,* PIP2, showed significantly increased expression after 24h when induced by ABA (Song et al. 2000). Another type I RIP, curcin-L, from *Jatropha curcas* displayed elevated expression in leaves (but not in stem, roots, or seeds) when treated with ABA (Qin et al. 2009). Lastly, saporin-L3 displayed about a 2-fold increase in expression 30min post ABA treatment correlated with stomatal closure while saporin-S6 displayed fluctuation upon ABA treatment reaching around 1.5-fold increase (Tartarini et al. 2010). When the treatment was extended to 12h, saporin-L3 increased 4-fold while saporin-S6 increased 5-fold. The literature indicates a variety of responses of RIPs to ABA. Previously described OSRIP1 (Jiang et al. 2008) displayed expression exclusively in young leaves. Upon ABA treatment, OSRIP1 did not show a significant expression change compared to wild-type (Wytynck et al. 2021) similar to KPAP. Unfortunately, the studies do not investigate further the reasons for RIPs to be involved in ABA responses but perhaps they could be related to other purposes in the cell. Our lab investigated the PAP isoforms in the first iteration of the pokeweed genome (Neller et al. 2019). PAP-I, and S1 were enriched in the JA/defense GO terms, while PAP-II and α were enriched in the proline/ornithine metabolic process, and putrescine biosynthetic process from arginine GO terms. Unfortunately, KPAP was not described, however, this evidence suggests that PAP isoforms may contribute to more widespread response in pokeweed. Perhaps the evidence presented here will contribute to elucidating the other functions of the PAP isoforms. In the meantime, more studies are necessary to complete the profile of RIPs beyond their stress response characteristics.

4.11 PAP isoforms under developmental stages

The tissue analysis of the PAP isoforms confirms the presence of all tested isoforms in a 4-leaf stage pokeweed plant (**Figure 19**). From the RT-qPCR comparing 4-leaf (~4 weeks) and 10-leaf stage (~8 weeks), PAP-α and PAP-S1 showed a significant increase ~3.5 and ~4 $\log_2$ fold change respectively whereas the other isoforms remained unchanged (**Figure 23**). A similar behavior has been noted in a rice study, where OSRIP 13, 14, 16 were upregulated in the mature

panicle (Jiang et al. 2008). The panicle refers to a loosely branched inflorescence bearing flowers. The only RIP upregulated in mature leaves compared to young leaves was OSRIP 15, in addition to being up regulated in mature panicle as well. In mature roots, OSRIP 12 was also shown to be upregulated (Jiang et al. 2008). The RIPs expression patterns seem to be tissue- and time-dependent in rice. In *Phytolacca dioica*, PD-L1-4 are four RIPs isolated from summer leaves in addition of diocin 1 and 2 isolated from young leaves (Parente et al. 2008). PD-L3 and L4 almost disappear in winter while they remain abundant in spring, summer, and decrease in autumn. PD-L1 is most abundant in winter and PD-L2 expression remains constant throughout the year. While seasonal expression may be related to temperature and related to stress, it also alludes to the age of the plant where spring plants are younger and autumn plants are older. In addition, diocin-2 is not seasonally or age-regulated. Interestingly, Parente and colleagues (2008) found that diocin 2 and PAP-II share 84.6% similarity in amino acid sequence. However, the behavior of PAP-II and diocin 2 do not correspond. While diocin 2 levels are low in a 10-day plant, they increase sharply in a 60-day plant. Here PAP-II seems to maintain its abundance from a 4-leaf to a 10-leaf stage plant. However, PAP-II has been isolated from mid-summer leaves (Barbieri et al. 1993) and has shown to be upregulated most between June and July (Poyet & Hoeveler 1997b), which approximates to 60 and 90 days. It remains possible that PAP-II is similarly upregulated like diocin 2 in older plants. PAP-α and PAP-S1 are briefly mentioned to be similar to PD-L4 and their upregulation in early stages of the plant is thought to protect new and vulnerable tissue from pathogens (Parente et al. 2008). Overall, transcript levels of PAP isoforms are different and may serve developmentally-related functions.

4.12 Conclusion and future initiatives

In this study, a broad investigation was undertaken to characterize the gene model, observe the expression in different tissues and abiotic stresses, predict the structure and simulate the binding of ligands and measure the enzymatic activity of KPAP. The gene model validated through our second iteration of the pokeweed genome (Dougherty & Hudak 2022 *unpublished*) showed a long intron thought to encode retrotransposon-like proteins, a first in PAP isoforms. Expression of PAP isoforms was tested in tissues and abiotic stress (drought, cold, ABA) where KPAP exclusively appeared in leaf tissue and was downregulated in PEG-treatment. PAP-II expression has also shown potential to function in abiotic stress. Despite its enzymatic activity as a RIP, KPAP may be involved in other pathways since its expression during phytohormones

treatments (JA, SA in Neller et al. 2019; and ABA) and stresses (wounding, drought in Neller et al. 2019; drought and cold) was atypical. The predicted structure of KPAP allowed to confirm a similar shape between KPAP and PAP-I despite low sequence similarity and its simulated docking on the SRL has allowed to investigate the potential role of other amino acids in RNA binding. Finally, KPAP's depurinating activity and ribosome inhibiting characteristics were also confirmed.

The investigated abiotic stresses showed different profiles across the PAP isoforms. So far PAP-I has been investigated the most, but we lack understanding towards the PAP isoforms. Investigating other abiotic stresses such as cadmium accumulation, silver nitrate toxicity, hypoxia, sustained darkness, salt, freezing or emulating stress with phytohormones such as gibberellic acid, or ethylene would complement our current knowledge. The investigation of KPAP would also benefit from a promoter study to confirm CREs involved in abiotic stress. The KPAP promoter sequence containing mutated CREs upstream of a GUS reporter gene could be agroinfiltrated in tobacco under abiotic stress and leaf discs turning blue in GUS substrate would indicate upregulation. In addition, using a fluorometric GUS assay would provide quantitative data. Moreover, combining RT-qPCR and RNAseq in short and long time courses would allow us to see how plants adapt to stress from several hours to several weeks. The use of high throughput techniques like RNAseq is more common which means a lot of data are available. For example, BioProject PRJNA649785 is the raw unassembled RNAseq data for Phytolacca americana in response to cadmium stress. Lastly, a hurdle in studying RIPs is the lack of antibodies for the isoforms. Mono- and poly-clonal antibodies could be used in various ways such as immunoprecipitation or immunofluorescence. In relation to KPAP, a lot remains to be known: its localization in the cell and leaf, regulation in more mature tissue, and its distinct contribution to a secondary activity beyond that of a RIP.

From the classical defense perspectives, RIPs are known to be induced by various stresses (Zhu et al. 2018), however, there is increased evidence that RIPs perform secondary functions in the cell (Helmy et al. 1999; Iglesias et al. 2015). A recent paper describes the known functional assays for RIPs and they fail to broaden the scope beyond the classical catalytic activity of RIPs that has always been studied (Zhou et al. 2018). When PAP-I is found to inhibit +1 ribosomal frameshift (Tumer et al. 1998), ricin to have phospholipase activity (Helmy et al. 1999), and

other RIPs showing SOD activity (Sharma et al. 2004; Li et al. 1997; Iglesias 2015; Zhu et al. 2016), this is a potential indication that RIPs are not bearing a single function in plant cells. To further understand the way these enzymes work, a combination of structural analysis (including prediction and simulation) and wet lab work would go hand in hand. With the global COVID-19 pandemic, the science industry has learnt to harness trial and error through simulations; and proof of concept through wet lab testing in order to produce an acceptable vaccine in a record time (Romero-Brufau et al. 2021). Although the scale is not comparable, the field of RIPs would benefit from using the same approach in identifying binding partners or testing different inhibitors. While it remains predictive, it allows to narrow the search for the wet lab portion and hopefully save time in the long run.

5 REFERENCES

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. Plant Cell. 15: 63-78.

Acharya BR, Assmann SM. 2009. Hormone interactions in stomatal function. Plant Mol Biol. 69: 451–462.

Agarwal M, Hao Y, Kapoor A, Dong CH, Fuji H, Zheng X, et al. 2006. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. J. Biol. Chem. 281: 37636-37645.

Agarwal S, Sairam RK, Srivastava GC, Tyagi A, Meena RC. 2005. Role of ABA, salicylic acid, calcium and hydrogen peroxide on antioxidant enzymes induction in wheat seedlings. Plant Sci. 169(3): 559-570.

Agurla S, Gahir S, Munemasa S, Murata Y, Raghavendra AS. 2018. Mechanism of Stomatal Closure in Plants Exposed to Drought and Cold Stress. Adv. Exp. Med. Biol. 1081: 215-232.

Alekseenko A, Barett D, Pareja-Sanchez Y, Howard RJ, Strandback E, et al. 2021. Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples. Sci. Rep. 11: 1820.

Aron GM, Irvin JD. 1980. Inhibition of Herpes Simplex Virus Multiplication by the Pokeweed Antiviral Protein. Antimicrob. Agents Chemother. 17(6): 1032-1033.

Bai Y, Monzingo AF, Robertus JD. 2009. The x-ray structure of ricin A chain with a novel inhibitor. Arch. Biochem. Biophys. 483: 23–28.

Banisa H, Bagaria S, Karande AA, Ramakumar S. 2019. Structural basis for neutralization of cytotoxic abrin by monoclonal antibody D6F10. FEBS J. 286: 1003-1029. doi:10.1111/febs.14716

Barbieri L, Aron GM, Irvin JD, Stirpe F. 1982. Purification and partial characterization of another form of the antiviral protein from the seeds of Phytolacca americana L. (pokeweed). Biochem. J. 203: 55–59.

Barbieri L, Bolognesi A, Cenini P, Falasca AI, Minghetti A, Garofano L, et al. 1989. Ribosome-inactivating proteins from plant cells in culture. Biochem. J. 257, 801–807.

Barbieri L, Battelli MG, Stirpe F. 1993. Ribosome-inactivating proteins from plants. Biochim. Biophys. Acta. 1154: 237–282.

Barbieri L, Valbones P, Bonora E, Gorini P, Bolognesi A, Stirpe F. 1997. Polynucleotide-adenosine glycosidase activity of ribosome-inactivating proteins-effect on DNA, RNA and Poly (A). Nuc. Acids Res. 25: 518-522.

Basu S, Ramegowda V, Kumar A, Pereira A. 2016. Plant adaptation to drought stress. F1000 Res. 5: 1554.

Bohnert HJ, Nelson DE, Jensen RG. 1995. Adaptations to Environmental Stresses. Plant Cell. 7(7): 1099–111.

Bolognesi A, Barbieri L, Abbondanza A, Falasca AI, Carnicelli D, Battelli MG, et al. 1990. Purification and properties of new ribosome-inactivating proteins with RNA Nglycosidase activity. Biochim. Biophys. Acta. 1087: 293–302.

Bonness MS, Ready MP, Irvin JD, Mabry TJ. 1994. Pokeweed antiviral protein inactivates pokeweed ribosomes; implications for the antiviral mechanism. Plant J. 5(2): 173-183.

Buchel AS, Brederode FT, Bol JF, Linthorst HJM. 1999. Mutation of GT-1 binding sites in the Pr-1a promoter influences the level of inducible gene expression in vivo. Plant Mol. Biol. 40: 387–96.

Cao B, Bai J, Wang X, Zhang Y, Yu X, Hu S, He Y. 2022. BnA.JAZ5 attenuates drought tolerance in rapeseed through mediation of ABA–JA crosstalk. Horticult. 8: 131.

Cawley DB, Hedblom ML, Houston LL. 1979. Protection and rescue of ribosomes from the action of ricin A chain. Biochem. 18: 2648–2654.

Chan DS, Chu LO, Lee KM, Too PH, Ma KW, Sze KH, Zhu G, Shaw PC, Wong KB. 2007. Interaction between trichosanthin, a ribosome-inactivating protein, and the ribosomal stalk protein P2 by chemical shift perturbation and mutagenesis analyses. Nucleic Acids Res. 35: 1660–167.

Chang N, Sun Q, Hu J, An C, Gao H. 2017. Large Introns of 5 to 10 Kilo Base Pairs Can Be Spliced out in Arabidopsis. Genes. 8: 200.

Chen X-Y, Berti PJ, Schramm VL. 2000. Transition-State Analysis for Depurination of DNA by Ricin A-Chain. J. Am. Chem. Soc. 122: 6527–6534.

Chen Y, Zhi J, Zhang H, Li J, Zhao Q, Xu J. 2017. Transcriptome analysis of Phytolacca americana L. in response to cadmium stress. PLOS ONE. 12(9): e0184681.

Chen Z, Antoniw JF, White RF. 1993. A possible mechanism for the antiviral activity of pokeweed antiviral protein. Physiol. Mol. Plant Path. 42: 249–258.

Daszkowska-Golec A, Szarejko I. 2013. Open or close the gate-stomata action under the control of phytohormones in drought stress conditions. Front. Plant Sci. 4: 138.

Depuydt M, Messens J, Collet J-F. 2011. How proteins form disulfide bonds. Antiox. Redox Sign. 15: 49-66.

Di R, Tumer NE. 2015. Pokeweed antiviral protein: it's cytotoxicity mechanism and applications in plant disease resistance. Toxins. 7: 755-772.

Di Maro A, Citores L, Russo R, Iglesias R, Ferreras JM. 2014. Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms. Plant Mol. Biol. 85, 575–588.

Ding Y, Shi Y, Yang S. 2019. Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. New Phytol. 222: 1690–1704.

Domashevskiy AV, Miyoshi H, Goss DJ. 2012. Inhibition of pokeweed antiviral protein (PAP) by turnip mosaic virus genome-linked protein (VPg). J. Biol. Chem. 287: 29729–29738.

Domashevskiy AV, Goss DJ. 2015. Pokeweed Antiviral Protein, a Ribosome Inactivating Protein: Activity, Inhibition and Prospects. Toxins. 7: 274-298.

Du H, Liu H, Xiong, L. 2013. Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. Front. Plant Sci. 4: 397.

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, et al. 2003. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J. 33: 751–63

Dunn MA, White AJ, Vural S, Hughes MA. 1998. Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (Hordeum vulgare L.). Plant Mol. Biol. 38: 551–64.

Fermani S, Tosi G, Farini V, Polito L, Falini G, Ripamonti A, Barbieri L, Chambery A, Bolognesi A. 2009. Structure/function studies on two type 1 ribosome inactivating proteins: Bouganin and lychnin. J. Struct. Biol. 168: 278-287.

Ferreras JM, Barbieri L, Girbes T, Batelli MG, Rojo MA, Arias FJ, Rocher MA, Soriano F, Mendez E, Stirpe F. 1993. Distribution and properties of major ribosome-inactivating proteins (28 S rRNA N-glycosidases) of the plant Saponaria officinalis L. (Caryophyllaceae). Bioch. Biophys. Acta. 1216: 31-42.

Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. 3: 329-341

Fu X, Dou C, Chen Y, Chen X, Shi J, Yu M, Xu J. 2011. Subcellular distribution and chemical forms of cadmium in Phytolacca americana L. J. Hazard. Mat. 186: 103-107.

Funatsu G, Islam MR, Minami Y, Sung-Sil K, Kimura M. 1991. Conserved amino acid residues in ribosome-inactivating proteins from plants. Bioch. 73: 1157-1161.

Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF. 2003 Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA. 9: 1422-1430.

Girbes T, Ferreras JM, Arias FJ, Stirpe F. 2004. Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. Mini Rev. Med. Chem. 4(5): 461-476.

Grela P, Szajwaj M, Horbowicz-Drozdzal P, Tchorzewski M. 2019. How ricin damages the ribosome. Toxins. 11(5): 241.

Gu YJ, Xia ZX. 2000. Crystal structures of the complexes of trichosanthin with four substrate analogs and catalytic mechanism of RNA N-glycosidase. Proteins 39(1): 37-46.

Gubler F, Kalla R, Roberts JK, Jacobsen JV. 1995. Gibberellin-Regulated Expression of a myb Gene in Barley Aleurone Cells: Evidence for Myb Transactivation of a High-pl alpha-Amylase Gene Promoter. Plant Cell. 7: 1879–91.

Havrila M, Réblová K, Zirbel CL, Leontis NB, Šponer J. 2013. Isosteric and nonisosteric base pairs in RNA motifs: Molecular dynamics and bioinformatics study of the sarcin-ricin internal loop. J. Phys. Chem. B. 117: 14302–14319

Hayashi S, Wu HC. 1990. Lipoproteins in bacteria. J. Bioenenerg. Biomembr. 22: 451-471.

He YW, Guo CX, Pan YF, Peng C, Weng ZH. 2008. Inhibition of hepatitis B virus replication by pokeweed antiviral protein in vitro. World J. Gastroenterol. 14: 1592–1597.

Helmy H, Lombard S, Pieroni G. 1999. Ricin RCA60: evidence of its phospholipase activity. Biochem. Biophys. Res. Commun. 258: 252–255.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science. 253: 191–194.

Honjo E, Dong D, Motoshima H, Watanabe K. 2002. Genomic clones encoding two isoforms of Pokeweed antiviral protein in seeds (PAP-S1 and S2) and the N-glycosidase activities of their recombinant proteins on ribosomes and DNA in comparison with other isoforms. J. Biochem. 131: 225–31.

Huang Q, Liu S, Tang Y, Jin S, Wang Y. 1995. Studies on crystal structures, active-centre geometry and depurinating mechanism of two ribosome-inactivating proteins. Biochem. J. 309: 285-298.

Hudak KA, Parikh BA, Di R, Baricevic M, Santana M, Seskar M, Tumer NE. 2004. Generation of pokeweed antiviral protein mutations in Saccharomyces cerevisiae: evidence that ribosome depurination is not sufficient for cytotoxicity. Nuc. Acids Res. 32(14): 4244-4256

Hur Y, Hwang D-J, Zoubenko O, Coetzer C, Uckun F, Turner NE. 1995. Isolation and characterization of pokeweed antiviral protein in Saccharomyces cerevisiae: identification of residues important for toxicity. Proc. Natl. Acad. Sci. USA. 92: 8448-8452.

Iglesias R, Citores L, Di Maro A, Ferreras JM. 2015. Biological activities of the antiviral protein BE27 from sugar beet (Beta vulgaris L.). Planta. 241: 421–433.

Irvin JD. 1975. Purification and partial characterization of the antiviral protein from Phytolacca americana which inhibits eukaryotic protein synthesis. Arch. Biochem. Biophys. 169: 522–528.

Irvin JD. 1983. Pokeweed antiviral protein. Pharmac. Ther. 21: 371-387.

Ishag HZ, Li C, Huang L, Sun MX, Ni B, Guo CX, Mao X. 2013. Inhibition of Japanese encephalitis virus infection in vitro and in vivo by pokeweed antiviral protein. Virus Res. 171: 89–96.

Jiang S-Y, Ramamoorthy R, Bhalla R, Luan H-F, Venkatesh PN, Cai M, Ramachandran S. 2008. Genome-wide survey of the RIP domain family in Oryza sativa and their expression profiles under various abiotic and biotic stresses. Plant Mol. Biol. 67: 603-614.

Jiang S-Y, Bhalla R, Ramamoorthy R, Luan H-F, Venkatesh P-N, Cai M, Ramachandran S. 2012. Over-expression of OSRIP18 increases drought and salt tolerance in transgenic rice plants. Transgenic Res. 21: 785-795.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature. 596(7873): 583-589.

Kaplan B, Davydov O, Knight H, Galon Y, Knight MR, Fluhr R, et al. 2006. Rapid Transcriptome Changes Induced by Cytosolic Ca2+ Transients Reveal ABRE-Related Sequences as Ca2+-Responsive cis Elements in Arabidopsis. Plant Cell 18: 2733–48.

Karran RA, Hudak KA. 2008. Depurination within the intergenic region of Brome mosaic virus RNA3 inhibits viral replication in vitro and in vivo. Nucleic Acids Res. 36: 7230–7239.

Kashkush, K., Feldman, M., and Levy, A.A. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. 33**:** 102–106.

Kataoka J, Habuka N, Masuta C, Miyano M, Kiowai A. 1992. Isolation and analysis of a genomic clone encoding a pokeweed antiviral protein. Plant Mol. Biol. 20: 879-886.

Kim HJ, Kim YK, Park JY, Kim J. 2002. Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in Arabidopsis thaliana. Plant J. 29: 693-704.

Kim SY, Chung HJ, Thomas TL. 1997. Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. Plant J. 11: 1237–51.

Kim W, Lee Y, Park J, Lee N, Choi G. 2013. HONSU, a protein phosphatase 2C, regulates seed dormancy by inhibiting ABA signaling in Arabidopsis. Plant Cell Physiol. 54: 555–572.

Kim Y, Mlsna D, Monzingo AF, Ready MP, Frankel A, Robertus JD. 1992. Structure of a ricin mutant showing rescue of activity by a noncatalytic residue. Biochem. 31(12): 3294-3296.

Kim YS, Lee M, Lee JH, Lee HJ, Park CM. 2015. The unified ICE–CBF pathway provides a transcriptional feedback control of freezing tolerance during cold acclimation in Arabidopsis. Plant Mol. Biol. 89: 187–201.

Kizis D, Pagès M. 2002. Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. Plant J. 30: 679–89.

Krivdova G, Hudak KA. 2015. Pokeweed antiviral protein restores levels of cellular APOBEC3G during HIV-1 infection by depurinating Vif mRNA. Antiviral Res. 122: 51–54.

Kung S, Kimura M, Funatsu G. 1990. The complete amino acid sequence of antiviral protein from the seeds of pokeweed (Phyto/acca americana). Agricult. Biol. Chem. 34: 3301.

Kurinov IV, Myers DE, Irvin JD, Uckun FM. 1999a. X-ray crystallographic analysis of the structural basis for the interactions of pokeweed antiviral protein with its active site inhibitor and ribosomal RNA substrate analogs. Protein Sci. 8: 1765-1772.

Kurinov IV, Rajamohan F, Venkatachalam TK, Uckun FM. 1999b. X-ray crystallographic analysis of the structural basis for the interaction of pokeweed antiviral protein with guanine residues of ribosomal RNA. Protein Sci. 8: 2399-2405.

Kurinov IV, Uckun FM. 2003. High resolution X-ray structure of potent anti-HIV pokeweed antiviral protein-III. Biochem. Pharmacol. 65: 1709–1717.

Kurinov IV, Myers DE, Uckun FM, Irvin JD. 2008. X-ray crystallographic analysis of the structural basis for the interactions of pokeweed antiviral protein with its active site inhibitor and ribosomal RNA substrate analogs. Prot. Sci. 8(9): 1765-1772.

Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, Kitamura S, Asami T, Hirai N, Koshiba T, Kamiya Y, Nambara E. 2004. The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: Key enzymes in ABA catabolism. EMBO J. 23: 1647–1656.

Lawlor DW, Cornic G. 2002. Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. Plant Cell Environ. 25(2): 275-294.

Leung J, Giraudat J. 1998. Abscisic acid signal transduction. Annu. Rev. Plant Biol. 49: 199–222.

Li XD, Chen WF, Liu WY, Wang GH. 1997. Large-scale preparation of two new ribosome inactivating proteins—cinnamomin and camphorin from the seeds of Cinnamomum camphora, Protein Expr. Purif. 10: 27–31

Li X-P, Kahn PC, Kahn JN, Grela P, Tumer NE. 2013. Arginine Residues on the Opposite Side of the Active Site Stimulate the Catalysis of Ribosome Depurination by Ricin A Chain by Interacting with the P-protein Stalk. 288(42): P30270-30284. J. Biol. Chem.

Liljas A, Ehrenberg M. 2012. Structural Aspects of Protein Synthesis, 2nd ed. World Scientific Publishing Company Incorporated: Hackensack. NJ, USA.

Lin Y, Xu L, Li Y, Wu X, Liu Y, Zhu H, Zhou H. 2021. Ribosome-Inactivating Proteins of Bougainvillea glabra Uncovered Polymorphism and Active Site Divergence. Toxins. 13: 331.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using realtime quantitative PCR and the 2-ΔΔCT method. Methods 25: 402–408.

Lodge JK, Kaniewski WK, Tumer NE. 1993. Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. Proc. Natl. Acad. Sci. USA. 90: 7089-7093

Lord JM, Roberts LM, Robertus JD. 1994. Ricin: structure, mode of action, and some current applications. FASEB J. 8: 201-208.

Mak AN-S, Wong Y-T, An Y-J, Cha S-S, Sze K-H, Au SW-N, Wong K-B, Shaw P-C. 2007. Structure-function study of maize ribosome-inactivating protein: implications for the internal inactivation region and the sole glutamate in the active site. Nuc. Acids Res. 35(18): 6259-6267.

Mansouri S, Choudhary G, Sarzala PM, Ratner L, Hudak KA. 2009. Suppression of human T-cell leukemia virus I gene expression by pokeweed antiviral protein. J. Biol. Chem. 284: 31453–31462.

Maracci C, Rodnina MV. 2016. Review: Translational GTPases. Biopolymers. 105: 463–475.

Marchant A, Hartley MR. 1995. The Action of Pokeweed Antiviral Protein and Ricin A-chain on Mutants in the α-Sarcin Loop of Escherichia coli 23S Ribosomal RNA. J. Mol. Biol. 254(5): 848-855.

McBride MB, Zhou Y. 2019. Cadmium and zinc bioaccumulation by Phytolacca americana from hydroponic media and contaminated soils. Intl. J. Phytoremed. 21(12): 1215-1224.

McCluskey AJ, Poon GM, Bolewska-Pedyczak E, Srikumar T, Jeram SM, Raught B, Gariepy J. 2008. The catalytic subunit of Shiga-like toxin 1 interacts with ribosomal stalk proteins and is inhibited by their conserved C-terminal domain. J. Mol. Biol. 378: 375–386.

McCluskey AJ, Bolewska-Pedyczak E, Jarvik N, Chen G, Sidhu SS, Gariepy J. 2012. Charged and hydrophobic surfaces on the a chain of Shiga-like toxin 1 recognize the C-terminal domain of ribosomal stalk proteins. PLoS ONE 7: e31191.

Melotto M, Underwood W, Koczan J, Nomura K, He SY. 2006. Plant stomata function in innate immunity against bacterial invasion. Cell. 126: 969–980.

Mena M, Cejudo F, Isabel-Lamoneda I, Carbonero P. 2002. A Role for the DOF Transcription Factor BPBF in the Regulation of Gibberellin-Responsive Genes in Barley Aleurone. Plant Physiol. 130: 111–9.

Mishra V, Mishra R, Shamra RS. 2022. Ribosome inactivating proteins – An unfathomed biomolecule for developing multi-stress tolerant transgenic plants. Intl. J. Biol. Macromol. 210: 107-122.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H. 2003. Target Site Specificity of the Tos17 Retrotransposon Shows a Preference for Insertion within Genes and against Insertion in Retrotransposon-Rich Regions of the Genome. Plant Cell. 15(8): 1771-1780.

Moazed D, Robertson JM, Noller HF. 1988. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature. 334: 362–36.

Mohanty B, Krishnan SPT, Swarup S, Bajic VB. 2005. Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann. Bot. 96: 669–81.

Munne-Bosch S, Alegre L. 2004. Die and let live: Leaf senescence contributes to plant survival under drought stress. Funct. Plant Biol. 31(3): 203-216.

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. NAC transcription factors in plant abiotic stress responses. Biochim. Biophys. Acta-Gene Regulatory Mechanisms 1819(2): 97–103.

Neller KCM, Klenov A, Hudak KA. 2016. The Pokeweed Leaf mRNA Transcriptome and Its Regulation by Jasmonic Acid. Front. Plant Sci. 7: 283.

Neller KCM, Diaz CA, Platts AE, Hudak KA. 2019. De novo Assembly of the Pokeweed Genome Provides Insight Into Pokeweed Antiviral Protein (PAP) Gene Expression. Front. Plant Sci. 10: 1002.

Ng Y-M, Yang Y, Sze K-H, Zhang X, Zheng Y-T, Shaw P-C. 2011. Structural characterization and anti-HIV-1 activities of arginine/glutamate-rich polypeptide Luffin P1 from the seeds of sponge gourd (Luffa cylindrica). J. Struct. Biol. 174: 164-172.

Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. 2003. Gibberellin Biosynthesis and Response during Arabidopsis Seed Germination. Plant Cell 15: 1591-604.

Olson MA, Cuff L. 1999. Free energy determination of binding the rRNA substrate and small ligands to ricin A-chain. Biophys. J. 76: 28-39.

Othman TA, Bakar NA, Abidin RAZ, Mahmmod M, Saidi NBB, Shaharuddin NA. 2019. Transcriptomic Profiling of Malaysia Rice Cultivars (Oryza sativa) under the Effect of Osmotic Stress Induced by PEG 6000. Intl. Agric. Biol. 21(2): 1560-8530.

Parente A, Conforto B, Di Maro A, Chambery A, De Luca P, Bologensi A, Iriti M, Faoro F. 2008. Type 1 ribosome-inactivating proteins from Phytolacca dioica L. leaves: diVerential seasonal and age expression, and cellular localization. Planta. 228: 963-975.

Park H, Kim M, Kang Y, Jeon J, Yoo J, Kim M, et al. 2004. Pathogen- and NaCl-Induced Expression of the SCaM-4 Promoter Is Mediated in Part by a GT-1 Box That Interacts with a GT-1-Like Transcription Factor. Plant Physiol. 135: 2150–61.

Parry MAJ, Andralojc PJ, Khan S, Lea PJ, Keys AJ. 2002. Rubisco Activity: Effects of Drought Stress. Annals of Botany. 89(7): 833-839.

Per TS, Khan MIR, Anjum NA, Masood A, Hussain SJ, Khan NA. 2018. Jasmonates in plants under abiotic stresses: Crosstalk with other phytohormones matters. Env. Exp. Botany. 145: 104-120.

Picard D, Kao CC, Hudak KA. 2005. Pokeweed antiviral protein inhibits brome mosaic virus replication in plant cells. J. Biol. Chem. 280: 20069–20075.

Poyet J-L, Hoeveler A. 1997a. cDNA cloning and expression of pokeweed antiviral protein from seeds in Escherichia coli, and its inhibition of protein synthesis in vitro. FEBS Letters. 406: 97-100.

Poyet J-L, Hoeveler A. 1997b. Presence of an Intron in a Gene of PAP II, the Ribosome-inactivating Protein from Summer Leaves of Phytolacca americana. Annals of Botany. 80: 685-688.

Qin W, Ming-Xing H, Ying X, Xin-Shen Z, Fang C. 2005. Expression of a ribosome inactivating protein (curcin 2) in Jatropha curcas is induced by stress. J. Biosci. 30(3): 351-357.

Qin X, Zheng X, Shao C, Gao J, Jiang L, Zhu X, Yan F, Tang L, Xu Y, Chen F. 2009. Stress-induced curcin-L promoter in leaves of Jatropha curcas L. and characterization in transgenic tobacco. Planta. 230: 387-395.

Rajamohan F, Venkatachalam TK, Irvin JD, Uckun FM. 1999a. Pokeweed antiviral protein isoforms PAP-I, PAP-II, and PAP-III depurinate RNA of human immunodeficiency virus (HIV)-1. Biochem. Biophys. Res. Commun. 260: 453–458.

Rajamohan F, Kurinov IV, Venkatachalam TK, Uckun FM. 1999b. Deguanylation of human immunodeficiency virus (HIV-1) RNA by recombinant pokeweed antiviral protein. Biochem. Biophys. Res. Commun. 263: 419–424

Rajamohan F, Pugmire MJ, Kurinov IV, Uckun FM. 2000. Modeling and alanine scanning mutagenesis studies of recombinant pokeweed antiviral protein. J. Biol. Chem. 275: 3382–3390.

Rajamohan F, Mao C, Uckun FM. 2001. Binding interactions between the active center cleft of recombinant pokeweed antiviral protein and the α-Sarcin/Ricin stem loop of ribosomal RNA. J. Biol. Chem. 276(26): 24075-24081.

Rock CD. 2000. Tansley Review No. 120: Pathways to abscisic acid-regulated gene expression. New Phytol. 148(3):357-396.

Romero-Brufau S, Chopra A, Raskar R, Subramanian J, Krishnamurthy B, et al. 2021. Public health impact of delaying second dose of BNT162b2 or mRNA-1273 covid-19 vaccine: simulation agent-based modeling study. BMJ. 373: n1087.

Saibi W, Brini F. 2018. Superoxide dismutase (SOD) and abiotic stress tolerance in plants: An overview. (ed.). Nova Science Publishers, Inc. ISBN: 978-53613-080-5.

Satoh R, Fujita Y, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. 2004. A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the ProDH gene in Arabidopsis. Plant Cell Physiol. 45: 309–17.

Schafer B, Gan L, Perlman P. 2003. Reverse Transcriptase and Reverse Splicing Activities Encoded by the Mobile Group II Intron COBI1 of Fission Yeast Mitochondrial DNA. 329(2): 191-206.

Schlossman D, Withers D, Welsh P, Alexander A, Robertus J, Frankel A. 1989. Role of glutamic acid 177 of the ricin toxin A chain in enzymatic inactivation of ribosomes. Mol. Cell Biol.9(11): 5012-5021.

Schulman A. 2013. Retrotransposon replication in plants. Curr. Op. Virol. 298: 1-11.

Schroeder JI, Kwak JM, Allen GJ. 2001. Engineering drought hardiness in plants. Nature. 410: 327–330.

Seo JS, Joo J, Kim MJ, Kim YK, Nahm BH, Song SI, Cheong JJ, Lee JS, Kim JK, Choi Y. 2011. Do OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. Plant J. 65: 907–921.

Sharma N, Park SW, Vepachedu R, Barbieri L, Ciani M, Stirpe F, Savary BJ, Vivanco JM. 2004. Isolation and characterization of an RIP (ribosome inactivating protein)-like protein from tobacco with dual enzymatic activity, Plant Physiol. 134: 171–181.

Shi W-W, Mak AN-S, Wong K-B, Shaw P-C. 2016. Structures and Ribosomal Interaction of Ribosome-Inactivating Proteins. Molecules 21(11). 1588-1601.

Simpson S D, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J. 33: 259–70.

Song S-K, Choi Y, Moon YH, Kim S-G, Choi YD, Lee JS. 2000. Systemic induction of a Phytolacca insularis antiviral protein gene by mechanical wounding, jasmonic acid and abscisic acid. Plant Mol. Biol. 43: 439-450.

Suoniemi A, Anamthawat-Jonsson K, Arna T, Schulman AH. 1996. Retrotransposon BARE-1 is a major, dispersed component of the barley (Hordeum vulgare L.) genome. Plant Mol. Biol. 30:1321–1329.

Sutoh K, Yamauchi D. 2003. Two cis-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. Plant J. 34: 635–45.

Tartarini A, Pittaluga E, Marcozzi G, Testone G, Rodrigues-Pousada RA, Gianno D, Spano L. 2010. Differential expression of saporin genes upon wounding, ABA treatment and leaf development. Phys. Plant. 140: 141-152.

Tezara W, Mitchell VJ, Driscoll SD et al. 1999. Water stress inhibits plant photosynthesis by decreasing coupling factor and ATP. Nature. 401: 914-917.

Tomlinson JA, Walker VM, Flewett TH, Barclay GR. 1974. The inhibition of infection by cucumber mosaic virus and influenza virus by extracts from Phytolacca americana. J. Gen. Virol. 22: 225-232.

Ton J, Flors V, Mauch-Mani B. 2009. The multifaceted role of ABA in disease resistance. Trends Plant Sci. 14: 310–317

Tumer NE, Hwang D-J, Bonness M. 1997. C-terminal deletion mutant of pokeweed antiviral protein inhibits viral infection but does not depurinate host ribosomes. Proc. Natl. Acad. Sci. 94: 3866-3871.

Tumer NE, Hudak KA, Di R, Coetzer C, Wang P, Zoubenko O. 2000. Pokeweed Antiviral Protein and Its Applications. J. Hammond et al. (eds.). Plant Biotech. Springer-Verlag. Berlin Heidelberg

Uckun FM, Rajamohan F, Pendergrass S, Ozer Z, Waurzyniak B, Mao C. 2003. Structure-based design and engineering of a nontoxic recombinant pokeweed antiviral protein with potent anti-human immunodeficiency virus activity. Antimicrob. Agents Chemother. 47: 1052–1061.

Uckun FM, Rustamova L, Vassilev AO, Tibbles HE, Petkevich AS. 2005. CNS activity of Pokeweed anti-viral protein (PAP) in mice infected with lymphocytic choriomeningitis virus (LCMV). BMC Infect. Dis. 5: 9.

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K. 2010. Molecular basis of the core regulatory network in ABA responses: Sensing, signaling and transport. Plant Cell Physiol. 51: 1821–1839.

Ussery MA, Irvin JD. 1977. Hardesty, B. Inhibition of poliovirus replication by a plant antiviral peptide. Ann. NY Acad. Sci. 284: 431–440.

Wahl MC, Moller W. 2002. Structure and function of the acidic ribosomal stalk proteins. Curr. Protein Pept. Sci. 3: 93–106.

Wang C, Yang A, Yin H. 2008. Influence of water stress on endogenous hormone contents and cell damage of maize seedlings. J. Integr. Plant Biol. 50(4): 427-434.

Wang P, Tumer NE. 1999. Pokeweed antiviral protein cleaves double-stranded supercoiled DNA using the same active site required to depurinate rRNA. Nuc. Acids Res. 27(8): 1900-1905.

Wang S, Li Z, Li S, Di R, Ho C-T, Yang G. 2016. Ribosome-inactivating proteins (RIPs) and their important health promoting property Roy. Soc. Chem. 6: 46794-46805.

Wang Y, Mostafa S, Zeng W, Jin B. 2021. Function and Mechanism of Jasmonic Acid in Plant Responses to Abiotic and Biotic Stresses. Mol. Sci. 22(16): 8568.

Waterhouse A, Bertoni M, Stefan Bienert, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nuc. Acids Res. 46: W296-303.

Wong JH, Bao H, Ng TB, Chan HHL, Ng CCW, Man GCW, Wang H, Guan S, Zhao S, Fang EF, Rolka K, Liu Q, Li C, Sha O, Xia L. 2020. New ribosome-inactivating proteins and other proteins with protein synthesis-inhibiting activities. Appl Microbiol Biotechnol. 104(10):4211-4226.

Wong K-B, Ke Y-B, Dong Y-C, Li, X-B, Guo Y-W, Yeung H-W, Shaw P-C. 1994. Structure/function relationship study of Gln156, Glu160 and Glu189 in the active site of trichosanthin. Eur. J. Biochem. 221: 787-791.

Xu J, Wang H, Fan J. 2007. Expression of a ribosome inactivating protein gene in bitter melon is induced by Sphaerotheca fuliginea and abiotic stimuli. 29: 1605-1610.

Xu J, Zhou Y, Xu Z, Chen Z, Duan L. 2020. Combining Physiological and Metabolomic Analysis to Unravel the Regulations of Coronatine Alleviating Water Stress in Tobacco (Nicotiana tabacum L.). Biomol. 10: 99.

Yadav DK, Barik BR, Pradhan G. 2019. Responses of crops plant to drought and its management for crop water availability: A review. J. Pharmacogn. Phytochem. 8: 167–172.

Yang Y, Mak AN, Shaw PC, Sze KH. 2010. Solution structure of an active mutant of maize ribosome-inactivating protein (MOD) and its interaction with the ribosomal stalk protein P2. J. Mol. Biol. 395: 897–907.

Yang J, Duan G, Li C, Liu L, Han G, Zhang Y, Wang C. 2019. The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. Front. Plant Sci. 10: 1349.

Ye N, Jia L, Zhang J. 2012. ABA signal in rice under stress conditions. Rice. 5: 1–9.

Yosefi A, Mozafari AA, Javadi T. 2020. Jasmonic acid improved in vitro strawberry (Fragaria × ananassa Duch.) resistance to PEG-induced water stress. Plant Cell, Tissue, and Org. Cult. (PCTOC). 142: 549-558.

Zeng Z-H, He Z-L, Li H-M, Hu Z, Wang D-C. 2003. Crystal structure of pokeweed antiviral protein well-defined sugars from seeds at 1.8Å resolution. J. Struct. Biol. 141: 171-178.

Zhao L, Sun YL, Cui SX, Chen M, Yang HM, Liu HM, et al. 2011. Cd-induced changes in leaf proteome of the hyperaccumulator plant Phytolacca americana. Chemosphere 85: 56–66.

Zhao Q, Xiang X, Liu D, Yang A, Wang Y. 2018. Tobacco Transcription Factor NtbHLH123 Confers Tolerance to Cold Stress by Regulating the NtCBF Pathway and Reactive Oxygen Species Homeostasis. Front. Plant Sci.9: 381.

Zhu F, Yuan S, Zhang Z-W, Qian K, Feng J-G, Yang Y-Z. 2016. Pokeweed antiviral protein (PAP) increases plant systemic resistance to tobacco mosaic virus infection in Nicotiana benthamiana. Eur. J. Plant Pathol. 146: 541-549.

Zhu F, Zhou YK, Ji ZL, Chen XR. 2018. The Plant Ribosome-Inactivating Proteins Play Important Roles in Defense against Pathogens and Insect Pest Attacks. Front. Plant Sci. 9: 146

6 APPENDIX A: Table of primers

Table 5: List of primers used in RT-qPCR for the stress treatments and developmental stages in pokeweed

Primer Names	Primer sequence (5'-3')
PAP-I RT REV	GATATGATTTGAATCACTCGAATTCACCAAGG
KPAP RT REV	GCTTTGAGGAGCATGTGATTATAGAATGGTG
PAP-II RT REV	CCTTGATGCCTCATTAACCATTTGAACG
PAP-α RT REV	CATCCACTCTCAGCACTATCCACTTGCTACCGTTTG
PAP-S1 RT REV	GGCATTTTGTTAAGTTGCTTGGCAAGTCCC
BDX RT REV	CTTCACTCTCAGGATCCGTTC
EF-1G RT REV	CTTCAAAAGGCTCCTGGTC
EF-1α RT REV	GTCGGATCCTTCTTCTCAACATTC
Tubulin-B6	CTCAGTGAACTCCATTTCATCCATAC
PAP-I-1-35- FOR	ATAACTGCATGTTCTCATAAAAAAGCCTCAGC
PAP-I-112-83 REV	CTTCCTACTAGCTAGTTCCAATACTTTCGC
KPAP-117-145-FOR	CGCCACGAAGCTTCGGCTATTATACCAG
KPAP-214-185-REV	CAAACCCATATTGCACATACAAGTGACACC
PAP-II-535-566-FOR	GAAAAGAGTTACAAAGGGATGGAATCAAAGG
PAP-II-644-618-REV	CGTTGCATCCTTGCCGTAGATTTTACC
PAP-α-20-47-FOR	CACATTGCATGTTCTCATAAAAAGCCTC
PAP-α-162-136-REV	CAAGTTGAAGGTGGTTTAAGAATGAGCC
PAP-S1-8-37-FOR	CAGAGTTATCCATCACATTGCATGCATG
PAP-S1-91-63-REV	CCGCTTCTTTCAACACTTTCACCTGTTAC
BDX-FOR	GGAGCACACTACACCTTCGCTC
BDX-REV	CCTGCAACCCTGATACAATGGAAG
EF-1G-FOR	ACCAACTTCCGTGAAGTAGCAATTAAAG
EF-1G-REV	CTTCCCAAATGCGTACTTGCGAG
EF-1α-FOR	ACTCCCAAGTACTCAAAGGCAAGGTATG
EF-1α-REV	CCAGAGATGGGAACGAATGGGATCTT
Tubulin-B6-FOR	TCTCTGTCTTCCCATCACCAAAGGTTTC
Tubulin-B6-REV	TCCAACACCATACACTCATCAGCATTCTC