Lancaster Environment Centre

Faculty of Science and Technology

Lancaster University

Exploring Genetic Diversity in Rubisco Activase

Duncan Bloemers

2023-07-17

This thesis is submitted for the degree of Doctor of Philosophy

Abstract

Via photosynthesis, plants are the most important point of entry for energy into the biosphere. The Calvin-Benson-Bassham cycle is the part of the photosynthetic process responsible for carbon assimilation, which is done by Rubisco. Rubisco activase (Rca) acts as an activator of Rubisco, removing inhibitory sugar phosphates. Rca is required for the maintenance of photosynthetic activity. Since high temperatures reduce and eventually eliminate Rca activity, this can threaten plant growth. This makes Rca an attractive target for crop improvement. **Chapter 1** introduces the literature surrounding Rubisco, Rca, and its applications in crop improvement. It introduces the crop species cowpea (*Vigna unguiculata* (L.) Walp) and how Rca may be targeted for safeguarding its production under changing climates.

Chapter 2 explores the genetic diversity of Rca, and how the natural diversity of Rca may be exploited for crop improvements - as more thermostable Rca may help to protect crop growth from the effects of climate change. First, the high-level diversity of Rca was investigated; transcriptomic data for Rca was collected for all species available in the One Thousand Plant Transcriptomes Project database. Then, the scope was narrowed to focus on *Fabaceae*, the family to which cowpea belongs, and identifying candidate species that may have Rca mutations relevant to thermotolerance.

Chapter 3 focuses on the expression and purification of recombinant, polyhistidine-tagged Rca. Some isoforms of Rca are challenging to purify and can have solubility issues during or after purification. Factors that can be changed to optimise purity, solubility, and expression levels were investigated. This work paves the way for later work that requires a considerable amount of pure Rca for characterisation of temperature and concentration response.

In **Chapter 4**, the temperature response of cowpea Rca isoforms was evaluated. There are four isoforms of Rca in cowpea, which differed in their temperature response despite limited differences in their protein sequence.

Finally, for **Chapter 5**, bespoke recombinant antibodies were produced targeting individual cowpea Rca isoforms. The antibodies were validated for their avidity and specificity for their target isoforms. The primary application of these antibodies is for western blotting, but potential further applications and research questions are discussed. Furthermore, the advantages and disadvantages of recombinant antibody technology compared to animal-derived antibodies are discussed.

Chapter 6 evaluates the results shown within, and assesses the possibility of crop improvement by targeting Rca. Additional research suggestions are posed to further this work, and suggestions for how to improve experimental methods are made.

In summary, this study provides a thorough examination of the thermal properties of cowpea Rca. It also improves capacity for future research into this area by optimising recombinant protein expression and purification protocols, and new tools in the form of isoform-specific antibodies.

Contents

List of Tables

List of Figures

Abbreviations

1KP One Thousand Plant Transcriptomes Project.

AA Amino acid.

AAA+ ATPases Associated with diverse cellular Activities.

BLAST Basic local alignment search tool.

CA1P 2-carboxy-D-arabinitol 1-phosphate.

CAM Crassulacean acid metabolism.

CBB Calvin-Benson-Bassham.

CCM Carbon-concentrating mechanism.

co-IP co-Immunoprecipitation.

Cowpea *Vigna unguiculata* (L.) Walp.

ER Endoplasmic reticulum.

GBIF Global Biodiversity Information Facility.

GmRca *Glycine max* Rubisco activase.

GMST Global mean surface temperature.

His-tag Polyhistidine tag.

HSD Honestly significant difference.

IgG Immunoglobulin G.

IMAC Immobilized Metal Affinity Chromatography.

IPTG Isopropyl β-D-1-thiogalactopyranoside.

LB Lysogeny broth.

MBP Maltose-binding protein.

OD Optical density.

ORI Origin of replication.

PDBP glycero-2,3-pentodiulose-1,5-bisphosphate.

PTM Posttranslational modification.

RbcL Rubisco large subunit.

RbcS Rubisco small subunit.

RBS Ribosome binding site.

Rca Rubisco activase.

Rubisco ribulose-1,5-bisphosphate carboxylase/oxygenase.

RuBP ribulose-1,5-bisphosphate.

SDS Sodium dodecyl sulphate.

SDS-PAGE Sodium dodecyl sulphate polyacrylamide gel electrophoresis.

TaRca *Triticum aestivum* Rubisco activase.

TEV Tobacco Etch Virus nuclear-inclusion-a endopeptidase.

TPU Triose phosphate use.

VH Variable heavy.

VL Variable light.

VuRca *Vigna unguiculata* Rubisco activase.

XuBP xylulose 1,5-Bisphosphate.

Chapter 1

Introduction

With a rapidly rising population, food security is an increasingly difficult and important problem to tackle. The 2018 prevalence of undernourishment was 10.8%, with over 800 million people going hungry worldwide (FAO et al., 2019). Furthermore, human population is expected to rise to 9.7 billion by 2050, a 26% increase over the 2019 population of 7.7 billion (UN, 2019), which would necessitate a proportionate increase in food production. As well as the increase in population, there are several other factors affecting food security, such as increased demand for more varied, higher quality food and an increase in per capita calorie consumption (Keating et al., 2014). There is inequality of food distribution, with Africa and West Asia having the greatest prevalence of undernourishment (FAO et al., 2019). Crop yields more than doubled during the green revolution, but since then there has been a decline in the rate of yield increase for staple crops (Ray et al., 2012). Finally, of critical importance, is stress to the current supply chains by climate change.

Climate change is impacting agriculture, ecosystem stability, ocean levels, and more. The primary impacts of climate change on agriculture are expected to come through decreases in available water and an increase in temperature past what crops can cope with. The global mean surface temperature (GMST) is believed to have warmed by ~1 °C compared to pre-industrial revolution temperatures as a result of human impact, and is predicted to rise further over the coming years, by 2 °C or more by 2050 (IPCC, 2018). The sensitivity of plants to this shift in temperature is a concern for crop production. Several crops that are relied on for food, fuel, or livestock feed are at risk of decreased production as a result of global warming. Global production of wheat, the most widely produced and consumed cereal crop globally (FAO, 2020) is estimated to be down ~5.5% compared to a projection without global warming (Lobell et al., 2011). Given that global production of wheat for 2019 is estimated at 763 million tons in 2019 (FAO, 2020), this amounts to a significant decrease in production and is a concern for food security. In order to meet the growing food and fuel demands in a sustainable and environmentally friendly fashion, more information is needed on how plants cope with environmental stress and what their limitations are.

Plants are responsible for almost all of the food for humans, either directly or indirectly through livestock. Being sessile by nature, plants are unable to avoid biotic and abiotic stressors and

must instead adapt or acclimate to environmental challenges. Heat stress is a common abiotic stressor plants are exposed to. To keep an operating leaf temperature under heat stress, plants open stomata to cool their leaves down by transpiration, a process driven by the difference in vapour pressure between the leaf and the air outside the leaf (Carmo-Silva et al., 2012). This comes at a heavy water cost, however, and under excessively high temperatures or low water availability stomata close to prevent excessive water loss. Further to this, plants possess several mechanisms such as complex transcriptional networks, and a cohort of enzymes and chaperones which enable them to survive heat stress. Heat stress is a particular issue in arid conditions where keeping stomata open is difficult due to lack of water availability. This is a serious consideration as climate change threatens to both raise temperature and destabilise water availability, with a net increase of precipitation and evaporation being predicted, as well as an increase in the variance of precipitation (IPCC, 2018; Konapala et al., 2020). The mechanisms plants employ to tolerate high temperatures, and the long term priming which enables them to better survive subsequent heat stress, are reviewed in Hasanuzzaman et al. (2013) and Bäurle (2016) respectively.

In this chapter I discuss in detail the mechanism of photosynthesis (Chapter 1.1) and the role of Rca in heat tolerance (Chapter 1.3.2). We will also explore the scope for crop improvement by targeting photosynthesis (Chapter 1.4), with a focus on Rca. Context is given surrounding the crop of interest - cowpea (*Vigna unguiculata* (L.) Walp) (Chapter 1.5) and the need to safeguard this crop for the future. Lastly, I discuss the production of recombinant proteins (Chapter 1.6), and the applications of antibodies in Rca research (Chapter 1.7).

1.1 Carbon Assimilation and Biomass Production

Photosynthesis is the process by which organisms utilise light energy to make energy stores and assimilate carbon. It is a vital process in plants which is especially sensitive to increases in temperature (Sage and Kubien, 2007). Through carbon fixation, photosynthesis is a major determinant of crop yield (Ainsworth and Long, 2005; Faralli and Lawson, 2020). Carbon is fixed during the light independent reaction of photosynthesis, the Calvin-Benson-Bassham (CBB) cycle (Benson and Calvin, 1947). The CBB cycle uses ATP and the reducing power of NADPH generated by the light dependent reactions to generate organic molecules which can be stored as starch or exported from the chloroplast, and to regenerate ribulose-1,5-bisphosphate (RuBP) which is consumed in the process. The key enzyme required for carbon fixation is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco is notorious for its catalytic inefficiency and unproductive oxygenase activity. It can fix either CO_2 or O_2 to RuBP, initiating either the CBB cycle or photorespiration respectively. The CBB cycle ultimately produces triose phosphates, metabolites used to build carbohydrates, while the oxygenation reaction produces phosphoglycolate, which is salvaged in an energy-intensive pathway (Andersson and Backlund, 2008). Rubisco's slow rate of carboxylation (~1-10 s^{-1}) (Iñiguez et al., 2020) means that Rubisco is often the limiting factor in photosynthesis, and requires substantial nitrogen investment to keep photosynthesis functioning

at acceptable levels (Parry et al., 2003). In fact, Rubisco is by far the most expressed protein in photosynthetic tissues and the most abundant protein on the planet. While the amount varies by species, Rubisco can constitute up to 50% of total soluble protein in leaves (Feller et al., 2008). These limitations and the global importance of Rubisco makes it an attractive target for improvement via genetic engineering. An increase in photosynthesis rates and a decrease in nitrogen investment in Rubisco has the potential to increase yield considerably (Spreitzer and Salvucci, 2002). However, most attempts at improving Rubisco have been largely unsuccessful, but modern approaches have shown some promise (Aigner et al., 2017; Batista-Silva et al., 2020; Erb and Zarzycki, 2018; Whitney et al., 2011; Wilson et al., 2019). Directed evolution and targeted mutagenesis has been able to improve the carboxylation rate and CO_2 specificity (Spreitzer et al., 2005; Zhou and Whitney, 2019).

Plant Rubisco is a large protein (~550 kDa holoenzyme), consisting of eight large subunits (RbcL) and eight small subunits (RbcS). RbcL is encoded in the chloroplast, and RcbS is encoded in the nucleus. Chloroplast transformation is more challenging than nuclear transformation, and techniques are not available for every species, which makes targeted mutagenesis of Rubisco large subunits difficult (Bock, 2014). Furthermore, Rubisco requires a suite of additional proteins for proper assembly. RbcL folding is mediated by Cpn60 and Cpn20 and further factors are required for assembly into an L_8S_8 complex such as the Rubisco Accumulation Factors (Raf1 & Raf2) and BSD2 (Wilson and Hayer-Hartl, 2018). These considerations complicate genetic manipulation of Rubisco, since editing the chloroplast genome is significantly more challenging than the nuclear genome; and makes heterologous expression of Rubisco a difficult task which was attempted for decades and only recently achieved by Aigner et al. (2017). For a review of the advancements and methods of chloroplast editing, as well as which species these techniques have been successfully applied in, see Yu et al. (2020).

While the limiting factor in photosynthesis is often Rubisco catalysis, other factors can limit photosynthesis rate depending on species, genotype and environmental conditions. The rate of electron transport in the light dependent reactions can bottleneck photosynthesis by supplying insufficient ATP and NADPH for regeneration of RuBP (Wise et al., 2004). Triose phosphate use (TPU), can also be limiting. TPU limitation causes build-up of triose phosphates which has both short-term and long term impacts on photosynthesis, including increase thylakoid proton motive force, and a reduction in RuBP regeneration capacity respectively (Yang et al., 2016). However some suggest this can be often ignored as TPU limitation is rare under physiological conditions (Sharkey, 2019). Stomatal conductance and mesophyll conductance can also limit photosynthesis by limiting the amount of CO_2 available for fixation (Flexas and Medrano, 2002). Light is another factor which can influence photosynthetic limitations. Rubisco activase (Rca) is the catalytic chaperone of Rubisco, and can be a limiting factor under certain light conditions. Rca α isoforms are modulated by the ATP/ADP ratio which changes with light intensity, causing a decrease in Rca activity and an increase in Rubisco inhibition in limiting light (Zhang et al., 2002). Rca is discussed in depth in 1.3. Under steady state light, Rubisco and mesophyll conductance limit photosynthesis

but under fluctuating light, the limiting factor can be stomatal conductance (Yamori et al., 2020). Plants often experience fluctuating light due to factors such as leaf movement, canopy dynamics and cloud cover so this is an important consideration and has been shown to vary significantly within species (Acevedo-Siaca et al., 2020; de Souza et al., 2020; Soleh et al., 2017). In fact, modelling work on wheat has shown that the induction and de-induction of photosynthesis in response to shade-sun and sun-shade transitions may cost up to 21% of photosynthetic productivity (Taylor and Long, 2017). In cowpea, rapid deactivation of Rubisco in response to fluctuating light causes a noteworthy loss in photosynthetic potential (Taylor et al., 2022). Targeted improvement of these processes may prove viable methods of increasing overall photosynthesis, depending on the environmental conditions of the plant.

1.2 Genetic Diversity in Photosynthesis

1.2.1 Carbon Concentrating Mechanisms

One potential method of improving photosynthesis is exploiting the genetic diversity of photosynthetic apparatus available in different species. One of the largest differences in photosynthetic machinery is the presence or absence of a carbon concentrating mechanism (CCM). CCMs compartmentalise photosynthetic apparatus spatially or temporally to provide a high concentration of Rubisco surrounded by a high concentration of CO_2. The majority of plants, termed C3 plants, do not possess a CCM (Yamori et al., 2014), however there are a small fraction termed C4 plants which physically separate the site of the CBB cycle into bundle sheath cells. Inside these cells, carbon dioxide is in a substantially higher concentration due to decarboxylation of compounds such as malate which are shuttled from the mesophyll cells (Burgess and Hibberd, 2015). This system creates a high CO_2 environment which favours the carboxylase activity of Rubisco and reduces photorespiration due to increased likelihood of Rubisco interacting with CO_2 compared to O_2. This, in turn, allows optimisation of Rubisco's catalytic properties to have an accelerated turnover rate in exchange for lower CO_2 specificity to accelerate photosynthesis without causing excessive photorespiration (Sage, 2002). In C4 plants, Rubisco is less abundant than in C3 plants, being only around 10-30% of total soluble protein (Carmo-Silva et al., 2015; Feller et al., 2008).

In addition to C3 and C4 pathways, there exist in plants the Crassulacean acid metabolism (CAM) and C2 photosynthetic pathways. CAM is a CCM which helps alleviate water loss. CAM plants close their stomata during the day, when water losses due to transpiration would be greatest, and open them during the night time to allow entry of CO_2 (Griffiths, 1989).

C2 photosynthesis is believed to be an evolutionary "stepping stone" between the ancestral C3 and C4 (Khoshravesh et al., 2016). C2 plants use decarboxylation of photorespiratory glycine in bundle sheath cells as a CCM. C2 plants do not generate intermediates for the purpose of raising CO_2 levels, but rather they more efficiently use photorespiratory byproducts. The intraplastidial CO_2 concentrations in C2 plants has been shown to be 3 times higher than in related C3 species

(Keerberg et al., 2014). Modelling predicts a broad environmental range in which C2 photosynthesis would result in greater carbon fixation compared to its C3 counterpart, and C2 has been suggested to be a viable method of crop improvement (Bellasio and Farquhar, 2019; Lundgren, 2020).

In addition to the photosynthetic diversity in plants, other photosynthetic organisms have unique CCMs such as the cyanobacterial carboxysome-based CCM. The carboxysome sequesters Rubisco and carbonic anhydrase inside of a protein shell, where carbonic anhydrase can convert carbonic acid to water and carbon dioxide to raise CO_2 concentration for Rubisco. Two different types of carboxysome exist from different evolutionary lineages and each with their cognate Rubiscos, termed α-carboxysomes and β-carboxysomes (Badger et al., 2002). There has been substantial interest in engineering cyanobacterial carboxysomes and their Rubiscos into C3 plants to improve photosynthetic efficiency (Hanson et al., 2016). Rubisco has been successfully encapsulated inside an α-carboxysome shell in tobacco chloroplasts (Long et al., 2018). The transgenic tobacco was capable of growth in elevated CO_2. This represents an important step towards full import of a functional carboxysome into C3 plants.

In many algae, there is a structure that is functionally similar to the carboxysome known as the pyrenoid, this is an important feature in algae as they live in an aqueous environment with lower CO_2 than is present in air. As with cyanobacterial carboxysomes, there is interest in engineering algal pyrenoids into higher plants to improve photosynthetic efficiency. Some important steps have been made towards this goal. Algal Rubiscos contain α helical loops on the small subunit required for recruitment to the pyrenoid. Atkinson et al. (2017) created chimeric *Chlamydomonas reinhardtii-Arabidopsis thaliana* SSUs containing these α helical loops to enable such recruitment of Rubisco to the microcompartment. This is an important step in engineering pyrenoids into higher plants (Sharwood, 2017).

1.2.2 Rubisco Diversity in Form and Function

In addition to the large scale differences between photosynthetic pathways which require anatomical and morphological changes, individual components of the photosynthetic apparatus can vary significantly between and within species. There is substantial diversity in the structural and functional properties of Rubiscos from different species (Tabita et al., 2008). The most abundant type of Rubisco is form I, this is found in algae (Taylor et al., 2001), plants and most photosynthetic bacteria, and is the Rubisco form focused on in primarily in this text. There are also three other forms of Rubisco: Form II Rubisco is found in some bacteria, and is composed solely of large subunits, as opposed to a hexadecamer composed of both large and small subunits (L_8S_8 in Form I). Form III is found in archaea and is notable for having exceptionally low substrate specificity, a feature which is not strongly selected for as archaea are rarely exposed to high concentrations of molecular oxygen. Form IV is structurally similar to canonical form I Rubisco but does not share traditional Rubisco carboxylase or oxygenase functions and is termed a Rubisco-like protein (Tabita et al., 2008). The lowest substrate specificity factor (a measure of the relative capacity

for carboxylation vs oxygenation reactions) is found in form III Rubisco and the highest in form I. The lowest carboxylation rate at 25 °C is found in form III and form II/III, an intermediate form of Rubisco from archaea (Iñiguez et al., 2020). This highlights the co-evolution of Rubisco with its respective CCMs and how the presence or absence of a CCM can influence Rubisco kinetics.

Within form I Rubisco isoforms there are catalytic differences. For example, Paniceae Rubisco has been shown to be diverse in its activity across a range of temperatures (Sharwood et al., 2016). Substrate specificity and catalytic velocity are also highly variable properties in Rubisco. It is often posited that a trade-off exists between these two catalytic properties, i.e. an increase in one will cause a concomitant decrease in the other. The degree to which the kinetics of Rubisco are constrained is a point of debate, with some arguing Rubisco can have changes in one kinetic factor without seriously impacting its other kinetics, and some arguing it cannot (Cummins et al., 2018; Flamholz et al., 2019; Tcherkez et al., 2006; Tcherkez et al., 2018). Distinguishing between CO_2 and O_2 is a difficult task for Rubisco as the two molecules are similar, and largely featureless. It is hence unsurprising that increasing the carboxylation rate of Rubisco is usually accompanied by a proportional increase in oxygenation rate. Analysis of over 300 Rubiscos revealed that over 95% of Rubiscos operate at between 1 and 10 rounds of catalysis per second, and specificity factor varies by only 30% even between C3 plants and plants containing a CCM (Flamholz et al., 2019). More recently, a Rubisco with an unusually fast carboxylation rate was found, with substantially higher activity than the next fastest known Rubisco, by interrogation of Rubiscos from all domains of life (Davidi et al., 2020).

It should be noted that this is not the only area of photosynthesis with considerable genetic diversity. Some other examples of photosynthetic traits with interesting and potentially exploitable diversity, the detailed discussion of which is beyond the scope of this review, are CBB cycle metabolites (Arrivault et al., 2019), and electron transport. By way of example, an alternative electron transport chain in *Chlamydomonas reinhardtii* produces molecular hydrogen in times of sulphur deficiency (Hemschemeier and Happe, 2011). Chapter 2 explores the genetic diversity of Rubisco activase in cowpea, another diverse photosynthetic enzyme.

1.3 Rubisco Activase

To perform its carboxylase or oxygenase functions, Rubisco must first have a carbamylated active site. Both Mg^{2+} and a non-substrate CO_2 are required to carbamylate Lys210 and hence prime Rubisco for a round of catalysis (Lorimer et al., 1976; Stec, 2012). Rubisco is prone to dead-end inhibition by sugar phosphates. When sugar phosphates bind to the active site of Rubisco, the enzyme can become inactivated. Examples of these inhibitors include RuBP (the substrate of Rubisco), xylulose 1,5-bisphosphate (XuBP), 2-carboxy-D-arabinitol 1-phosphate (CA1P), fructose-1,6-bisphosphate and glycero-2,3-pentodiulose-1,5-bisphosphate (PDBP) (Andralojc et al., 2012; Zhu and Jensen, 1991). RuBP is inhibitory only when bound to a non-carbamylated Rubisco active site. Other inhibitors more commonly bind to, or arise from, a carbamylated active site, but

can bind when non-carbamylated. Rubisco inhibitors can result from catalytic misfires (such as PDBP and XuBP) but can also be actively synthesised to serve a functional role, such as the nocturnal inhibition of Rubisco by CA1P (Gutteridge et al., 1986; Orr et al., 2023; Servaites, 1990). While Rubisco can reactivate spontaneously, this happens at low rates. Sustained Rubisco activity requires the action of Rubisco's catalytic chaperone - Rubisco activase (Rca) (Lilley and Portis, 1990; Portis et al., 1986).

1.3.1 Structure and Function

Rca is an AAA+ protein which uses energy from ATP hydrolysis to remodel the active site of Rubisco and allow release of the inhibitor. It activates active sites of Rubisco one at a time by interfacing with both the large and small subunits of Rubisco (Bhat et al., 2017). There is evidence to suggest the functional oligomeric state of Rca is a toroid-shaped hexamer (Waheeda et al., 2023). The reactivation activity of Rca is coupled to the ATP hydrolysis-dependent dissociation of the hexamer form into a tetramer and a dimer (Serban et al., 2018). The oligomerisation state of Rca is ATP dependent, with higher concentrations of ATP, or ATP analogues such as ATPγS, leading to larger oligomers and lowering the rate of subunit exchange (Wang et al., 2018). Figure 1.1 shows the dual activities of Rca and important structural features.

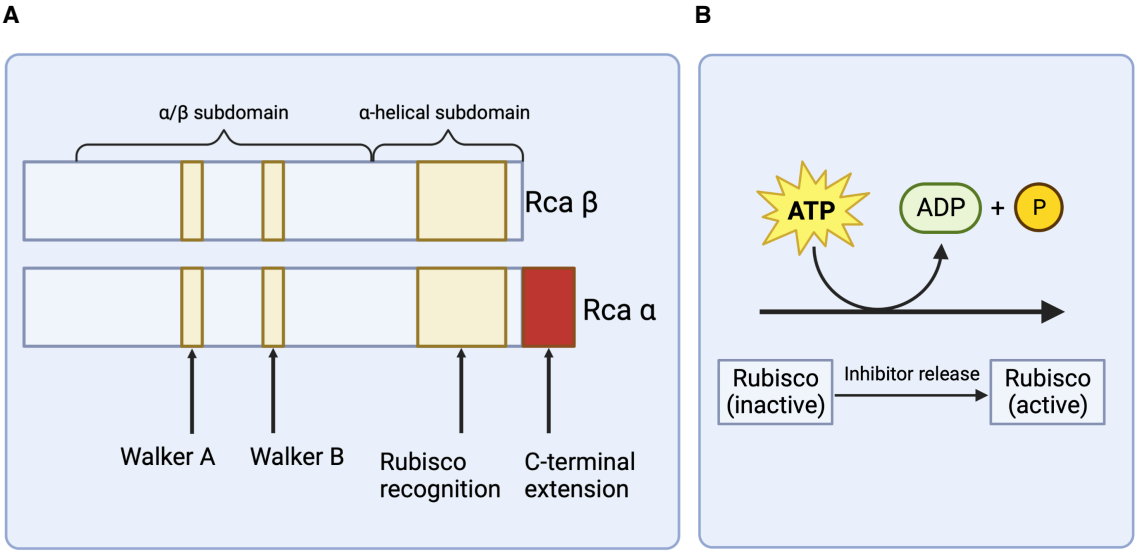


Figure 1.1: *Simplified schematics of Rca structure and function.* ***A*** *Important motifs in Rca. The Walker A & B motifs are involved in nucleotide binding. The Rubisco recognition motif binds Rubisco and confers specificity. In α isoforms, the C-terminal extension controls Rca activity via a disulfide bridge. The chloroplast transit peptide is not shown as this is cleaved from the mature protein.* ***B*** *Two activities of Rca. Rca hydrolyses ATP into ADP and inorganic phosphate. Using the energy of ATP hydrolysis, Rca remodels Rubisco to release its bound inhibitory sugar phosphate. Created with BioRender.com*

Rca is well conserved, being found in all known photosynthetic species, and often with multi-

ple unique isoforms present in the same genome. These different isoforms often show different regulatory and induction patterns (Degen et al., 2021). Rca is broadly categorised into two distinct types, made by alternative splicing or encoded on different genes. There is the shorter β isoform and the longer α isoform which contains a redox sensitive C-terminal extension which can be oxidised by thioredoxin-f to inhibit activase activity (Zhang and Portis, 1999). Rca oligomers can form between different isoforms of Rca, and exchange these subunits. Mutated subunits can ”poison” an Rca hexamer and make it nonfunctional, this was used to investigate rice Rca-Rubisco interactions by Shivhare et al. (2019). It was found that mutating residues closer to the periphery of the toroid resulted in less poisoning than mutating residues closer to the centre of the toroid. This implies that residues closer to the periphery are necessary for docking onto Rubisco, and those closer to the centre are important for Rubisco remodelling. Residues in α4-β4, a loop in the Rca nucleotide binding motif, showed some poisoning, implying they were of importance in transmitting of forces but not key in Rubisco remodelling.

ADP usually acts as an inhibitor of Rca function, however activity varies between Rca isoforms. For example, the different isoforms of wheat Rca respond differently to ADP in terms of their Rubisco activation activities. Wheat Rca 2β Rubisco activation is insensitive to ADP concentration while 1β and 2α are inhibited (Perdomo et al., 2019).

Rca is usually capable of reactivating not only its cognate Rubisco but also Rubiscos from other species. The curious exception to this rule is Rca from plants of the Solanaceae family. Activase from Solanaceae species is a poor activator of non-Solanaceae Rubisco and vice-versa (Li et al., 2005; Wang et al., 1992). This knowledge has been used to investigate the interaction model between Rubisco and Rca. A hybrid Rubisco composed of tobacco (Solanaceae) small subunits and *Chlamydomonas reinhardtii* large subunits can be efficiently activated by non-Solanaceae Rca but inefficiently by Solanaceae Rca, implying that the small subunit does not play an important role in Rubisco-Rca recognition and docking (Wachter et al., 2013).

Under steady state and saturating light conditions, Rca is usually not a limiting factor, but under fluctuating light, Rca can be limiting. Rca may provide an opportunity for improving sun-shade and shade-sun transitions (Carmo-Silva and Salvucci, 2013; Mott and Woodrow, 2000).

Interestingly, Rubisco activation state correlates negatively with Rubisco amount, and Rca expression levels correlate negatively with Rubisco content (Carmo-Silva et al., 2017). Fukayama et al. (2018) investigated this phenomenon and found no difference in expression of other CBB cycle proteins in transgenic rice overexpressing Rca. They attributed the lower Rubisco content to interference in the synthesis stage of Rubisco, and also noted the slightly higher polysome loading of RbcL mRNA. By contrast, Suganami et al. (2020) investigated the same phenomenon in rice grown under different nitrogen availabilities. The decrease of Rubisco in Rca-overexpressing lines was attributed to a decrease in polysome loading.

1.3.2 Role in Thermotolerance

Rca plays an important role in the thermotolerance of plants. Rubisco itself is rather tolerant to high temperatures, being an active carboxylase in temperatures at least as high as 50 °C (Salvucci and Crafts-Brandner, 2004a), but Rca is thermolabile and will aggregate and fail to reactivate Rubisco at much lower temperatures (Feller et al., 1998). The thermal optimum of Rca and the highest temperature it will tolerate is dependent on species and the environment that species inhabits (Salvucci and Crafts-Brandner, 2004b; Salvucci et al., 2001). *Camelina sativa* and *Arabidopsis* for example are plants adapted to cooler climates and have thermal optima at lower temperatures than *Nicotiana tabacum* and *Gossypium hirsutum*, plants adapted to hotter climates (Carmo-Silva and Salvucci, 2011). This thermolability can make Rca one of the limiting factors for photosynthetic activity, and capacity for plant growth at higher temperatures (Carmo-Silva and Salvucci, 2012; Scafaro et al., 2012).

Thermotolerance in some species can be due to activase adapted to withstand higher temperatures. For example, the rice wild relative *Oryza australiensis* from the northern Australian savannah maintains much higher leaf elongation rate and Rubisco activation state than *Oryza sativa* at higher temperatures. These characteristics and the ability of *O. australiensis* to survive and actively grow in such hot conditions are attributable to its activase being able to stay active at these temperatures (Scafaro et al., 2016). The use of chimeric activases allowed discovery of the amino acid region which gives *Agave tequilana*, a CAM plant, its thermotolerant attributes. *A. tequilana* activase remains functional even at 45 °C, and this thermotolerance maps to the N terminal 244 amino acids. A chimeric activase comprised of the N terminal 244 amino acids of *A. tequilana* and the C terminal region of *Oryza sativa* activase displayed thermostability like *Agave tequilana* while the inverse had the thermostability of *O. sativa* (Shivhare and Mueller-Cajar, 2017).

Plants with multiple isoforms of Rca can induce a more thermostable isoform in response to heat stress (Degen et al., 2021). Wheat (*Triticum aestivum*) induced Rca 1β and suppress expression of Rca 2α and 2β in response to heat stress. TaRca1β is much more tolerant of heat than the TaRca2 isoforms, with a thermal midpoint (the temperature at which activity is half of maximum) of 42 °C instead of 36 °C for the TaRca2 isoforms. This difference in thermotolerance can be attributed to just 11 amino acids, which can be mutated without impacting catalytic velocity at 25 °C (Scafaro et al., 2019a). A single amino acid difference from wheat Rca 1β vs 2β (M159I) has been found to increase the temperature optimum by 5 °C (Degen et al., 2020). It is peculiar that Rca in wheat has evolved to induce a thermostable isoform when under heat stress instead of constitutively expressing this isoform. The aforementioned thermostable *Agave tequilana* Rca does not compromise catalytic efficiency for thermostability, even having an above average rate of Rubisco activation (Shivhare and Mueller-Cajar, 2017). An activase isoform with a broad thermal optimum may enable plant adaptation to high temperatures and a larger temperature range. This could aid in crop climate resilience and increase growth in unstable or non-optimal climate conditions.

There is some debate as to the cause of the thermolabile nature of Rca and the role of this

in limiting photosynthesis at higher temperatures (discussed in Sage et al., 2008). Some argue that the thermolabile nature is a flaw of the enzyme and the cause of photosynthetic limitation at higher temperatures due to decreased Rubisco activation state (Crafts-Brandner and Salvucci, 2000). Another hypothesis is that it is a protective mechanism to inactivate Rubisco when having it active would have deleterious effects (Sharkey, 2005). Photosynthetic electron transport decreases at moderate heat stress and the rate of photorespiration increases compared to carboxylation at higher temperatures. This makes active Rubisco potentially dangerous for the cell as active Rubisco without electron transport leads to energy imbalance and photorespiration can lead to build up of H_2O_2, a reactive oxygen species which can damage cells at high concentrations (Berry and Bjorkman, 1980; Song et al., 2014). Furthermore, there is a potential role in Rubisco inhibition for Rubisco protection. Overexpressing CA1Pase, a phosphatase which dephosphorylates the Rubisco inhibitor CA1P was anticipated to increase net photosynthesis by increasing Rubisco activation state in wheat. However, it was found that this overexpression decreased Rubisco abundance, implying a mechanism by which Rubisco inhibition may protect Rubisco, and provide a net fitness benefit (Lobo et al., 2019).

Single amino acid residue changes can have substantial impact on Rca function. Table 1.1 shows some of the known residues which are important for Rca functional diversity.

Table 1.1: *Important Rubisco activase isoforms from various species and their functional roles in the protein, from select publications.*

Relevant residue	Function	Citation(s)
Tobacco: 311D; Spinach: 311K; Creosote: 313K	Rubisco specificity	Henderson et al., 2011; Li et al., 2005
Tobacco: 312K; non-solanaceae: 312K(R)	Rubisco interaction	Li et al., 2005; Portis et al., 2007
Tobacco: 314L; Spinach: 316V; Creosote: 316V	Rubisco specificity	Henderson et al., 2011; Li et al., 2005
Tobacco: 16W	Rubisco binding	van de Loo and Salvucci, 1996
Arabidopsis: 390E	ATP dissociation	Wang and Portis, 2006
Arabidopsis: 401D	ATP dissociation	Wang and Portis, 2006
Wheat 1β: 159I; Wheat 2β: 159M	Temperature response, ADP sensitivity	Degen et al., 2020; Perdomo et al., 2019
Wheat 1β: 161N; Wheat 2β: 161K	ADP sensitivity	Perdomo et al., 2019
Wheat 1β: 359K; Wheat 2β: 359T	ADP sensitivity	Perdomo et al., 2019
Wheat 1β: 362E; Wheat 2β: 362Q	ADP sensitivity	Perdomo et al., 2019
Rice: 148K	ATPase - activase activity coupling	Shivhare et al., 2019
Rice: 321R	ATPase - activase activity coupling	Shivhare et al., 2019
Rice: 152Q	ATPase & activase activity	Shivhare et al., 2019
Rice: 156E	ATPase & activase activity	Shivhare et al., 2019
Rice: 159D	Activase activity	Shivhare et al., 2019
Tobacco: 241R	ATPase activity	Li et al., 2006
Tobacco: 244R	ATPase activity	Li et al., 2006
Tobacco: 294R	ATPase activity	Li et al., 2006
Tobacco: 296R	ATPase activity	Li et al., 2006
Rice: 215N	Activase activity	Shivhare and Mueller-Cajar, 2017
Rice: 216K	Activase activity	Shivhare and Mueller-Cajar, 2017
Rice: 217E; Agave: 217Q	Activase activity	Shivhare and Mueller-Cajar, 2017
Arabidopsis: 78T	Light induction (phosphorylation site)	Kim et al., 2016
Arabidopsis: 172S	Phosphorylation site	Boex-Fontvieille et al., 2013
Tobacco: 109W	ATP association	Wang and Portis, 2006
Tobacco: 250W	ATP association	Wang and Portis, 2006
Arabidopsis: 392C	Redox regulation	Zhang et al., 2002
Arabidopsis: 411C	Redox regulation	Zhang et al., 2002
Arabidopsis: 111Q	Phosphate binding	Kallis et al., 2000
Wheat 1β: 404K	Temperature response	Scafaro et al., 2019a
Wheat 1β: 406L	Temperature response	Scafaro et al., 2019a
Wheat 1β: 408E	Temperature response	Scafaro et al., 2019a

1.4 Scope for Photosynthetic Improvements

There have been numerous attempts at improving rates of photosynthesis with the goal of improving crop yields (Long et al., 2015). This is seen by many as the next solution to the need for increased food production. However, there is some debate as to whether or not increased photosynthesis rates will necessarily improve yields. This issue appears situation specific, and the impact of increased photosynthesis on yield depends on the crop species, environment and photosynthetic trait being enhanced (Ainsworth and Long, 2005; Wu et al., 2019).

Engineering Rubisco with the goal of improving photosynthesis has proven a significant challenge. There is substantial genetic diversity between Rubiscos but most share similar catalytic properties and the difficulty of genetic transformations makes this difficult to utilise. Furthermore while the commonly stated axiom is that Rubisco is a sluggish and inefficient enzyme, there is some debate around this. Bathellier et al. (2018) reviewed Rubisco in the context of other en-

zymes, in light of which its 1-10 rounds of catalysis per second seemed a lot more reasonable, with the median of over 2,000 enzymes being 10 per second. Furthermore, in light of the complexity of reactions required by Rubisco, the catalytic velocity and substrate affinity seem rather reasonable, and its oxygenase activity is not unique to Rubisco, but rather a quite common feature amongst carboxylases. Nonetheless, they do conclude that there remains potential for improvement of Rubisco (Bathellier et al., 2018 and references within).

Simple overexpression of Rubisco is not a trivial task, requiring both overexpression of the large and small subunits and RAF1, a Rubisco assembly factor. It has also been met with varying results. Overexpression of Rubisco in maize has been met by a concomitant decrease in activation state, but maintained a net increase in carbon assimilation (Salesse-Smith et al., 2018). Whereas in rice, an increase in photosynthetic efficiency and an increase in yield due to overexpression of Rubisco was reported (Yoon et al., 2020).

To date, little is known about the genetic diversity of Rca outside of crops and model organisms. Rca is an important factor in photosynthesis, in terms of thermotolerance, response to changing light conditions and controlling abundance and activation state of Rubisco. Further research into activase genetic diversity from species from a wide range of environmental niches may provide useful information on adapting plants to tolerate high temperatures or more quickly adapt to light fluctuations.

1.5 *Vigna unguiculata*

Cowpea (*Vigna unguiculata* (L.) Walp) is a legume crop - a member of the *Fabaceae* family, which is noteworthy for harbouring nitrogen-fixing bacteria in their root nodules. Domestic cowpea is an annual crop, and has considerable morphological diversity between cultivars. It originated, and was domesticated in, western sub-Saharan Africa (Herniter et al., 2020). Cowpea is critical to food security in large regions of Africa, especially in Nigeria where it is produced in the largest quantities, and its leaf and seed are a major source of caloric intake (Carvalho et al., 2017; Muñoz-Amatriaín et al., 2017). Cowpea seeds are valued for their diverse nutrients and high protein content (Jayathilake et al., 2018); the leaves are also rich in nutrients such as β-carotene but preservation of leaves is challenging (Owade et al., 2020).

1.5.1 Trends in Production and Consumption

Cowpea production in Nigeria has been trending upwards, but the rate of yield increase is slowing. The area in which cowpea is cultivated has also been increasing since the 1980s (Fig. 1.2).

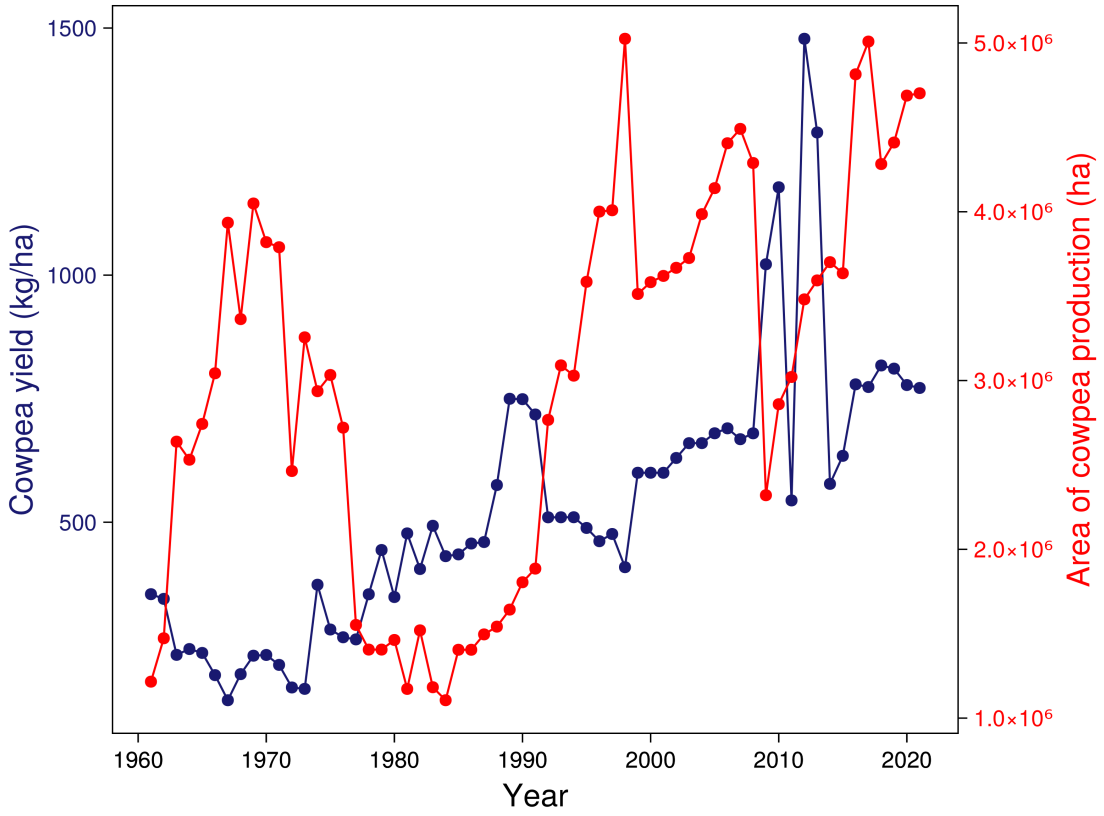


Figure 1.2: *Trends in cowpea production in Nigeria. Data from FAOSTAT.*

The productivity of cowpea is relatively low compared to the theoretical potential grain yield of 2,000 kg/ha (Fig. 1.2, Horn et al., 2022 and references therein). Some possible reasons for this are soil quality, drought and pests (Nkomo et al., 2021). The relatively low per area production of cowpea and the threat of climate change call for improvements to existing cowpea cultivars and production techniques.

1.5.2 Growth Conditions

Cowpea is considered a stress tolerant crop. It is resistant to heat, salt, and drought stress. (Agbicodo et al., 2009; Carvalho et al., 2019) with an optimal temperature for germination in the range of 30-35 °C (Barros et al., 2020). Cowpea is capable of growth in a broad range of temperatures, from as low as 17 °C to as high as 37 °C. However, temperatures of 30 °C and above have been shown to reduce crop yield (Barros et al., 2021; Barros et al., 2022 and references therein). This suggests that Rca may be targeted as a method of increasing cowpea yield under a broad range of temperatures (Parry et al., 2013).

1.6 Recombinant Protein Production

Heterologously expressed proteins are commonly used in Rca research. This allows for selective analysis of individual isoforms as opposed to all isoforms purified from a plant source. In this project, isoforms of cowpea Rca are purified from *Escherichia coli* by IMAC (Immobilized Metal Affinity Chromatography) purification (Block et al., 2009). To purify recombinant proteins by IMAC, the gene for the desired protein is inserted into a bacterial host. This gene will contain a polyhistidine, aka His, tag - a sequence of typically six contiguous histidine residues. The host is cultured in a shake flask or fermenter, and expression is induced. The host is then harvested, typically centrifugated, and lysed to expose intercellular contents, including the target protein. Cellular lysate is clarified by centrifugation, and loaded onto an IMAC column. This column contains a metal ion resin, which will bind to the His tag. The protein can then be eluted in pure form by the addition of a competitive binder such as imidazole (Spriestersbach et al., 2015).

Although a routine procedure, there are several steps in this process that can be optimised to improve the final protein yield and quality. Bacterial growth medium, length of expression and pre-expression growth stages, and growth temperature can all be optimised to improve protein yield and solubility (Ahmad et al., 2018). Composition of buffers used in purification, and lysis parameters can affect protein recovery and quality. The effects of these factors on Rca purification is not well characterised in current literature. The optimisation of cowpea Rca expression and purification forms the basis of Chapter 3.

1.7 Antibodies

Antibodies are proteins produced by animals as part of the adaptive immune system. These are large molecules (~150 kDa), which are composed of two heavy chains and two light chains. Together, these form a Y-shaped molecule, with a variable domain that controls the specific antigen to which the antibody binds (Stanfield and Wilson, 2014). The variability of antibody structure through a process known as V(D)J recombination allows for antibodies to be produced against almost any target antigen (Chi et al., 2020). Antibodies can bind their target with nanomolar affinity and high specificity, qualities that have made them an attractive and indispensable tool in research.

1.7.1 Uses of Antibodies

Antibodies have found many uses as research tools in molecular biology. Some of the most common uses include:

- Selective purification of an antibody's target by immunoprecipitation
- Detection or quantification of an antigen by western blot/dot blot/ELISA
- *In vivo* detection and localisation of a target by immunohistochemistry

- Cell labelling in flow cytometry

Care must be taken when using antibodies in research, as antibodies are not guaranteed to bind their intended target specifically, and they require careful validation (Edfors et al., 2018). Antibodies are valuable in Rca research (see Salesse-Smith et al. (2018) and Yin et al. (2014) for examples of anti-Rca antibodies in research). As previously discussed, multiple, similar isoforms of Rca are often found in the same species. Antibodies can aid in identifying and quantifying isoforms that cannot be distinguished by other methods, such as electrophoresis.

1.7.2 Antibody Production

There are three commonly employed types of antibodies used as research tools - polyclonal, monoclonal, and recombinant antibodies. Figure 1.3 details these different methods of producing antibodies.

Polyclonal antibodies are produced in animals, commonly small mammals such as mice and rabbits. The blood serum of an immunised animal is extracted, and the antibodies are purified from this. This will contain all the antibody clones in the serum, hence the term polyclonal. This may include antibodies against targets other than the intended antigen, and will target several sites on the antigen, which can result in a higher signal.

The beginning stages of monoclonal antibody production are the same as those of polyclonal antibodies. An animal is repeatedly exposed to a target antigen to invoke an immune response. B cells producing antibodies are fused with a myeloma to produce an immortalised cell line termed a hybridoma that continues to produce antibodies. These antibodies can be reproduced as the hybridoma can be indefinitely grown, but the precise antibody sequence may change due to genetic drift. See Lipman et al. (2005) for information on the differences between monoclonal and polyclonal antibodies.

Recombinant antibodies are produced *in vitro* from a library of synthetic antibody genes. These genes are produced synthetically and inserted into viral capsids. The library is initially derived from an animal source, but no animals are needed after the initial library construction. Antibodies that bind the intended antigen are selected using *in vitro* selection methods, such as phage display (Shukra et al., 2014). After sequencing, recombinant antibodies are able to be produced indefinitely.

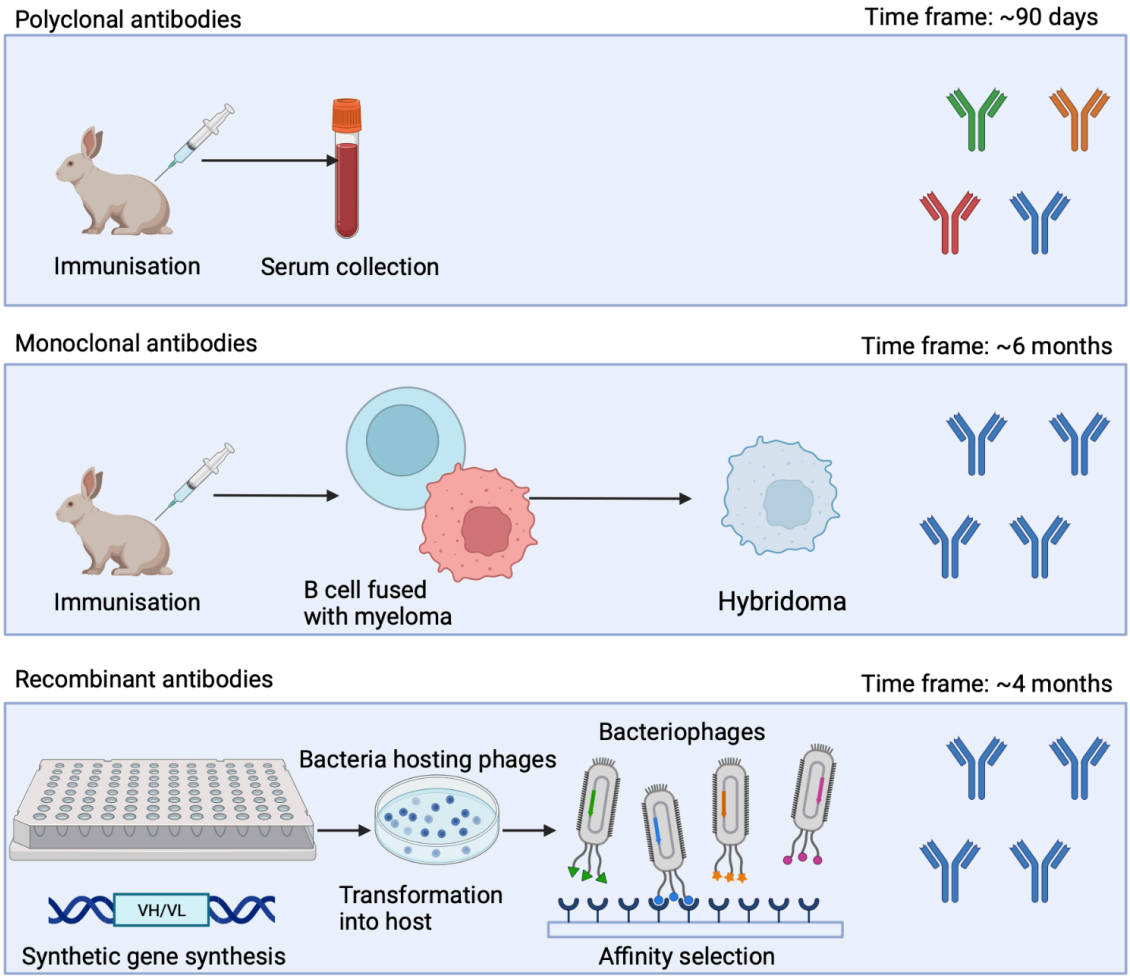


Figure 1.3: *Comparison of antibody production methods.* ***Polyclonal antibodies:*** *an animal such as a mouse or rabbit is exposed to adjuvants and the target antigen by injection. The blood serum is extracted, containing polyclonal antibodies against the antigen.* ***Monoclonal antibodies:*** *an animal such as a mouse or rabbit is exposed to adjuvants and the target antigen by injection. B cells producing antibodies against the target are fused with myeloma cells to produce a hybridoma expressing the desired antibody. This can be cultured indefinitely.* ***Recombinant antibodies:*** *a library of antibody variable heavy/light chain genes are produced by synthetic gene synthesis. These are cultured in bacteria and used for* in vitro *affinity selection by methods such as phage display. The selected antibodies are sequenced and produced* in vitro*. VH: variable heavy chain; VL: variable light chain. Created with BioRender.com*

1.8 Thesis Objectives

The primary objectives of this thesis are to characterise Rca in cowpea, and to explore the genetic diversity of Rca in diverse plant species. While cowpea is the main focus of this study, Rca from soybean (*Glycine max*) is included in some experiments. Soybean is another *Fabaceae* crop, that is grown more globally than cowpea is, and can serve as a reference or comparison. **Chapter 2**

investigates Rca sequences of different *Fabaceae* species. By taking sequence data in tandem with climatological data, three species were identified as candidates with potentially thermotolerant Rca isoforms.

Chapter 3 is an optimisation study which improves on methods of recombinant Rca expression and purification. This work was essential to obtain sufficient recombinant cowpea Rca for *in vitro* experiments.

Chapter 4 investigates the sequence diversity and catalytic properties of Rca from cowpea. Individual isoforms were characterised for their temperature response and, where possible, concentration response.

Chapter 5 discusses the development and validation of new tools for research into cowpea Rca in the form of antibodies. Antibodies were produced which were intended to bind all isoforms of Rca, and which are intended to bind only specific isoforms of cowpea Rca.

Introduction Bibliography

Acevedo-Siaca, L. G., Coe, R., Wang, Y., Kromdijk, J., Quick, W. P., & Long, S. P. (2020). Variation in photosynthetic induction between rice accessions and its potential for improving productivity. *The New Phytologist*. https://doi.org/10.1111/nph.16454

Agbicodo, E. M., Fatokun, C. A., Muranaka, S., Visser, R. G. F., & Linden van der, C. G. (2009). Breeding drought tolerant cowpea: Constraints, accomplishments, and future prospects. *Euphytica*, *167*(3), 353–370. https://doi.org/10.1007/s10681-009-9893-8

Ahmad, I., Nawaz, N., Darwesh, N. M., ur Rahman, S., Mustafa, M. Z., Khan, S. B., & Patching, S. G. (2018). Overcoming challenges for amplified expression of recombinant proteins using Escherichia coli. *Protein Expression and Purification*, *144*, 12–18. https://doi.org/10.1016/j.pep.2017.11.005

Aigner, H., Wilson, R. H., Bracher, A., Calisse, L., Bhat, J. Y., Hartl, F. U., & Hayer-Hartl, M. (2017). Plant RuBisCo assembly in *E. Coli* with five chloroplast chaperones including BSD2. *Science (New York, N.Y.)*, *358*(6368), 1272–1278. https://doi.org/10.1126/science.aap9221

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. *New Phytologist*, *165*(2), 351–372. https://doi.org/10.1111/j.1469-8137.2004.01224.x

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiology and Biochemistry*, *46*(3), 275–291. https://doi.org/10.1016/j.plaphy.2008.01.001

Andralojc, P. J., Madgwick, P. J., Tao, Y., Keys, A., Ward, J. L., Beale, M. H., Loveland, J. E., Jackson, P. J., Willis, A. C., Gutteridge, S., & Parry, M. A. (2012). 2-Carboxy-D-Arabinitol 1-Phosphate (CA1P) phosphatase: Evidence for a wider role in plant Rubisco regulation. *Biochemical Journal*, *442*(3), 733–742. https://doi.org/10.1042/BJ20111443

Arrivault, S., Alexandre Moraes, T., Obata, T., Medeiros, D. B., Fernie, A. R., Boulouis, A., Ludwig, M., Lunn, J. E., Borghi, G. L., Schlereth, A., Guenther, M., & Stitt, M. (2019). Metabolite profiles reveal interspecific variation in operation of the Calvin–Benson cycle in both C4 and C3 plants. *Journal of Experimental Botany*, *70*(6), 1843–1858. https://doi.org/10.1093/jxb/erz051

Atkinson, N., Leitão, N., Orr, D. J., Meyer, M. T., Carmo-Silva, E., Griffiths, H., Smith, A. M., & McCormick, A. J. (2017). Rubisco small subunits from the unicellular green alga Chlamydomonas complement Rubisco-Deficient mutants of Arabidopsis. *New Phytologist*, *214*(2), 655–667. https://doi.org/10.1111/nph.14414

Badger, M. R., Hanson, D., & Price, G. D. (2002). Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Functional Plant Biology*, *29*(3), 161–173. https://doi.org/10.1071/pp01213

Barros, J. R. A., Angelotti, F., Santos, J. D. O., e Silva, R. M., Dantas, B. F., & de Melo, N. F. (2020). Optimal temperature for germination and seedling development of cowpea seeds. *Revista Colombiana de Ciencias Hortícolas*, *14*(2), 231–239. https://doi.org/10.17584/rcch.2020v14i2.10339

Barros, J. R. A., Guimarães, M. J. M., e Silva, R. M., Rêgo, M. T. C., de Melo, N. F., de Melo Chaves, A. R., & Angelotti, F. (2021). Selection of cowpea cultivars for high temperature tolerance: Physiological, biochemical and yield aspects. *Physiology and Molecular Biology of Plants*, *27*(1), 29–38. https://doi.org/10.1007/s12298-020-00919-7

Barros, J. R. A., Guimarães, M. J. M., Simões, W. L., de Melo, N. F., & Angelotti, F. (2022). Temperature: A major climatic determinant of cowpea production. *Acta Scientiarum. Agronomy*, *45*, e56812. https://doi.org/10.4025/actasciagron.v45i1.56812

Batista-Silva, W., da Fonseca-Pereira, P., Martins, A. O., Zsögön, A., Nunes-Nesi, A., & Araújo, W. L. (2020). Engineering Improved Photosynthesis in the Era of Synthetic Biology. *Plant Communications*, *1*(2), 100032. https://doi.org/10.1016/j.xplc.2020.100032

Bäurle, I. (2016). Plant Heat Adaptation: Priming in response to heat stress. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.7526.1

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Bellasio, C., & Farquhar, G. D. (2019). A Leaf-level biochemical model simulating the introduction of C $_2$ and C $_4$ photosynthesis in C $_3$ rice: Gains, losses and metabolite fluxes. *New Phytologist*, *223*(1), 150–166. https://doi.org/10.1111/nph.15787

Benson, A., & Calvin, M. (1947). The Dark Reductions of Photosynthesis. *Science*, *105*(2738), 648–649. https://doi.org/10.1126/science.105.2738.648

Berry, J., & Bjorkman, O. (1980). Photosynthetic Response and Adaptation to Temperature in Higher Plants. *Annual Review of Plant Physiology*, *31*(1), 491–543. https://doi.org/10.1146/annurev.pp.31.060180.002423

Bhat, J., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J. R., Hartl, F. U., Wendler, P., & Hayer-Hartl, M. (2017). Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Molecular Cell*, *67*(5), 744–756.e6. https://doi.org/10.1016/j.molcel.2017.07.004

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). In *Methods in Enzymology* (pp. 439–473). Elsevier. https://doi.org/10.1016/S0076-6879(09)63027-5

Bock, R. (2014). Genetic engineering of the chloroplast: Novel tools and new applications. *Current Opinion in Biotechnology*, *26*, 7–13. https://doi.org/10.1016/j.copbio.2013.06.004

Burgess, S. J., & Hibberd, J. M. (2015). Insights into C 4 metabolism from comparative deep sequencing. *Current Opinion in Plant Biology*, *25*, 138–144. https://doi.org/10.1016/j.pbi.2015.05.017

Carmo-Silva, E., Andralojc, P. J., Scales, J. C., Driever, S. M., Mead, A., Lawson, T., Raines, C. A., & Parry, M. A. J. (2017). Phenotyping of field-grown wheat in the UK highlights contribution of light response of photosynthesis and flag leaf longevity to grain yield. *Journal of Experimental Botany*, *68*(13), 3473–3486. https://doi.org/10.1093/jxb/erx169

Carmo-Silva, E., Gore, M. A., Andrade-Sanchez, P., French, A. N., Hunsaker, D. J., & Salvucci, M. E. (2012). Decreased CO2 availability and inactivation of Rubisco limit photosynthesis in cotton plants under heat and drought stress in the field. *Environmental and Experimental Botany*, *83*, 1–11. https://doi.org/10.1016/j.envexpbot.2012.04.001

Carmo-Silva, E., & Salvucci, M. E. (2012). The temperature response of CO2 assimilation, photochemical activities and Rubisco activation in Camelina sativa, a potential bioenergy crop with limited capacity for acclimation to heat stress. *Planta*, *236*(5), 1433–1445. https://doi.org/10.1007/s00425-012-1691-1

Carmo-Silva, E., & Salvucci, M. E. (2013). The Regulatory Properties of Rubisco Activase Differ among Species and Affect Photosynthetic Induction during Light Transitions. *Plant Physiology*, *161*(4), 1645–1655. https://doi.org/10.1104/pp.112.213348

Carmo-Silva, E., Scales, J. C., Madgwick, P. J., & Parry, M. a. J. (2015). Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant, Cell & Environment*, *38*(9), 1817–1832. https://doi.org/10.1111/pce.12425

Carvalho, M., Castro, I., Moutinho-Pereira, J., Correia, C., Egea-Cortines, M., Matos, M., Rosa, E., Carnide, V., & Lino-Neto, T. (2019). Evaluating stress responses in cowpea under drought stress. *Journal of Plant Physiology*, *241*, 153001. https://doi.org/10.1016/j.jplph.2019.153001

Carvalho, M., Lino-Neto, T., Rosa, E., & Carnide, V. (2017). Cowpea: A legume crop for a challenging environment: Cowpea for a challenging environment. *Journal of the Science of Food and Agriculture*, *97*(13), 4273–4284. https://doi.org/10.1002/jsfa.8250

Crafts-Brandner, S. J., & Salvucci, M. E. (2000). Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO2. *Proceedings of the National Academy of Sciences*, *97*(24), 13430–13435. https://doi.org/10.1073/pnas.230451497

Cummins, P. L., Kannappan, B., & Gready, J. E. (2018). Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All. *Frontiers in Plant Science*, *9*, 183. https://doi.org/10.3389/fpls.2018.00183

Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y. M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D. G., Antonovsky, N., Pins, B., Shachar, L., Hochhauser, D., Peleg, Y., Albeck, S., Sharon, I., Mueller-Cajar, O., & Milo, R. (2020). Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *The EMBO Journal*. https://doi.org/10.15252/embj.2019104081

de Souza, A. P., Wang, Y., Orr, D. J., Carmo-Silva, E., & Long, S. P. (2020). Photosynthesis across African cassava germplasm is limited by Rubisco and mesophyll conductance at steady state, but by stomatal conductance in fluctuating light. *New Phytologist*, *225*(6), 2498–2512. https://doi.org/10.1111/nph.16142

Degen, G., Worrall, D., & Carmo-Silva, E. (2020). An isoleucine residue acts as a thermal and regulatory switch in wheat Rubisco activase: Rubisco activase thermal and regulatory switch. *The Plant Journal*. https://doi.org/10.1111/tpj.14766

Erb, T. J., & Zarzycki, J. (2018). A short history of RubisCO: The rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Current Opinion in Biotechnology*, *49*, 100–107. https://doi.org/10.1016/j.copbio.2017.07.017

FAO. (2020). Crop Prospects and Food Situation - Quarterly Global Report No. 1, March 2020. Rome.

FAO, IFAD, UNICEF, WFP, & WHO. (2019). The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns. Rome, FAO.

FAOSTAT. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data Date of Access: 05-03-2023. (n.d.).

Faralli, M., & Lawson, T. (2020). Natural genetic variation in photosynthesis: An untapped resource to increase crop yield potential? *The Plant Journal*, *101*(3), 518–528. https://doi.org/10.1111/tpj.14568

Feller, U., Anders, I., & Mae, T. (2008). Rubiscolytics: Fate of Rubisco after its enzymatic function in a cell is terminated. *Journal of Experimental Botany*, *59*(7), 1615–1624. https://doi.org/10.1093/jxb/erm242

Feller, U., Crafts-Brandner, S. J., & Salvucci, M. E. (1998). Moderately High Temperatures Inhibit Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) Activase-Mediated Activation of Rubisco. *Plant Physiology*, *116*(2), 539–546. https://doi.org/10.1104/pp.116.2.539

Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y. M., Oltrogge, L. M., Alves, R., Savage, D., & Milo, R. (2019). Revisiting Trade-Offs between Rubisco Kinetic Parameters. *Biochemistry*, *58*(31), 3365–3376. https://doi.org/10.1021/acs.biochem.9b00237

Flexas, J., & Medrano, H. (2002). Drought-inhibition of photosynthesis in C3 plants: Stomatal and non-stomatal limitations revisited. *Annals of Botany*, *89*(2), 183–189. https://doi.org/10.1093/aob/mcf027

Fukayama, H., Mizumoto, A., Ueguchi, C., Katsunuma, J., Morita, R., Sasayama, D., Hatanaka, T., & Azuma, T. (2018). Expression level of Rubisco activase negatively correlates with Rubisco content in transgenic rice. *Photosynthesis Research*, *137*(3), 465–474. https://doi.org/10.1007/s11120-018-0525-9

Griffiths, H. (1989). Crassulacean Acid Metabolism: A Re-Appraisal of Physiological Plasticity in Form and Function. In J. A. Callow (Ed.), *Advances in Botanical Research* (pp. 43–92). Academic Press. https://doi.org/10.1016/S0065-2296(08)60044-0

Gutteridge, S., Parry, M. a. J., Burton, S., Keys, A. J., Mudd, A., Feeney, J., Servaites, J. C., & Pierce, J. (1986). A nocturnal inhibitor of carboxylation in leaves. *Nature*, *324*(6094), 274–276. https://doi.org/10.1038/324274a0

Hanson, M. R., Lin, M. T., Carmo-Silva, A. E., & Parry, M. A. J. (2016). Towards engineering carboxysomes into C3 plants. *The Plant Journal: For Cell and Molecular Biology*, *87*(1), 38–50. https://doi.org/10.1111/tpj.13139

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. (2013). Physiological, Biochemical, and Molecular Mechanisms of Heat Stress Tolerance in Plants. *International Journal of Molecular Sciences*, *14*(5), 9643–9684. https://doi.org/10.3390/ijms14059643

Hemschemeier, A., & Happe, T. (2011). Alternative photosynthetic electron transport pathways during anaerobiosis in the green alga Chlamydomonas reinhardtii. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1807*(8), 919–926. https://doi.org/10.1016/j.bbabio.2011.02.010

Henderson, J. N., Kuriata, A. M., Fromme, R., Salvucci, M. E., & Wachter, R. M. (2011). Atomic Resolution X-ray Structure of the Substrate Recognition Domain of Higher Plant Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase. *Journal of Biological Chemistry*, *286*(41), 35683–35688. https://doi.org/10.1074/jbc.C111.289595

Herniter, I. A., Muñoz-Amatriaín, M., & Close, T. J. (2020). Genetic, textual, and archeological evidence of the historical global spread of cowpea (*Vigna unguiculata* [L.] Walp.) *Legume Science*, *2*(4). https://doi.org/10.1002/leg3.57

Horn, L. N., Nghituwamata, S. N., & Isabella, U. (2022). Cowpea Production Challenges and Contribution to Livelihood in Sub-Saharan Region. *Agricultural Sciences*, *13*(01), 25–32. https://doi.org/10.4236/as.2022.131003

Iñiguez, C., Capó-Bauçà, S., Niinemets, Ü., Stoll, H., Aguiló-Nicolau, P., & Galmés, J. (2020). Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating

mechanisms. *The Plant Journal: For Cell and Molecular Biology*, *101*(4), 897–918. https://doi.org/10.1111/tpj.14643

IPCC. (2018). Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.

Jayathilake, C., Visvanathan, R., Deen, A., Bangamuwage, R., Jayawardana, B. C., Nammi, S., & Liyanage, R. (2018). Cowpea: An overview on its nutritional facts and health benefits: Nutritional and health properties of cowpea. *Journal of the Science of Food and Agriculture*, *98*(13), 4793–4806. https://doi.org/10.1002/jsfa.9074

Kallis, R. P., Ewy, R. G., & Portis, A. R. (2000). Alteration of the Adenine Nucleotide Response and Increased Rubisco Activation Activity of Arabidopsis Rubisco Activase by Site-Directed Mutagenesis. *Plant Physiology*, *123*(3), 1077–1086. https://doi.org/10.1104/pp.123.3.1077

Keating, B. A., Herrero, M., Carberry, P. S., Gardner, J., & Cole, M. B. (2014). Food wedges: Framing the global food demand and supply challenge towards 2050. *Global Food Security*, *3*(3), 125–132. https://doi.org/10.1016/j.gfs.2014.08.004

Keerberg, O., Pärnik, T., Ivanova, H., Bassüner, B., & Bauwe, H. (2014). C2 photosynthesis generates about 3-Fold elevated leaf CO2 levels in the C3–C4 intermediate species Flaveria pubescens. *Journal of Experimental Botany*, *65*(13), 3649–3656. https://doi.org/10.1093/jxb/eru239

Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., & Sage, T. L. (2016). C $_3$ –C $_4$ intermediacy in grasses: Organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C $_2$ photosynthesis. *Journal of Experimental Botany*, *67*(10), 3065–3078. https://doi.org/10.1093/jxb/erw150

Kim, S. Y., Bender, K. W., Walker, B. J., Zielinski, R. E., Spalding, M. H., Ort, D. R., & Huber, S. C. (2016). The Plastid Casein Kinase 2 Phosphorylates Rubisco Activase at the Thr-78 Site but Is Not Essential for Regulation of Rubisco Activation State. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00404

Konapala, G., Mishra, A. K., Wada, Y., & Mann, M. E. (2020). Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications*, *11*(1), 3044. https://doi.org/10.1038/s41467-020-16757-w

Li, C., Salvucci, M. E., & Portis, A. R. (2005). Two Residues of Rubisco Activase Involved in Recognition of the Rubisco Substrate. *Journal of Biological Chemistry*, *280*(26), 24864–24869. https://doi.org/10.1074/jbc.M503547200

Lilley, R. M., & Portis, A. R. (1990). Activation of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) by Rubisco Activase: Effects of Some Sugar Phosphates. *Plant Physiology*, *94*(1), 245–250. https://doi.org/10.1104/pp.94.1.245

Lipman, N. S., Jackson, L. R., Trudel, L. J., & Weis-Garcia, F. (2005). Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources. *ILAR Journal*, *46*(3), 258–268. https://doi.org/10.1093/ilar.46.3.258

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate Trends and Global Crop Production Since 1980. *Science (New York, N.Y.)*, *333*(6042), 616–620. https://doi.org/10.1126/science.1204531

Lobo, A. K. M., Orr, D. J., Gutierrez, M. O., Andralojc, P. J., Sparks, C., Parry, M. A., & Carmo-Silva, E. (2019). Overexpression of *ca1pase* Decreases Rubisco Abundance and Grain Yield in Wheat. *Plant Physiology*, *181*(2), 471–479. https://doi.org/10.1104/pp.19.00693

Long, B., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y.-L., Nguyen, N. D., Massey, B., Bala, S., von Caemmerer, S., Badger, M. R., & Price, G. D. (2018). Carboxysome encapsulation of the CO 2 -Fixing enzyme Rubisco in tobacco chloroplasts. *Nature Communications*, *9*(1), 1–14. https://doi.org/10.1038/s41467-018-06044-0

Long, S. P., Marshall-Colon, A., & Zhu, X.-G. (2015). Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell*, *161*(1), 56–66. https://doi.org/10.1016/j.cell.2015.03.019

Lorimer, G. H., Badger, M. R., & Andrews, T. J. (1976). The activation of Ribulose-1,5-Bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry*, *15*(3), 529–536. https://doi.org/10.1021/bi00648a012

Lundgren, M. R. (2020). C $_2$ photosynthesis: A promising route towards crop improvement? *New Phytologist*, nph.16494. https://doi.org/10.1111/nph.16494

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., & Mijakovic, I. (2019). Protein post-translational modifications in bacteria. *Nature Reviews Microbiology*, *17*(11), 651–664. https://doi.org/10.1038/s41579-019-0243-0

Mott, K. A., & Woodrow, I. E. (2000). Modelling the role of Rubisco activase in limiting Non-steady-state photosynthesis. *Journal of Experimental Botany*, *51*(suppl_1), 399–406. https://doi.org/10.1093/jexbot/51.suppl_1.399

Muñoz-Amatriaín, M., Mirebrahim, H., Xu, P., Wanamaker, S. I., Luo, M., Alhakami, H., Alpert, M., Atokple, I., Batieno, B. J., Boukar, O., Bozdag, S., Cisse, N., Drabo, I., Ehlers, J. D., Farmer, A., Fatokun, C., Gu, Y. Q., Guo, Y.-N., Huynh, B.-L., Jackson, S. A., Kusi, F., Lawley, C. T., Lucas, M. R., Ma, Y., Timko, M. P., Wu, J., You, F., Barkley, N. A., Roberts, P. A., Lonardi, S., & Close, T. J. (2017). Genome resources for Climate-resilient cowpea, an essential crop for food security. *The Plant Journal*, *89*(5), 1042–1054. https://doi.org/10.1111/tpj.13404

Nkomo, G. V., Sedibe, M. M., & Mofokeng, M. A. (2021). Production Constraints and Improvement Strategies of Cowpea (Vigna unguiculata L. Walp.) Genotypes for Drought Tolerance (C. Patan, Ed.). *International Journal of Agronomy*, *2021*, 1–9. https://doi.org/10.1155/2021/5536417

Orr, D. J., Robijns, A. K. J., Baker, C. R., Niyogi, K. K., & Carmo-Silva, E. (2023). Dynamics of Rubisco regulation by sugar phosphate derivatives and their phosphatases (T. Lawson, Ed.). *Journal of Experimental Botany*, *74*(2), 581–590. https://doi.org/10.1093/jxb/erac386

Owade, J. O., Abong', G., Okoth, M., & Mwang'ombe, A. W. (2020). A review of the contribution of cowpea leaves to food and nutrition security in East Africa. *Food Science & Nutrition*, *8*(1), 36–47. https://doi.org/10.1002/fsn3.1337

Parry, M. A. J., Andralojc, P. J., Mitchell, R. a. C., Madgwick, P. J., & Keys, A. J. (2003). Manipulation of Rubisco: The amount, activity, function and regulation. *Journal of Experimental Botany*, *54*(386), 1321–1333. https://doi.org/10.1093/jxb/erg141

Perdomo, J. A., Degen, G. E., Worrall, D., & Carmo-Silva, E. (2019). Rubisco activation by wheat Rubisco activase isoform 2β is insensitive to inhibition by ADP. *Biochemical Journal*, *476*(18), 2595–2606. https://doi.org/10.1042/BCJ20190110

Portis, A. R., Li, C., Wang, D., & Salvucci, M. E. (2007). Regulation of Rubisco activase and its interaction with Rubisco. *Journal of Experimental Botany*, *59*(7), 1597–1604. https://doi.org/10.1093/jxb/erm240

Portis, A. R., Salvucci, M. E., & Ogren, W. L. (1986). Activation of Ribulosebisphosphate Carboxylase/Oxygenase at Physiological CO_2 and Ribulosebisphosphate Concentrations by Rubisco Activase. *Plant Physiology*, *82*(4), 967–971. https://doi.org/10.1104/pp.82.4.967

Sage, R. F. (2002). Variation in the kcat of Rubisco in C3 and C4 plants and some implications for photosynthetic performance at high and low temperature. *Journal of Experimental Botany*, *53*(369), 609–620. https://doi.org/10.1093/jexbot/53.369.609

Sage, R. F., & Kubien, D. S. (2007). The temperature response of C_3 and C_4 photosynthesis. *Plant, Cell & Environment*, *30*(9), 1086–1106. https://doi.org/10.1111/j.1365-3040.2007.01682.x

Salesse-Smith, C. E., Sharwood, R. E., Busch, F. A., Kromdijk, J., Bardal, V., & Stern, D. B. (2018). Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nature Plants*, *4*(10), 802–810. https://doi.org/10.1038/s41477-018-0252-4

Salvucci, M. E., & Crafts-Brandner, S. J. (2004a). Inhibition of photosynthesis by heat stress: The activation state of Rubisco as a limiting factor in photosynthesis. *Physiologia Plantarum*, *120*(2), 179–186. https://doi.org/10.1111/j.0031-9317.2004.0173.x

Salvucci, M. E., & Crafts-Brandner, S. J. (2004b). Relationship between the Heat Tolerance of Photosynthesis and the Thermal Stability of Rubisco Activase in Plants from Contrasting Thermal Environments. *Plant Physiology*, *134*(4), 1460–1470. https://doi.org/10.1104/pp.103.038323

Salvucci, M. E., Osteryoung, K. W., Crafts-Brandner, S. J., & Vierling, E. (2001). Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo. *Plant Physiology*, *127*(3), 1053–1064.

Salvucci, M. E., van de Loo, F. J., & Stecher, D. (2003). Two isoforms of Rubisco activase in cotton, the products of separate genes not alternative splicing. *Planta*, *216*(5), 736–744. https://doi.org/10.1007/s00425-002-0923-1

Scafaro, A., Bautsoens, N., den Boer, B., Van Rie, J., & Gallé, A. (2019a). A Conserved Sequence from Heat-Adapted Species Improves Rubisco Activase Thermostability in Wheat. *Plant Physiology*, *181*(1), 43–54. https://doi.org/10.1104/pp.19.00425

Scafaro, A., Gallé, A., Rie, J. V., Carmo-Silva, E., Salvucci, M. E., & Atwell, B. J. (2016). Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytologist*, *211*(3), 899–911. https://doi.org/10.1111/nph.13963

Scafaro, A., Yamori, W., Carmo-Silva, A. E., Salvucci, M. E., von Caemmerer, S., & Atwell, B. J. (2012). Rubisco activity is associated with photosynthetic thermotolerance in a wild rice (Oryza meridionalis). *Physiologia Plantarum*, *146*(1), 99–109. https://doi.org/10.1111/j.1399-3054.2012.01597.x

Serban, A. J., Breen, I. L., Bui, H. Q., Levitus, M., & Wachter, R. M. (2018). Assembly–Disassembly is coupled to the ATPase cycle of tobacco Rubisco activase. *The Journal of Biological Chemistry*, *293*(50), 19451–19465. https://doi.org/10.1074/jbc.RA118.005047

Servaites, J. C. (1990). Inhibition of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase by 2-Carboxyarabinitol-1-Phosphate. *Plant Physiology*, *92*(4), 867–870. https://doi.org/10.1104/pp.92.4.867

Sharkey, T. D. (2005). Effects of moderate heat stress on photosynthesis: Importance of thylakoid reactions, rubisco deactivation, reactive oxygen species, and thermotolerance provided by isoprene. *Plant, Cell & Environment*, *28*(3), 269–277. https://doi.org/10.1111/j.1365-3040.2005.01324.x

Sharkey, T. D. (2019). Is triose phosphate utilization important for understanding photosynthesis? *Journal of Experimental Botany*, *70*(20), 5521–5525. https://doi.org/10.1093/jxb/erz393

Sharwood, R. E. (2017). A step forward to building an algal pyrenoid in higher plants. *New Phytologist*, *214*(2), 496–499. https://doi.org/10.1111/nph.14514

Sharwood, R. E., Ghannoum, O., Kapralov, M. V., Gunn, L. H., & Whitney, S. M. (2016). Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C3 photosynthesis. *Nature Plants*, *2*(12), 16186. https://doi.org/10.1038/nplants.2016.186

Shivhare, D., & Mueller-Cajar, O. (2017). In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiology*, *174*(3), 1505–1516. https://doi.org/10.1104/pp.17.00554

Shivhare, D., Ng, J., Tsai, Y.-C. C., & Mueller-Cajar, O. (2019). Probing the rice Rubisco–Rubisco activase interaction via subunit heterooligomerization. *Proceedings of the National Academy of Sciences*, *116*(48), 24041–24048. https://doi.org/10.1073/pnas.1914245116

Shukra, A. M., Sridevi, N. V., Chandran, D., & Maithal, K. (2014). Production of recombinant antibodies using bacteriophages. *European Journal of Microbiology and Immunology*, *4*(2), 91–98. https://doi.org/10.1556/EuJMI.4.2014.2.1

Soleh, M. A., Tanaka, Y., Kim, S. Y., Huber, S. C., Sakoda, K., & Shiraiwa, T. (2017). Identification of large variation in the photosynthetic induction response among 37 soybean [Glycine max (L.) Merr.] genotypes that is not correlated with steady-state photosynthetic capacity. *Photosynthesis Research*, *131*(3), 305–315. https://doi.org/10.1007/s11120-016-0323-1

Song, Y., Chen, Q., Ci, D., Shao, X., & Zhang, D. (2014). Effects of high temperature on photosynthesis and related gene expression in poplar. *BMC Plant Biology*, *14*, 111. https://doi.org/10.1186/1471-2229-14-111

Spreitzer, R. J., Peddi, S. R., & Satagopan, S. (2005). Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proceedings of the National Academy of Sciences*, *102*(47), 17225–17230. https://doi.org/10.1073/pnas.0508042102

Spreitzer, R. J., & Salvucci, M. E. (2002). RUBISCO: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme. *Annual Review of Plant Biology*, *53*(1), 449–475. https://doi.org/10.1146/annurev.arplant.53.100301.135233

Spriestersbach, A., Kubicek, J., Schäfer, F., Block, H., & Maertens, B. (2015). Purification of His-Tagged Proteins. In *Methods in Enzymology* (pp. 1–15). Elsevier. https://doi.org/10.1016/bs.mie.2014.11.003

Stanfield, R. L., & Wilson, I. A. (2014). Antibody Structure (J. E. Crowe Jr., D. Boraschi, & R. Rappuoli, Eds.). *Microbiology Spectrum*, *2*(2), 2.2.06. https://doi.org/10.1128/microbiolspec.AID-0012-2013

Stec, B. (2012). Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 18785–18790. https://doi.org/10.1073/pnas.1210754109

Suganami, M., Suzuki, Y., Kondo, E., Nishida, S., Konno, S., & Makino, A. (2020). Effects of Overproduction of Rubisco Activase on Rubisco Content in Transgenic Rice Grown at Different N Levels. *International Journal of Molecular Sciences*, *21*(5), 1626. https://doi.org/10.3390/ijms21051626

Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., & Scott, S. S. (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and Structure/Function relationships. *Journal of Experimental Botany*, *59*(7), 1515–1524. https://doi.org/10.1093/jxb/erm361

Taylor, S., & Long, S. P. (2017). Slow induction of photosynthesis on shade to sun transitions in wheat may cost at least 21% of productivity. *Philosophical Transactions of the Royal*

Society B: Biological Sciences, *372*(1730), 20160543. https://doi.org/10.1098/rstb.2016.0543

Taylor, S. H., Gonzalez-Escobar, E., Page, R., Parry, M. A. J., Long, S. P., & Carmo-Silva, E. (2022). Faster than expected Rubisco deactivation in shade reduces cowpea photosynthetic potential in variable light conditions. *Nature Plants*, *8*(2), 118–124. https://doi.org/10.1038/s41477-021-01068-9

Taylor, T., Backlund, A., Bjorhall, K., Spreitzer, R. J., & Andersson, I. (2001). First Crystal Structure of Rubisco from a Green Alga, *Chlamydomonas Reinhardtii*. *Journal of Biological Chemistry*, *276*(51), 48159–48164. https://doi.org/10.1074/jbc.M107765200

Tcherkez, G. G. B., Farquhar, G. D., & Andrews, T. J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences*, *103*(19), 7246–7251. https://doi.org/10.1073/pnas.0600605103

Tcherkez, G. G., Bathellier, C., Farquhar, G. D., & Lorimer, G. H. (2018). Commentary: Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All. *Frontiers in Plant Science*, *9*, 929. https://doi.org/10.3389/fpls.2018.00929

UN. (2019). United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights. ST/ESA/SER.A/423.

van de Loo, F. J., & Salvucci, M. E. (1996). Activation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Involves Rubisco Activase Trp16. *Biochemistry*, *35*(25), 8143–8148. https://doi.org/10.1021/bi9604901

Wachter, R., Salvucci, M. E., Carmo-Silva, A. E., Barta, C., Genkov, T., & Spreitzer, R. J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco–Rubisco activase interaction. *Photosynthesis Research*, *117*(1), 557–566. https://doi.org/10.1007/s11120-013-9827-0

Waheeda, K., Kitchel, H., Wang, Q., & Chiu, P.-L. (2023). Molecular mechanism of Rubisco activase: Dynamic assembly and Rubisco remodeling. *Frontiers in Molecular Biosciences*, *10*, 1125922. https://doi.org/10.3389/fmolb.2023.1125922

Wang, D., & Portis, A. R. (2006). Two conserved tryptophan residues are responsible for intrinsic fluorescence enhancement in Rubisco activase upon ATP binding. *Photosynthesis Research*, *88*(2), 185–193. https://doi.org/10.1007/s11120-006-9051-2

Wang, Q., Serban, A. J., Wachter, R. M., & Moerner, W. E. (2018). Single-molecule diffusometry reveals the nucleotide-dependent oligomerization pathways of Nicotiana tabacum Rubisco activase. *The Journal of Chemical Physics*, *148*(12), 123319. https://doi.org/10.1063/1.5005930

Wang, Z.-Y., Snyder, G. W., Esau, B. D., Portis, A. R., & Ogren, W. L. (1992). Species-Dependent Variation in the Interaction of Substrate-Bound Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase

(Rubisco) and Rubisco Activase. *Plant Physiology*, *100*(4), 1858–1862. https://doi.org/10.1104/pp.100.4.1858

Whitney, S. M., Houtz, R. L., & Alonso, H. (2011). Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco1[W]. *Plant Physiology*, *155*(1), 27–35. https://doi.org/10.1104/pp.110.164814

Wilson, R. H., Thieulin-Pardo, G., Hartl, F.-U., & Hayer-Hartl, M. (2019). Improved recombinant expression and purification of functional plant Rubisco. *FEBS Letters*, *593*(6), 611–621. https://doi.org/10.1002/1873-3468.13352

Wise, R. R., Olson, A. J., Schrader, S. M., & Sharkey, T. D. (2004). Electron transport is the functional limitation of photosynthesis in field-grown Pima cotton plants at high temperature. *Plant, Cell & Environment*, *27*(6), 717–724. https://doi.org/10.1111/j.1365-3040.2004.01171.x

Wu, A., Hammer, G. L., Doherty, A., von Caemmerer, S., & Farquhar, G. D. (2019). Quantifying impacts of enhancing photosynthesis on crop yield. *Nature Plants*, *5*(4), 380–388. https://doi.org/10.1038/s41477-019-0398-8

Yamori, W., Hikosaka, K., & Way, D. A. (2014). Temperature response of photosynthesis in C 3 , C 4 , and CAM plants: Temperature acclimation and temperature adaptation. *Photosynthesis Research*, *119*(1), 101–117. https://doi.org/10.1007/s11120-013-9874-6

Yamori, W., Kusumi, K., Iba, K., & Terashima, I. (2020). Increased stomatal conductance induces rapid changes to photosynthetic rate in response to naturally fluctuating light conditions in rice. *Plant, Cell & Environment*, *43*(5), 1230–1240. https://doi.org/10.1111/pce.13725

Yang, J. T., Preiser, A. L., Li, Z., Weise, S. E., & Sharkey, T. D. (2016). Triose phosphate use limitation of photosynthesis: Short-term and long-term effects. *Planta*, *243*(3), 687–698. https://doi.org/10.1007/s00425-015-2436-8

Yin, Z., Zhang, Z., Deng, D., Chao, M., Gao, Q., Wang, Y., Yang, Z., Bian, Y., Hao, D., & Xu, C. (2014). Characterization of Rubisco Activase Genes in Maize: An α-Isoform Gene Functions alongside a β-Isoform Gene. *Plant Physiology*, *164*(4), 2096–2106. https://doi.org/10.1104/pp.113.230854

Yoon, D.-K., Ishiyama, K., Suganami, M., Tazoe, Y., Watanabe, M., Imaruoka, S., Ogura, M., Ishida, H., Suzuki, Y., Obara, M., Mae, T., & Makino, A. (2020). Transgenic rice overproducing Rubisco exhibits increased yields with improved nitrogen-use efficiency in an experimental paddy field. *Nature Food*, *1*(2), 134–139. https://doi.org/10.1038/s43016-020-0033-x

Yu, Y., Yu, P.-C., Chang, W.-J., Yu, K., & Lin, C.-S. (2020). Plastid Transformation: How Does it Work? Can it Be Applied to Crops? What Can it Offer? *International Journal of Molecular Sciences*, *21*(14), 4854. https://doi.org/10.3390/ijms21144854

Zhang, N., Kallis, R. P., Ewy, R. G., & Portis, A. R. (2002). Light modulation of Rubisco in Arabidopsis requires a capacity for redox regulation of the larger Rubisco activase isoform. *Proceedings of the National Academy of Sciences*, *99*(5), 3330–3334. https://doi.org/10.1073/pnas.042529999

Zhang, N., & Portis, A. R. (1999). Mechanism of light regulation of Rubisco: A specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 9438–9443.

Zhou, Y., & Whitney, S. (2019). Directed Evolution of an Improved Rubisco; In Vitro Analyses to Decipher Fact from Fiction. *International Journal of Molecular Sciences*, *20*(20), 5019. https://doi.org/10.3390/ijms20205019

Zhu, G., & Jensen, R. G. (1991). Xylulose 1,5-Bisphosphate Synthesized by Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase during Catalysis Binds to Decarbamylated Enzyme. *Plant Physiology*, *97*(4), 1348–1353. https://doi.org/10.1104/pp.97.4.1348

Chapter 2

Identifying Candidate Plants for Rca Analysis

2.1 Introduction

Little is known of the natural genetic diversity of Rubisco activase. Despite its importance in photosynthesis, Rca remains poorly characterised outside of important model and crop species, such as tobacco (*Nicotiana tabacum*) and *Arabidopsis thaliana*. Characterising the properties of Rca from less well-studied species will increase knowledge of the diversity of Rca. Rca is required for removal of inhibitory sugar phosphates from inactivated Rubisco, its thermal optimum is considerably below that of Rubisco, meaning that the Rubisco is inactivated at higher temperatures despite its thermostability (Lilley and Portis, 1990; Parry et al., 2007; Salvucci et al., 2001). Amongst the known and characterised Rca isoforms, there is considerable difference in the catalytic properties - including ADP sensitivity, thermal optimum, rate of ATP hydrolysis, rate of light induction, and specificity for Rubisco (Carmo-Silva and Salvucci, 2013; Carmo-Silva et al., 2015; Perdomo et al., 2019; Wachter et al., 2013). It is known that small changes in the amino acid sequence of Rca, even single amino acid residues, can significantly alter the catalytic properties of the enzyme (Degen et al., 2020; Scafaro et al., 2019b). Rca is growing in importance as a target for crop improvement as its role in maintaining photosynthesis becomes more understood, and its role in thermotolerance makes it an attractive target in light of climate change, when temperatures are warming and becoming progressively more unstable (Hayer-Hartl and Hartl, 2020; Parry et al., 2013). An isoform of Rca with a broader thermal optimum may aid climate resilience in crops, as photosynthesis can remain active over a broader range of temperatures (Ogbaga et al., 2018).

Some *Fabaceae* species such as soybean and cowpea are an important part of the global food supply. Cowpea (*Vigna unguiculata* (L.) Walp), also known as black-eyed pea, is an important legume crop in sub-Saharan Africa. Cowpea is cultivated as a forage crop, for direct consumption, and as a subsistence crop. Both the grain in its dry and fresh form, and the leaves are eaten (OECD, 2019). Nigeria is the world's largest producer and consumer of cowpea, producing an

estimated 3,576,361 tonnes in 2019 (FAO et al., 2019, Figure 1.2). The climate of Nigeria covers several different biomes, with the Northern, more inland regions being arid desert and the Southern, coastal regions being tropical (Beck et al., 2018). The temperature in Nigeria can vary significantly throughout the country. These drastically different environments make cowpea a prime candidate for improvements in photosynthesis through Rca activity. An improvement of this kind could lead to improved yield under unfavourable conditions (Wijewardene et al., 2021). Furthermore, climate change is impacting Nigeria, further destabilising temperature and water availability, which varies considerably over the year through rainfall (Adenodi, 2018; Daramola et al., 2017). A broader thermal optimum for Rca may improve the climate resilience of cowpea under these changing environments by enabling photosynthesis to remain active under a larger temperature range.

The process of obtaining Rca sequences is complicated by multiple factors. Rca is often present in multiple copies in a given genome, either arising from gene duplication or alternative splicing. Each isoform can have varying expression levels, which can change under different environmental stimuli (Degen et al., 2021; Salvucci et al., 2003; Werneke et al., 1989). This is especially relevant when the sequence data source is transcriptomic instead of genomic, as genes with especially low mRNA expression levels may not be represented in transcriptomic datasets with insufficient coverage (Wang et al., 2009). Less expressed Rca genes can be of adaptive relevance. For example, the less expressed Rca 1β gene in wheat (*Triticum aestivum*) is upregulated under heat stress conditions, and has a higher thermal optimum compared to Rca 2β which is more expressed under favourable conditions (Degen et al., 2021). Therefore, focusing solely on the most expressed gene does not give a holistic view of the functional diversity of Rca in a given species. Furthermore, BLAST of Rca genes will often return non-genuine Rca-like genes which can be difficult to identify at early stages of analysis due to the low E-value (number of sequences with the same or higher alignment score that would be expected to be seen by chance), and regions of near-perfect conservation with genuine Rca. By nature of this project's focus on less well studied species, there is often only one source of available genomic or transcriptomic data, increasing the chances of spurious reads or sequencing errors. These considerations necessitate a reliable method of quality control for sequence data in order to obtain reliable sequences for further investigation.

Selecting candidate species for Rca with a broad thermal optimum has multiple considerations. Data for plant growing locations are inherently prone to sampling bias and misidentification, as well as false location records (Maldonado et al., 2015). While there are methods for quality controlling location records, due to these factors, data will not be completely reliable. Photosynthetic type is an important consideration, with C4 and CAM plants being naturally more resilient to high temperatures than C3 plants are (Sage and Kubien, 2007), but C3 being more acclimated to a broad temperature range (Yamori et al., 2014). Crop species do not make ideal candidates as they may be purposefully bred to be climate resilient in a way unrelated to Rca. Invasive species are not ideal candidates as while they spread rapidly, they may not necessarily maintain high lev-

els of photosynthetic competence that would be desired in crops. However, invasive species are difficult to avoid in this work as species which are capable of adaptation to broad climates will naturally coincide with invasiveness. Air temperature data is available for most of the globe but due to stomatal activity and other meteorological factors such as wind speed and air temperature, often differs from the leaf temperature which will be experienced by Rca (Haseba and Ito, 1975, 1982). This is especially pertinent in areas with high water supply, which can be partially estimated through precipitation rates but is complicated by availability of groundwater. For these reasons, ecoregion data are investigated alongside temperature data.

This chapter will focus on the steps taken to obtain reliable Rca sequence data for less well characterised species and correlate this with geographical and climatological data. The aim is to find candidate Rca sequences for subsequent analysis which are hypothesised to retain activity over a broad range of temperatures, for its applicability to climate resilience in *Fabaceae* crops. Throughout the chapter, *Arabidopsis thaliana* Rca and *Triticum aestivum* Rca will be used as representatives of non-crop and crop Rca respectively as these are both well studied. Two separate analysis are included. Analysis of all plants in the 1KP (1000 plant transcriptomes) shows the challenges related with obtaining reliable Rca sequences. The scope then narrowed to *Fabaceae* species, which are more relevant to *Fabaceae* crop improvement. Climatological data of locations occupied by various plant species has been analysed in search of species which grow and remain photosynthetically active under a broad range of temperatures. Three candidate species and five amino acid residue polymorphisms have been identified by this analysis.

2.2 Materials and Methods

2.2.1 Cowpea Sequences

Cowpea Rca sequences were identified in the genome of the `IT97K-499-35` cultivar of cowpea using Phytozome.

Table 2.1: *Cowpea Rca & Rca-like sequence accession numbers*

Isoform	Accession number
VuRca1β	Vigun01g219300.1
VuRca8α	Vigun08g150700.1
VuRca8α	Vigun08g150700.2
VuRca10α	Vigun10g051600.1
VuRca10β	Vigun10g051600.2
VuRca-like	Vigun05g302000.1.p

2.2.2 Obtaining *Fabaceae* Rca Sequences

Fabaceae Rca sequences were obtained from the 1KP (One Thousand Plant Transcriptomes Initiative, 2019) with a consensus Rca sequence derived from publicly available *Fabaceae* genomes. Diamond BLAST (Buchfink et al., 2021) was used to find matching sequences and Hisat2 (Kim et al., 2019) to map these onto the consensus. The sequences with highest amount of reads were selected.

2.2.3 Bioinformatics

DNA and amino acid sequences were filtered and analysed using Bioconductor 3.13 (Morgan, 2022) packages in R. Multiple sequence alignments were performed using the ClustalW algorithm in msa 1.24.0 (Bodenhofer et al., 2015). Full taxonomic data was obtained for 1KP species through Taxize 0.9.99 (Chamberlain et al., 2020).

Putative Rca sequences were obtained from the 1KP project by BLAST (https://db.cngb.org/blast/blast/blastp/?project=onekp) with the *Arabidopsis thaliana* short isoform of Rca.

Sequences were quality controlled by the following steps:

- Remove sequences with an E-value of $>10^{-10}$.
- Remove sequences where the scaffold contains multiple matching sequences
- Remove sequences without associated metadata
- Filter based on sequence length (AtRcaβ is of length 388 AA, A conservative threshold is set at no less than 5% (20AA) shorter than AtRcaβ. α isoforms should be targeted by this BLAST search so a more lenient upper bound is set at 460)
- Removed duplicate AA sequences
- Remove sequences where % identity to VuRca1β < % identity to VuRca-like

E-value represents the number of sequences with the same or higher alignment score that would be expected to be seen by chance. Scripts and raw data are available at https://github.com/DuncanBloemers/RcaAAAnalysis

Percentage identity was calculated using the `PID()` function of Biostrings (Pagès et al., 2022) with the following equation:

$$Percentage\,identity = 100 \cdot \frac{Identical\,Positions}{Average\,Sequence\,Length}$$

Chloroplast transit peptides were identified using TargetP v.2.0 (Almagro Armenteros et al., 2019).

2.2.4 Growing region and Whittaker biome data

All growth accessions for each species were obtained via GBIF and tested for the following common quality issues: “capitals”, “centroids”, “equal”, “gbif”, “institutions”, “zeros”, “seas” using the CoordinateCleaner R package (Zizka et al., 2019).

Locations which were not flagged as erroneous were used to extract climatological data from WorldClim using raster with the simple extract method. These climatic data are used to calculate temperature ranges and approximate Whittaker biomes.

Data were analysed in R 4.1 (R Core Team, 2013) and RStudio 1.4.1717 (RStudio Team, 2020) on macOS Big Sur 11.4. All data were obtained on 2023-03-09. Growth locations of species of interest were obtained and quality controlled using rgbif 3.6.0 and CoordinateCleaner 2.0-18 respectively. sp 1.4-5 (Bivand et al., 2013) and rgdal 1.5-23 (Bivand et al., 2022) were used to manage spatial data. raster 3.4-13 (Hijmans, 2022) was used to obtain climate data for each location from WorldClim (Fick and Hijmans, 2017). Maps were plotted using ggmap 3.0.0 (Kahle and Wickham, 2013). Approximate Whittaker biomes were plotted using plotbiomes 0.0.0.9001 (Ștefan and Levin, 2018). Data were managed and plotted using tidyverse 1.3.1 (Wickham et al., 2019). Documents for each species of interest were synthesised from this information using knitr 1.33 (Xie, 2022).

Scripts and data are available at https://github.com/DuncanBloemers/2019GBIF

2.2.5 Terrestrial Ecoregions

Ecoregions were obtained from the WWF’s Terrestrial Ecoregions of the World dataset (Olson et al., 2001). Plant growth locations (see Chapter 2.2.4) were plotted onto these ecoregions using CairoMakie v0.8.12 and GeoMakie v0.4.2 (Danisch and Krumbiegel, 2021).

Delimited data were handled with CSV v0.10.4 (Quinn et al., 2022) and DataFrames v1.3.4 (Kamiński et al., 2022). Shapefiles were handled with Shapefile v0.7.4.

Data were analysed using Julia v1.7.3 (Bezanson et al., 2017) in VSCodium version 1.69.2 on MacOS 12.5 Monterey.

Due to hosting limitations, this large volume of data and associated scripts are available upon request.

2.3 Results

2.3.1 Reliable Rca Sequences From Public Transcriptome Data

Potential Rca sequences were obtained from the 1000 plant transcriptomes (One Thousand Plant Transcriptomes Initiative, 2019) database using BLAST with *A. thaliana* Rca β isoform as a query. The returned sequences were subject to quality control measures to help attain reliable sequence data. To aid in identifying quality of sequences, sequence length and alignment length were used

as indicators of quality. A strict cutoff of sequence length was set at 5% below the sequence length of the shorter *A. thaliana* Rca β isoform and 5% above the longer α isoform. Figure 2.1 shows how this filtering step changes the distribution of amino acid sequence length for obtained sequences.

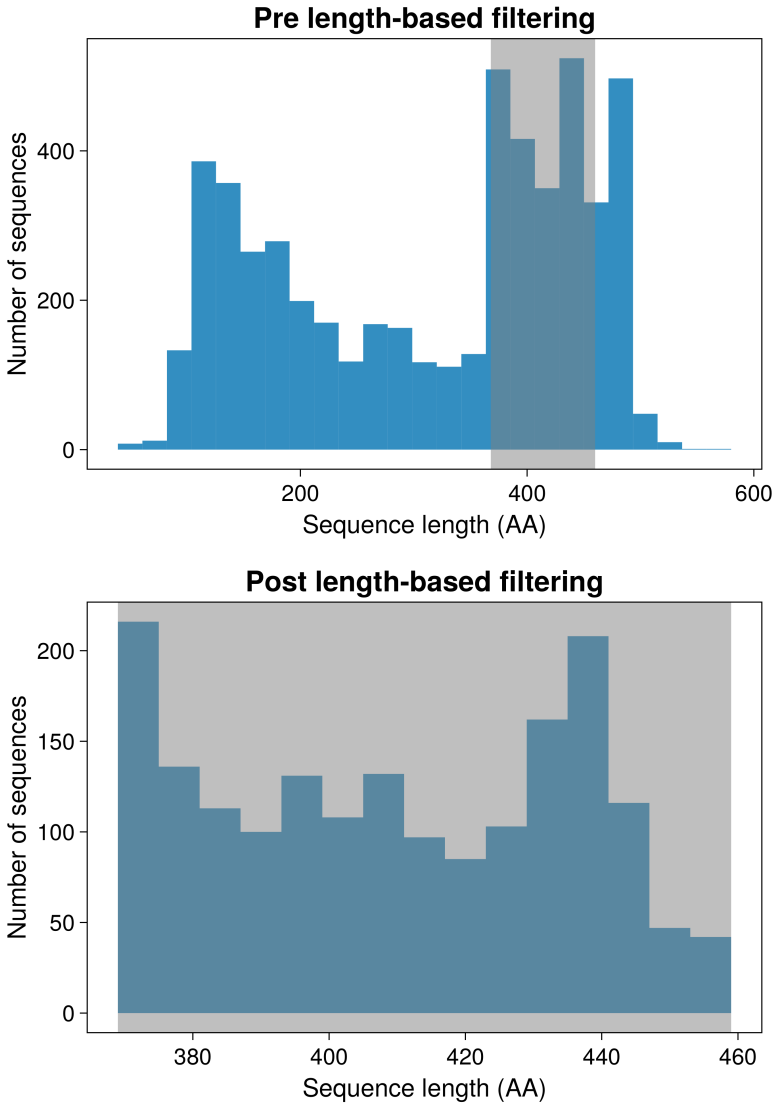
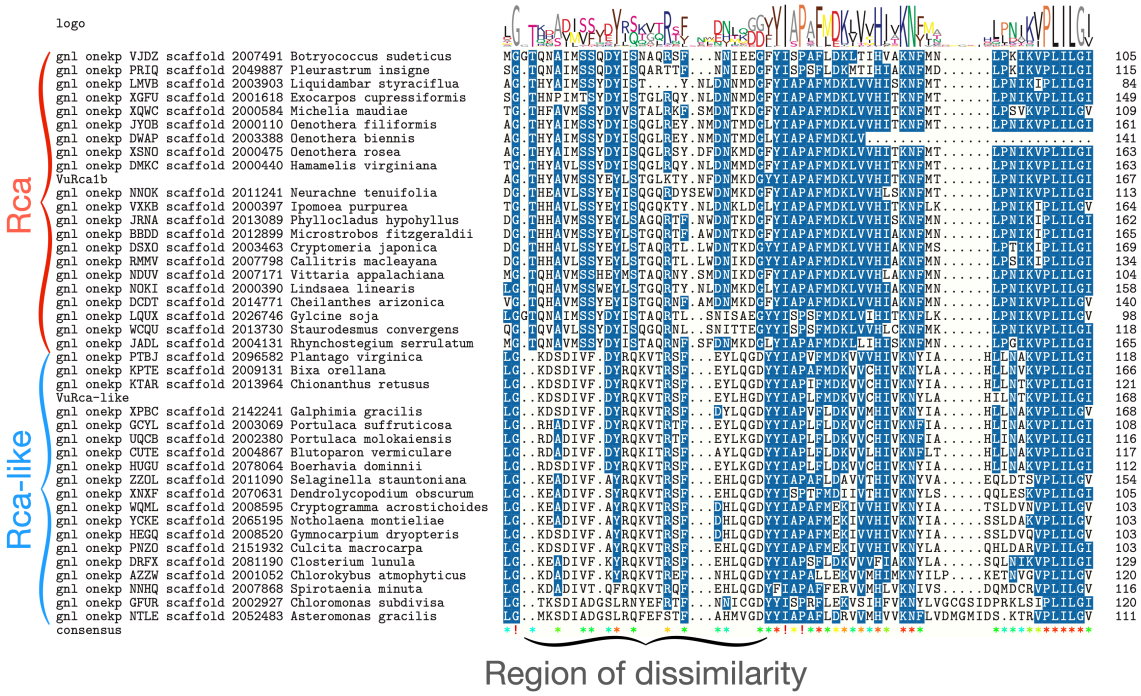


Figure 2.1: *Distribution of amino acid sequence length of Rca obtained from the 1000 plant transcriptomes database, before and after filtering based on sequence length.*

As shown in Figure 2.1, a significant proportion (66%) of unfiltered sequences were of lengths too short or too long to be reasonable Rca sequences. Additionally, several of the sequences were non-genuine Rca-like transcripts, judged by pairwise alignment to Rca-like sequences being more similar than to Rca, and should be removed from analysis. Rca-like proteins share long stretches of sequence which are almost identical to genuine Rca, and regions of dissimilarity (figure 2.2A) which can be used to distinguish them. Furthermore, a proportion of sequences lack complete metadata such as species of origin and must be removed from analysis. After initial quality control steps, 1,620 sequences out of 10,038 sequences were retained.

Figure 2.2B shows the % identity of the remaining sequences to both VuRca1β and VuRca-like. Sequences tend to align more closely with one gene over the other. This implies that the sequences which align poorly with Rca are ”contaminating” Rca-like sequences.

A



B

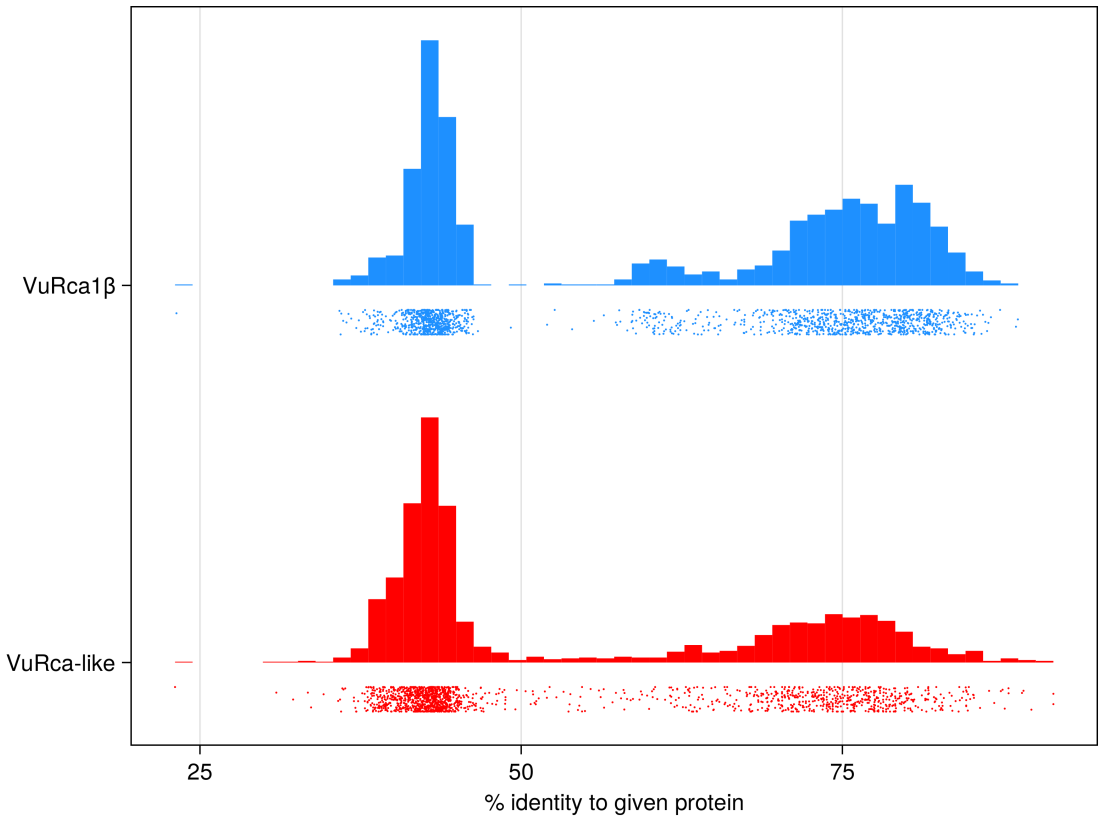


Figure 2.2: *Identification of Rca-like gene sequences.* ***A*** *Multiple sequence alignment of sequences from the 1000 plant transcriptomes database showing a region of dissimilarity between putative Rca and Rca-like amino acid sequences. VuRca1β and VuRca-like are included as controls. Sequences are a random subset of the full dataset.* ***B*** *Sequences from the 1000 plant transcriptomes database are compared to VuRca1β and VuRca-like proteins by percentage identity.*

Of these remaining sequences, 922 aligned more closely to VuRca1β than VuRca-like and were retained. The species these sequences are derived from span five phyla: *Streptophyta*, *Chlorophta*, *Ciliophora*, *Prasinodermophyta* & *Mucoromycota*. Table 2.2 shows the number of species with obtained Rca sequences at different taxonomic ranks. Figure 2.3 compares a consensus Rca sequence derived from the remaining sequences to VuRca1β. There are 33 amino acid differences between the sequences.

Table 2.2: *Number of species at different taxonomic ranks for which Rca species have been obtained from the One Thousand Plant Transcriptomes project.*

Taxonomic rank	Number of species
Phylum	5
Class	36
Order	130
Family	285

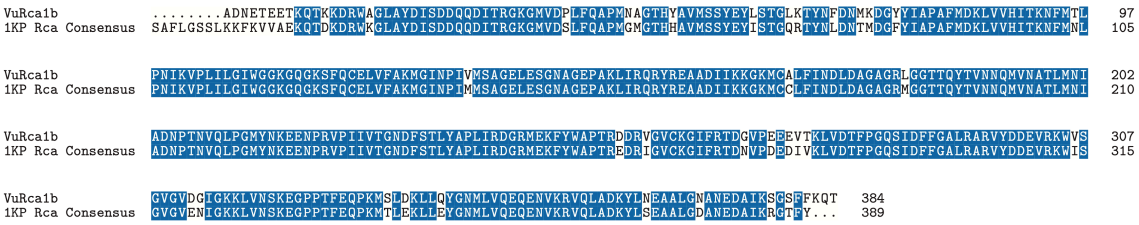


Figure 2.3: *Pairwise sequence alignment between VuRca1β and a consensus sequence derived from One Thousand Plant Transcriptomes project sequences. All sequences which passed quality control measures were used to calculate consensus. Consensus sequences are calculated by the most common amino acid in each position which is in ≥25% of sequences.*

2.3.2 Select *Fabaceae* Species Grow in Diverse Biomes

For the 25 species of *Fabaceae* represented in the 1KP, data on growing regions were obtained from the Global Biodiversity Information Facility (GBIF). Climatological data was then obtained via Worldclim for each of the coordinates where each of the 25 species grow, as recognised from GBIF data. These data were then analysed to estimate the range of growing temperature and the spread of biomes in which each species grows. Prior to further analysis, GBIF records were quality controlled to remove spurious records such as those which may have been incorrectly entered or those near botanical gardens or similar human establishments which would not be representative of natural growing regions. The amount of GBIF records varied substantially by species, and consequentially so does the amount of climatological data that can be obtained.

Figures 2.4, 2.5, & 2.6 show examples of the data collected for each species. For each species, data availability and resolution vary depending on growing locations. Several biologically relevant

variables are available for geolocations with relevant data. Mean temperature of wettest season was used as a proxy for growing season temperature.

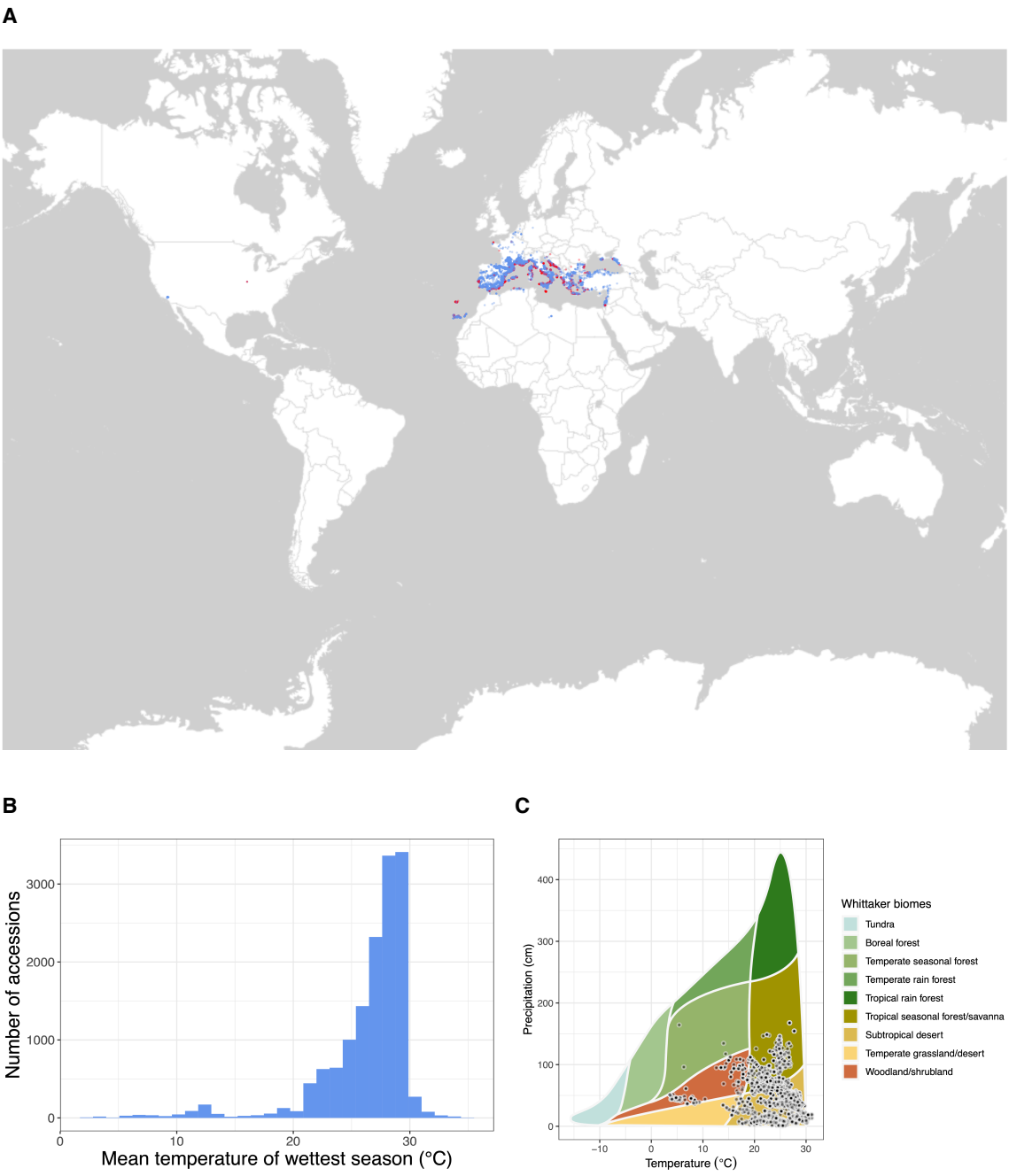


Figure 2.4: *Plant growth-relevant climatological metrics for geolocations containing* Bituminaria bituminosa *GBIF accessions.* ***A*** *Growth locations from the Global Biodiversity Information Facility for* Bituminaria bituminosa. *Accessions excluded from subsequent analysis are coloured in red.* ***B*** *The mean temperature of the wettest season in the areas in which* Bituminaria bituminosa *grows.* ***C*** *Estimated Whittaker biomes in the areas in which* Bituminaria bituminosa *grows.*

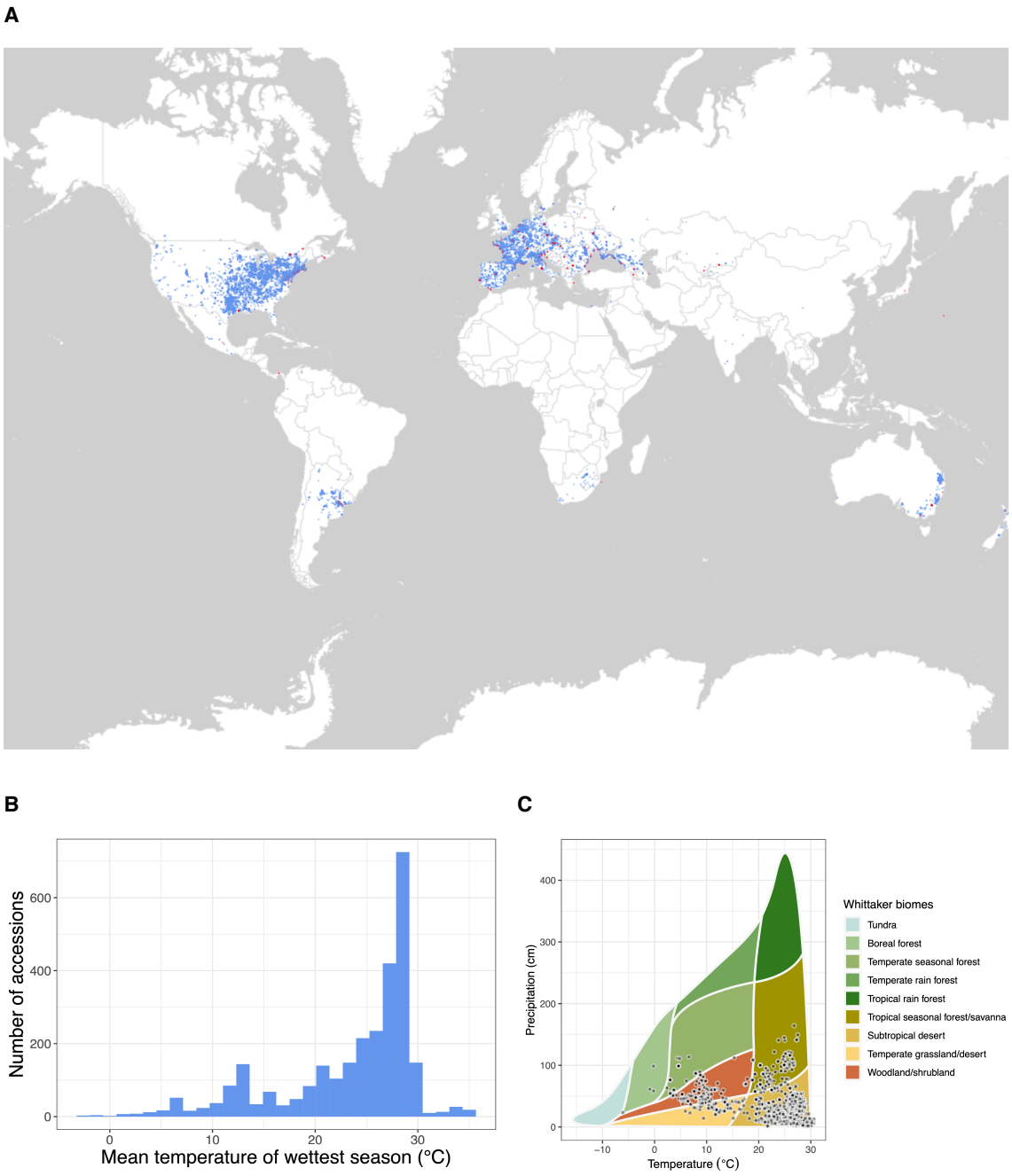


Figure 2.5: *Plant growth-relevant climatological metrics for geolocations containing* Gleditsia triacanthos *GBIF accessions.* ***A*** *Growth locations from the Global Biodiversity Information Facility for* Gleditsia triacanthos*. Accessions excluded from subsequent analysis are coloured in red.* ***B*** *The mean temperature of the wettest season in the areas in which* Gleditsia triacanthos *grows.* ***C*** *Estimated Whittaker biomes in the areas in which* Gleditsia triacanthos *grows.*

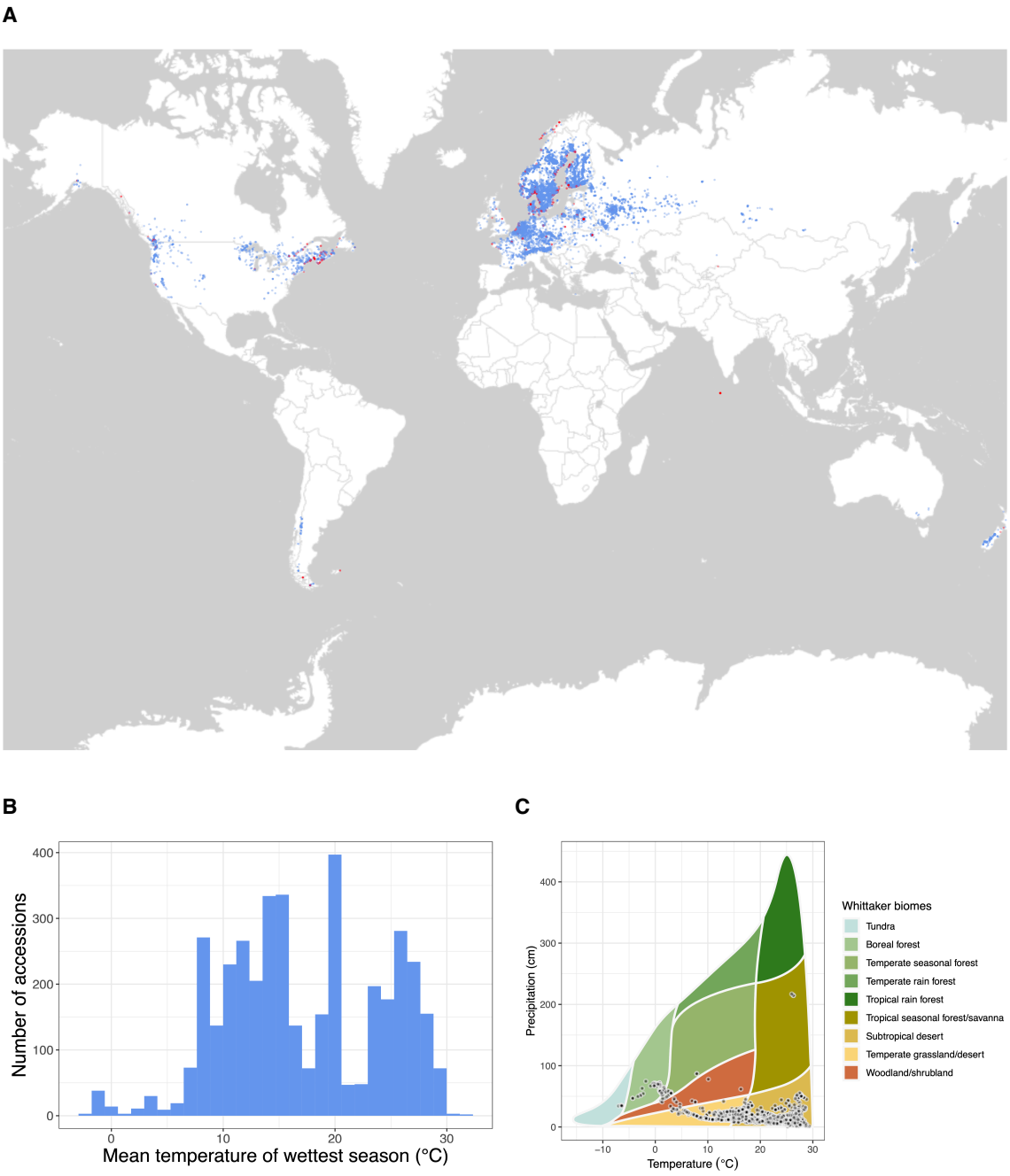


Figure 2.6: *Plant growth-relevant climatological metrics for geolocations containing* Lupinus polyphyllus *GBIF accessions.* ***A*** *Growth locations from the Global Biodiversity Information Facility for* Lupinus polyphyllus*. Accessions excluded from subsequent analysis are coloured in red.* ***B*** *The mean temperature of the wettest season in the areas in which* Lupinus polyphyllus *grows.* ***C*** *Estimated Whittaker biomes in the areas in which* Lupinus polyphyllus *grows.*

In addition to the sequences obtained across the species represented by the 1KP, high quality Rca sequences were obtained for the most abundant isoforms of Rca in *Fabaceae* by Diamond BLAST. Of the 25 *Fabaceae* species, sequences were obtained for 24 of them, no sequence is available for *Glycine soja*. *Fabaceae* was the focus of interest due to its relevance to cowpea,

the crop of interest. Table 2.3 shows the number of GBIF accessions which were not flagged as erroneous for each of the *Fabaceae* species represented in the 1KP as well as the range of mean temperatures of the wettest season for these. Of the 25 species, no climatological data was available for the growing regions of 7 of them and no GBIF accessions were present for 2. The highest mean temperature of wettest season range is *Gleditsia triacanthos* at 42.1 °C and the lowest non-zero temperature range is *Acacia pycnantha* at 7.1 °C. Most species are found in only a few continents with 3 species having accessions in all non-antarctic continents - *Lathyrus sativus*, *Lupinus angustifolius* & *Gleditsia triacanthos*.

Table 2.3: *Metrics of growth diversity for* Fabaceae *species. Aggregate climate data is shown for the locations at which* Fabaceae *species grow. Growth data from the Global Biodiversity Information Facility. NA: not available. Red text indicates data not available or insufficient to calculate reliable statistics.*

Species	Number of non-flagged GBIF accessions	Difference between highest and lowest daily mean temperature of wettest season (°C)	Number of continents with accessions
Acacia argyrophylla	399	NA	1
Acacia pycnantha	18,281	7.1	5
Apios americana	3,294	9.1	3
Astragalus membranaceus	NA	NA	NA
Astragalus propinquus	152	NA	NA
Bauhinia tomentosa	227	36	3
Bituminaria bituminosa	13,193	36	4
Cercis canadensis	16,611	24.3	3
Cladrastis lutea	28	0	3
Codariocalyx motorius	172	0	2
Copaifera officinalis	34	NA	2
Desmanthus illinoensis	2,690	NA	2
Gleditsia sinensis	189	11.6	2
Gleditsia triacanthos	15,812	42.1	6
Glycine soja	399	0	1
Glycyrrhiza glabra	1,112	28.7	3
Glycyrrhiza lepidota	2,675	NA	2
Gompholobium polymorphum	NA	NA	NA
Gymnocladus dioicus	1,735	29.5	2
Lathyrus sativus	2,041	38	6
Lupinus angustifolius	7,358	28.9	6
Lupinus polyphyllus	18,512	40.8	5
Senna hebecarpa	788	NA	2
Wisteria floribunda	1,428	20.8	4
Xanthocercis zambesiaca	109	0	2

Figure 2.7 shows an alignment of the *Fabaceae* Rca sequence obtained from the 1KP. There is diversity at the 159th and 161st (*Triticum aestivum* numbering) amino acid residues which are associated with temperature response (Degen et al., 2020). The 159th residue in *Bituminaria bituminosa* is leucine, and is methionine in *Gymnocladus dioicus* and *Gleditsia sinensis*. The most conserved residue at this point amongst the selected species is isoleucine, which implies adaptation to higher temperatures. Similar amino acid substitutions can be found amongst the

sequences including substitutions between methionine and amino acids with higher hydrophobicity which has been suggested to improve thermal tolerance. *Gleditsia triacanthos*, the species with the highest range of wettest season mean temperature, does not have an Rca isoform which is exceptionally different from consensus. There are methionine-leucine & methionine-isoleucine substitutions in the protein which may contribute to temperature resilience, and other substitutions which could impact protein function.

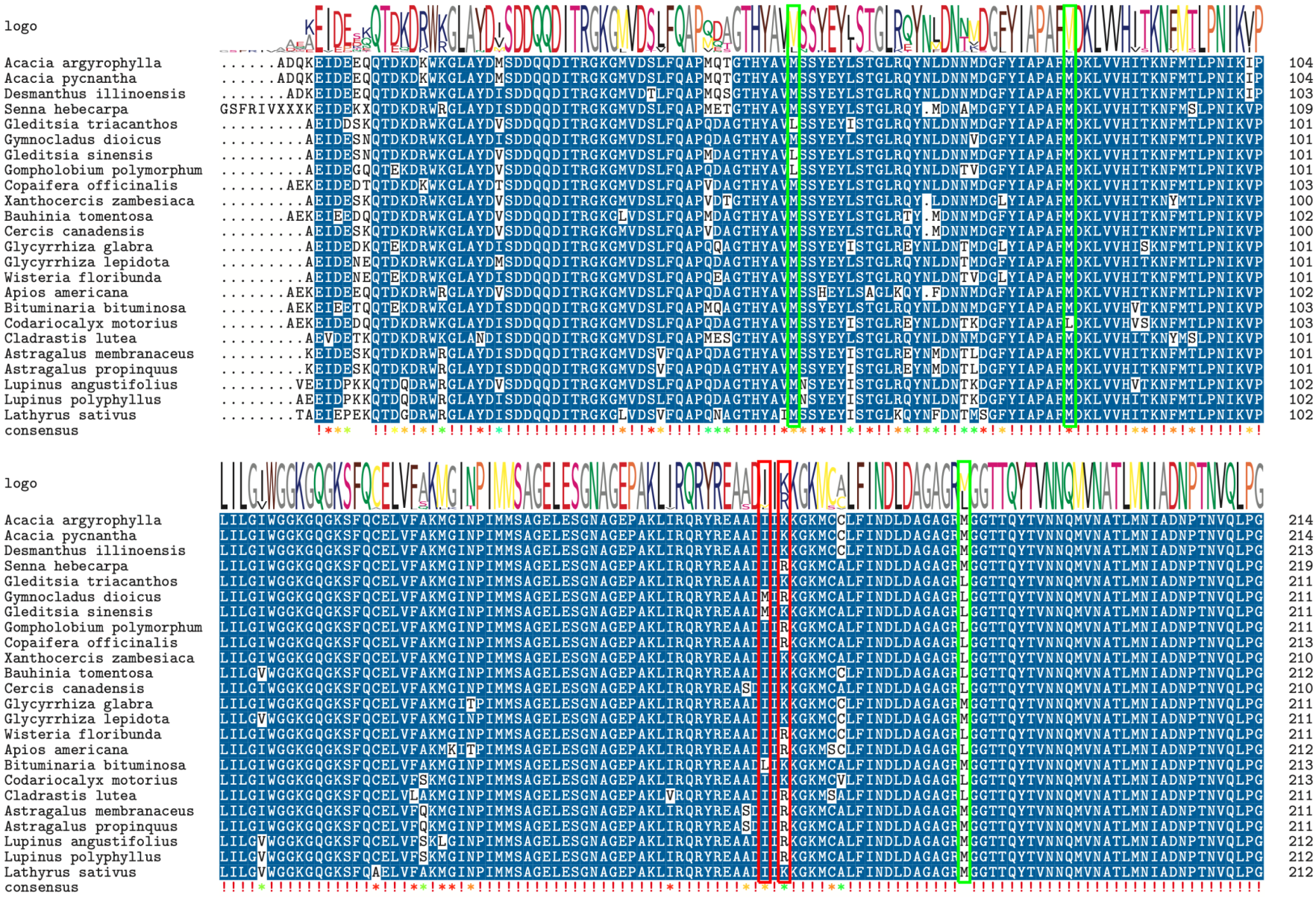
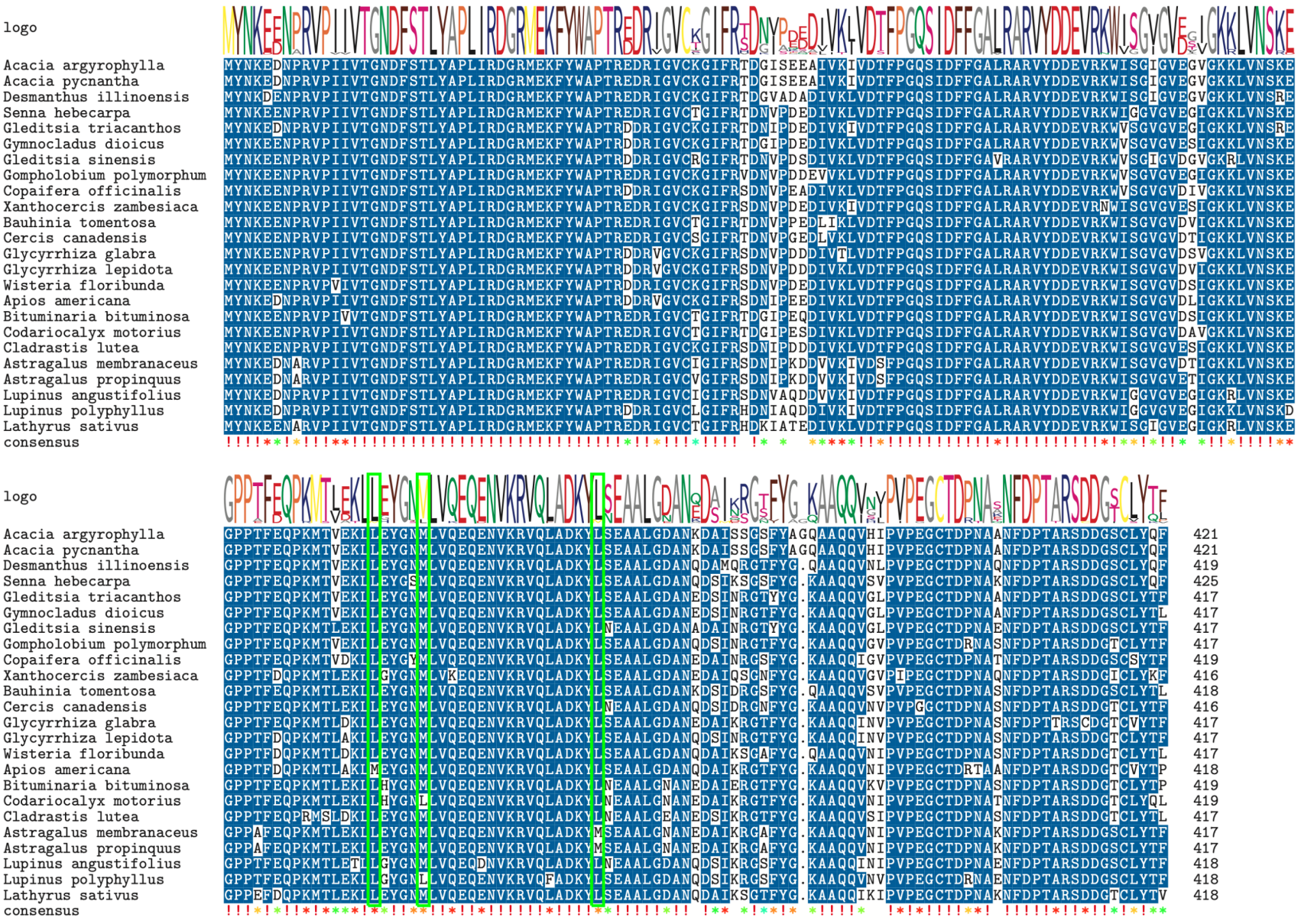
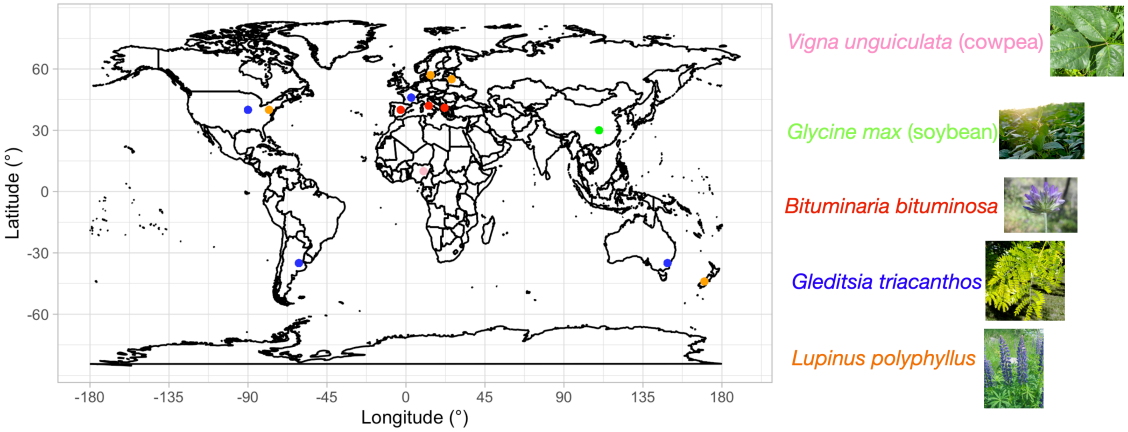


Figure 2.7: *Multiple sequence alignment of the Rca isoforms of represented* Fabaceae *species from the One Thousand Plant Transcriptomes project. Highlighted in red are the 159th and 161st amino acid residues (*Triticum aestivum *numbering) which are relevant to thermal optima of the enzyme. Highlighted in green are similar amino acid substitutions. Sequence order is determined algorithmically. Text boxes coloured blue indicates > 50% conservation. Alignment continues on next page.*


logo
Acacia argyrophylla MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDGISEEAIVKIVDTFPGQSIDFFGALRARVYDDEVRKWISGIGVEGVGKKLVNSKE 324
Acacia pycnantha MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDGISEEAIVKIVDTFPGQSIDFFGALRARVYDDEVRKWISGIGVEGVGKKLVNSKE 324
Desmanthus illinoensis MYNKDENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDGVADADIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGIGVEGVGKKLVNSRE 323
Senna hebecarpa MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCTGIFRTDNVPDEDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWIGGVGVEGIGKKLVNSKE 329
Gleditsia triacanthos MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFRTDNIPDEDIVKIVDTFPGQSIDFFGALRARVYDDEVRKWVSGVGVEGIGKKLVNSRE 321
Gymnocladus dioicus MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFRTDGIPDEDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWVSGVGVESIGKKLVNSKE 321
Gleditsia sinensis MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCRGIFRTDNVPDSDIVKLVDTFPGQSIDFFGAVRARVYDDEVRKWVSGIGVDGVGKRLVNSKE 321
Gompholobium polymorphum MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRVDNVPDDEVVKLVDTFPGQSIDFFGALRARVYDDEVRKWVSGVGVEGIGKKLVNSKE 321
Copaifera officinalis MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFRSDNVPEADIVKLVDTFPGQSIDFFGALRARVYDDEVRKWVSGVGVDIVGKKLVNSKE 323
Xanthocercis zambesiaca MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRSDNVPDEDIVKIVDTFPGQSIDFFGALRARVYDDEVRNWISGVGVESIGKKLVNSKE 320
Bauhinia tomentosa MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCTGIFRTDNVPPEDLIKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDVIGKKLVNSKE 322
Cercis canadensis MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCSGIFRTDNVPGEDLVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDTIGKKLVNSKE 320
Glycyrrhiza glabra MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRVGVCKGIFRSDNVPDDDIVTLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDSVGKKLVNSKE 321
Glycyrrhiza lepidota MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRVGVCKGIFRSDNVPDDDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDVIGKKLVNSKE 321
Wisteria floribunda MYNKEENPRVPVIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFRSDNVPDEDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDSIGKKLVNSKE 321
Apios americana MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRVGVCKGIFRSDNIPEEDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDLIGKKLVNSKE 322
Bituminaria bituminosa MYNKEENPRVPIVVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCTGIFRTDGIPEQDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDSIGKKLVNSKE 323
Codariocalyx motorius MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCTGIFRTDGIPESDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVDAVGKKLVNSKE 323
Cladrastis lutea MYNKEENPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRSDNIPDDDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGVGVESIGKKLVNSKE 321
Astragalus membranaceus MYNKEDNARVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCIGIFRSDNIPKDDVVKIVDSFPGQSIDFFGALRARVYDDEVRKWISGVGVDTIGKKLVNSKE 321
Astragalus propinquus MYNKEDNARVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCVGIFRSDNIPKDDVVKIVDSFPGQSIDFFGALRARVYDDEVRKWISGVGVETIGKKLVNSKE 321
Lupinus angustifolius MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCIGIFRSDNVAQDDVVKIVDTFPGQSIDFFGALRARVYDDEVRKWIGGVGVEGIGKRLVNSKE 322
Lupinus polyphyllus MYNKEDNPRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCLGIFRHDNIAQDDIVKIVDTFPGQSIDFFGALRARVYDDEVRKWIGGVGVEGIGKKLVNSKD 322
Lathyrus sativus MYNKEENARVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCTGIFRHDKIATEDIVKLVDTFPGQSIDFFGALRARVYDDEVRKWISGIGVEGIGKRLVNSKE 322
consensus
logo
Acacia argyrophylla GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANKDAISSGSFYAGQAAQQVHIPVPEGCTDPNAANFDPTARSDDGSCLYQF 421
Acacia pycnantha GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANKDAISSGSFYAGQAAQQVHIPVPEGCTDPNAANFDPTARSDDGSCLYQF 421
Desmanthus illinoensis GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDAMQRGTFYG.QAAQQVNLPVPEGCTDPNAQNFDPTARSDDGSCLYQF 419
Senna hebecarpa GPPTFEQPKMTVEKILEYGSMLVQEQENVKRVQLADKYLSEAALGDANQDSIKSGSFYG.KAAQQVSVPVPEGCTDPNAKNFDPTARSDDGSCLYQF 425
Gleditsia triacanthos GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANEDSINRGTYYG.KAAQQVGLPVPEGCTDPNAANFDPTARSDDGSCLYTF 417
Gymnocladus dioicus GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANEDSINRGTFYG.KAAQQVGLPVPEGCTDPNAANFDPTARSDDGSCLYTL 417
Gleditsia sinensis GPPTFEQPKMTLEKILEYGNMLVQEQENVKRVQLADKYLNEAALGDANADAINRGTYYG.KAAQQVGLPVPEGCTDPNAENFDPTARSDDGSCLYTF 417
Gompholobium polymorphum GPPTFEQPKMTVEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDSINRGTFYG.KAAQQVGVPVPEGCTDRNASNFDPTARSDDGTCLYTF 417
Copaifera officinalis GPPTFEQPKMTVDKILEYGYMLVQEQENVKRVQLADKYLSEAALGDANEDAINRGSFYG.KAAQQIGVPVPEGCTDPNATNFDPTARSDDGSCSYTF 419
Xanthocercis zambesiaca GPPTFDQPKMTLEKILGYGNMLVKEQENVKRVQLADKYLSEAALGDANEDAIQSGNFYG.KAAQQVGVPIPEGCTDPNAQNFDPTARSDDGICLYKF 416
Bauhinia tomentosa GPPTFEQPKMTLEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANKDSIDRGSFYG.QAAQQVSVPVPEGCTDPNASNFDPTARSDDGSCLYTL 418
Cercis canadensis GPPTFEQPKMTLEKILEYGNMLVQEQENVKRVQLADKYLNEAALGDANQDSIDRGNFYG.KAAQQVNVPVPGCTDPNASNFDPTARSDDGTCLYTF 416
Glycyrrhiza glabra GPPTFEQPKMTLDKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANEDAIKRGTFYG.KAAQQINVPVPEGCTDPNASNFDPTTRSCDGTCVYTF 417
Glycyrrhiza lepidota GPPTFDQPKMTLAKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDSINRGTFYG.KAAQQINVPVPEGCTDPNASNFDPTARSDDGTCLYTF 417
Wisteria floribunda GPPTFDQPKMTLDKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDAIKSGAFYG.QAAQQVNIPVPEGCTDPNASNFDPTARSDDGTCLYTL 417
Apios americana GPPTFDQPKMTLAKIMEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDAIKRGTFYG.KAAQQVNIPVPEGCTDRTAANFDPTARSDDGTCVYTP 418
Bituminaria bituminosa GPPTFEQPKMTLEKILHYGNMLVQEQENVKRVQLADKYLNEAALGNANEDAIERGTFYG.KAAQQVKVPVPEGCTDPNASNFDPTARSDDGTCLYTP 419
Codariocalyx motorius GPPTFEQPKMTLEKILHYGNLLVQEQENVKRVQLADKYLNEAALGNANEDAIKRGTFYG.KAAQQVNIPVPEGCTDPNATNFDPTARSDDGTCLYQL 419
Cladrastis lutea GPPTFEQPRMSLDKILEYGNMLVQEQENVKRVQLADKYLNEAALGEANEDSIKRGTFYG.KAAQQVSIPVPEGCTDPNASNFDPTARSDDGTCLYTL 417
Astragalus membranaceus GPPAFEQPKMTLEKILEYGNMLVQEQENVKRVQLADKYMSEAALGNANEDAIKRGAFYG.KAAQQVNIPVPEGCTDPNAKNFDPTARSDDGSCLYTF 417
Astragalus propinquus GPPAFEQPKMTLEKILEYGNMLVQEQENVKRVQLADKYMSEAALGNANEDAIKRGAFYG.KAAQQVNIPVPEGCTDPNAKNFDPTARSDDGSCLYTF 417
Lupinus angustifolius GPPTFEQPKMTLETILGYGNMLVQEQDNVKRVQLADKYLNEAALGDANQDSIKRGSFYG.KAAQQINIPVPEGCTDPNAENFDPTARSDDGSCLYTF 418
Lupinus polyphyllus GPPTFEQPKMTLEKILLGYGNLLVQEQENVKRVQFADKYLSEAALGDANQDSIKRGSFYG.KAAQQVNVPVPEGCTDRNAENFDPTARSDDGSCLYTF 418
Lathyrus sativus GPPEFDQPKMTLEKILEYGNMLVQEQENVKRVQLADKYLSEAALGDANQDAIKSGTFYG.KAAQQIKIPVPEGCTDPNAKNFDPTARSDDGTCLYTV 418
consensus

Three *Fabaceae* species and five amino acid residues have been identified as potentially relevant for thermostability based on sequence and analysis of geographical distribution Figure 2.8. The mutations at these locations in relation to VuRca8α are: I220L (*Bituminaria bituminosa*), L241M (*Bituminaria bituminosa* & *Lupinus polyphyllus*), E399H (*Bituminaria Bituminosa*), E399G (*Lupinus polyphyllus*), N422S (*Gleditsia triacanthos* & *Lupinus polyphyllus*), K435E (*Bituminaria Bituminosa*), K435N (*Gleditsia triacanthos*).

A



B

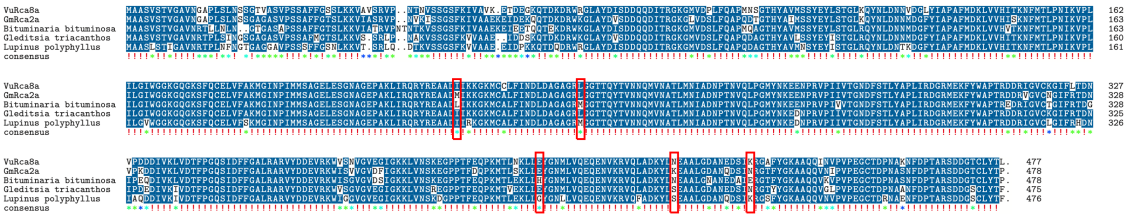


Figure 2.8: *Growth regions and sequence comparisons of select* Fabaceae *species. **A** The most common global growing locations of selected species. **B** Sequence comparison of Rca from selected species, and 8α & 2α isoforms from cowpea and soybean respectively. Text boxes coloured blue indicates > 50% conservation. Image sources are in citations.*

Every selected species grows in multiple ecoregions (see Figure 2.9). There is a tendency for *Lupinus polyphyllus* to cluster in temperate European regions, and *Bituminaria bituminosa* to grow in mediterranean regions. The ability for species to grow in a range of climates is a relevant consideration due to the interaction between leaf temperature and water availability. A plant in a warm, wet climate will be able to keep its leaves cooler than one in a warm, arid biome through use of transpiration.

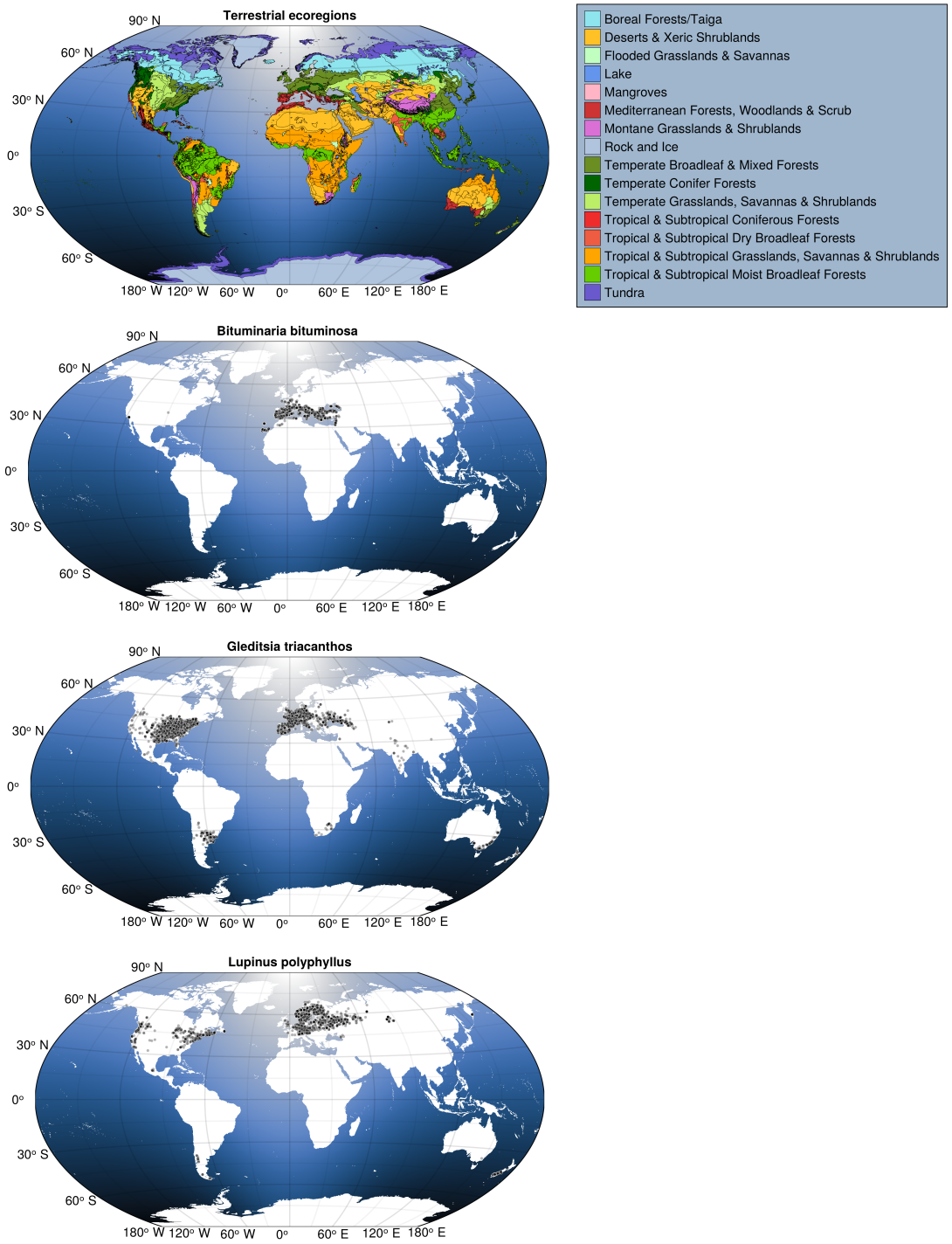


Figure 2.9: *Terrestrial ecoregions in which selected species grow. Each point is an accession from the Global Biodiversity Information Facility.*

2.4 Discussion

Sequence data has been obtained for Rubisco activase from multiple different species. These sequences were compared with climatological data at the growth locations they're derived from to find species which fulfil the following criteria:

- Rca amino acid sequence diversity
- Species grows in a broad temperature range
- Species grown in a range of climates

Ultimately, three candidate species were found to fit these criteria - *Bituminaria bituminosa*, *Gleditsia triacanthos* & *Lupinus polyphyllus*.

Sequence data can be of variable quality, this is especially visible with the presence of short/incomplete sequences. The reason behind erroneous sequences is not known but may be the result of poor mRNA integrity or premature transcription termination. Multiple quality control steps can help increase confidence in sequence quality. The sequences obtained provide a view of the genetic diversity of Rca across the plant kingdom. This process, however, highlights the issues with the use of BLAST to obtain Rca sequences from large datasets. Of the 10,000 initial sequences, only 920 appear sufficiently identical to known Rca to remain confident in the sequence. The other sequences have issues such as sequence length or aligning closer to Rca-like genes than genuine Rca. Another method of obtaining sequences from direct sequence reads using Diamond BLAST and hisat2 provides high quality sequence data but is more intensive and only provides reliable sequences for the most expressed sequences.

The sequences obtained for *Fabaceae* Rca show potentially interesting amino acid substitutions. It has been suggested that the increase in hydropathy index from methionine to isoleucine at the 159th amino acid residue of Rca is the reason for the raised thermal optimum (Degen et al., 2020). Similar amino acid substitutions can be found in these sequences, including substitutions from methionine to leucine (hydropathy index of 1.9 and 3.8 respectively) and vice versa (Kyte and Doolittle, 1982). Other potentially interesting sequence changes include single amino acid deletions, and replacement of a more reactive amino acid (such as cysteine, which contains a thiol group that can form disulfide bonds) with an inert one (such as alanine, which is structurally similar to cysteine but does not contain a reactive side-chain) (Morrison and Weiss, 2001). There are also regions of more diversity, including residues at which there is no single amino acid with over 50% prevalence. Residues may show diversity for a number of different reasons - these residues could functional adaptation to local environment, such as temperature response, or could have little functional relevance and thus be able to be mutated without harming functionality.

Climatological data has been collected for the *Fabaceae* species for which there are sequence data. The main variable of interest is the range of temperatures at which the species will grow, as this implies active photosynthesis and thus active Rca. Growing months can be predicted with holistic view of temperature, precipitation and soil moisture data but these factors interact to define growing season in a species-dependent manner (Lasky et al., 2012; Lasky et al., 2015). This makes it difficult to properly predict growing season in less well understood species. Because of this, wettest season was chosen as a proxy for growing season to simplify analysis. However, this means that the growing season may not be entirely accurate for each species, and a more robust analysis should aim to integrate knowledge of the precise growing season of the species.

While analysis of wettest season temperature metrics has not yielded an obviously divergent Rca isoform, the sequence diversity that is present could be functionally relevant.

As sequence data is only available for one genotype per species, the subspecies diversity of each Rca isoform is not known. Rca may be adapted within the species to the environment it inhabits, rather than having evolved an isoform which is more robust to temperature. Furthermore, since sequences for isoforms other than the most expressed isoform are not represented, the entire range of sequence diversity is not known.

Three species of interest have been identified as having a broad growing season temperature range and a broad range of biomes in which they grow, these species are *Gleditsia triacanthos*, *Lupinus polyphyllus*, and *Bituminaria bituminosa*. These species vary in the regions they grow in as well as their life cycle and morphology. *Bituminaria bituminosa* is a perennial herb which is mostly found in the Mediterranean and Near East, but has been naturalised in Australia (Cardin and Moury, 2009). It is used as a forage crop and for stabilisation of arid soils, and is resilient to cold and drought stress (Jones et al., 2012). *Lupinus polyphyllus* is a perennial herb species native to North America but also found throughout Europe, Central and Southern Asia, South American and Oceania (Ramula, 2022). It is commonly grown for ornamental and landscaping purposes, and has been investigated for grazing purposes (Scott and Tesfaye, 2000). *Gleditsia triacanthos*, also known as honey locust, is a deciduous tree which can be found in North and South America, throughout Europe, and in Southern Asia and Southern Australia (CABI, 2022). It is considered a hardy plant and despite being capable of growth in tropical environments, it is cold adapted (Calkins and Swanson, 1998). *Lupinus polyphyllus* and *Gleditsia triacanthos* are considered invasive species. Both are native to North America but have been naturalised in other continents. The photosynthetic type of these species is not known but given the exceptional rarity of C4 photosynthesis in trees (Young et al., 2020), *Gleditsia triacanthos* is likely to be a C3 species.

In the candidate species, five amino acid residues and seven unique substitutions have been identified as likely relevant to the catalytic properties or thermostability of the enzyme. Some of these substitutions fit into the category of aforementioned hydropathy index changes which has been associated with temperature response previously (Degen et al., 2020). Other substitutions such as K435E/K435N could provide adaptation to local environment. Another hypothesis is that these residues are not important to the function of the enzyme and are therefore permitted more variation without causing a fitness loss. It is further worth noting that the species that have been selected for analysis due to their broad temperature range are amongst the species with the most GBIF accessions. It is not possible to confirm if the apparent wettest season temperature range is genuine or due to the larger amount of data increasing probability of accessions being in locations with different climates.

The sequence and amino acid substitutions identified here expand on the knowledge of diversity of Rubisco activase, which has historically been largely limited to crop and model species. Some of the sequences identified may provide useful variants of Rca which could aid in identifying targets for crop breeding or genetic manipulation for improving photosynthesis.

2.5 Acknowledgements

The author acknowledges the kind help of Matt Stata from the University of Toronto in obtaining high quality Rca sequence data for *Fabaceae* species. He has helped in obtaining Rca sequence data for the *Fabaceae* species represented in the 1KP.

Chapter 2 Bibliography

Adenodi, R. (2018). A centurial analysis of rainfall variability in Nigeria. *Nigerian Journal of Technology*, *37*(2), 543. https://doi.org/10.4314/njt.v37i2.34

Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, *2*(5), e201900429. https://doi.org/10.26508/lsa.201900429

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Bivand, R., Keitt, T., & Rowlingson, B. (2022). *Rgdal: Bindings for the 'geospatial' data abstraction library* (tech. rep.).

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY.

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. (2015). Msa: An R package for multiple sequence alignment. *Bioinformatics*, btv494. https://doi.org/10.1093/bioinformatics/btv494

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

CABI. (2022). Gleditsia triacanthos (honey locust). *CABI Compendium*, *CABI Compendium*, 25272. https://doi.org/10.1079/cabicompendium.25272

Calkins, J. B., & Swanson, B. T. (1998). Plant cold acclimation, hardiness, and winter injury in response to bare soil and groundcover-based nursery field management systems. *Journal of Environmental Horticulture*, *16*(2), 82–89. https://doi.org/10.24266/0738-2898-16.2.82

Cardin, L., & Moury, B. (2009). First Report of Passiflora chlorosis virus in Bituminaria bituminosa in Europe. *Plant Disease*, *93*(2), 196–196. https://doi.org/10.1094/pdis-93-2-0196a

Carmo-Silva, E., & Salvucci, M. E. (2013). The Regulatory Properties of Rubisco Activase Differ among Species and Affect Photosynthetic Induction during Light Transitions. *Plant Physiology*, *161*(4), 1645–1655. https://doi.org/10.1104/pp.112.213348

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon,

M., Li, G., & Grenié, M. (2020). *Taxize: Taxonomic information from around the web* (tech. rep.).

Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software*, *6*(65), 3349. https://doi.org/10.21105/joss.03349

Daramola, M., Eresanya, E., & Erhabor, S. (2017). Analysis of rainfall and temperature over climatic zones in nigeria. *Journal of Geography, Environment and Earth Science International*, *11*(2), 1–14. https://doi.org/10.9734/jgeesi/2017/35304

Degen, G., Worrall, D., & Carmo-Silva, E. (2020). An isoleucine residue acts as a thermal and regulatory switch in wheat Rubisco activase: Rubisco activase thermal and regulatory switch. *The Plant Journal*. https://doi.org/10.1111/tpj.14766

Degen, G. E., Orr, D. J., & Carmo-Silva, E. (2021). Heat-induced changes in the abundance of wheat Rubisco activase isoforms. *New Phytologist*, *229*(3), 1298–1311. https://doi.org/10.1111/nph.16937

FAO, IFAD, UNICEF, WFP, & WHO. (2019). The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns. Rome, FAO.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

File:Bituminaria bituminosa flower.jpg — Wikimedia Commons, the free media repository. (2020).

File:BlackEyedPeas.JPG — Wikimedia commons, the free media repository. (2020).

File:Gleditsia triacanthos Maryhill Museum 01.jpg — Wikimedia Commons, the free media repository. (2020).

File:Lupinus polyphyllus.JPG — Wikimedia commons, the free media repository. (2022).

File:Soybean.USDA.jpg — Wikimedia commons, the free media repository. (2022).

Haseba, T., & Ito, D. (1975). Studies of transpiration in relation to the environment. *Journal of Agricultural Meteorology*, *30*(4), 173–182. https://doi.org/10.2480/agrmet.30.173

Haseba, T., & Ito, D. (1982). Leaf temperature in relation to meteorological factors. *Journal of Agricultural Meteorology*, *38*(3), 269–277. https://doi.org/10.2480/agrmet.38.269

Hayer-Hartl, M., & Hartl, F. U. (2020). Chaperone machineries of rubisco – the most abundant enzyme. *Trends in Biochemical Sciences*, *45*(9), 748–763. https://doi.org/10.1016/j.tibs.2020.05.001

Hijmans, R. J. (2022). *Raster: Geographic data analysis and modeling* (tech. rep.).

Jones, R. A. C., Real, D., Vincent, S. J., Gajda, B. E., & Coutts, B. A. (2012). First report of alfalfa mosaic virus infecting tedera (bituminaria bituminosa (l.) C.H. stirton var. albomarginata and crassiuscula) in australia. *Plant Disease*, *96*(9), 1384–1384. https://doi.org/10.1094/pdis-04-12-0378-pdn

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

Kamiński, B., White, J. M., Bouchet-Valat, M., Powerdistribution, Garborg, S., Quinn, J., Kornblith, S., Cjprybol, Stukalov, A., Bates, D., Short, T., DuBois, C., Harris, H., Squire, K., Arslan, A., Pdeffebach, Anthoff, D., Kleinschmidt, D., Noack, A., Shah, V. B., Mellnik, A., Arakaki, T., Mohapatra, T., Peter, Karpinski, S., Lin, D., Chagas, R. A. J., Timema, ExpandingMan, & Oswald, F. (2022). JuliaData/DataFrames.jl: V1.4.3. https://doi.org/10.5281/ZENODO.7316260

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation ofArabidopsis thaliana: The roles of geography and climate. *Molecular Ecology*, *21*(22), 5512–5529. https://doi.org/10.1111/j.1365-294x.2012.05709.x

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, *1*(6), e1400218. https://doi.org/10.1126/sciadv.1400218

Lilley, R. M., & Portis, A. R. (1990). Activation of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) by Rubisco Activase: Effects of Some Sugar Phosphates. *Plant Physiology*, *94*(1), 245–250. https://doi.org/10.1104/pp.94.1.245

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of B ig D ata: To what extent can we trust public databases? *Global Ecology and Biogeography*, *24*(8), 973–984. https://doi.org/10.1111/geb.12326

Morgan, M. (2022). *BiocManager: Access the bioconductor project package repository* (tech. rep.).

Morrison, K. L., & Weiss, G. A. (2001). Combinatorial alanine-scanning. *Current Opinion in Chemical Biology*, *5*(3), 302–307. https://doi.org/10.1016/S1367-5931(00)00206-4

OECD. (2019). *Safety Assessment of Foods and Feeds Derived from Transgenic Crops, Volume 3: Common bean, Rice, Cowpea and Apple Compositional Considerations*. https://doi.org/10.1787/f04f3c98-en

Ogbaga, C. C., Stepien, P., Athar, H.-U.-R., & Ashraf, M. (2018). Engineering Rubisco activase from thermophilic cyanobacteria into high-temperature sensitive plants. *Critical Reviews in Biotechnology*, *38*(4), 559–572. https://doi.org/10.1080/07388551.2017.1378998

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R.

(2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–685. https://doi.org/10.1038/s41586-019-1693-2

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2022). Biostrings: Efficient manipulation of biological strings. https://doi.org/10.18129/B9.bioc.Biostrings

Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E., & Andralojc, P. J. (2007). Rubisco regulation: A role for inhibitors. *Journal of Experimental Botany*, *59*(7), 1569–1580. https://doi.org/10.1093/jxb/ern084

Parry, M. A. J., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., & Whitney, S. M. (2013). Rubisco activity and regulation as targets for crop improvement. *Journal of Experimental Botany*, *64*(3), 717–730. https://doi.org/10.1093/jxb/ers336

Perdomo, J. A., Degen, G. E., Worrall, D., & Carmo-Silva, E. (2019). Rubisco activation by wheat Rubisco activase isoform 2β is insensitive to inhibition by ADP. *Biochemical Journal*, *476*(18), 2595–2606. https://doi.org/10.1042/BCJ20190110

Quinn, J., Bouchet-Valat, M., Robinson, N., Kamiński, B., Newman, G., Stukalov, A., Vogt, C., Cjprybol, Noack, A., Kelman, T., Davies, E., ExpandingMan, Ian, Piibeleht, M., Finnegan, R., Evalparse, Silberstein, A. M., Anthony Blaom, P., Lungwitz, B., König, B., De Graaf, C., Woodfield, C., Barton, D., Aluthge, D., Saba, E., Noronha, F., Kragol, Rosario, H. D., Ranocha, H., & Butterworth, I. (2022). JuliaData/CSV.jl: V0.10.7. https://doi.org/10.5281/ZENODO.7250547

R Core Team. (2013). *R: A language and environment for statistical computing* (tech. rep.). Vienna, Austria.

Ramula, S. (2022). Lupinus polyphyllus (garden lupin). *CABI Compendium*, *CABI Compendium*, 31710. https://doi.org/10.1079/cabicompendium.31710

RStudio Team. (2020). *RStudio: Integrated development environment for R* (tech. rep.). Boston, MA.

Sage, R. F., & Kubien, D. S. (2007). The temperature response of C $_3$ and C $_4$ photosynthesis. *Plant, Cell & Environment*, *30*(9), 1086–1106. https://doi.org/10.1111/j.1365-3040.2007.01682.x

Salvucci, M. E., Osteryoung, K. W., Crafts-Brandner, S. J., & Vierling, E. (2001). Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo. *Plant Physiology*, *127*(3), 1053–1064.

Salvucci, M. E., van de Loo, F. J., & Stecher, D. (2003). Two isoforms of Rubisco activase in cotton, the products of separate genes not alternative splicing. *Planta*, *216*(5), 736–744. https://doi.org/10.1007/s00425-002-0923-1

Scafaro, A., De Vleesschauwer, D., Bautsoens, N., Hannah, M. A., den Boer, B., Gallé, A., & Van Rie, J. (2019b). A single point mutation in the C-Terminal extension of wheat Rubisco

activase dramatically reduces ADP inhibition via enhanced ATP binding affinity. *Journal of Biological Chemistry*, *294*(47), 17931–17940. https://doi.org/10.1074/jbc.RA119.010684

Scott, D., & Tesfaye, M. (2000). Development of a breeding pool for a grazingLupinus polyphyllus. *New Zealand Journal of Agricultural Research*, *43*(1), 27–33. https://doi.org/10.1080/00288233.2000.9513405

Ștefan, V., & Levin, S. (2018). Plotbiomes: R package for plotting Whittaker biomes with ggplot2. https://doi.org/10.5281/ZENODO.7145245

Wachter, R., Salvucci, M. E., Carmo-Silva, A. E., Barta, C., Genkov, T., & Spreitzer, R. J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco–Rubisco activase interaction. *Photosynthesis Research*, *117*(1), 557–566. https://doi.org/10.1007/s11120-013-9827-0

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Werneke, J. M., Chatfield, J. M., & Ogren, W. L. (1989). Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and Arabidopsis. *The Plant Cell*, *1*(8), 815–825. https://doi.org/10.1105/tpc.1.8.815

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wijewardene, I., Shen, G., & Zhang, H. (2021). Enhancing crop yield by using Rubisco activase to improve photosynthesis under elevated temperatures. *Stress Biology*, *1*(1), 2. https://doi.org/10.1007/s44154-021-00002-5

Xie, Y. (2022). *Knitr: A general-purpose package for dynamic report generation in R* (tech. rep.).

Yamori, W., Hikosaka, K., & Way, D. A. (2014). Temperature response of photosynthesis in C 3 , C 4 , and CAM plants: Temperature acclimation and temperature adaptation. *Photosynthesis Research*, *119*(1), 101–117. https://doi.org/10.1007/s11120-013-9874-6

Young, S. N. R., Sack, L., Sporck-Koehler, M. J., & Lundgren, M. R. (2020). Why is C4 photosynthesis so rare in trees? (C. Raines, Ed.). *Journal of Experimental Botany*, *71*(16), 4629–4638. https://doi.org/10.1093/jxb/eraa234

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner : Standardized cleaning of occurrence records from biological collection databases (T. Quental, Ed.). *Methods in Ecology and Evolution*, *10*(5), 744–751. https://doi.org/10.1111/2041-210X.13152

Chapter 3

Optimising Expression and Purification of Recombinant Rubisco activase

3.1 Introduction

Rubisco activase (Rca) is a critical enzyme in photosynthesis, required for the maintenance of Rubisco activity. Rca is often present in multiple isoforms in plant species. In order to characterise the catalytic properties of these isoforms individually, each isoform must be separately purified. This makes purification from leaves non-viable for assaying characteristics of individual isoforms. Biochemical characterisation of the properties of Rca requires milligram quantities of pure and active protein. This task can be subdivided into multiple smaller requirements which enable purification of adequate Rca quantities:

- Sufficient protein yield
- Protein solubility & stability
- Activity of purified protein
- Protein purity
- Time to attain end product
- Cost-effectiveness

Each of these can be affected by multiple factors such as *E. coli* growth media and growing conditions, construct design and protein purification method. There is currently no consensus of the best conditions for expression and purification of Rca.

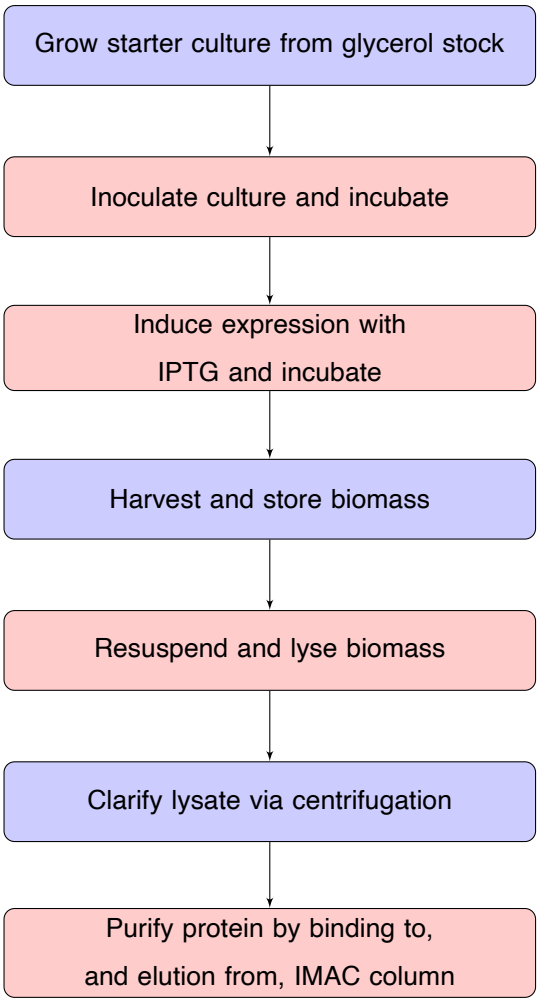


Figure 3.1: *Outline of the basic steps of recombinant expression and purification of proteins using IMAC (Immobilized Metal Affinity Chromatography). Red blocks indicate those targeted for optimisation. IPTG: Isopropyl β-D-1-thiogalactopyranoside*

Figure 3.1 shows the steps involved in expression and purification of heterologously expressed protein. Every step has factors which can be optimised for particular proteins to improve overall yield and quality of final product. Some of the more important factors to consider and how they are used in previously published Rca expression methods are shown in Table 3.1 and Table 3.2.

Table 3.1: *Growth conditions of* E. coli *cultures for heterologous expression of Rca from select publications. NA: not available; IPTG: isopropyl β-D-1-thiogalactopyranoside; LB: lysogeny broth.*

Medium	Expression time / OD	Culture stirring (RPM)	Growth temperature (°C)	Expression temperature (°C)	Expression time (hr)	Flask volume (L)	Medium volume (L)	Induction conditions	Citation
LB	Mid-log phase	NA	37	23	20	NA	NA	0.5 mM IPTG	Stotz et al., 2011
LB	OD_{600} of 0.6-0.8	225	37	25	12	15	6	0.5 mM IPTG	Barta et al., 2011a
2xYT	OD_{600} of 0.4	NA	37	30	4	NA	8	0.1 µg/mL anhydrotetracycline	Hasse et al., 2015
LB	OD_{600} of 0.4-0.6	225	37	25	12	1	0.4	0.5 mM IPTG	Scafaro et al., 2016
LB	OD_{600} of 0.6	200	37	25	8	NA	3	1 mM IPTG	Peterson-Forbrook et al., 2017

Publications with Rca purification methods tend to exclude at least a few pieces of information which are known to be relevant to yield and quality during protein production, making comparisons of methods challenging and incomplete. LB is the most common medium for production of Rca, with one research group favouring 2xYT, which contains the same components in different proportions. Aside from medium, various growth conditions are employed for aiding solubility. Hasse et al. (2015) use heat and salt shock and multiple groups use lower temperature for expression than for growth (Table 3.1), which has been shown to aid solubility due to a decrease in hydrophobic interactions which can lead to aggregation.

Table 3.2: *Lysis, purification, and storage methods for Rca expressed heterologously in* E. coli *from select publications.*

Lysis conditions	Lysis buffer	Purification conditions	Storage conditions	Citation
Lysed by ultrasonication	50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 1 mg/mL lysozyme, 1 mM PMSF, pH 8.0	Ni-NTA gravity column His tag purification, eluted with 200mM imidazole, ubiquitin moiety cleaved overnight at 23 °C	Dialysed against 50 mM Tris-HCl pH8.0 1mM NaCl, eluted with a linear salt gradient up to 500mM on MonoQ column and stored after concentration in 5% glycerol at -80 °C	Stotz et al., 2011
Five cycles of sonication, 1 min each, at 50% of maximum power. 30 s pause between each cycle	50 mM HEPES-KOH, pH 7.0, 5 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) Triton X-100, 2 mM ATP, 5 mM DTT, 20 mM ascorbate, 1 mM PMSF, and 10 mM leupeptin	Ammonium sulfate precipitation and anion-exchange chromatography	Flash frozen in 20 mM HEPES-KOH, pH 7.2, 5 mM $MgCl_2$, 0.2 mM ATP and 2 mM DTT. Stored at -80 °C	Barta et al., 2011a
Five cycles of sonication, 1 min each, 30 s pause between each cycle	40 mL of 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole pH 8.0, 30% glycerol, 1 mM DTT, 0.1 mM ADP, 0.1 mM EDTA pH 8.0, 1 mM PMSF and 20 mg hen egg white lysozyme	Ni-NTA column (Qiagen), eluted with imidazole step gradient from 10 to 500 mM, ubiquitin moiety cleaved overnight at 16 °C during dialysis	Flash frozen in 25 mM HEPES pH 7.2, 250 mM KCl, $MgCl_2$, 1% glycerol, and 2 mM ATPγS. Stored at -80 °C	Peterson-Forbrook et al., 2017
Lysed by sonication	20 mM bis-tris propane pH 8, 200 mM KCl, 10% glycerol, 5 mM ATP, 5 mM $MgCl_2$, 5 mM TCEP	Purified using a three-step protocol comprising affinity, ion-exchange and size-exclusion chromatography	Not stored, used for crystallisation	Hasse et al., 2015

As with expression, a lot of variation exists in the methods for purification of Rca. The community has largely shifted in favour of tagged protein (usually His-tags) due to the relative ease of purification with IMAC, but the methods of Barta et al. (2011a) are still employed by some researchers. The principle of IMAC purification is that histidine residues will bind strongly but reversibly to metal ions (Spriestersbach et al., 2015). A polyhistidine tag is inserted on either the N or C terminus of the protein of interest, which will allow it to bind to the IMAC resin. The protein of interest can then be eluted using imidazole, which competes with the protein of interest for the resin's binding space. For purification of cowpea Rca, N-terminally His-tagged Rca genes were transformed into *E. coli*. A T7 expression system enables induction by IPTG (Studier and Moffatt, 1986), and a TEV protease cleavage site allows for removal of the His-tag. An example map of the plasmids transformed into *E. coli* can be seen in Figure 3.2.

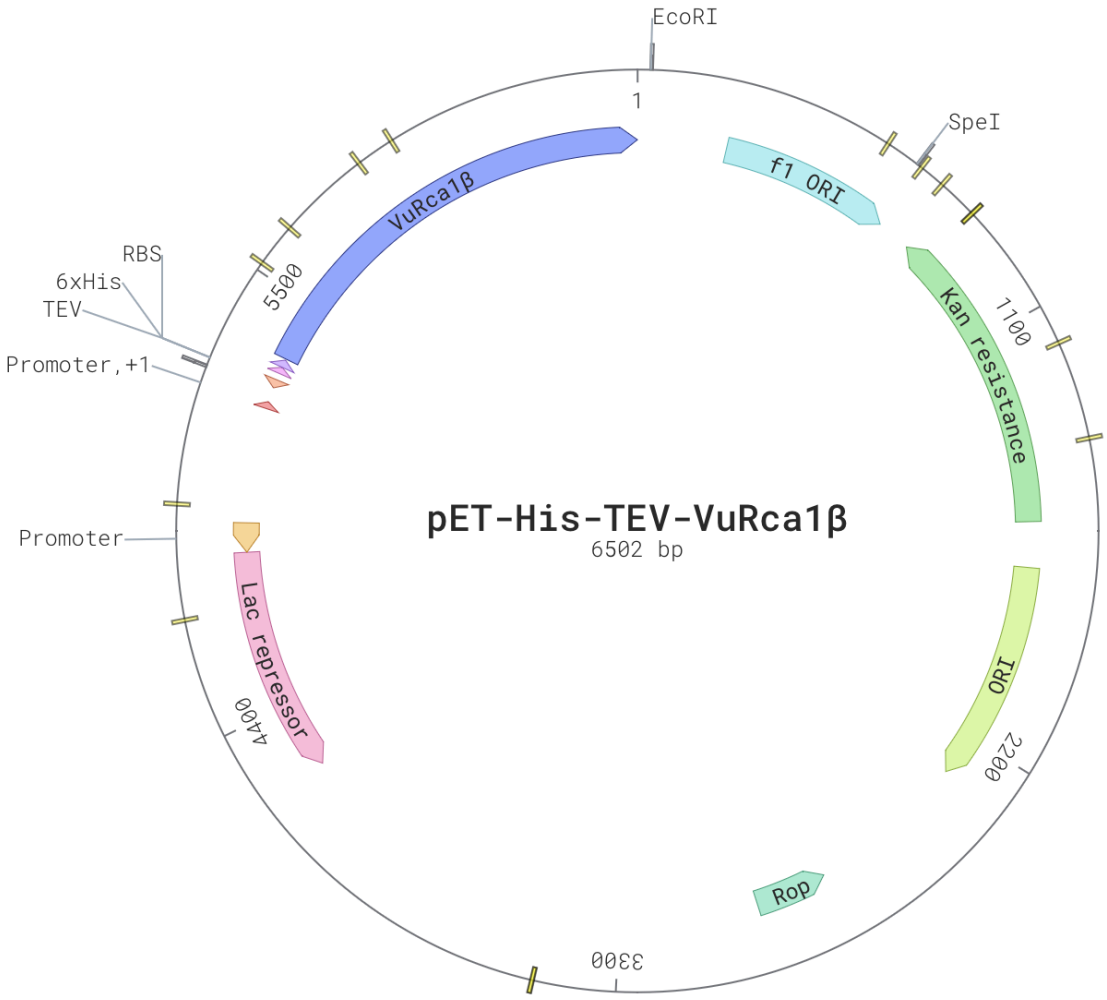


Figure 3.2: *Plasmid map of pET-His-TEV-VuRca1β. This circular plasmid is represented with labelled features indicating polyhistidine tag (6xHis), TEV cleavage site (TEV), origins of replication (ORI), Ribosome binding site (RBS), and additional functional elements.*

This chapter will focus on the methods developed to obtain adequate Rca in a robust and high-throughput manner, with simple shake-flask methodology and no need for reactor scale expression. It also aims to detail considerations for further developments, and assessments of more difficult to purify isoforms. The developed protocol was tested on ten Rca isoforms - four from *Vigna unguiculata* (cowpea) and six from *Glycine max* (soybean).

3.1.1 Optimising Media to Improve Protein Production

Three factors are important for heterologous purification of protein - biomass, protein produced per cell, and solubility of protein, as proteins cannot be purified from insoluble inclusion bodies without time consuming and laborious methods. Different growth media will affect these factors. Furthermore, cost effectiveness of these media is important as multiple isoforms are needed which will be a substantial investment if the media is costly. Lysogeny broth (LB) is a growth medium

used since the 1950s (Bertani, 1951) and is a standard growth medium for molecular biology work for bacteria without strict growth requirements. LB and M9ZB are first generation media which serve as a benchmark for media developed later. Second generation media such as 4xYT and Terrific broth grow to higher biomass than first generation, and third generation media such as Dynamite and Magic media are developed to increase per cell productivity and biomass gains over second generation (Taylor et al., 2017). Magic medium is an autoinducible medium which begins to express protein of interest upon temperature shift to 20 °C rather than upon addition of IPTG. Autoinduction was first utilised by Studier *et al.* in 2005, based on the concept that as cultures grow, the preferred carbon source of *E. coli* shifts from glucose to lactose, an inducer of the T7 promoter system (Studier, 2005). This has the advantage of removing the need to monitor cultures and manually add IPTG, making protocols more convenient and improving reproducibility. However, autoinducible media are more complex than regular growth media and must be more carefully produced (Studier, 2018). The optimal medium for a given project will vary based on cost, convenience, yield of protein required and the number of different proteins which must be produced.

3.1.2 Optimising Induction Parameters

The time of induction has been shown to have a significant impact on recombinant protein production. Traditionally, the time of induction is based on OD_{600} as a metric of cell biomass. OD_{600} is problematic as a proxy for cell biomass due to its narrow linear range and instrument-to-instrument variance (Beal et al., 2020). Furthermore, the traditional point of induction of OD_{600} = 0.8-1.0 is suggested to be too low, and later induction may result in increased protein production (Collins et al., 2013). The concentration of IPTG used to induce protein synthesis has additionally been shown to change recombinant protein production. Traditionally, a concentration of 0.1-3.0 mM IPTG has been used to induce protein synthesis. However, the optimal IPTG concentration for induction varies with other parameters such as growth temperature, and it has been demonstrated that the traditional concentrations may be too high (Mühlmann et al., 2017; Slouka et al., 2018).

3.1.3 Optimising Solubility

Protein solubility is a long-standing issue with protein purification. Insoluble proteins in *E. coli* form inclusion bodies made of partially folded protein aggregates which lack activity and proper 3D structure. While insoluble proteins are not the dead end waste products they were thought to be, it is challenging and laborious to isolate active proteins from the insoluble fraction (Ramón et al., 2014; Singh and Panda, 2005). Inclusion bodies are an interesting area of research as they can provide a high density source of high purity protein, assuming they can be isolated and refolded to regain proper activity. With current technology, however, if it can be avoided, it is simpler to keep the protein of interest soluble for purification. The mechanisms governing protein aggregation and inclusion body formation are poorly understood but some evidence points to it

being a concentration-dependent process (Pei et al., 2021). A balance between expression and proper folding must be maintained but since prokaryotic cells translate mRNA much faster than eukaryotic cells, insufficient folding factors or excessive translation can cause some proteins to remain unfolded (Gatti-Lafranconi et al., 2011; Wurm et al., 2017). This can be ameliorated with concomitant expression of chaperones or other folding factors, or induction of heat shock proteins such as via heat shock or addition of ethanol (Thomas and Baneyx, 1996). Factors such as medium pH and temperature can affect inclusion body formation (Strandberg and Enfors, 1991). The amino acid interactions leading towards insolubility are complex and difficult to characterise, even single amino acid changes can drive proteins into insolubility with little recourse for recovery (Izard et al., 1994). While software exists that can predict the solubility of a protein it is imperfect and often incorrectly predicts solubility (Hou et al., 2018; Raimondi et al., 2020; Trevino et al., 2007; Yang et al., 2021). Likewise, they employ sequence data primarily and do not typically account for the exact system of expression such as pH and temperature. The addition of various supplements to lysis buffer in the preparation of cell lysate can aid in protein solubility prior to purification. L-Arginine and L-Glutamine are common additives but there is no ”silver bullet” which will work on all proteins and screening experiments are required to find a suitable supplement (Golovanov et al., 2004; Hamada et al., 2009; Leibly et al., 2012).

3.1.4 Optimising Purification Parameters

Before proteins can be purified, the cells which express the protein of interest must be lysed to release intracellular content. There are several methods of cell lysis available. The most common cell lysis methods are sonication, the French press, and chemical lysis (Shehadul Islam et al., 2017). Simple freeze-thaw cycling can also be sufficient to extract recombinant protein (Johnson and Hecht, 1994). Sonication is the most prevailing method due to its simplicity and low cost. It further has the advantage of randomly shearing DNA (Elsner and Lindblad, 1989), while other methods typically require DNAse treatment to reduce viscosity caused by release of DNA. Sonication is deceptively complex, and results can be affected by factors that are not often considered such as volume (Feliu et al., 1998), power (James et al., 1972), time of sonication, and cell concentration (Pchelintsev et al., 2016). Additionally, sonication produces a lot of heat so adequate temperature control is necessary (Chisti and Moo-Young, 1986), especially when working with enzymes which can be quite fragile and prone to heat degradation. Sufficient time of sonication is needed to release adequate protein, but excessive sonication can cause damage due to heating, and protein aggregation (Stathopulos et al., 2008). The method also scales poorly so while it is adequate for small scale purifications it is inefficient for industrial use (Geciova et al., 2002). Chemical lysis can be done with commercially available reagents or with non-commercial extraction solutions which often consist of detergents and chaotropic agents (Danilevich et al., 2008; Peach et al., 2015).

For protein purification, IMAC purification is a rapid and reliable method of purifying a protein containing a polyhistidine tag (Block et al., 2009). This method can attain a pure and active pro-

tein with minimal steps without relying on the expressed protein's biochemical properties such as size or isoelectric point. The usual eluent of IMAC purification is imidazole, which competes for binding sites with the polyhistidine tag of the protein of interest (Bornhorst and Falke, 2000). It is common to encounter some degree of impurity from contaminating proteins during IMAC purification. This is typically from proteins with multiple histidine residues which will bind the metal resin. Some specialised *E. coli* strains have been developed to mitigate this issue (Andersen et al., 2013; Robichon et al., 2011). Increasing imidazole during elution will increase protein yield, but may increase contaminating protein. Imidazole can interfere with some assays (Bhat et al., 2018), and there is conflicting information on whether it improves (Hamilton et al., 2003) or reduces (Mezzasalma et al., 2007) protein stability and solubility. Thus, it is recommended to remove imidazole post-purification, and to optimise the concentration of imidazole in the wash and elution stages to balance protein stability, purity and yield.

During and after purification, (i) buffer choice and composition is an important consideration. (ii) pH is relevant to activity, stability (Talley and Alexov, 2010), and solubility (Kramer et al., 2012; Pelegrine and Gasparetto, 2005; Zbacnik et al., 2017) of protein. (iii) the choice of buffer depends on downstream uses of the protein of interest, and any specific protein-buffer interactions. For example, tris can interfere with some protein activity (Desmarais et al., 2002), and phosphate buffers are incompatible if measurements of ATPase activity are needed. Various additives may be needed such as reducing agents for redox sensitive proteins.

Post purification, protein must be frozen to ensure stability during long term storage. Protein can be lyophilised and stored dry, this is common for therapeutic purposes, but it is simpler to store protein in solution. Traditionally, protein is flash frozen and stored at -80 °C with small concentrations of cryoprotectants (e.g. ~5-10% glycerol) for long term storage but there is conflicting information over whether this is the best method of protein storage (Bhatnagar et al., 2007; Cao et al., 2003). Fast freezing such as in liquid nitrogen has been associated with protein unfolding and loss of enzyme activity, and low temperatures have a large, and buffer-dependent, effect on system pH (Jiang and Nail, 1998; Park et al., 2021).

This chapter describes an optimisation process to improve yield and solubility of cowpea Rca isoforms. As early attempts at purification showed VuRca10α as challenging to purify, this was chosen as an optimisation target. Not every isoform benefits equally from the optimisation process, but improvement was shown across all isoforms. The improved protocol developed by this work has enabled purification of sufficient protein for later characterisation of cowpea Rca isoforms.

3.2 Materials and Methods

3.2.1 Cell Transformation

500 ng of plasmid was mixed with 50 µL samples of competent BL21 *E. coli* cells and left on ice 30 minutes. Competent cells were then heat shocked for 60 s at 42 °C and returned to ice for 60 s.

1 mL of LB was added to competent cells and they were left to express for 60 minutes before being streaked on agar plates containing relevant antibiotics (kanamycin and chloramphenicol). Successful transformants were used to make glycerol stocks for long term storage.

3.2.2 Protein Concentration Determination

Protein concentration was determined using a modified version of the Bradford method (Bradford, 1976; Ernst and Zor, 2010). 0-50 µg of bovine serum albumin was used to construct the standard curve. 1 µL of sample, 99 µL of milliQ water and 100 µL of Bradford reagent were added in triplicate to wells of a 96-well plate. Protein concentration was estimated using the standard curve comparing protein amount to the ratio of absorbance at 590 and 450 nm.

3.2.3 SDS-PAGE

Protein samples were mixed 1:1 with 2x reducing laemmli buffer (20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.004% bromophenol blue, 0.125 M Tris-HCl, pH 6.8) and boiled at 95 °C for 5 minutes. Samples were separated by electrophoresis on 12% Mini-PROTEAN TGX Stain-Free gels (Bio-Rad) at 300 V until the dye front ran off the gel (approximately 40 minutes). SDS-PAGE Running buffer contains (125 mM Tris, 200 mM glycine, 0.1% SDS) in a 1x stock. After separation, stain-free gels were imaged on the GelDoc Go gel imaging system (Bio-Rad).

3.2.4 Western Blotting

Stain-free gels were cut down to membrane size and transferred to nitrocellulose membranes using the iBlot Gel Transfer Device (ThermoFisher) as per manufacturer instructions. Membranes were rinsed twice in milliQ water and blocked in 4% milk in TBS (tris-buffered saline) (20 mM Tris, 150 mM NaCl, pH 7.5) for 30 minutes. After blocking, membranes were rinsed in milliQ water and probed with primary antibody with gentle shaking (15 RPM) either overnight, or for 2 h. Overnight staining and 2 h staining in antibody were found equivalent, and one may be preferred over another for logistical reasons. Membranes were washed 3x 5 minutes in TBST (TBS + 500 µL/L Tween-20) and visualised using the Odyssey Fc (LI-COR) at 700 nm (molecular weight marker) and 800 nm (antibody). A mouse anti-histidine tag antibody (Bio-Rad Cat# MCA1396 D800GA, RRID:AB_11152596) was used for visualisation of His-tagged Rca. For more information on quantification of photosynthetic proteins by western blotting see Perdomo et al., 2018.

3.2.5 ATPase Analysis

ATPase activity was measured by the methods of Chifflet et al. (1988). The reaction was initiated by addition of 5 µg of Rca to reaction mix, and terminated by addition of 50 µL of 12% SDS after 5 minutes of reaction time. Blanks were produced by addition of SDS to the reaction mix prior to addition of Rca. A calibration curve was constructed from 0 to 50 nmol of inorganic phosphate.

Absorbance at 850 nm was measured using a SPECTROstar Nano microplate spectrophotometer (BMG Labtech)

3.2.6 Solubility Analysis

0.5 mL of cell lysate was centrifugated at room temperature in a miniSpin (Eppendorf) centrifuge at 13.4k RPM for 2 minutes. The supernatant was reserved as the soluble fraction and the pellet was washed in 0.5 mL of milliQ water and centrifugated under the same conditions as previously stated. This wash process was repeated once and the pellet was then resuspended in 0.5 mL of 8 M urea. This was reserved as the insoluble fraction. These samples can be used in gel and western blot analysis.

3.2.7 Rca Expression

Expression conditions as they changed are specified in text. Under the optimized protocol, a glycerol stock of relevant BL21 *E. coli* was used to inoculate 15 mL of LB (Lennox, supplemented with 50 µg/mL of kanamycin and chloramphenicol) in a 50 mL falcon tube. This was then left to grow overnight with 225 RPM shaking at 37°C. 7 mL of overnight medium was added to 1 L of dynamite medium (see 3.2.8).

The cultures were then left to express protein overnight with 250 RPM shaking at 24 °C. After 20 h, the cultures were centrifugated in a JLA-10.500 rotor (Beckman) at 5,000 RPM for 20 minutes at 4 °C. For ease of storage, the biomass can be further resuspended and centrifugated in 50 mL falcon tubes. The biomass was snap frozen in liquid nitrogen and stored at -80 °C until purification.

For optimisation experiments, smaller 50 mL cultures in 250 mL flasks were used instead of 1 L cultures in 3 L flasks to enable larger sample sizes.

3.2.8 Media Preparation

Proper medium preparation is crucial to enable bacterial growth. During Dynamite medium preparation it was noted that a precipitate would form after autoclave sterilisation. The recipe for Dynamite medium as described in Taylor et al. (2017) is quoted verbatim here:

Dissolve 12 g of tryptone, 24 g of yeast extract, 6.3 mL of glycerol, 3.8 g of KH_2PO_4, 12.5 g of K_2HPO_4, 5 g of glucose, and 0.195 g of $MgSO_4$ (anhydrous) in a total volume of 1 L of water, autoclave to sterilize.

It is known that preparation and sterilisation of individual components can improve specific protein productivity (Collins et al., 2013; Tanaka et al., 2014), and that autoclaving phosphates in the presence of metals can lead to precipitation of medium components. Given this, a modified method of preparation of Dynamite medium has been used.

- Dissolve 12 g tryptone, 24 g yeast extract, 0.195 g $MgSO_4$, 6.3 mL glycerol in a total volume of 950 mL of water, autoclave to sterilise.

- Dissolve 5 g of glucose in 10 mL of water and filter sterilise through a 0.22 µm filter into a sterile glass bottle.
- Dissolve 3.8 g of KH_2PO_4, 12.5 g of K_2HPO_4 each in 20 mL of water and autoclave to sterilise.
- In a sterile environment, combine all components.

No precipitation of medium components was noted by this method. By the original method of Taylor et al. (2017), this issue was only noted with a P2222 (Sigma) stock of K_2HPO_4 and not with a 60353 (Sigma) stock.

LB (Lennox) was prepared as per manufacturer instructions.

3.2.9 Cell Lysis and Lysate Preparation

Cell lysis conditions as they changed are specified in text. Under the optimized protocol, fresh biomass was resuspended in lysis buffer (50 mM HEPES, 5 mM $MgCl_2$, 3 M NaCl, 50 mM imidazole, 1% (v/v) Triton X-100, 20 mM sodium ascorbate 0.2 mM ATP, pH 7.2 with HCl) in a plastic beaker. 7 mL of lysis buffer was added for each gram of fresh pellet. The cell suspension was sonicated for 5 minutes (10 minutes including cooling time: 30s on, 30s off), at 50% power in a VCX 130 Ultrasonic Processor (Sonics & Materials). Ice-water was used as a cooling solution and the sonicator probe was immersed near the bottom of the beaker. A 500 µL sample of lysate was taken for solubility analysis and the remaining lysate was centrifugated at 4 °C in a JA25.50 rotor (Beckman) for 30 minutes at 23,000 x *g*. The resultant pellet was discarded and the supernatant was filtered through a 0.45 µm filter prior to purification.

When measuring the change of temperature during lysis, RO water was used as a proxy for lysate as this was shown to have no significant difference on rate of temperature change.

To measure cellular humidity, the mass of an aliquot of cell pellet was measured before and after drying by heat.

3.2.10 Manual Rca Purification

Purification conditions as they changed are specified in text. Under the optimized protocol, 2 mL of Profinity IMAC resin (Bio-Rad) per litre of cell culture was equilibrated as per manufacturers instruction with sodium phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM $MgCl_2$, 50 mM imidazole). The clarified lysate was left to mix with gentle shaking at 4 °C for 30 minutes. The mixed slurry was applied to an AGC250 gravity column (Expedeon) and the flowthrough collected for western blot analysis. The column was washed twice with 5 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM $MgCl_2$, 100 mM imidazole) per wash. The column was then eluted into 1 mL fractions with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM $MgCl_2$, 300 mM imidazole) until the fractions no longer developed colour with a qualitative Bradford assay. 50 µL of each aliquot was taken for gel and blot analysis and the rest was concentrated in an amicon spin column with a 30 kDa molecular weight cut-off then stored.

3.2.11 Automated Rca Purification

Purification conditions as they changed are specified in text. Under the optimized protocol, clarified lysate was loaded onto the Bio-Rad NGC system. The lysate was applied to a 1 mL HiTrap FF Crude column (Cytiva). The column was then washed with 10 column volumes (CVs) of wash buffer (50 mM HEPES, 300 mM NaCl, 5 mM $MgCl_2$, 100 mM imidazole) and eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 5 mM $MgCl_2$, 300 mM imidazole) onto a Bio-Scale Mini Bio-Gel P-6 Desalting Cartridge (Bio-Rad). Rca was then eluted from the desalting column with storage buffer (50 mM HEPES, 5 mM $MgCl_2$, 5% (v/v) glycerol, pH 7.2) in 1 mL fractions. 50 µL of each aliquot was taken for gel and blot analysis and the rest was stored.

3.2.12 Rca Storage

Purified Rca aliquots were combined and supplemented with 2 mM DTT and 0.2 mM ATP then aliquoted into 50 µL aliquots. The aliquots were pipetted into tubes pre-cooled in liquid nitrogen and stored at -80 °C long term.

3.2.13 Models and Figure Preparation

Analysis of variance followed by Tukey's post-hoc tests were used to test for significance of the effects of optimisation parameters on Rca expression and solubility. Tukey-style box plots are plotted with summary statistics as described in the ggplot2 documentation.

Data were analysed in R (R Core Team, 2013) 4.1 and RStudio (RStudio Team, 2020) 1.4.1717 on macOS Big Sur 11.4. The Tidyverse (Wickham et al., 2019) 1.3.1 collection of packages and patchwork (Pedersen, 2022) 1.1.1 was used for plotting of data and general data wrangling. rstatix 0.7.0 and ggpubr 0.4.0 were used for calculating and plotting significance values. Compact letter display of significance values were calculated using multcomp (Hothorn et al., 2008) 1.4-17.

Plasmid map was visualised with SeqViz (Lattice-Automation). pLannotate (McGuffie and Barrick, 2021) was used for plasmid annotation.

3.3 Results

3.3.1 Expression Optimisation

Growth Medium Affects Rca Production and Biomass

Optimisation of expression using the one factor at a time method is a multi-step process. The first step in the process was to optimise the growth medium used to grow *E. coli*. Other factors will have different optima depending on the medium used. LB was initially selected as growth medium for culture of *E. coli*. It is a simple and cheap medium which can grow *E. coli* to moderate biomass.

However, it quickly became apparent during preliminary experiments that LB would produce insufficient Rca for characterisation without scaling up to extreme amounts, and thus different media were investigated which have been known to increase heterologous protein production. Six media were selected for experimentation to find the highest protein production - LB, M9ZB, 4xYT, Terrific broth, Dynamite medium and Magic medium. As MagicMedia is an autoinducible medium, it does not require induction with IPTG and instead is often temperature shifted for the expression stage. Fresh biomass and raw protein expression was obtained for the six different media after induction with 0.5 mM IPTG (where necessary) at OD_{600} of 0.6-0.8 followed by 20 h of expression at 25 °C.

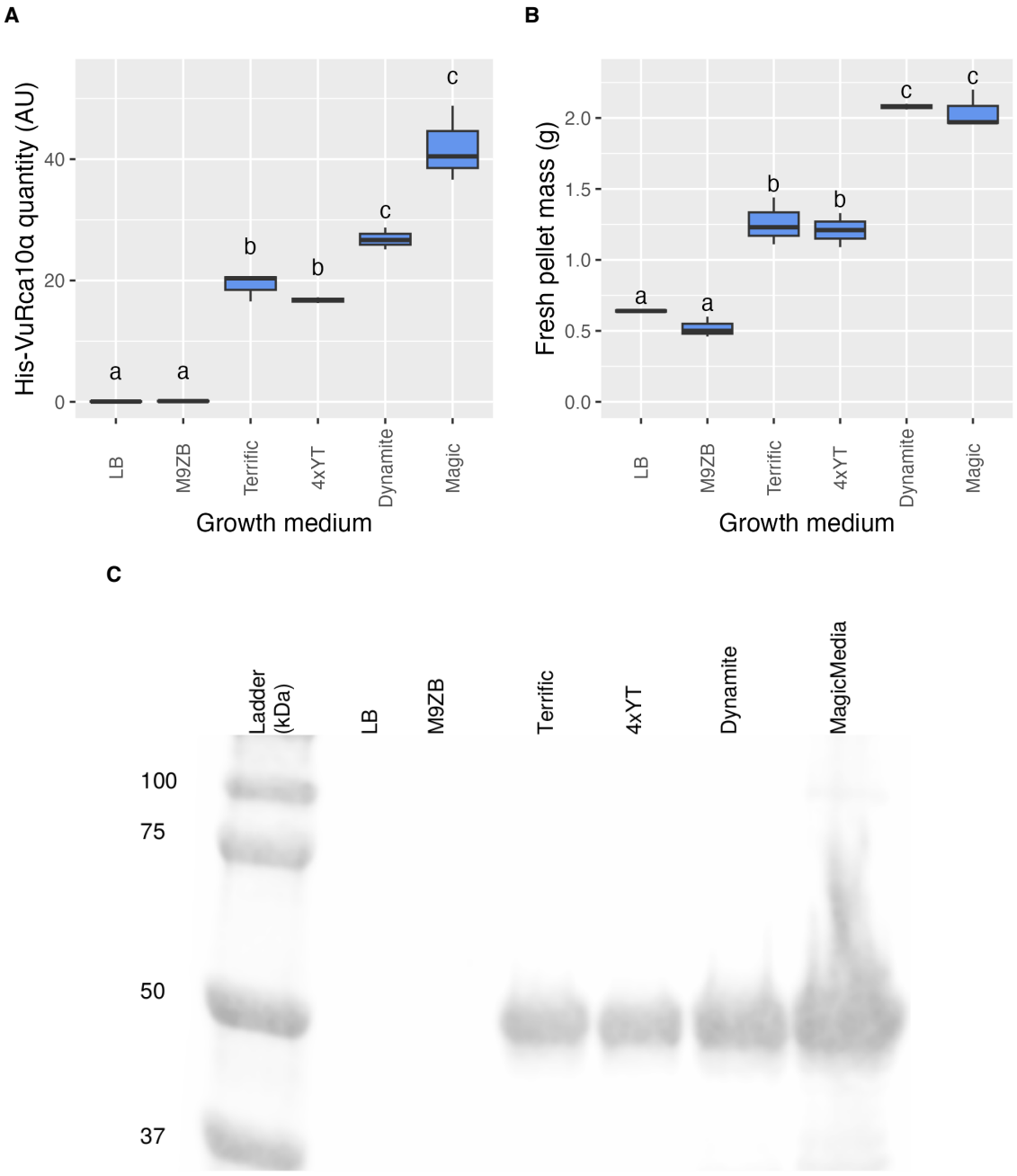


Figure 3.3: *The effect of growth medium on recombinant protein production and fresh cell biomass. Comparison of the expression* ***A*** *of His-VuRca10α and fresh cell biomass* ***B*** *after expression in six different growth media.* ***C*** *Western blot used to quantify Rca expression. Manually induced cultures were grown for four hours at 37 °C then induced with IPTG and grown for a further 20 hours after induction at 24 °C. MagicMedia, an auto-inducing growth medium, was temperature shifted to 24 °C instead of induced with IPTG. Expression was estimated by band intensity on a western blot probed with anti-His antibody. n = 3 replicates of cultures in each growth medium. Boxes which do not share a letter indicates significant difference in a pairwise comparison by Tukey HSD ($p < 0.05$).*

MagicMedia produced the highest volumetric protein production, while LB and M9ZB produced

comparatively poor amounts of protein (Fig. 3.3A). Note that while no band can be seen for LB and M9ZB in Figure 3.3C, these become visible under longer exposure, but at the cost of oversaturation of more prominent bands. Being autoinducible, MagicMedia is also highly reproducible and convenient compared to IPTG inducible media. MagicMedia produced a statistically significantly higher amount of His-Rca than the next closest amount (Dynamite). However, MagicMedia is a commercially available product which is comparatively expensive (77.50 GBP/L), while Dynamite medium can be made quickly from inexpensive components. Because of this, Dynamite medium was chosen as the growth medium for subsequent optimisation and expression experiments. Dynamite contains glucose and glycerol as cellular energy sources. Excessive glucose can cause a reduction in cell growth (Kazan et al., 1995). The potassium phosphates in the medium serve both to buffer the pH of the system and as a source of bioavailable phosphates. The amount of fresh biomass was not significantly different between MagicMedia and Dynamite, but the humidity of cellular biomass can change from culture to culture and by medium. This makes the dry biomass a more accurate comparison of cell biomass. The dry biomass of other media was not tested, but subsequent measurements with Dynamite medium put the humidity of cultures grown in Dynamite medium at 78%±8%.

Induction Time Affects Rca Production

With a medium selected, the time of induction was the next factor optimised. Cultures were grown in Dynamite medium at 37 °C then induced at various times of growth. After induction the cultures were temperature shifted to 25 °C and grown for a further 20 h before collection of biomass. The amount of soluble recombinant His-Rca produced depends significantly on the time of induction. Four hours seems to be the optimal time of growth before induction (Fig. 3.4) and induction after this time results in significantly reduced expression.

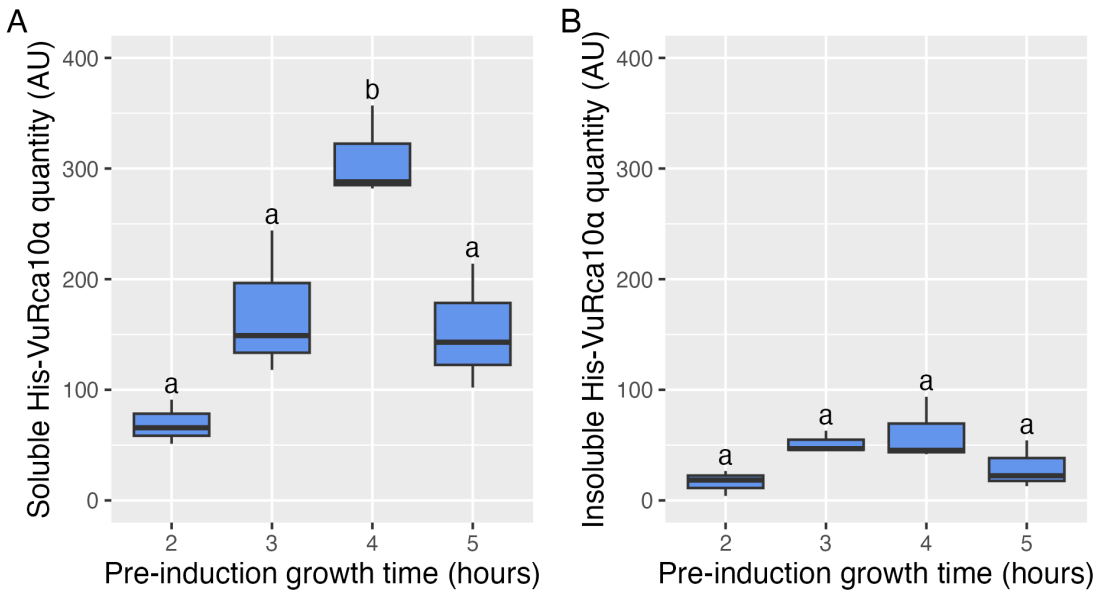


Figure 3.4: *The effect of pre-induction growth time on **A** soluble Rca production, and **B** insoluble Rca production. Cultures were fermented at 37 °C for 2, 3, 4, or 5 hours then expression was induced by addition of IPTG. Cultures were grown for a further 20 hours after induction and expression was estimated by band intensity of lysate on a western blot probed with anti-His antibody. Insoluble proteins in the resultant pellet were resuspended in 8 M urea. n = 3 replicates of each culture condition. Boxes which do not share a letter indicates significant difference in a pairwise comparison by Tukey HSD ($p < 0.05$).*

After eight hours, the amount of protein produced had dropped to less than the amount produced at 2 hours (data not shown). The amount of insoluble His-Rca did not depend significantly on time of induction. A change in the amount of pre-induction growth time is a simple way to improve the amount of recombinant protein produced.

Expression Temperature Affects Rca Production and Solubility

The next expression factor to be optimised was the temperature of expression. Expression temperature is typically at 37 °C (optimal for growth of *E. coli*) in single day expression protocols, but lower for overnight expression. Reducing expression temperature can aid in solubility of recombinant proteins.

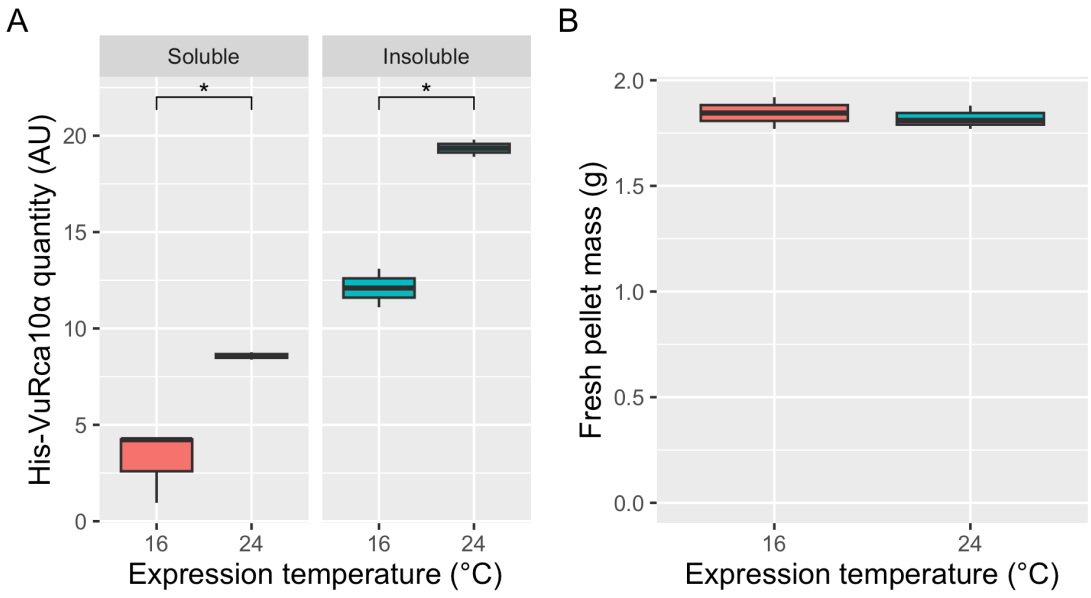


Figure 3.5: *The effect of expression temperature on **A** recombinant protein production and **B** fresh cell biomass. Cultures were grown for 4 hours at 37 °C then induced with IPTG and grown for a further 20 hours after induction at either 16 °C or 24 °C. Expression was estimated by band intensity of lysate on a western blot probed with anti-His antibody. Insoluble proteins in the resultant pellet were resuspended in 8 M urea. n = 3 replicates of cultures at 24 °C and n = 2 at 16 °C. Asterisks indicate significant differences between pairwise comparisons by Tukey HSD (* $p < 0.05$).*

The solubility of His-Rca was greater in cultures grown at 24 °C (31% of protein was soluble) than cultures grown at 16 °C (21% of protein was soluble). Moreover, the total amount of His-Rca was greater at 24 °C than at 16 °C, meaning more soluble protein was produced at 24 °C. The drastic difference in solubility compared to previous optimisation experiments highlights issues with attaining repeatable results from expression or downstream processing of biomass with recombinant protein.

3.3.2 Purification Optimisation

After increasing protein available for purification, the process of purification was optimised, with the aim to maximise Rca yield and quality, while minimising the amount of contaminating protein.

Imidazole Concentration Can Be Optimised to Be Improve Rca Purity

As a competitor for the binding sites on an IMAC column, an increase in the concentration of imidazole in binding buffer will result in a more pure protein. Proteins which loosely bind the column will be out-competed by the imidazole in the buffer. However, this will also reduce the amount of target protein which binds the column and thus the amount of eluted protein. The concentration of imidazole in binding, wash, and elution buffer was optimised to increase purity without excessive

loss of protein yield.

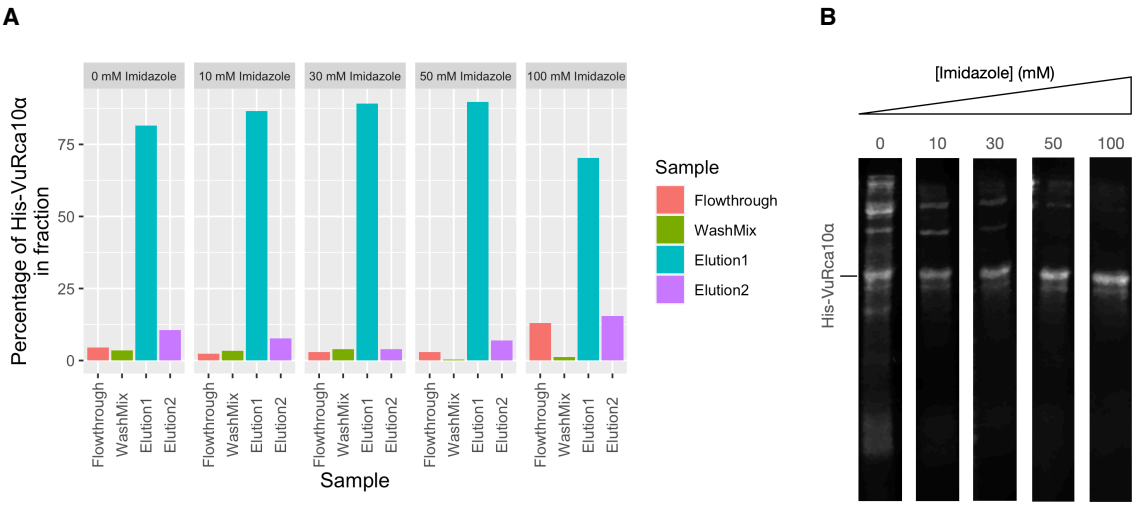


Figure 3.6: *The effect of binding buffer imidazole concentration on elution of Rca from IMAC resin. Column was washed with 150 mM imidazole and eluted with 500 mM imidazole.* ***A*** *The effect of binding buffer imidazole concentration on elution of His-Rca in each fraction. Amount of His-Rca was estimated by band intensity on a western blot probed with anti-His antibody.* ***B*** *SDS-PAGE gel image analysis of the effect of binding buffer imidazole concentration on purity of first elution.*

No imidazole in the binding buffer results in an impure protein (Fig. 3.6) and increasing the amount of imidazole increases purity substantially. However, at 100 mM, an appreciable amount of target protein does not bind the column and was lost to the flowthrough. 50 mM imidazole was selected as the binding buffer concentration to balance purity and recovery.

Imidazole in the wash buffer will similarly strip away loosely bound proteins, reducing the amount of contaminating protein in the final product. In Figure 3.7, the imidazole concentration was kept the same in the binding and elution buffer, and changed in wash buffer.

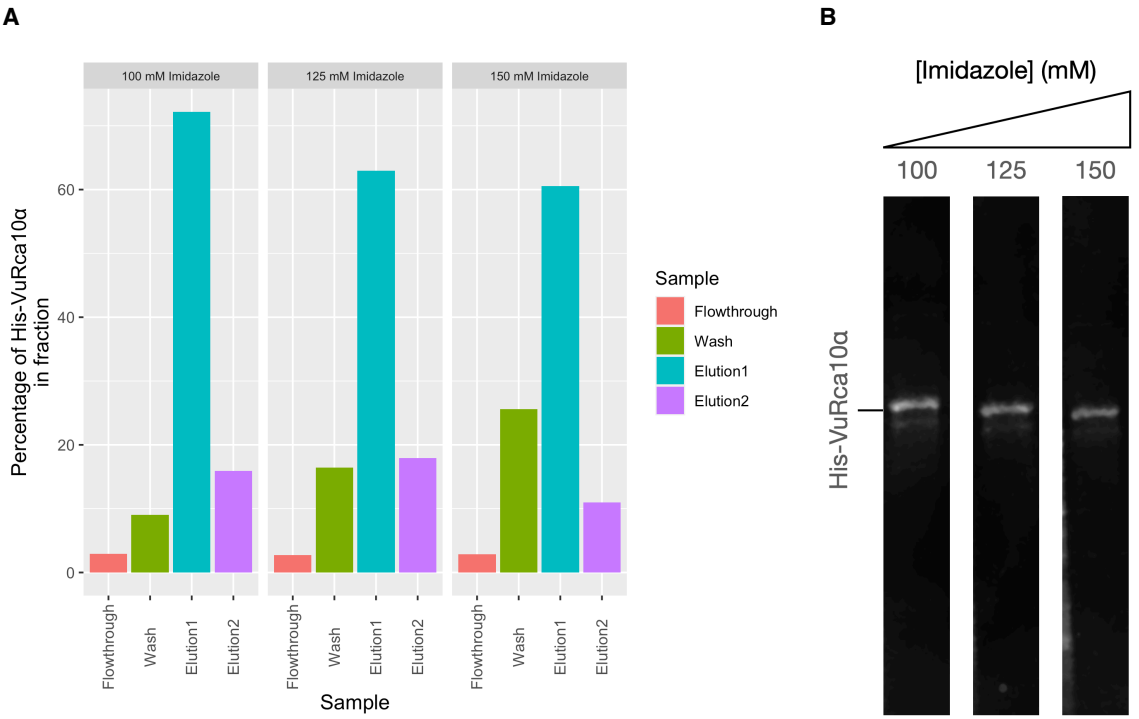


Figure 3.7: *The effect of wash buffer imidazole concentration on elution of Rca from IMAC resin. Column was eluted with 500 mM imidazole. 50 mM imidazole was used in binding buffer.* ***A*** *The effect of wash buffer imidazole concentration on elution of His-Rca in each fraction. Amount of His-Rca was estimated by band intensity on a western blot probed with anti-His antibody.* ***B*** *SDS-PAGE gel image analysis of the effect of wash buffer imidazole concentration of purity of first elution.*

Every investigated concentration of imidazole in wash buffer resulted in a quite pure product (Fig. 3.7). Concomitantly with increased imidazole concentration, the amount of His-Rca lost to the wash step increased. 125 mM imidazole was used in subsequent purification as this slightly improved purity without a large loss of final protein. Finally, the optimal concentration of imidazole in the elution buffer was investigated, with the concentration of imidazole in binding and wash buffer staying the same.

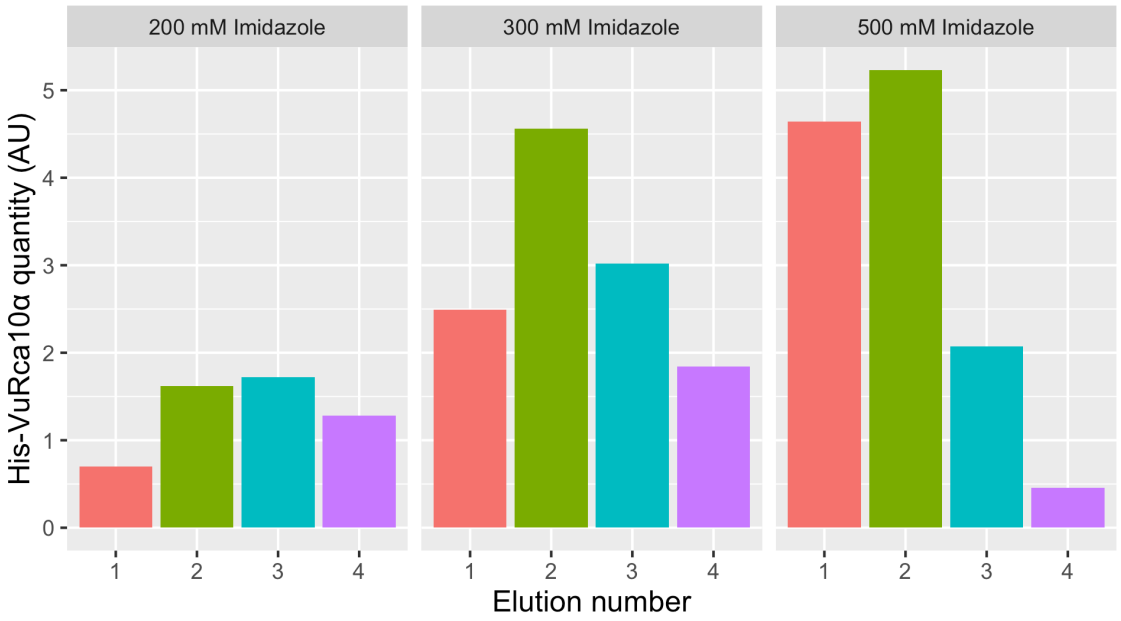


Figure 3.8: *The effect of elution buffer imidazole concentration on quantity of His-Rca eluted from IMAC resin. Clarified lysate expressing His-tagged VuRca10α was incubated with IMAC resin at 4 °C with agitation for 30 minutes before. The lysate was then split into three identical samples which were washed with 125 mM imidazole wash buffer before elution at varying concentrations of imidazole. Elutions were of 1 mL volume. Amount of His-Rca was estimated by band intensity on a western blot probed with anti-His antibody.*

An increase in the concentration of imidazole in elution buffer effects an increase in the amount of protein eluted from the column (Fig. 3.8). 200 mM was insufficient to elute most of the His-Rca bound to a column. 500 mM imidazole was sufficient to elute most if not all bound protein in a tight peak. However, it is not beneficial to expose proteins to high concentrations of imidazole as this can affect function and stability. 300 mM was chosen as the concentration to use for elution buffer, which seems to elute almost all protein, albeit in a wider peak. If necessary, the purified protein can be pooled and concentrated which can solve the issue of insufficient concentration in some eluates. These experiments decided the concentration of imidazole in buffers throughout the purification process - 50 mM in binding buffer, 125 mM in wash buffer, and 300 mM in elution buffer.

Sonication Parameters Can Be Optimised to Improve Temperature Control and Rca Solubility

Prior to protein purification, cells must be lysed to expose intracellular contents. Lysis by sonication is quick and effective but causes temperature rise which may be damaging to thermolabile proteins. To find a set of lysis conditions that does not excessively heat sample, and retains soluble protein, a set of optimisation experiments were conducted (Figure 3.9). Different cooling options (water & ice-water) were tested for their effects on temperature control, and time of sonication and lysis

buffer volume were tested for their effects on Rca solubility. The heat produced during sonication can be damaging to proteins, especially to heat-sensitive enzymes such as Rca. Ice-water controls temperatures much better than ice (Fig. 3.9A), acting as a simple and serviceable replacement for more complex cooling solutions such as a cooling jacket. The volume of buffer the cell pellet is resuspended in was also found to have an important effect on the amount of soluble protein released during sonication (Fig. 3.9C). A ratio of 2 mL buffer per g of fresh biomass meant only around 5% of protein was soluble, this improved to almost a 1:1 ratio by doubling the amount of buffer per g of fresh biomass. More protein was not found to be released by increasing the time of sonication past 5 minutes (Fig. 3.9B). It can be assumed at this point that almost all if not all of the cellular protein content has been released.

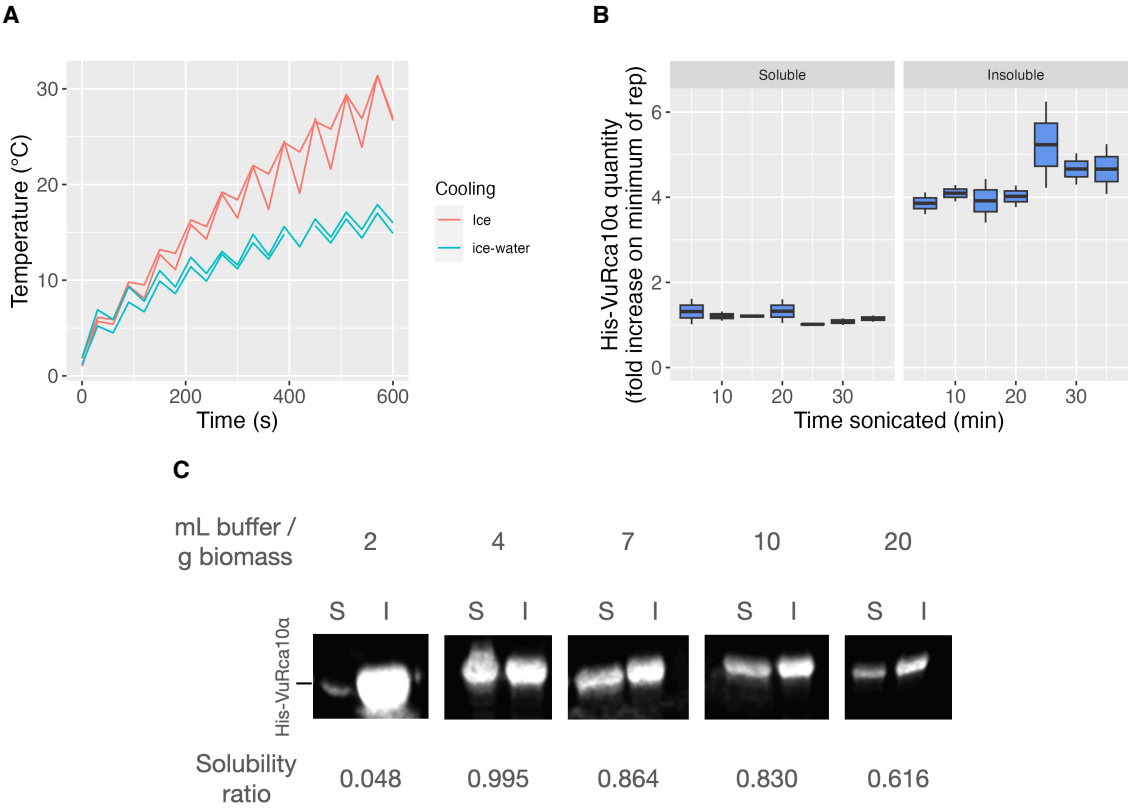


Figure 3.9: *Optimisation of cell lysis via sonication.* ***A*** *Comparison of change of temperature during sonication with ice vs ice-water as a cooling solution. Sonication was on for 30 s and off for 30 s for 10 cycles at 50% power.* ***B*** *The effect of increased time of sonication on His-Rca release and solubility. Every 5 minutes of lysis (10 minutes including cooling), a sample of lysate was taken for solubility analysis.* ***C*** *The effect of volume of binding buffer on the solubility of His-Rca after sonication. Fresh biomass was resuspended in various volumes of the same buffer. Amount of His-Rca was estimated by band intensity on a western blot probed with anti-His antibody. S: Soluble; I: Insoluble.*

From this optimisation study, a 5 minute sonication time, 6 mL of buffer per gram of biomass and an ice-water bath cooling solution was chosen for further cell lysis work. While 4 mL of buffer

per g of biomass seems sufficient for improved solubility, it is not known if this is adequate in other proteins so a larger volume was chosen.

3.3.3 Rca Activity Validation & Quality Control

Quality of pure protein samples was assessed by SDS-PAGE analysis and the change in OD_{280} during purification. Figure 3.10 shows examples of these quality control measures.

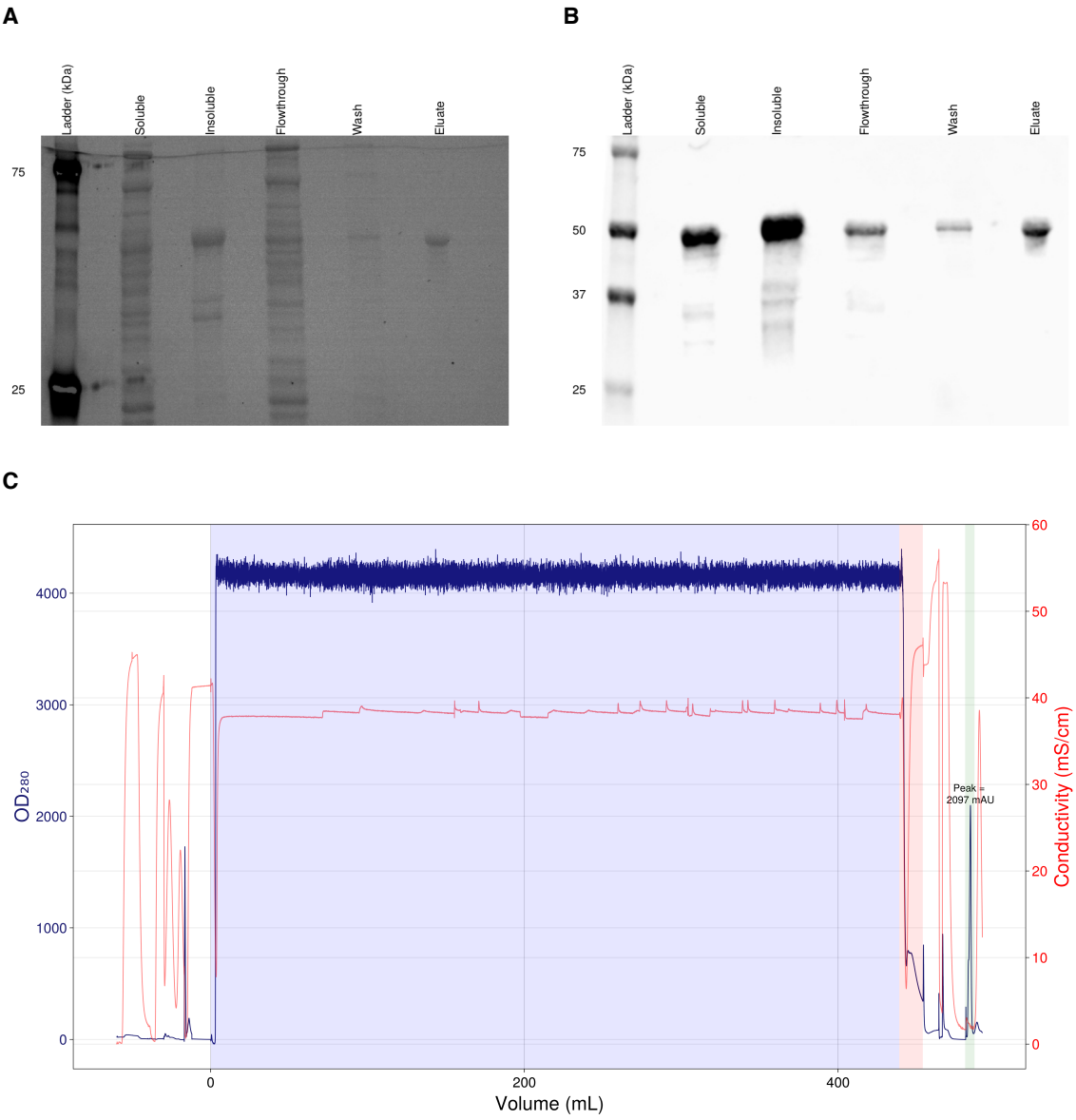


Figure 3.10: *Quality control of purified proteins.* ***A*** *SDS-PAGE gel of VuRca10a purification product and byproducts.* ***B*** *The gel in* ***A*** *was used for immunoblotting with anti-His-tag antibody.* ***C*** *Chromatogram showing OD_{280} and conductivity through the course of purification. Vertical shading indicates fraction collection: blue: flowthrough; red: wash; green: elution.*

Figure 3.10A shows only one clearly defined band in the eluate lane, indicating a high purity of final product. The sample is not entirely pure and free of contaminants as some higher

molecular weight species can be seen. In Figure 3.10B, a smear can be seen beneath the His-VuRca10α band, indicating possible degradation of the target. The slight asymmetry in the peak in Figure 3.10C is indicative of small amounts of impurities in the final product. Impurities tend to manifest as a ”shoulder” or ”slope” to the elution fraction.

The quality control measures shown can help assess the quality of the final protein, but do not assess enzymatic activity. The process of purification can be damaging to proteins and reduce or eliminate enzymatic activity. To confirm that Rca remains active after purification, the ATPase activity of the purified isoforms was measured (Fig. 3.11).

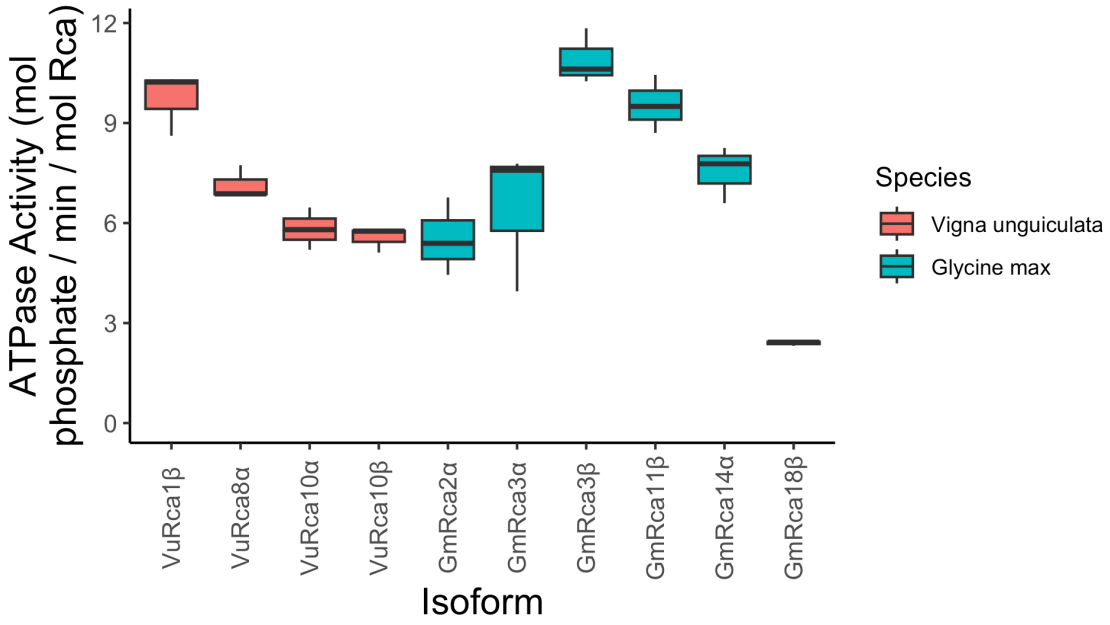


Figure 3.11: *Validation of Rca ATPase activity. Rca was expressed and purified according to the optimised protocol described herein - expression in* E. coli *grown in Dynamite medium at 24 °C after 4 h induction. Rate of ATP hydrolysis by His-tagged Rca isoforms of* Vigna unguiculata *and* Glycine max*. Box plots are mean and standard error of n = 3 technical replicates. Assays were carried out at 30 °C for 5 minutes and terminated by addition of 12% SDS.*

As expected from previous studies, there is variability in the levels of ATPase activity between different isoforms of Rca from cowpea and soybean. Every isoform appears to be active with the exception of GmRca18β which shows very weak activity if any. Note that the scaling of units makes small differences in absorption appear larger, and that raw blank-corrected absorption values are only slightly above 0.

Activity of Rca in this study, and future studies, was measured without cleavage of the His tag. This makes purification simpler and quicker, and reduces the time that the Rca must be unfrozen, when it is liable to encounter issues with degradation and precipitation. An early experiment (Figure 3.12) showed no significant difference between the ATPase activity with and without cleavage of the His-tag. Note that this experiment was conducted using Rca from *Triticum aestivum* as this was readily available at this time of experimentation.

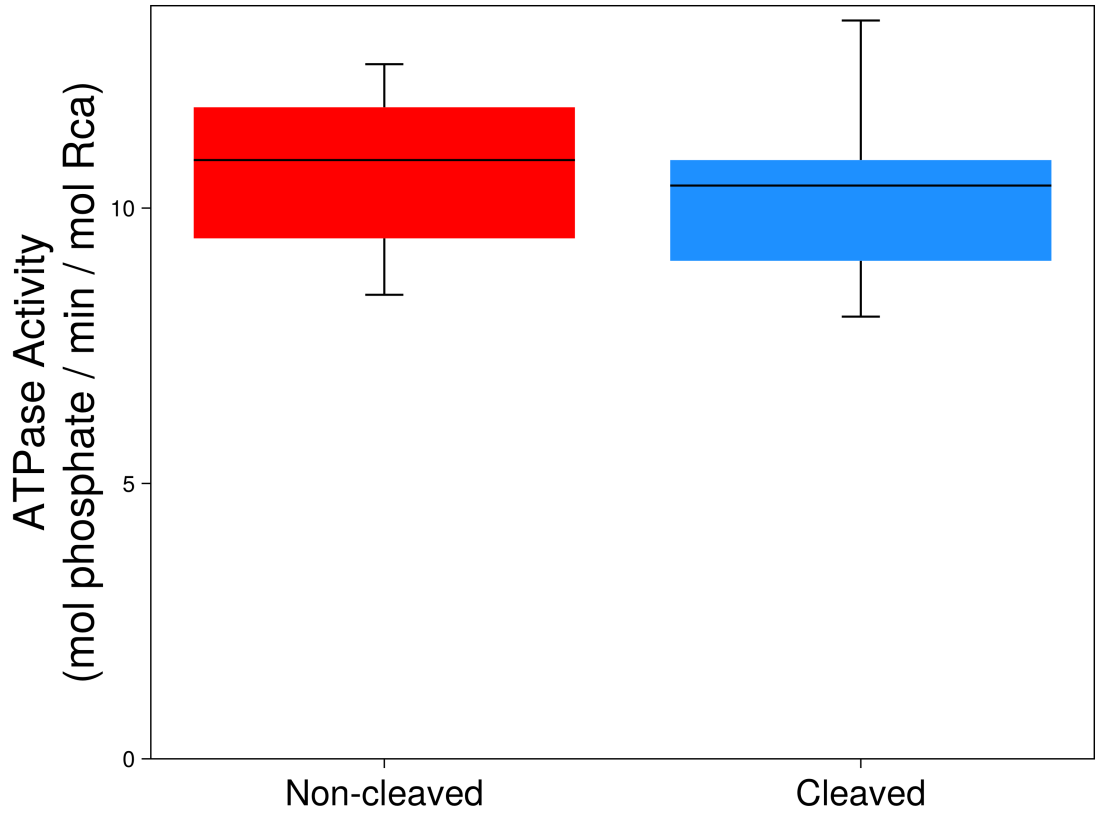


Figure 3.12: *Effect of tag cleavage on Rca ATPase activity. Rate of ATP hydrolysis by* Triticum aestivum *Rca 1β which has been left with the His-tag used for purification, or which had this tag cleaved by TEV protease. Box plots are mean and standard error of n = 3 technical replicates. Assays were carried out at 25 °C for 5 minutes and terminated by addition of 12% SDS.*

3.4 Discussion

Described here are the methods used to optimise expression and purification of His-Rca expressed in *E. coli*. Expression has been improved many-fold compared to methods employed prior to optimisation, by a change in medium and expression time. The purity of the final protein was increased by improvement of stringency of binding and wash buffers, without a drastic loss of yield. In spite of the improvements to methods, VuRca10α & VuRca10β remain difficult to purify, while VuRca1β and VuRca8α can be easily obtained in milligram scale. All isoforms that were tested appeared to be active except for GmRca18β. The lack of ATPase may be a consequence of the method of purification, or activity of GmRca18β may be naturally low or may require changing conditions such as temperature or concentration.

The optimisation work was carried out focusing on one isoform of Rca (the α isoform on chromosome 10 of *Vigna unguiculata*). Although successful for 9/10 attempted isoforms, the factors identified as optimal for VuRca10α may not be optimal for other isoforms, even a single amino acid change can affect expression of recombinant proteins so more optimisation work may be needed

to optimally obtain other isoforms. Further to this, the parameters investigated here are not an exhaustive list of factors known to affect recombinant protein expression. Other factors include:

- Inducer concentration (IPTG)
- Culture:flask volume ratio
- Flask architecture
- Expression time
- Rate of agitation
- Growth temperature

The one-factor-at-a-time optimisation methods used here become increasingly time consuming and laborious as increasingly more factors are included, and cannot account for interactions between these factors, which are known to exist, so truly optimal values can be missed (Czitrom, 1999). A more thorough and robust analysis could use response surface methodology such as Box-Behnken design which has been used in other optimisation experiments to great effect (Díaz-Dinamarca et al., 2018; Giordano et al., 2010; Gutiérrez-González et al., 2019). This methodology changes multiple factors in a single experiment and can interrogate the interactions between factors to find the optimal values of all factors.

Further optimisation work may also look to change expression vector. The choice of expression vector is an important consideration in protein expression. Bacterial, yeast, mammalian, algal, insect, and cell-free expression systems are all viable methods of recombinant protein productions with their own advantages and disadvantages. Table 3.3 provides an overview of some common expression vectors and their use cases.

Table 3.3: *Comparisons of different systems for expression of recombinant proteins. PTM: post-translational modification.*

Expression vector	Advantages	Disadvantages	Citation
Bacterial	Cost effective, scalable	No PTMs, solubility, disulfide bonds can form incorrectly	Rosano and Ceccarelli, 2014
Yeast	Eukaryotic PTMs, scalable, correct disulfide bond formation, high cell densities	Plasmid instability, low protein yields	Gündüz Ergün et al., 2019
Insect	Near mammalian protein processing	Complex culture requirements, glycosylation differs from mammalian	Yee et al., 2018
Mammalian	Advanced posttranslational processing	Complex medium requirements, risk of pathogen infection	O'Flaherty et al., 2020
Cell-free	Can withstand toxic compounds	Cost effectiveness, production of cell-free expression system can be challenging	Khambhati et al., 2019

Escherichia coli was used in this study as an expression vector for heterologous expression of Rca. *E. coli* has multiple advantages as an expression vector such as low cost, ease of cloning

and scalability. *E. coli* can grow well in large volumes to high biomass without need for specialised equipment. Despite its advantages, *E. coli* is not ideal for every application. One noteworthy drawback for production of recombinant Rca is that *E. coli* will not express Rca with native post-translational modifications (PTMs), as bacteria lack the same PTM machinery as is in plants (Macek et al., 2019). Rca has been shown to have functionally relevant PTMs, such as phosphorylation of Thr78 or Ser172 in Arabidopsis (Boex-Fontvieille et al., 2013). The specific strain of *E. coli* used in this study is BL21 (DE3). This strain is commonly used for heterologous protein expression as it enables induction of protein expression upon addition of IPTG through the T7 expression system. Furthermore, this strain lacks lon protease and OmpT which reduce expression of heterologous proteins. To make use of the IPTG-inducible T7 RNA polymerase of the BL21 (DE3) cell lines, a pET expression vector was chosen, which contains a T7 promoter sequence. This promoter controls expression of the desired isoform of Rca, fused N-terminally to a 6xHistidine tag and a cleavage site for TEV protease. The His tag is chosen to enable rapid and high purity purification by use of IMAC. The cleavage site enables TEV protease to restore the N terminus to near-native (leaving a 2 non-native amino acid overhand) in a highly site specific manner. It is often useful to be able to restore the N terminus to a native, or near-native, state as this can impede enzyme activity. No ATPase activity difference was seen between cleaved and non-cleaved Rca, however any potential effect on Rubisco reactivation was not tested due to logistical constraints. Furthermore, this experiment was conducted using wheat Rca 1β, as this protein was readily available. A more robust analysis would test all isoforms of cowpea Rca for both activities. Other common fusion constructs include ubiquitin, which is known to increase expression and can be easily removed by a deubiquitylating enzyme (Baker, 1996), and maltose-binding protein (MBP), which can increase solubility of the protein of interest (Kapust and Waugh, 1999). The downsides of these fusions is that the protein can become insoluble after MBP cleavage, and the addition of a fusion partner can interfere with cleavage. While MBP works to enhance the solubility of many fusion partners, it does not work for every protein and there is no known tag which will enable purification and solubility of any given protein.

Recombinant protein expression is a field that is notoriously difficult to obtain repeatable results in. The reasons for this can be challenging to identify, and often lie in factors that are infrequently considered. Many expression protocols (including the one described herein) use a saturated overnight starter culture to initiate growth in the full scale culture or in creation of a glycerol stock. A saturated culture may cause plasmid instability with unforeseen repercussions on the quality of subsequent cultures (Sivashanmugam et al., 2009). This serves as an example of how simple and routine procedures can cause issues with reproducibility in the field. Furthermore, the methods used to analyse and quantify effects have notably high variance. The OD_{600} method of cell density diverges from linearity at values as low as 0.25 (Studier, 2018). The Bradford method of soluble protein determination relies upon dye binding to basic amino acids and thus shows large protein-to-protein variability, and can show quite low repeatability and sensitivity (Brady and Macnaughtan, 2015; Zor and Selinger, 1996). Western blotting to quantify protein amounts, while

simple and cost effective, is hamstrung by a narrow linear range, differences in transfer efficiency per protein, blot to blot variability, and saturation at quite low protein concentrations (Pillai-Kastoori et al., 2020b). This variability, and non-standardised units, make it challenging to compare results between experiments. For example, Figures 3.4 & 3.5 having notably different quantification values with a similar experimental setup. Quantification from SDS-PAGE gels likewise suffers from similar issues, such as narrow linear range and dye binding differing on a protein-to-protein basis (Alonso Villela et al., 2020).

VuRca1β and VuRca8α had high expression and solubility and were thus easily purified. VuRca10α and VuRca10β were challenging to purify. Despite high expression levels, they were largely located in the insoluble fraction and thus had low yield. Furthermore, issues arose with precipitation either immediately after purification, or during spin concentration. Solubilising additives, solubility-enhancing tags, or a separate optimisation experiment may aid in the solubility of these isoforms. Unfortunately, the field is far from a catch-all solution to protein purification, and there can be little help in purifying some proteins.

To the author's knowledge this is the first study on optimising expression and purification parameters in Rca. While the parameters identified here will not be optimal for all isoforms of Rca, it provides a useful starting point in further experiments requiring recombinant Rca, as well as identifying key areas that can be used to improve expression.

3.5 Acknowledgements

Rubisco activase expression and purification optimisation was done in collaboration with Dr. Peter Gould. Primary data on the effect of His-tag cleavage was produced by Dr. Peter Gould. Plasmids containing codon-optimised His-Rca were designed and provided by Dr. Mike Page.

Chapter 3 Bibliography

Alonso Villela, S. M., Kraïem, H., Bouhaouala-Zahar, B., Bideaux, C., Aceves Lara, C. A., & Fillaudeau, L. (2020). A protocol for recombinant protein quantification by densitometry. *MicrobiologyOpen*, *9*(6), 1175–1182. https://doi.org/10.1002/mbo3.1027

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Structure, Function, and Bioinformatics*, *81*(11), 1857–1861. https://doi.org/10.1002/prot.24364

Baker, R. T. (1996). Protein expression using ubiquitin fusion and cleavage. *Current Opinion in Biotechnology*, *7*(5), 541–546. https://doi.org/10.1016/s0958-1669(96)80059-0

Barta, C., Carmo-Silva, A. E., & Salvucci, M. E. (2011a). Purification of Rubisco Activase from Leaves or after Expression in Escherichia coli. In R. Carpentier (Ed.), *Photosynthesis Research Protocols* (pp. 363–374). Humana Press. https://doi.org/10.1007/978-1-60761-925-3_28

Beal, J., Farny, N. G., Haddock-Angelli, T., Selvarajah, V., Baldwin, G. S., Buckley-Taylor, R., Gershater, M., Marken, J., Sanchania, V., Sison, A., Workman, C. T., Pehlivan, M., Roige, B. B., Aarnio, T., Kivisto, S., Koski, J., Lehtonen, L., Pezzutto, D., Rautanen, P., Bian, W., Hu, Z., Liu, Z., Liu, Z., Ma, L., Pan, L., Qin, Z., Wang, H., Wang, X., Xu, H., Xu, X., El Moubayed, Y., Dong, S., Fang, C., He, H., He, H., Huang, F., Shi, R., Tang, C., Tang, C., Xu, S., Yan, C., Bartzoka, N., Kanata, E., Kapsokefalou, M., Katopodi, X.-L., Kostadima, E., Kostopoulos, I. V., Kotzastratis, S., Koutelidakis, A. E., Krokos, V., Litsa, M., Ntekas, I., Spatharas, P., Tsitsilonis, O. E., Zerva, A., Annem, V., Cone, E., Elias, N., Gupta, S., Lam, K., Tutuianu, A., Mishler, D. M., Toro, B., Akinfenwa, A., Burns, F., Herbert, H., Jones, M., Laun, S., Morrison, S., Smith, Z., Peng, Z., Ziwei, Z., Deng, R., Huang, Y., Li, T., Ma, Y., Shen, Z., Wang, C., Wang, Y., Zhao, T., Lang, Y., Liang, Y., Wang, X., Wu, Y., Aizik, D., Angel, S., Farhi, E., Keidar, N., Oser, E., Pasi, M., Kalinowski, J., Otto, M., Ruhnau, J., Cubukcu, H., Hoskan, M. A., Senyuz, I., Chi, J., Sauter, A. P., Simona, M. F., Byun, S., Cho, S., Kim, G., Lee, Y., Lim, S., Yang, H., Xin, T., Yaxi, Z., Zhao, P., Han, W., He, F., He, Y., Li, N., Luo, X., Boxuan, C., Jiaqi, H., Liangjian, Y., Wanji, L., Xinguang, C., Xinyu, L., Wu, Z., Xi, Y., Yang, X., Yang, Y., Yang, Z., Zhang, Y., Zhou, Y., Peng, Y., Yadi, L., Yang, S., Yuanxu, J., Zhang, K., Abraham, D., Heger, T., Leach, C., Lorch, K., Luo, L., Gaudi, A., Ho, A., Huang, M., Kim, C., Kugathasan, L., Lam, K., Pan, C., Qi, A., Yan, C., Schaaf, K., Sillner, C., Coates, R., Elliott, H., Heath, E., McShane, E., Parry, G., Tariq, A., Thomas, S., Chen, C.-W., Cheng, Y.-H., Hsu, C.-W., Liao, C.-H., Liu, W.-T., Tang, Y.-C., Tang, Y.-H., Yang, Z. E., Jian, L., Li, C., Lin, C., Ran, G., Run, Z., Ting, W., Yong, Z., Yu, L., Lind, A. C., Norberg, A., Olmin, A., Sjolin, J., Torell, A., Trivellin, C., Zorrilla, F., de Vries, P. G., Cheng, H., Peng, J., Xiong, Z., Altarawneh, D., Amir, S. S., Hassan, S., Vincent, A., Costa, B., Gallegos, I., Hale, M., Sonnier, M., Whalen, K., Elikan, M., Kim, S., You, J., Rambhatla, R., Viswanathan, A., Tian, H., Xu, H., Zhang, W., Zhou, S., Jiamiao, L., Jiaqi, X., Craw, D., Goetz, M., Rettedal, N., Yarbrough, H., Ahlgren, C., Guadagnino, B., Guenther, J., Huynh, J., He, Z., Liu, H., Liu, Y., Qu, M., Song, L., Yang, C., Yang, J., Yin, X., Zhang, Y., Zhou, J., Zi, L., Jinyu, Z., Kang, X., Xilei, P., Xue, H., Xun, S., Babu, P., Dogra, A., Thokachichu, P., Faurdal, D., Jensen, J. H., Mejlsted, J., Nielsen, L., Rasmussen, T., Denter, J., Husnatter, K., Longo, Y., Luzuriaga, J. C., Moncayo, E., Moreira, N. T., Tapia, J., Dingyue, T., Jingjing, Z., Wenhao, X., Xinyu, T., Xiujing, H., DeKloe, J., Astles, B., Baronaite, U., & Grazulyte, I. (2020). Robust estimation of bacterial cell count from optical density. *Communications Biology*, *3*(1), 512. https://doi.org/10.1038/s42003-020-01127-5

Bertani, G. (1951). STUDIES ON LYSOGENESIS I: The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *Journal of Bacteriology*, *62*(3), 293–300. https://doi.org/10.1128/jb.62.3.293-300.1951

Bhat, E., Abdalla, M., & Rather, I. (2018). Key factors for successful protein purification and crystallization. *Global Journal of Biotechnology and Biomaterial Science*, *4*(1), 001–007. https://doi.org/10.17352/gjbbs.000010

Bhatnagar, B. S., Bogner, R. H., & Pikal, M. J. (2007). Protein stability during freezing: Separation of stresses and mechanisms of protein stabilization. *Pharmaceutical Development and Technology*, *12*(5), 505–523. https://doi.org/10.1080/10837450701481157

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). In *Methods in Enzymology* (pp. 439–473). Elsevier. https://doi.org/10.1016/S0076-6879(09)63027-5

Boex-Fontvieille, E., Daventure, M., Jossier, M., Hodges, M., Zivy, M., & Tcherkez, G. (2013). Phosphorylation pattern of rubisco activase in Arabidopsis leaves (W. Adams, Ed.). *Plant Biology*, *16*(3), 550–557. https://doi.org/10.1111/plb.12100

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of proteins using polyhistidine affinity tags. Elsevier. https://doi.org/10.1016/s0076-6879(00)26058-8

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2), 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Brady, P. N., & Macnaughtan, M. A. (2015). Evaluation of colorimetric assays for analyzing reductively methylated proteins: Biases and mechanistic insights. *Analytical Biochemistry*, *491*, 43–51. https://doi.org/10.1016/j.ab.2015.08.027

Cao, E., Chen, Y., Cui, Z., & Foster, P. R. (2003). Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnology and Bioengineering*, *82*(6), 684–690. https://doi.org/10.1002/bit.10612

Chifflet, S., Torriglia, A., Chiesa, R., & Tolosa, S. (1988). A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: Application to lens ATPases. *Analytical Biochemistry*, *168*(1), 1–4. https://doi.org/10.1016/0003-2697(88)90002-4

Chisti, Y., & Moo-Young, M. (1986). Disruption of microbial cells for intracellular products. *Enzyme and Microbial Technology*, *8*(4), 194–204. https://doi.org/10.1016/0141-0229(86)90087-6

Collins, T., Azevedo-Silva, J., da Costa, A., Branca, F., Machado, R., & Casal, M. (2013). Batch production of a silk-elastin-like protein in *E. coli* BL21(DE3): Key parameters for optimisation. *Microbial Cell Factories*, *12*(1), 21. https://doi.org/10.1186/1475-2859-12-21

Czitrom, V. (1999). One-factor-at-a-time versus designed experiments. *The American Statistician*, *53*(2), 126. https://doi.org/10.2307/2685731

Danilevich, V. N., Petrovskaya, L. E., & Grishin, E. V. (2008). A highly efficient procedure for the extraction of soluble proteins from bacterial cells with mild chaotropic solutions. *Chemical Engineering & Technology*, *31*(6), 904–910. https://doi.org/10.1002/ceat.200800024

Desmarais, W. T., Bienvenue, D. L., Bzymek, K. P., Holz, R. C., Petsko, G. A., & Ringe, D. (2002). The 1.20 å resolution crystal structure of the aminopeptidase from aeromonas proteolytica complexed with tris. *Structure (London, England : 1993)*, *10*(8), 1063–1072. https://doi.org/10.1016/s0969-2126(02)00810-9

Díaz-Dinamarca, D. A., Jerias, J. I., Soto, D. A., Soto, J. A., Díaz, N. V., Leyton, Y. Y., Villegas, R. A., Kalergis, A. M., & Vásquez, A. E. (2018). The optimisation of the expression of recombinant surface immunogenic protein of group b streptococcus in *escherichia coli* by response surface methodology improves humoral immunity. *Molecular Biotechnology*, *60*(3), 215–225. https://doi.org/10.1007/s12033-018-0065-8

Elsner, H. I., & Lindblad, E. B. (1989). Ultrasonic degradation of DNA. *DNA (Mary Ann Liebert, Inc.)*, *8*(10), 697–701. https://doi.org/10.1089/dna.1989.8.697

Ernst, O., & Zor, T. (2010). Linearization of the bradford protein assay. *Journal of Visualized Experiments*, (38). https://doi.org/10.3791/1918

Feliu, J. X., Cubarsi, R., & Villaverde, A. (1998). Optimized release of recombinant proteins by ultrasonication of *E. coli* cells. *Biotechnology and Bioengineering*, *58*(5), 536–540. https://doi.org/10.1002/(sici)1097-0290(19980605)58:5{$<$}536::aid-bit10{$>$}3.0.co;2-9

Gatti-Lafranconi, P., Natalello, A., Ami, D., Doglia, S. M., & Lotti, M. (2011). Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology. *FEBS Journal*, *278*(14), 2408–2418. https://doi.org/10.1111/j.1742-4658.2011.08163.x

Geciova, J., Bury, D., & Jelen, P. (2002). Methods for disruption of microbial cells for potential use in the dairy industry-a review. *International Dairy Journal*, *12*(6), 541–553. https://doi.org/10.1016/s0958-6946(02)00038-9

Giordano, P. C., Martínez, H. D., Iglesias, A. A., Beccaria, A. J., & Goicoechea, H. C. (2010). Application of response surface methodology and artificial neural networks for optimization of recombinant *Oryza sativa* non-symbiotic hemoglobin 1 production by *Escherichia coli* in medium containing byproduct glycerol. *Bioresource Technology*, *101*(19), 7537–7544. https://doi.org/10.1016/j.biortech.2010.05.001

Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., & Lian, L.-Y. (2004). A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society*, *126*(29), 8933–8939. https://doi.org/10.1021/ja049297h

Gündüz Ergün, B., Hüccetoğulları, D., Öztürk, S., Çelik, E., & Çalık, P. (2019). Established and Upcoming Yeast Expression Systems. In B. Gasser & D. Mattanovich (Eds.), *Recombinant Protein Production in Yeast* (pp. 1–74). Springer New York. https://doi.org/10.1007/978-1-4939-9024-5_1

Gutiérrez-González, M., Farías, C., Tello, S., Pérez-Etcheverry, D., Romero, A., Zúñiga, R., Ribeiro, C. H., Lorenzo-Ferreiro, C., & Molina, M. C. (2019). Optimization of culture conditions for the expression of three different insoluble proteins in *Escherichia coli*. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-53200-7

Hamada, H., Arakawa, T., & Shiraki, K. (2009). Effect of additives on protein aggregation. *Current Pharmaceutical Biotechnology*, *10*(4), 400–407. https://doi.org/10.2174/138920109788488941

Hamilton, S., Odili, J., Pacifico, M. D., Wilson, G. D., & Kupsch, J.-M. (2003). Effect of imidazole on the solubility of a his-tagged antibody fragment. *Hybridoma and Hybridomics*, *22*(6), 347–355. https://doi.org/10.1089/153685903771797048

Hasse, D., Larsson, A. M., & Andersson, I. (2015). Structure of *Arabidopsis thaliana* Rubisco activase. *Acta Crystallographica Section D Biological Crystallography*, *71*(4), 800–808. https://doi.org/10.1107/S1399004715001182

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*(3), 346–363.

Hou, Q., Bourgeas, R., Pucci, F., & Rooman, M. (2018). Computational analysis of the amino acid interactions that promote or decrease protein solubility. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-32988-w

Izard, J., Parker, M. W., Chartier, M., Ducheé, D., & Baty, D. (1994). A single amino acid substitution can restore the solubility of aggregated colicin A mutants in *Escherichia coli*. *"Protein Engineering, Design and Selection"*, *7*(12), 1495–1500. https://doi.org/10.1093/protein/7.12.1495

James, C. J., Coakley, W. T., & Hughes, D. E. (1972). Kinetics of protein release from yeast sonicated in batch and flow systems at 20 kHz. *Biotechnology and Bioengineering*, *14*(1), 33–42. https://doi.org/10.1002/bit.260140105

Jiang, S., & Nail, S. L. (1998). Effect of process conditions on recovery of protein activity after freezing and freeze-drying. *European Journal of Pharmaceutics and Biopharmaceutics*, *45*(3), 249–257. https://doi.org/10.1016/S0939-6411(98)00007-1

Johnson, B. H., & Hecht, M. H. (1994). Recombinant proteins can be isolated from *e. coli* cells by repeated cycles of freezing and thawing. *Nature Biotechnology*, *12*(12), 1357–1360. https://doi.org/10.1038/nbt1294-1357

Kapust, R. B., & Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science*, *8*(8), 1668–1674. https://doi.org/10.1110/ps.8.8.1668

Kazan, D., Çamurdan, A., & Hortaçsu, A. (1995). The effect of glucose concentration on the growth rate and some intracellular components of a recombinant E. coli culture. *Process Biochemistry*, *30*(3), 269–273. https://doi.org/10.1016/0032-9592(95)85008-2

Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., & Singh, V. (2019). Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Frontiers in Bioengineering and Biotechnology*, *7*, 248. https://doi.org/10.3389/fbioe.2019.00248

Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., & Scholtz, J. M. (2012). Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility. *Biophysical Journal*, *102*(8), 1907–1915. https://doi.org/10.1016/j.bpj.2012.01.060

Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K., & Van Voorhis, W. C. (2012). Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins (P. D. Riggs, Ed.). *PLoS ONE*, *7*(12), e52482. https://doi.org/10.1371/journal.pone.0052482

Lilley, R. M., & Portis Jr, A. R. (1997). ATP Hydrolysis Activity and Polymerization State of Ribulose-1,5-Bisphosphate Carboxylase Oxygenase Activase (Do the Effects of Mg2+, K+, and Activase Concentrations Indicate a Functional Similarity to Actin?) *Plant Physiology*, *114*(2), 605–613. https://doi.org/10.1104/pp.114.2.605

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., & Mijakovic, I. (2019). Protein post-translational modifications in bacteria. *Nature Reviews Microbiology*, *17*(11), 651–664. https://doi.org/10.1038/s41579-019-0243-0

McGuffie, M. J., & Barrick, J. E. (2021). pLannotate: Engineered plasmid annotation. *Nucleic Acids Research*, *49*(W1), W516–W522. https://doi.org/10.1093/nar/gkab374

Mezzasalma, T. M., Kranz, J. K., Chan, W., Struble, G. T., Schalk-Hihi, C., Deckman, I. C., Springer, B. A., & Todd, M. J. (2007). Enhancing recombinant protein quality and yield by protein stability profiling. *Journal of Biomolecular Screening*, *12*(3), 418–428. https://doi.org/10.1177/1087057106297984

Mühlmann, M., Forsten, E., Noack, S., & Büchs, J. (2017). Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microbial Cell Factories*, *16*(1). https://doi.org/10.1186/s12934-017-0832-4

O'Flaherty, R., Bergin, A., Flampouri, E., Mota, L. M., Obaidi, I., Quigley, A., Xie, Y., & Butler, M. (2020). Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnology Advances*, *43*, 107552. https://doi.org/10.1016/j.biotechadv.2020.107552

Park, H., Park, J.-Y., Park, K.-M., & Chang, P.-S. (2021). Effects of freezing rate on structural changes in l-lactate dehydrogenase during the freezing process. *Scientific Reports*, *11*(1), 13643. https://doi.org/10.1038/s41598-021-93127-6

Pchelintsev, N. A., Adams, P. D., & Nelson, D. M. (2016). Critical parameters for efficient sonication and improved chromatin immunoprecipitation of high molecular weight proteins (M. Wu, Ed.). *PLOS ONE*, *11*(1), e0148023. https://doi.org/10.1371/journal.pone.0148023

Peach, M., Marsh, N., Miskiewicz, E. I., & MacPhee, D. J. (2015). Solubilization of proteins: The importance of lysis buffer choice. Springer New York. https://doi.org/10.1007/978-1-4939-2694-7_8

Pedersen, T. L. (2022). *Patchwork: The Composer of Plots*.

Pei, S., Swayne, T. C., Morris, J. F., & Emtage, L. (2021). Threshold concentration and random collision determine the growth of the huntingtin inclusion from a stable core. *Communications Biology*, *4*(1). https://doi.org/10.1038/s42003-021-02460-z

Pelegrine, D., & Gasparetto, C. (2005). Whey proteins solubility as function of temperature and pH. *LWT - Food Science and Technology*, *38*(1), 77–80. https://doi.org/10.1016/j.lwt.2004.03.013

Perdomo, J. A., Sales, C. R. G., & Carmo-Silva, E. (2018). Quantification of Photosynthetic Enzymes in Leaf Extracts by Immunoblotting. In S. Covshoff (Ed.), *Photosynthesis* (pp. 215–227). Springer New York. https://doi.org/10.1007/978-1-4939-7786-4_12

Peterson-Forbrook, D. S., Hilton, M. T., Tichacek, L., Henderson, J. N., Bui, H. Q., & Wachter, R. M. (2017). Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry*, *56*(36), 4906–4921. https://doi.org/10.1021/acs.biochem.7b00574

Pillai-Kastoori, L., Schutz-Geschwender, A. R., & Harford, J. A. (2020b). A systematic approach to quantitative Western blot analysis. *Analytical Biochemistry*, *593*, 113608. https://doi.org/10.1016/j.ab.2020.113608

R Core Team. (2013). *R: A language and environment for statistical computing* (tech. rep.). Vienna, Austria.

Ramón, A., Señorale-Pose, M., & Marín, M. (2014). Inclusion bodies: Not that bad… *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00056

Robichon, C., Luo, J., Causey, T. B., Benner, J. S., & Samuelson, J. C. (2011). Engineering *Escherichia coli* BL21(DE3) Derivative Strains To Minimize *E. coli* Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. *Applied and Environmental Microbiology*, *77*(13), 4634–4646. https://doi.org/10.1128/AEM.00119-11

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00172

RStudio Team. (2020). *RStudio: Integrated development environment for R* (tech. rep.). Boston, MA.

Scafaro, A., Gallé, A., Rie, J. V., Carmo-Silva, E., Salvucci, M. E., & Atwell, B. J. (2016). Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytologist*, *211*(3), 899–911. https://doi.org/10.1111/nph.13963

Shehadul Islam, M., Aryasomayajula, A., & Selvaganapathy, P. (2017). A review on macroscale and microscale cell lysis methods. *Micromachines*, *8*(3), 83. https://doi.org/10.3390/mi8030083

Singh, S. M., & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*, *99*(4), 303–310. https://doi.org/10.1263/jbb.99.303

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., & Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science*, *18*(5), 936–948. https://doi.org/10.1002/pro.102

Slouka, C., Kopp, J., Spadiut, O., & Herwig, C. (2018). Perspectives of inclusion bodies for biobased products: Curse or blessing? *Applied Microbiology and Biotechnology*, *103*(3), 1143–1153. https://doi.org/10.1007/s00253-018-9569-1

Stathopulos, P. B., Scholz, G. A., Hwang, Y.-M., Rumfeldt, J. A., Lepock, J. R., & Meiering, E. M. (2008). Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Science*, *13*(11), 3017–3027. https://doi.org/10.1110/ps.04831804

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nature Structural & Molecular Biology*, *18*(12), 1366–1370. https://doi.org/10.1038/nsmb.2171

Strandberg, L., & Enfors, S. O. (1991). Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Applied and Environmental Microbiology*, *57*(6), 1669–1674. https://doi.org/10.1128/aem.57.6.1669-1674.1991

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016

Studier, F. W. (2018). T7 Expression Systems for Inducible Production of Proteins from Cloned Genes in *E. Coli*: T7 expression Systems for Inducible Production of Proteins. *Current Protocols in Molecular Biology*, *124*(1), e63. https://doi.org/10.1002/cpmb.63

Studier, F., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130. https://doi.org/10.1016/0022-2836(86)90385-2

Talley, K., & Alexov, E. (2010). On the pH-optimum of activity and stability of proteins. *Proteins: Structure, Function, and Bioinformatics*, n/a–n/a. https://doi.org/10.1002/prot.22786

Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., Tanaka, M., Nakatsu, C. H., & Kamagata, Y. (2014). A hidden pitfall in the preparation of agar media undermines microorganism cultivability (H. Nojiri, Ed.). *Applied and Environmental Microbiology*, *80*(24), 7659–7666. https://doi.org/10.1128/aem.02741-14

Taylor, T., Denson, J.-P., & Esposito, D. (2017). Optimizing Expression and Solubility of Proteins in *E. coli* Using Modified Media and Induction Parameters. In N. A. Burgess-Brown (Ed.), *Heterologous Gene Expression in E.coli* (pp. 65–82). Springer New York. https://doi.org/10.1007/978-1-4939-6887-9_5

Thomas, J. G., & Baneyx, F. (1996). Protein misfolding and inclusion body formation in recombinant *escherichia coli* cells overexpressing heat-shock proteins. *Journal of Biological Chemistry*, *271*(19), 11141–11147. https://doi.org/10.1074/jbc.271.19.11141

Trevino, S. R., Scholtz, J. M., & Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, glu, and ser contribute more favorably than the other hydrophilic amino acids in RNase sa. *Journal of Molecular Biology*, *366*(2), 449–460. https://doi.org/10.1016/j.jmb.2006.10.026

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wurm, D. J., Quehenberger, J., Mildner, J., Eggenreich, B., Slouka, C., Schwaighofer, A., Wieland, K., Lendl, B., Rajamanickam, V., Herwig, C., & Spadiut, O. (2017). Teaching an old pET new tricks: Tuning of inclusion body formation and properties by a mixed feed system in

E. coli. *Applied Microbiology and Biotechnology*, *102*(2), 667–676. https://doi.org/10.1007/s00253-017-8641-6

Yang, Y., Zeng, L., & Vihinen, M. (2021). PON-Sol2: Prediction of effects of variants on protein solubility. *International Journal of Molecular Sciences*, *22*(15), 8027. https://doi.org/10.3390/ijms22158027

Yee, C. M., Zak, A. J., Hill, B. D., & Wen, F. (2018). The Coming Age of Insect Cells for Manufacturing and Development of Protein Therapeutics. *Industrial & Engineering Chemistry Research*, *57*(31), 10061–10070. https://doi.org/10.1021/acs.iecr.8b00985

Zbacnik, T. J., Holcomb, R. E., Katayama, D. S., Murphy, B. M., Payne, R. W., Coccaro, R. C., Evans, G. J., Matsuura, J. E., Henry, C. S., & Manning, M. C. (2017). Role of buffers in protein formulations. *Journal of Pharmaceutical Sciences*, *106*(3), 713–733. https://doi.org/10.1016/j.xphs.2016.11.014

Zor, T., & Selinger, Z. (1996). Linearization of the bradford protein assay increases its sensitivity: Theoretical and experimental studies. *Analytical Biochemistry*, *236*(2), 302–308. https://doi.org/10.1006/abio.1996.0171

Chapter 4

The Temperature Profiles of Cowpea Rubisco Activase

4.1 Introduction

Cowpea (*Vigna unguiculata* (L.) Walp) is a legume crop primarily grown in sub-Saharan Africa. It is grown for its seeds and leaves, for primary human consumption, and as animal feed (OECD, 2019). Nigeria is the world's largest producer and consumer of cowpea (Boukar et al., 2015). Nigeria is home to a broad range of climates, with a tropical climate in the south near the coast, and a more arid climate in the northeast near the Sahara (Beck et al., 2018). The climates of Nigeria are changing - with higher temperatures, more frequent extreme weather events, and water supply becoming destabilised (Adenodi, 2018; Olaniyi et al., 2019). The changing climate of Nigeria threatens cowpea production, and other agriculture. Safeguarding future cowpea production will require targeted improvements to improve its resistance to abiotic stress (Lucas et al., 2013; Srivastava et al., 2022).

A potential target for cowpea crop improvement under heat stress is Rubisco activase (Rca) (Wijewardene et al., 2021). Rca is an enzyme critical to photosynthesis (see Chapter 1.3 for details). It is required to maintain the catalytic competence of Rubisco (Parry et al., 2007; Portis et al., 1986). Rubisco is a carboxylating enzyme that initiates the Calvin-Benson-Bassham (CBB) cycle by addition of CO_2 to its substrate ribulose-1,5-bisphosphate (RuBP). However, sugar phosphate inhibitors bind to Rubisco, inactivating it. These inhibitors include Ca1P, XuBP, and even RuBP if bound to a non-carbamylated active site (Lilley and Portis, 1990). Rca has two distinct but related activities - Rubisco reactivation and ATP hydrolysis. The energy released by ATP hydrolysis is used to remodel Rubisco and remove inhibitors (Portis, 1990). ATP hydrolysis can still occur without reactivation of Rubisco, even in the absence of Rubisco, so analysis of both activities is required to fully characterise the properties of Rca.

Rca can be broadly divided into two groups - α and β isoforms. α isoforms are longer and contain a redox-sensitive C-terminal extension. β isoforms are shorter, they do not contain this

extension and are thus insensitive to changes in the chloroplast's redox state. Plants often, but not always, have more than one copy of Rca, usually at least a β and an α isoform. These can be encoded on separate genes or produced by alternative splicing (Waheeda et al., 2023).

Rca is a quite thermolabile enzyme, and as such is implicated in photosynthetic limitation under heat stress (Salvucci and Crafts-Brandner, 2004a). There is substantial diversity in the response of Rca to temperature between plants, and between different isoforms in the same plant. In wheat (*Triticum aestivum*), the Rca1β isoform has been demonstrated to have a higher temperature optimum than the other two wheat Rca isoforms. This can be attributed partially to a single amino acid substitution (Degen et al., 2020). Domesticated rice (*Oryza sativa*) has been shown to have much less thermotolerant Rca than *Oryza australiensis* - a wild relative. This confers some degree of heat tolerance to *Oryza australiensis*, allowing for higher net photosynthesis rates at supraoptimal temperatures (Scafaro et al., 2016). *Agave tequilana*, a CAM plant, has also been shown to have Rca with a higher thermal optimum than is typical for C3 plants. The temperature optimum of this Rca is around 10 °C higher than a typical C3 plant (Shivhare and Mueller-Cajar, 2017). Despite this, even in warm adapted plants, Rca is active at considerably lower temperatures than Rubisco is (Crafts-Brandner and Salvucci, 2000; Salvucci et al., 2001), and this can make Rubisco activation rate-limiting for photosynthesis under heat stress. Rca is typically reported to have an optimal concentration, with lower per molar activity above or below this concentration, which varies by species and isoform (Keown et al., 2013; Lilley and Portis Jr, 1997; Scafaro et al., 2016). In light of the known diversity of catalytic properties of Rca, it was hypothesised that cowpea Rca isoforms would show diversity in their response to temperature and concentration.

This chapter focuses on the characterisation of different isoforms of Rca from cowpea. The effect of Rca concentration on ATP hydrolysis rate is investigated for the more highly expressing isoforms. The dual activities of ATP hydrolysis and Rubisco reactivation are analysed across a range of temperatures. Keeping Rca active at a broader range of temperatures may help keep plants actively photosynthesising and growing under changing climates. Research into the diversity of Rca within cowpea may inform future breeding or genetic engineering efforts for improvement of this important crop.

4.2 Materials and Methods

4.2.1 Recombinant Expression and Purification of His-Rca

Pure, recombinant, His-tagged Rca was produced using the optimised protocol developed in Chapter 3.2.

4.2.2 Protein Concentration Determination

Protein concentration was determined using the Bradford method (Bradford, 1976). 0-50 µg of bovine serum albumin was used to construct the standard curve. 1 µL of sample and 200 µL of

Bradford reagent were added in triplicate to wells of a 96-well plate. Protein concentration was estimated using the standard curve comparing protein amount to the absorbance at 595 nm.

4.2.3 ATPase activity

ATPase activity was measured using an adapted version of the method of Chifflet et al. (1988). To initiate reactions, 10 µL of 10 µM Rca stock was added to 40 µL of assay buffer (2 µM final concentration). After 5 minutes, assays are terminated by addition of 50 µL of 12% (w/v) SDS. To prepare blanks, assay mix was prepared identically, but SDS was added to the reaction mix prior to Rca. This ensures that Rca is inactivated upon addition to the assay mix.

Prior to reactions, assay mix was allowed to reach temperature for 10 minutes in a pre-warmed heat block. All reactions at the same temperature were run at the same time for logistical practicality. The order at which temperatures are run, and the order of isoforms at each temperature, are randomised.

For colour development, 100 µL of 3% (w/v) L-ascorbate in 0.5 N HCl and 0.5% (w/v) ammonium molybdate was added to sample tubes, mixed, and left at room temperature for 3 minutes. Room temperature was monitored and averages around 25 °C.

After 20 minutes, 150 µL of development solution (1% (v/v) acetic acid, 4% (w/v) sodium (meta)arsenite, 4% (w/v) sodium citrate) was added, mixed, and left to develop colour for 20 minutes at room temperature. Absorbance of 200 µL of sample was measured at 850 nm. To calculate phosphate release, a calibration curve was prepared using 0-50 mM KPi stocks.

The following number of unique preps were used for analysis of ATPase activity: VuRca1β: 2; VuRca8α: 3; VuRca10α: 3; VuRca10β: 3

4.2.4 Rubisco Reactivation Activity

Rubisco reactivation activity was measured using the methods of Barta et al. (2011b). Rubisco was purified from leaves and carbamylated by incubation with Activation Mix (100 mM Tricine–NaOH, pH 8.0, 10 mM $MgCl_2$, 10 mM $NaHCO_3$, and 50 mM DTT). Carbamylated Rubisco was centrifugated at 400x *g* and 4 °C through a 2 mL Sephadex G-50-fine column which was equilibrated with Rubisco ER Desalting Buffer (50 mM Tricine-NaOH, pH 8.0, and 0.1 mM EDTA, prepared CO_2-free by purging with N_2 prior to adjusting the pH with 60% (w/v) NaOH at 400x *g* and 4 °C). This was then incubated with 0.5 mM RuBP for at least 4 h at 4 °C to form a decarbamylated Rubisco-RuBP complex termed ER. For the assay, ER was incubated with Rca for various length of time (30, 60, 90, 360 s). The Rubisco which was reactivated in this time was then added to Rubisco assay mix containing $^{14}CO_2$, allowing Rubisco to fix carbon for 30 s. Rubisco activity was quenched with 4 N formic acid, and the assay mix was dried at 100 °C. The vial contents were rehydrated in milliQ water, and mixed with 3.6 mL of scintillation cocktail. $^{14}CO_2$ contents were measured using a liquid scintillation counter.

To account for spontaneous reactivation of Rubisco, another assay was run without addition of Rca.

4.2.5 SDS-PAGE & Western Blotting

Rca quality and purity was assessed by SDS-PAGE. The methods are described in Chapter 3.2.3.

4.2.6 Cowpea Rca Sequences

Cowpea Rca sequences were identified in the genome of the `IT97K-499-35` cultivar of cowpea using Phytozome. Chloroplast transit peptides were identified using ChloroP v.1.1 (Emanuelsson et al., 1999).

Accession accession numbers are available in 2.2.1.

4.2.7 Structure Prediction and Visualisation

I-TASSER (Yang and Zhang, 2015) was used for structure prediction using the amino acid sequence of mature Rca. PyMol (Schrödinger, LLC, 2015) v2.5 was used for 3D visualisation. The Rubisco recognition motif was determined by alignment with known Rubisco recognition motif in *Triticum aestivum* Rca.

4.2.8 *E. coli* Transformation

20 ng of plasmid vector was added 50 µL of competent *E. coli*, and incubated for 30 minutes on ice. The competent cells were heat shocked for 60 s at 42 °C then left on ice for 60 s. 1 mL of LB was added to competent cells, and this mixture was shaken for 1 h at 37 °C. For selection, the transformed cells were streaked onto an agar plate containing kanamycin and chloramphenicol. Successful transformants were streaked to single colonies, and these were used to make glycerol stocks for long-term storage.

4.2.9 Models and Figure Preparation

2nd and 3rd order polynomial models were fit to ATPase and Rubisco reactivation data and the quality of fit was assessed by Akaike's Information Criterion (AIC). Models with the lowest AIC value were selected for figure preparation and further data analysis. These models were used to predict activity values and 95% confidence interval across the tested temperature range. Linear regression models were used to predict ATPase activity values across the tested concentration range. Temperature optima and T_{80} were calculated using the highest interpolated value of ATPase and Rubisco reactivation activity, and the temperature values which corresponded to ≥80% of this. A K-nearest neighbours machine in OutlierDetections.jl (Muhr et al., 2022) v0.3.3 was used to filter outliers. Prior to this, negative blank corrected values were removed.

Data were analysed using Julia 1.8.5 (Bezanson et al., 2017) on VSCodium 1.74.3 running on macOS Ventura 13.2 (22D49). CSV (Quinn et al., 2022) v0.10.9 and DataFrames (Kamiński et al., 2022) v1.4.4 were used for data wrangling. CairoMakie (Danisch and Krumbiegel, 2021) v0.10.1 was used for plotting. GLM v1.8.1 was used for statistical modelling.

Multiple sequence alignments were made using ggmsa (Zhou et al., 2022) v1.0.3 in R 4.2.2.

Percentage identity was calculated using the `PID()` function of Biostrings (Pagès et al., 2022) with the following equation:

$$Percentage\,identity = 100 \cdot \frac{Identical\,Positions}{Average\,Sequence\,Length}$$

4.3 Results

4.3.1 Cowpea Rca is Similar on a Sequence Level

In cowpea, there are four Rca isoforms - two α isoforms and two β isoforms. These were identified by BLAST search of the `IT97K-499-35` cultivar of cowpea. These isoforms are termed VuRca1β, VuRca8α, VuRca10α & VuRca10β, indicating the chromosome on which the gene is encoded, and whether the isoform is α or β. The VuRca10β isoform is produced by alternative splicing of the VuRca10 gene. Other isoforms are encoded by different genes. In addition, there is an extra splice variant of chromosome 8 VuRca (for convenience termed VuRca8β) that is shorter than typical β Rca isoforms (295 amino acids vs 384 amino acids for VuRca1β). Although this splice variant contains high sequence similarity to genuine Rca, it is missing a critical motif required for recognition and binding of Rubisco. Figure 4.1 shows a sequence and structural comparison between VuRca1β and this isoform, highlighting the lack of Rubisco recognition motif. For this reason, the alternative splice variant of VuRca8α was excluded from further analysis. Any genuine function of this gene is not at present known. However, it is likely to be an ATPase as it contains Walker A and Walker B motifs which are well conserved in ATP-binding proteins. Below is a sequence (Fig. 4.1A), and structure (Fig. 4.1B) comparison of VuRca1β and VuRca8β, highlighting Walker motifs and the missing Rubisco recognition motif.

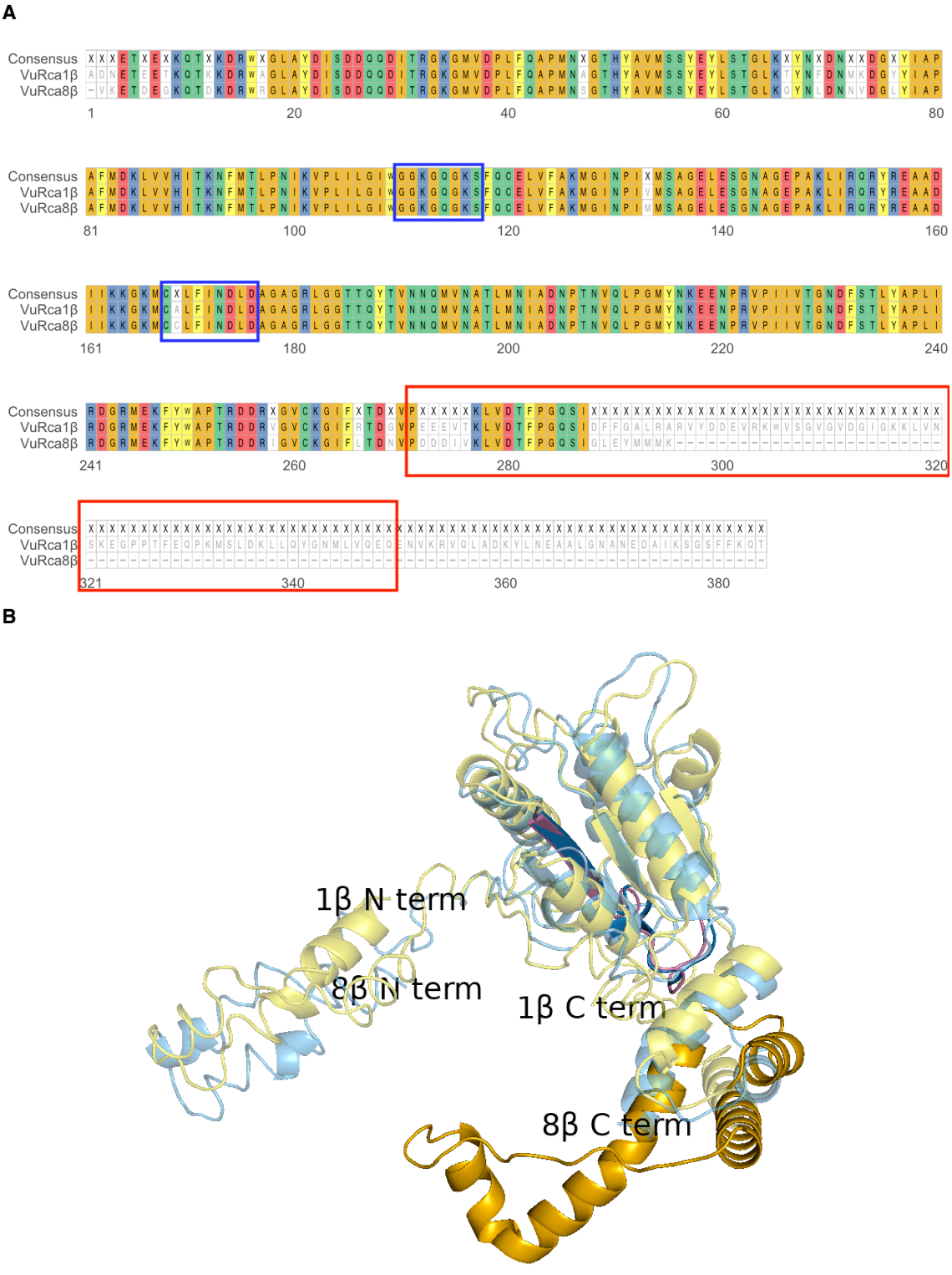


Figure 4.1: *Comparison between VuRca1β and VuRca8β sequence and structure, demonstrating that VuRca8β is likely not a genuine Rca.* ***A*** *Pairwise sequence alignment of VuRca1β and VuRca8β. Residue colour indicates side-chain chemistry. Uncoloured residues indicate that these do not conform to consensus. The region indicated by the red box is the Rubisco recognition motif. The regions indicated by blue boxes are the Walker A & Walker B motifs.* ***B*** *Superimposition of predicted structures of VuRca1β and VuRca8β.* ***Colour scheme*** *- VuRca1β: yellow; VuRca8β: sky blue; Rubisco recognition motif: orange; VuRca1β Walker motifs: purple; VuRca8β Walker motifs: blue.*

The four isoforms of Rca from cowpea share high sequence similarity. As shown in Table 4.1, the least similar isoforms are VuRca1β and VuRca10α with 84.2% identity. The most similar isoforms are VuRca10α and VuRca10β which, being splice variants, differ only in the C-terminal extension that VuRca10β lacks. The differences in these isoforms are explored in Figure 4.2. Amino acid polymorphisms tend to cluster to the N and C termini, with the centre-most amino acids having higher identity. Alignment residues 161, 163, 360, 362 & 364 are implicated in temperature response in wheat (Degen et al. (2020) and Scafaro et al. (2019a)). There is no diversity at residues 161, 360 & 364 of cowpea Rca. At these locations, all cowpea Rca isoforms have the residue that is associated with higher thermotolerance. At the 163rd residue, VuRca1β and VuRca8α have a lysine residue, while the chromosome 10 isoforms have an arginine, both of which are linear, positively charged residues. There is low diversity in known thermostability-related residues in cowpea Rca sequences. Other residues which are not presently known to affect thermostability show more diversity (Table 4.1). The sequence differences between these isoforms provide scope for differing catalytic properties.

Table 4.1: *Pairwise percentage identities for amino acid sequences of cowpea Rca isoforms.*

Isoform	VuRca1β	VuRca8α	VuRca10α	VuRca10β
VuRca1β	-	87.4	84.2	87.8
VuRca8α	87.4	-	89.0	85.9
VuRca10α	84.2	89.0	-	95.4
VuRca10β	87.8	85.9	95.4	-

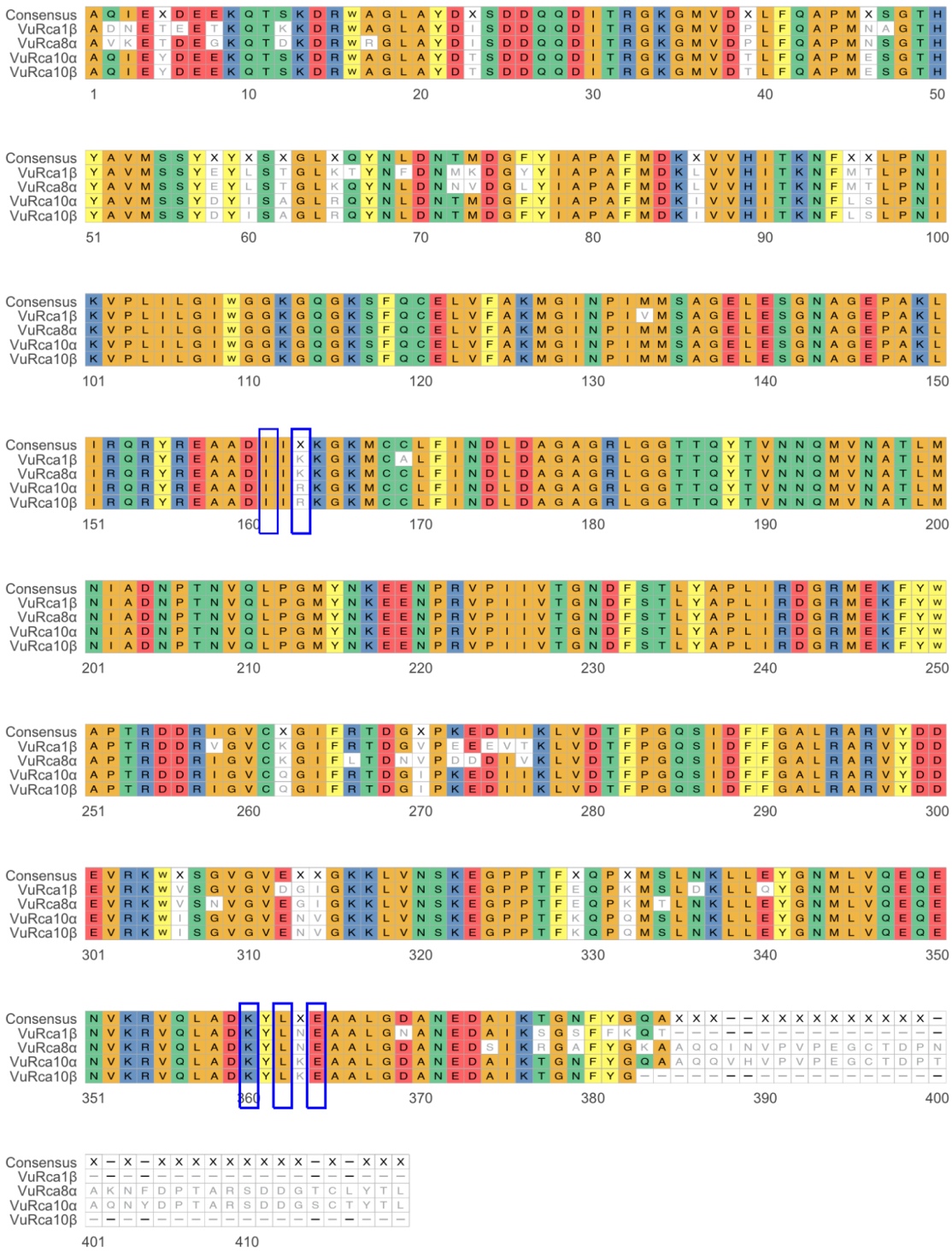


Figure 4.2: *Multiple sequence alignment of the amino acid sequence of cowpea Rca isoforms. Residue colour indicates side-chain chemistry. Uncoloured residues indicate that these do not conform to consensus. Blue boxes indicate residues implicated in temperature response in other species.*

4.3.2 Linearity in Concentration Response of Cowpea Rca

To investigate the concentration response of cowpea Rca, the rate of ATP hydrolysis was assessed at 30 °C for VuRca1β and VuRca8α under a range of concentrations. Recombinant, His-tagged cowpea Rca was used for characterisation of the catalytic properties of cowpea Rca. Figure 4.3 shows the rate of ATPase activity of those isoforms from 0.5-10 µM of Rca.

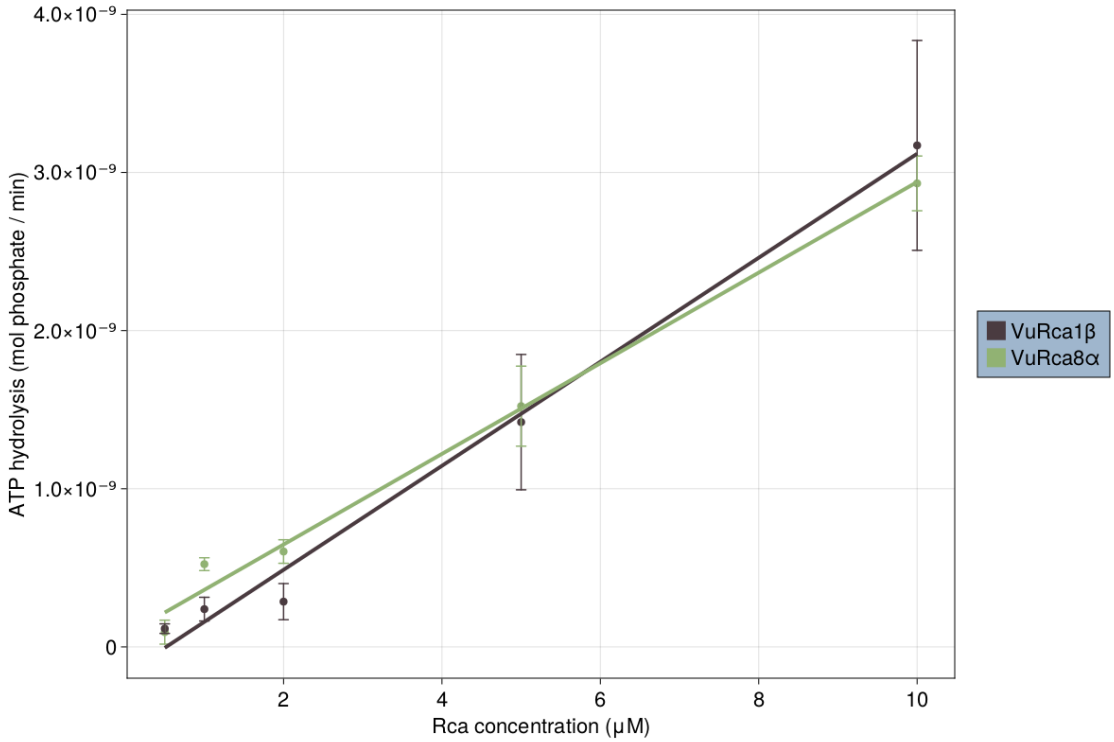


Figure 4.3: *Concentration response of two cowpea Rca isoforms (VuRca1β and VuRca8α). The relevant isoform of Rca was incubated at 30 °C for 5 minutes with 2 mM ATP and 5 mM $MgCl_2$. The reactions were terminated by addition of SDS to a final concentration of 6%. Error bars indicate mean ± standard deviation of n = 9 technical replicates.*

The tested cowpea Rca isoforms did not show the expected concentration response with an optimal concentration for per molar activity. The effect of concentration is more similar to classic Michaelis-Menten kinetics, in which per molar activity is irrespective of enzyme concentration under substrate saturating conditions. This knowledge guided temperature response experiments which were conduced with 2 µM of Rca, an amount which produced a measurable response while avoiding wastefulness.

4.3.3 Cowpea Rca Vary in Their Temperature Profile

To investigate if there are differences in the thermal optima of cowpea Rca isoforms, the rates of ATPase activity and Rubisco reactivation were measured across a range of temperatures for the four isoforms of cowpea Rca. The rates of Rubisco reactivation by Rca were calculated by the

difference between the rate of activity of Rubisco in the presence of Rca, and in the absence of Rca. To calculate ATP hydrolysis rate, Rca was incubated in a reaction mix containing 2 mM ATP for 5 minutes before quenching and measurement of phosphate release in a colourimetric assay.

A 2nd or 3rd order polynomial model was fit to the temperature profile of each isoform, for each activity (Figure 4.4). Absolute ATPase activity across the temperature range is shown in Table 4.2. The calculated temperature optima are shown in Table 4.3.

Table 4.2: *Absolute ATPase activity of VuRca isoforms from 25-55 °C. All units are mol ATP released min^{-1} mol Rca^{-1}. μ: sample mean; σ: sample standard deviation.*

Isoform	20		25		30		35		40		45		50		55	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
VuRca1β	5	2	12	3	15	5	17	4	12	5	7	3	7	2	7	4
VuRca8α	17	9	36	14	42	11	44	11	44	13	26	13	12	2	10	2
VuRca10α	80	35	137	44	185	53	261	22	291	34	300	27	269	23	74	39
VuRca10β	85	64	114	85	156	140	193	147	188	144	248	167	213	154	82	44

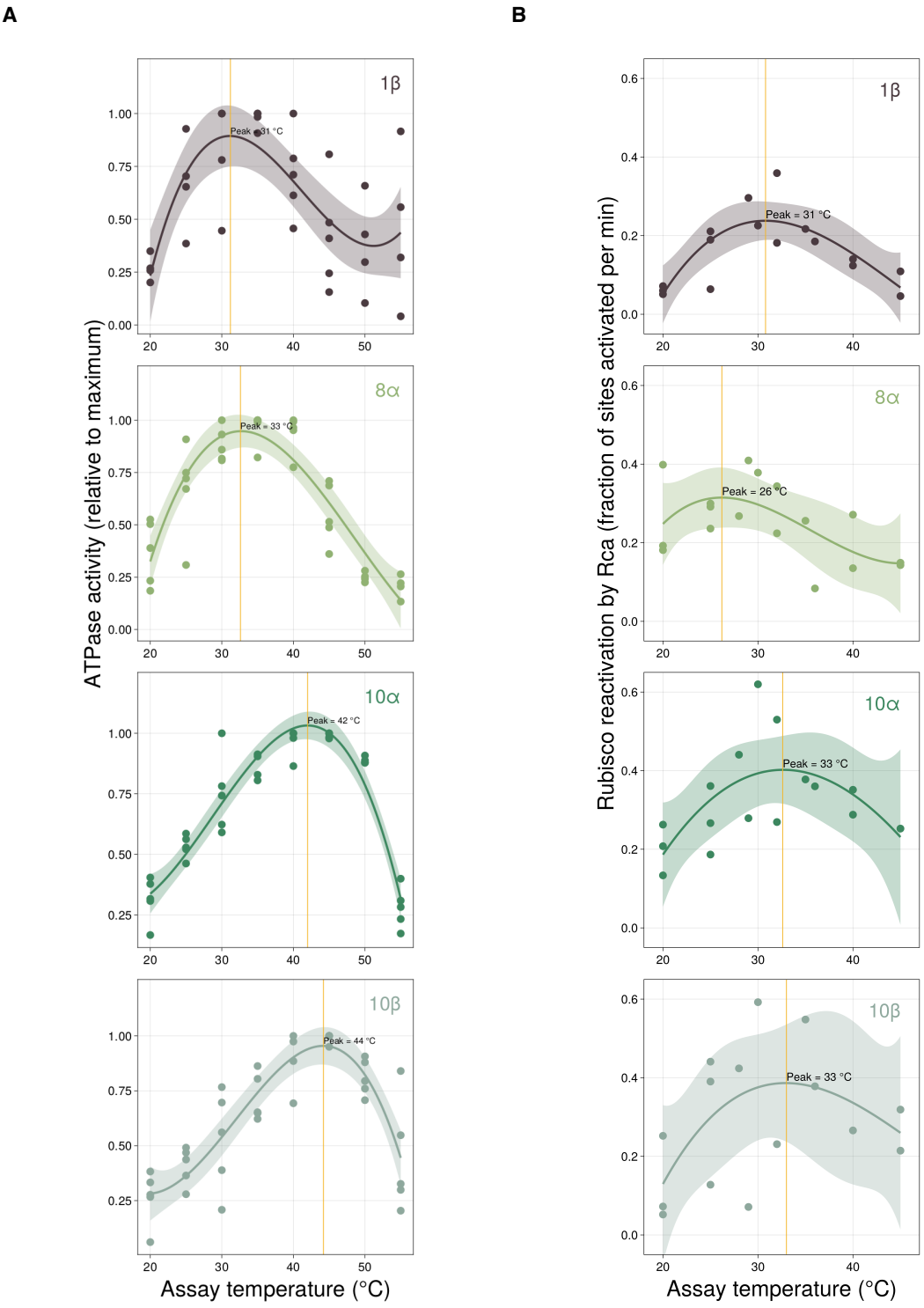


Figure 4.4: *Temperature optima of cowpea Rca.* ***4.4A*** *Temperature response of ATPase activity of cowpea Rca. The relevant isoform of Rca was incubated at a 2 µM concentration for 5 minutes with 2 mM ATP and 5 mM $MgCl_2$. The reactions were terminated by addition of SDS to a final concentration of 6%. ATPase activity values are expressed relative to maximum.* ***4.4B*** *Temperature response of Rubisco reactivation activity of cowpea Rca. In a two-stage assay, inhibited Rubisco was activated by Rca for set periods of time. The reactivated Rubisco was then allowed to fix $^{14}CO_2$. In both figures, points are mean of n = 3 technical replicates, and shaded regions indicate 95% confidence interval. Coloured lines within shaded regions represent calculated values from best-fitting regression models. Vertical yellow lines represent the calculated temperature optimum.*

Table 4.3: *Temperature optima of ATPase activity and Rubisco reactivation for cowpea Rca isoforms. T_{opt} is the temperature at which activity is highest. T_{80} is the range of temperatures at which activity is above 80% of maximum. Values are calculated from the data shown in Figure 4.4.*

Isoform	ATP hydrolysis		Rubisco reactivation	
	T_{opt} (°C)	T_{80} (°C)	T_{opt} (°C)	T_{80} (°C)
VuRca1β	31	25-39	31	25-37
VuRca8α	33	25-41	26	20-34
VuRca10α	42	33-49	33	26-42
VuRca10β	44	36-51	33	25-41

In all four isoforms, the optimal temperature for ATPase activity is equal to or higher than the optimal temperature for Rubisco reactivation. VuRca1β and VuRca8α both have noticeably lower temperature optima than the chromosome 10 isoforms. Furthermore, there are differences in the range of temperatures at which these isoforms keep highest activity. The T_{80} for ATPase activity spans 16 °C at the highest (VuRca10α, VuRca8α), and 14 °C at the lowest (VuRca1β). For Rubisco reactivation activity, the broadest T_{80} is also VuRca10α at 16 °C, and the narrowest is also VuRca1β at 12 °C. The higher temperature profile of these isoforms may imply an adaptation to the warm climate in which cowpea grows.

4.4 Discussion

Cowpea Rca has been shown to have a wide range of responses to temperature. The chromosome 10 isoforms have higher temperature optima than the other isoforms. These temperature (42 & 44 °C) optima resemble the Rca from warm adapted plants such as cotton (*Gossypium hirsutum*) (~42 °C) (Crafts-Brandner and Salvucci, 2000). VuRca1β and VuRca8α more closely resemble temperate plants such as *Triticum aestivum* (TaRca2α ATPase temperature optimum = 30.8 °C) (Degen et al., 2020). Accounting for all isoforms, the T_{opt} spans 25-51 °C for ATPase activity, and 25-42 °C for Rubisco reactivation. Only two isoforms of cowpea Rca were measured for concentration response due to difficulties in obtaining sufficient pure protein for this work. A more complete analysis would include the concentration response of VuRca10α & VuRca10β. The VuRca10 isoforms also showed notable variability in absolute rate of ATP hydrolysis, but the same trend in temperature response. This may be an issue with purity differences between purification preps, and could be normalised for by purity estimation. In spite of the variance, the rate of ATPase hydrolysis is higher in VuRca10α and VuRca10β than VuRca1β and VuRca8α.

The short splice variant of VuRca8 was excluded from analysis but may be photosynthetically relevant despite likely not activating Rubisco. Rca is known to form heteroligomers between different isoforms (Shivhare et al., 2019), so it is possible this may serve to regulate activity of other Rca isoforms by direct interaction with them. Experiments which mix different proportions of Rca iso-

forms could validate this (see Shivhare et al. (2019) for examples of subunit mixing experiments).

It is unusual that cowpea Rca displays a linear response of per molar ATPase activity with changes in concentration. The concentration of Rca used in these experiments is lower than chloroplastic Rca concentrations (Portis and Jr., 2003; Robinson et al., 1988), but is typical of *in vitro* Rca experiments. Rca is believed to act as a hexamer, and aggregates in a concentration-dependent manner, implying that there is an optimal concentration for it to be in its active form. In a paper published by Scafaro et al. (2016), two different temperature profiles can be seen in *Oryza australiensis* Rca at different concentrations. This implies an interplay between concentration and temperature on the rate of Rca's activities. A possible explanation for this effect is that both temperature and concentration are factors which affect aggregation of proteins. I.e. the concentration which is optimal for aggregation of Rca to a hexamer will not be the same at every temperature, and vice-versa. The concentration of ATP is also known to affect Rca thermal stability (Scafaro et al., 2019a). This interaction could be investigated with a multifactor experiment design (Selvamuthu and Das, 2018) which changes all variables in the same assay, and measures the two activities of Rca.

There are further characteristics of these isoforms that could be investigated. The effect of concentration on Rubisco reactivation activity was not investigated due to time constraints, and lack of sufficient pure protein. Different Rca isoforms respond differently to changing ratios of ADP:ATP. The ATPase activity, and Rubisco reactivation activity can be independently sensitive or insensitive to ADP (Perdomo et al., 2019). Preliminary data suggest VuRca1β and VuRca8α are more sensitive to ADP inhibition of ATPase activity than the cowpea 10 Rca isoforms are (Peter Gould, 2022, personal communications). The degree of ADP inhibition of cowpea Rca Rubisco reactivation is not known. Furthermore, the effects of any posttranslational modifications (PTM) on cowpea Rca would not be covered by this work, as recombinant expression in *E. coli* would not include native PTMs. Analysis of Rca purified from cowpea leaf extract would give information of activity of native Rca, including PTMs, but would not allow characterisation of individual isoforms. As Rca is known to have functionally relevant PTMs (Boex-Fontvieille et al., 2013), this would give a more complete picture of Rca activity in cowpea. *In vitro* isoform mixing experiments would allow for interactions between isoforms to be investigated. Another unknown factor is the protein levels of each isoform *in planta*. Quantitative RT-PCR experiments suggest that VuRca1β & VuRca8α are expressed in much higher abundance than VuRca10α & VuRca10β. (Rhiannon Page, 2022, personal communications). However, mRNA expression does not necessarily correlate with protein expression (Liu et al., 2016), and they may be induced under conditions of stress, in a similar manner to the less expressed isoform of wheat Rca (Degen et al., 2021). Chapter 5 details the development and validation of antibodies specific to cowpea Rca isoforms which would enable quantification of protein expression levels for these isoforms.

To the author's knowledge, this body of work represents the first characterisation of the biochemical characteristics of cowpea Rca. The information contained herewith may be of value for crop improvement of cowpea. For example, if a more heat-tolerant cowpea cultivar is desired,

upregulation of the chromosome 10 isoforms may help to keep photosynthesis active at higher temperatures.

4.5 Acknowledgements

Rubisco reactivation assays were done by Professor Elizabete Carmo-Silva. Plasmids containing codon-optimised His-Rca were designed and provided by Dr. Mike Page. This work would not have been possible without the expression and purification optimisation done in collaboration with Dr. Peter Gould, described in Chapter 3

Chapter 4 Bibliography

Adenodi, R. (2018). A centurial analysis of rainfall variability in Nigeria. *Nigerian Journal of Technology*, *37*(2), 543. https://doi.org/10.4314/njt.v37i2.34

Barta, C., Carmo-Silva, A. E., & Salvucci, M. E. (2011b). Rubisco Activase Activity Assays. In R. Carpentier (Ed.), *Photosynthesis Research Protocols* (pp. 375–382). Humana Press. https://doi.org/10.1007/978-1-60761-925-3_29

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Boex-Fontvieille, E., Daventure, M., Jossier, M., Hodges, M., Zivy, M., & Tcherkez, G. (2013). Phosphorylation pattern of rubisco activase in Arabidopsis leaves (W. Adams, Ed.). *Plant Biology*, *16*(3), 550–557. https://doi.org/10.1111/plb.12100

Boukar, O., Fatokun, C. A., Roberts, P. A., Abberton, M., Huynh, B. L., Close, T. J., Kyei-Boahen, S., Higgins, T. J., & Ehlers, J. D. (2015). Cowpea. In A. M. De Ron (Ed.), *Grain Legumes* (pp. 219–250). Springer New York. https://doi.org/10.1007/978-1-4939-2797-5_7

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2), 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Chifflet, S., Torriglia, A., Chiesa, R., & Tolosa, S. (1988). A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: Application to lens ATPases. *Analytical Biochemistry*, *168*(1), 1–4. https://doi.org/10.1016/0003-2697(88)90002-4

Crafts-Brandner, S. J., & Salvucci, M. E. (2000). Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO2. *Proceedings of the National Academy of Sciences*, *97*(24), 13430–13435. https://doi.org/10.1073/pnas.230451497

Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software*, *6*(65), 3349. https://doi.org/10.21105/joss.03349

Degen, G., Worrall, D., & Carmo-Silva, E. (2020). An isoleucine residue acts as a thermal and regulatory switch in wheat Rubisco activase: Rubisco activase thermal and regulatory switch. *The Plant Journal*. https://doi.org/10.1111/tpj.14766

Degen, G. E., Orr, D. J., & Carmo-Silva, E. (2021). Heat-induced changes in the abundance of wheat Rubisco activase isoforms. *New Phytologist*, *229*(3), 1298–1311. https://doi.org/10.1111/nph.16937

Emanuelsson, O., Nielsen, H., & Heijne, G. V. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, *8*(5), 978–984. https://doi.org/10.1110/ps.8.5.978

Kamiński, B., White, J. M., Bouchet-Valat, M., Powerdistribution, Garborg, S., Quinn, J., Kornblith, S., Cjprybol, Stukalov, A., Bates, D., Short, T., DuBois, C., Harris, H., Squire, K., Arslan, A., Pdeffebach, Anthoff, D., Kleinschmidt, D., Noack, A., Shah, V. B., Mellnik, A., Arakaki, T., Mohapatra, T., Peter, Karpinski, S., Lin, D., Chagas, R. A. J., Timema, ExpandingMan, & Oswald, F. (2022). JuliaData/DataFrames.jl: V1.4.3. https://doi.org/10.5281/ZENODO.7316260

Lilley, R. M., & Portis, A. R. (1990). Activation of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) by Rubisco Activase: Effects of Some Sugar Phosphates. *Plant Physiology*, *94*(1), 245–250. https://doi.org/10.1104/pp.94.1.245

Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, *165*(3), 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Lucas, M. R., Ehlers, J. D., Huynh, B.-L., Diop, N.-N., Roberts, P. A., & Close, T. J. (2013). Markers for breeding heat-tolerant cowpea. *Molecular Breeding*, *31*(3), 529–536. https://doi.org/10.1007/s11032-012-9810-z

Muhr, D., Affenzeller, M., & Blaom, A. D. (2022). OutlierDetection.jl: A modular outlier detection ecosystem for the Julia programming language. https://doi.org/10.48550/ARXIV.2211.04550

OECD. (2019). *Safety Assessment of Foods and Feeds Derived from Transgenic Crops, Volume 3: Common bean, Rice, Cowpea and Apple Compositional Considerations*. https://doi.org/10.1787/f04f3c98-en

Olaniyi, O., Olutimehin, I., & Funmilayo, O. (2019). Review of Climate Change and Its effect on Nigeria Ecosystem. *International journal of Rural Development, Environment and Health Research*, *3*(3), 92–100. https://doi.org/10.22161/ijreh.3.3.3

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2022). Biostrings: Efficient manipulation of biological strings. https://doi.org/10.18129/B9.bioc.Biostrings

Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E., & Andralojc, P. J. (2007). Rubisco regulation: A role for inhibitors. *Journal of Experimental Botany*, *59*(7), 1569–1580. https://doi.org/10.1093/jxb/ern084

Perdomo, J. A., Degen, G. E., Worrall, D., & Carmo-Silva, E. (2019). Rubisco activation by wheat Rubisco activase isoform 2β is insensitive to inhibition by ADP. *Biochemical Journal*, *476*(18), 2595–2606. https://doi.org/10.1042/BCJ20190110

Portis, A. R. (1990). Rubisco activase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1015*(1), 15–28. https://doi.org/10.1016/0005-2728(90)90211-L

Portis, A. R., & Jr. (2003). Rubisco activase – Rubisco's catalytic chaperone. *Photosynthesis Research*, *75*(1), 11–27. https://doi.org/10.1023/A:1022458108678

Portis, A. R., Salvucci, M. E., & Ogren, W. L. (1986). Activation of Ribulosebisphosphate Carboxylase/Oxygenase at Physiological CO $_2$ and Ribulosebisphosphate Concentrations by Rubisco Activase. *Plant Physiology*, *82*(4), 967–971. https://doi.org/10.1104/pp.82.4.967

Quinn, J., Bouchet-Valat, M., Robinson, N., Kamiński, B., Newman, G., Stukalov, A., Vogt, C., Cjprybol, Noack, A., Kelman, T., Davies, E., ExpandingMan, Ian, Piibeleht, M., Finnegan, R., Evalparse, Silberstein, A. M., Anthony Blaom, P., Lungwitz, B., König, B., De Graaf, C., Woodfield, C., Barton, D., Aluthge, D., Saba, E., Noronha, F., Kragol, Rosario, H. D., Ranocha, H., & Butterworth, I. (2022). JuliaData/CSV.jl: V0.10.7. https://doi.org/10.5281/ZENODO.7250547

Robinson, S. P., Streusand, V. J., Chatfield, J. M., & Portis, A. R. (1988). Purification and Assay of Rubisco Activase from Leaves. *Plant Physiology*, *88*(4), 1008–1014. https://doi.org/10.1104/pp.88.4.1008

Salvucci, M. E., Osteryoung, K. W., Crafts-Brandner, S. J., & Vierling, E. (2001). Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo. *Plant Physiology*, *127*(3), 1053–1064.

Salvucci, M. E., van de Loo, F. J., & Stecher, D. (2003). Two isoforms of Rubisco activase in cotton, the products of separate genes not alternative splicing. *Planta*, *216*(5), 736–744. https://doi.org/10.1007/s00425-002-0923-1

Scafaro, A., Gallé, A., Rie, J. V., Carmo-Silva, E., Salvucci, M. E., & Atwell, B. J. (2016). Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytologist*, *211*(3), 899–911. https://doi.org/10.1111/nph.13963

Schrödinger, LLC. (2015). *The PyMOL molecular graphics system, version 2.5*.

Selvamuthu, D., & Das, D. (2018). Multifactor Experimental Designs. In *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control* (pp. 265–317). Springer Singapore. https://doi.org/10.1007/978-981-13-1736-1_8

Shivhare, D., & Mueller-Cajar, O. (2017). In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiology*, *174*(3), 1505–1516. https://doi.org/10.1104/pp.17.00554

Shivhare, D., Ng, J., Tsai, Y.-C. C., & Mueller-Cajar, O. (2019). Probing the rice Rubisco–Rubisco activase interaction via subunit heterooligomerization. *Proceedings of the National Academy of Sciences*, *116*(48), 24041–24048. https://doi.org/10.1073/pnas.1914245116

Srivastava, R., Kobayashi, Y., Koyama, H., & Sahoo, L. (2022). Cowpea NAC1/NAC2 transcription factors improve growth and tolerance to drought and heat in transgenic cowpea through combined activation of photosynthetic and antioxidant mechanisms. *Journal of Integrative Plant Biology*, jipb.13365. https://doi.org/10.1111/jipb.13365

Taylor, T., Denson, J.-P., & Esposito, D. (2017). Optimizing Expression and Solubility of Proteins in *E. coli* Using Modified Media and Induction Parameters. In N. A. Burgess-Brown (Ed.), *Heterologous Gene Expression in E.coli* (pp. 65–82). Springer New York. https://doi.org/10.1007/978-1-4939-6887-9_5

Waheeda, K., Kitchel, H., Wang, Q., & Chiu, P.-L. (2023). Molecular mechanism of Rubisco activase: Dynamic assembly and Rubisco remodeling. *Frontiers in Molecular Biosciences*, *10*, 1125922. https://doi.org/10.3389/fmolb.2023.1125922

Wijewardene, I., Shen, G., & Zhang, H. (2021). Enhancing crop yield by using Rubisco activase to improve photosynthesis under elevated temperatures. *Stress Biology*, *1*(1), 2. https://doi.org/10.1007/s44154-021-00002-5

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174–W181. https://doi.org/10.1093/nar/gkv342

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L., Guan, Y., Dai, Z., & Yu, G. (2022). Ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics*, *23*(4), bbac222. https://doi.org/10.1093/bib/bbac222

Chapter 5

Production and validation of recombinant anti-cowpea Rca antibodies

5.1 Introduction

Rubisco activase (Rca) is an important enzyme in photosynthesis. Rca is responsible for maintaining catalytic competence of Rubisco, and has important implications as a photosynthetic rate limiter under conditions such as high temperature and fluctuating light. In many species, Rca is present as several isoforms, which can have different catalytic properties. Research into these individual isoforms can be challenging as they tend to be highly similar on a sequence level. Alpha and beta isoforms can be distinguished on the basis of molecular weight (alpha isoforms are ~4 kDa larger as they contain a C-terminal extension). However, alphas cannot be distinguished from other alphas, nor betas from other betas, on the basis of charge or molecular weight. This is relevant in the context of cowpea, as it contains two α isoforms and two β isoforms, which have different catalytic properties (Chapter 4). To solve the problem of distinguishing isoforms, antibodies specific to certain isoforms could be used. This would would enable detection and quantification of the protein abundance of each isoform. In Chapter 4, the temperature response of the four isoforms of cowpea Rca was investigated, and in VuRca-specific antibodies would enable further characterisation of their unique properties, particularly *in planta*.

Antibodies are near-ubiquitous tools in molecular biology research (see Chapter 1.7 for more information on antibodies). They are Y-shaped molecules which are found both on the surface of B-cells, and in soluble form. They originate in the adaptive immune response, where they bind to specific cognate molecules known as their antigen. The evolutionary requirement for antibodies to bind to a large array of molecules with high specificity and avidity makes them excellent tools for detection of specific molecules. The ability for antibodies to bind a target with high specificity

and avidity makes them invaluable for detection, identification, and quantification of biomolecules. It is possible for antibodies to distinguish between very similar molecules such as proteins with a single amino acid substitution, over even a posttranslational modification (Hattori and Koide, 2018; Hebbes et al., 1989). Despite their many advantages, a lot of irreproducible literature can be traced back to improper use of antibodies (Baker, 2015). It is possible for antibodies to bind molecules other than their target (non-specific binding), and to bind their target weakly or not at all. This can lead to misinterpretation of results, and antibodies require extensive validation to be reliable as detection, purification, and quantification tools. Various methods can be used to validate antibodies. Heterologous expression of a target protein provides strong evidence that the antibody binds its target properly, but requires additional validation for non-specific binding in the context the antibodies will be used. Correlation with genetic strategies such as upregulation or downregulation of target protein is a strong indicator of specificity, but at a low throughput. Multiple antibodies can be tested in tandem as a validation strategy, but it is possible they could fail in the same way (Edfors et al., 2018; Taussig et al., 2018; Voskuil et al., 2020).

Traditional antibody production requires exposure of an animal (usually a small mammal such as a rabbit or mouse) to an immune challenge with the antigen one wishes to identify. The blood serum can be harvested directly - producing polyclonal antibodies. Alternatively, a single B cell producing antibodies can be fused to an immortalised cell line - producing a hybridoma which can be grown and harvested for monoclonal antibodies (Zaroff and Tan, 2019). Polyclonal and monoclonal antibodies have their pros and cons. Monoclonal antibodies have a higher up front cost to produce, but are a somewhat repeatable source of antibody production. Due to genetic drift, the antibody sequence can change and be lost over time, and monoclonal antibodies are not always monospecific - frequently expressing more than one antibody clone (Bradbury et al., 2018; Goodman, 2018). By contrast, polyclonal antibodies are a finite resource - once it is depleted, the same polyclonal antibody cannot be made again. Moreover, polyclonal antibodies can also have a higher sensitivity for low abundance antigen as it can bind multiple targets, but this same characteristic typically results in higher non-specific binding. More information on polyclonal vs monoclonal antibodies can be found in Lipman et al. (2005).

Due to the cost and ethical concern associated with rearing animals for antibody production, there is interest in animal-free antibody production. Production of antibodies in bacterial systems is more challenging than with most proteins due to the machinery required for antibody synthesis. Antibodies are glycoproteins produced in the endoplasmic reticulum (ER) of eukaryotic cells (Feige et al., 2010). Because of this, expression in bacterial hosts is extremely difficult, though not impossible (Eaglesham et al., 2021; Robinson et al., 2015). Alternative methods of recombinant antibody production exist such as phage display. Phage display is an *in vitro* technology in which antibody genes are inserted into the phage genome to create an antibody library. These phage-antibody fusion proteins are expressed on the phage, and the library is panned against the desired antibody target. Antibodies which bind strongly and specifically can be sequenced and tailored into the desired format (Roth et al., 2021). This creates a truly repeatable source of monospecific

antibody.

This chapter details the design process and production of recombinant antibodies against specific isoforms of cowpea Rca. Phage display was used to generate these antibodies, and the chosen antibody format is mouse IgG 2a. The results of validation of these antibodies are shown, and considerations for robust antibody validation are detailed. In the discussion, I discuss the pros and cons of recombinant antibodies vs traditionally produced antibodies, and explore potential research questions enabled by specific anti-Rca antibodies for cowpea.

5.2 Materials and Methods

5.2.1 Cowpea Rca Sequences

Cowpea Rca sequences were identified in the genome of the `IT97K-499-35` cultivar of cowpea using Phytozome. Chloroplast transit peptides were identified using ChloroP v.1.1 (Emanuelsson et al., 1999).

Table 5.1: *Cowpea Rca sequence accession numbers*

Isoform	Accession number
VuRca1β	Vigun01g219300.1
VuRca8α	Vigun08g150700.1
VuRca10α	Vigun10g051600.1
VuRca10β	Vigun10g051600.2

5.2.2 Recombinant Rca Expression & Purification

Pure Rca used for antibody validation was produced using the optimised method described in Chapter 3.2.

5.2.3 Leaf Extract Preparation

Crude leaf extracts were extracted from 15 day old *Vigna unguiculata* plants. Four leaf discs of leaf area 0.55 cm^2 were ground in ice-cooled mortars containing 600 µL of extraction buffer (50 mM tricine-NaOH pH 8.0, 10 mM EDTA, 1% (w/v) PVP_{40}, 20 mM β-mercaptoethanol, 1 mM PMSF, 10 µM leupeptin). Leaf homogenate was centrifugated at 4 °C and 14,000x *g*. Supernatant was mixed with laemmli buffer to make SDS-PAGE samples as described in 5.2.5.

5.2.4 Protein Concentration Determination

Protein concentration was determined using the Bradford method (Bradford, 1976). 0-50 µg of bovine serum albumin was used to construct the standard curve. 1 µL of sample and 200 µL of Bradford reagent were added in triplicate to wells of a 96-well plate. Protein concentration was estimated using the standard curve comparing protein amount to the absorbance at 595 nm.

5.2.5 SDS-PAGE

Protein samples were mixed 1:1 with 2x reducing laemmli buffer (20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.004% bromophenol blue, 125 mM Tris-HCl, pH 6.8) and boiled at 95 °C for 5 minutes. Samples were separated by electrophoresis on 12% Mini-PROTEAN TGX Stain-Free gels (Bio-Rad) at 150 V until the dye front ran off the gel (approximately 70 minutes). SDS-PAGE Running buffer contains (125 mM Tris, 200 mM glycine, 0.1% SDS) in a 1x stock. After separation, stain-free gels were imaged on the GelDoc Go gel imaging system (Bio-Rad).

For specificity determination, 5 µL of 0.5 µg/µL of each protein sample was loaded. For linear range analysis, a range of 0.5-3 µg of Rca was loaded. 2 µL of molecular weight marker (Bio-Rad Precision Plus Protein Dual Color Standards) was loaded on each gel.

5.2.6 Western Blotting

Stain-free gels were cut down to membrane size and transferred to nitrocellulose membranes using the iBlot Gel Transfer Device (ThermoFisher) as per manufacturer instructions. Membranes were rinsed twice in milliQ water and blocked in 4% milk in TBS (tris-buffered saline) (20 mM Tris, 150 mM NaCl, pH 7.5) for 30 minutes. After blocking, membranes were rinsed in milliQ water and probed overnight with 20 mL of primary antibody (Abcalis) against the relevant Rca isoform at 10 nM concentration with gentle shaking (15 RPM). Membranes were washed 6x 15 minutes in TBST (TBS + 500 µL/L Tween-20), then rinsed with milliQ water and incubated with 20 mL goat anti-mouse secondary antibody (LI-COR Biosciences Cat# 926-32351, RRID:AB_2782998), diluted 1:10000, for 2 h. Membranes were then further washed 4x 15 min in TBST, rinsed in milliQ water, and visualised using the Odyssey Fc (LI-COR) at 700 nm (molecular weight marker) and 800 nm (antibody).

The mass concentration of anti-Rca antibody solutions was supplied by Abcalis. This was converted to amount concentration using the average molecular weight of an IgG molecule (150 kDa).

5.2.7 Band Quantification

Band intensity was quantified using the quantification feature of Image Studio (Li-COR).

5.2.8 Models and Figure Preparation

Linear regression models were fit to subsets of antibody binding data. Model quality was assessed by Akaike information criterion (AIC) and the coefficient of determination (R^2).

Data were analysed using Julia 1.8.5 (Bezanson et al., 2017) on VSCodium 1.74.3 running on macOS Ventura 13.2 (22D49). CSV (Quinn et al., 2022) v0.10.9 and DataFrames (Kamiński et al., 2022) v1.4.4 were used for data wrangling. CairoMakie (Danisch and Krumbiegel, 2021) v0.10.1 was used for plotting and image annotation. GLM v1.8.1 and StatsBase v0.33.21 were used for statistical modelling and calculation of model quality metrics.

Multiple sequence alignments were made using ggmsa (Zhou et al., 2022) v1.0.3 in R 4.2.2.

5.3 Results

5.3.1 Design of Antibodies

As the sequences of cowpea Rca (see Chapter 4) are highly similar, antibodies to distinguish them must be carefully designed to target regions of dissimilarity. The primary application of these antibodies is in western blotting, so α and β isoforms can be distinguished on the basis of molecular weight. However, antibodies targeting α isoforms must not cross-react with other α isoforms, and likewise for β isoforms. Alongside isoform-specific antibodies, antibodies were produced targeting all Rca isoforms (termed pan-Rca), ideally from all plant species. In Figure 5.1, a multiple sequence alignment of cowpea Rca isoforms is shown highlighting regions that were targeted for antibody production. These peptides were synthesised and used for selection or deselection of antibodies. Two peptides were targeted for pan-Rca antibodies - `GNAGEPAKLIRQRYREAAD` (Box 1) & `IRDGRMEKFYWAPTREDR` (Box 2). An important note is that the cowpea Rca sequence in Box 2 is one amino acid different from the sequence used for antibody selection. This antibody is intended to target Rca from a wide range of species, and the chosen sequence is more common. The regions that were targeted are conserved in cowpea Rca, and show little diversity across broader plant phylogenies. Two peptides were targeted which are specific to VuRca10 isoforms - `RTDGIPKEDIIKL` (Box 3) & `KTGNFYGQAAQQVHVP` (Box 4). One peptide was targeted against VuRca1β and VuRca8α, respectively - `RTDGVPEEEVTKL` (Box 3) & `KRGAFYGKAAQQINVP` (Box 4). These regions were selected for their sequence diversity between cowpea Rca antibodies that cannot be distinguished on the basis of molecular weight.

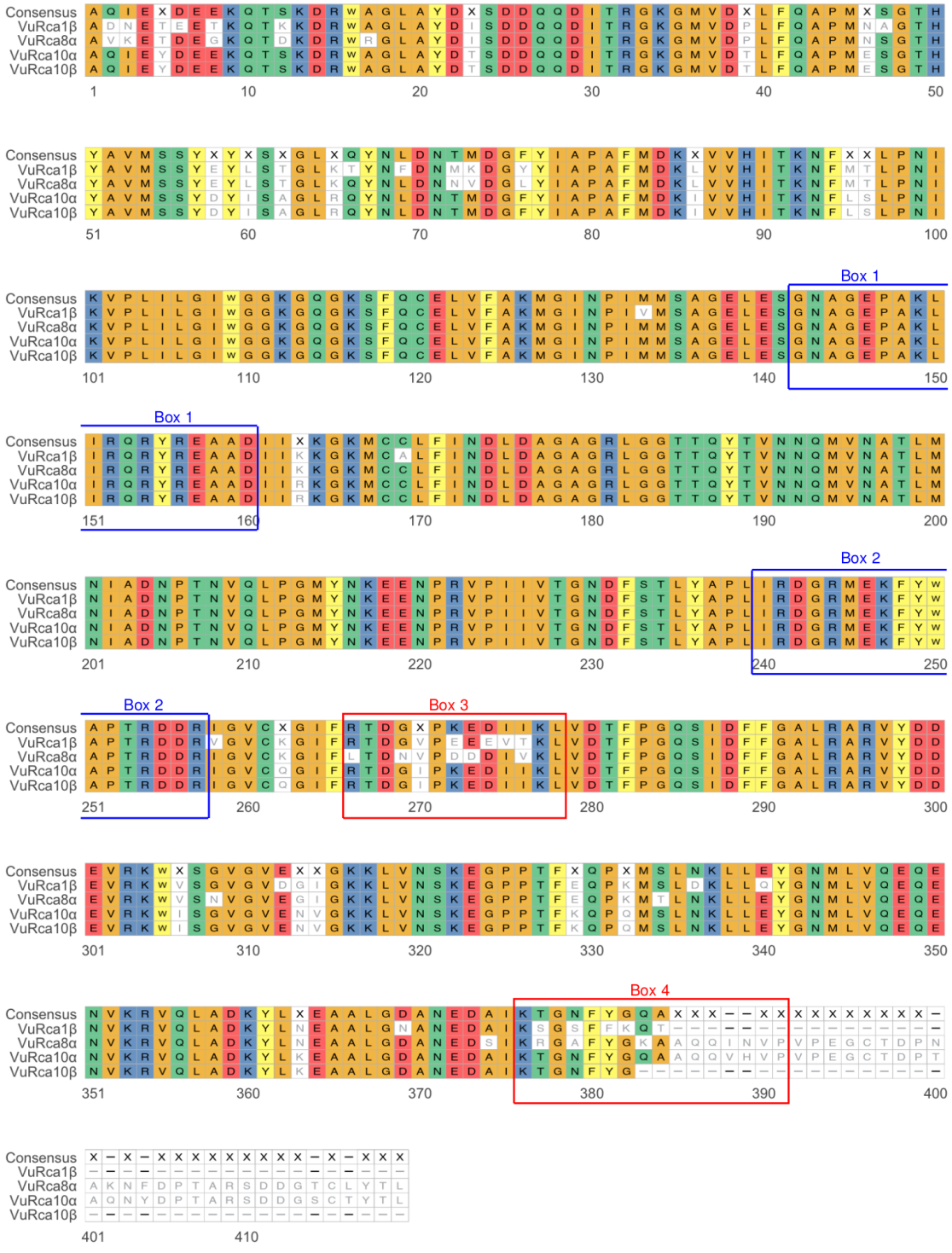


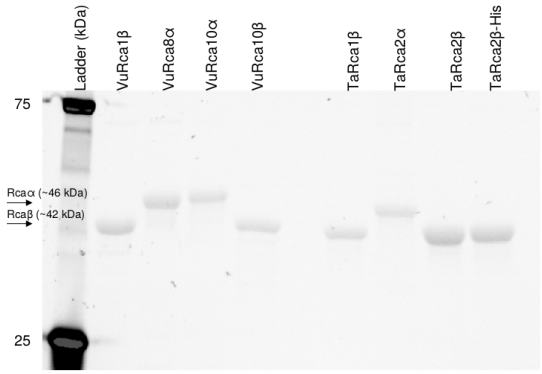
Figure 5.1: *Sequence comparison of the Rca isoforms of cowpea. Blue boxes indicate regions targeted by pan-Rca antibodies. Red boxes indicate regions targeted by VuRca isoform-specific antibodies.*

5.3.2 Antibody Validation for Target Specificity

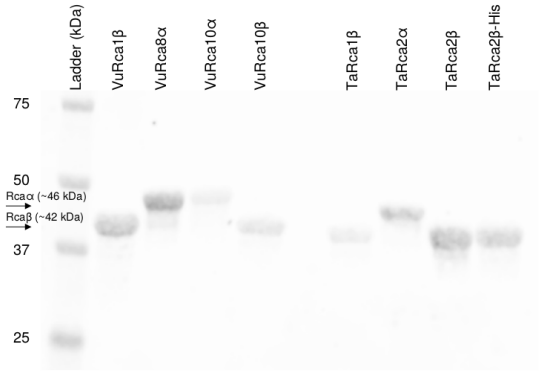
Multiple antibody clones against specific isoforms of cowpea Rca, or pan-Rca, were produced by Abcalis (Table 5.2). To confirm specificity of these antibodies, western blots were run with heterologously expressed, pure protein for each isoform of cowpea Rca. Alongside these, recombinant wheat Rca isoforms were loaded as a control and as additional tests for the pan-Rca antibodies. After washing, each blot was incubated with 20 mL of goat anti-mouse IgG fluorescent secondary antibody, as the recombinant antibodies are formatted to mouse IgG.

A blot was run for each antibody clone, and the aggregate results of these experiments can be seen in Table 5.2. The blots used for validation of successful antibodies and the unsuccessful antibody targeting VuRca1β are shown in figures below. Figures 5.2B, 5.2D, 5.2F, 5.2H, & 5.2J show the results of successful pan-Rca antibodies. Figure 5.3B shows the results of the unsuccessful VuRca1β-specific antibody. Figures 5.3D & 5.3F show the results of the VuRca8α-specific antibodies, and lastly Figure 5.3H is the successful VuRca10-specific antibody. Note that both α and β isoforms are targeted by this antibody.

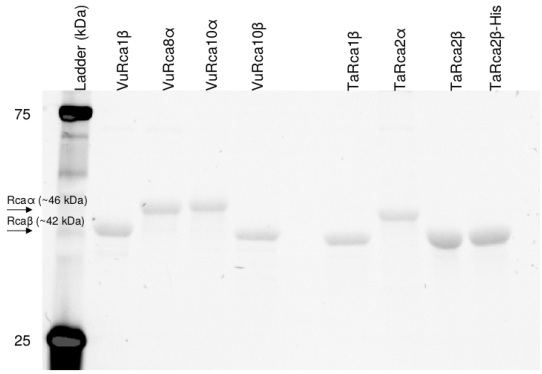
A *ABK286-1-D03-M I pan-Rca*



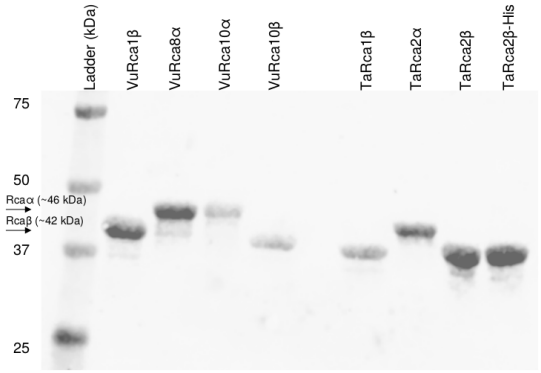
B *ABK286-1-D03-M I pan-Rca*



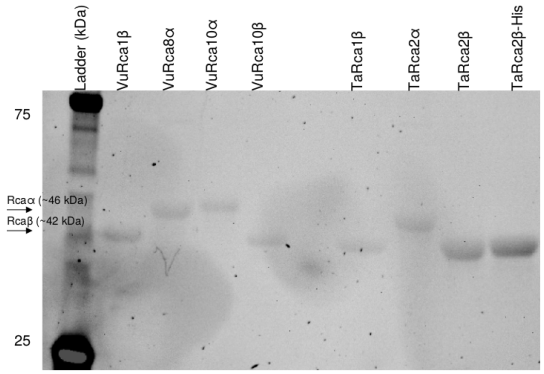
C *ABK286-1-D12-M I pan-Rca*



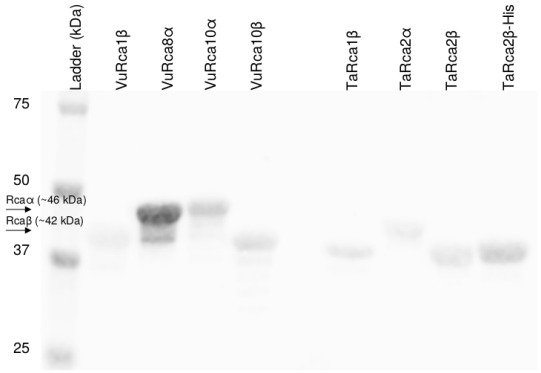
D *ABK286-1-D12-M I pan-Rca*



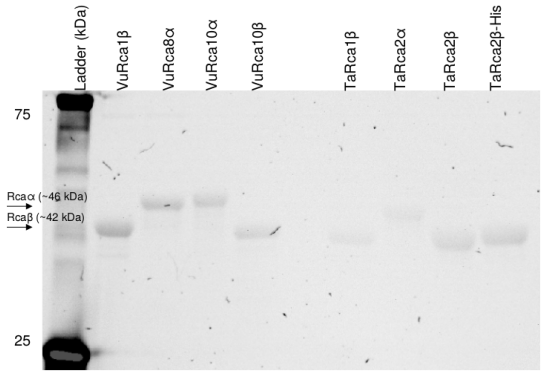
E *ABK286-1-B01-M l pan-Rca*



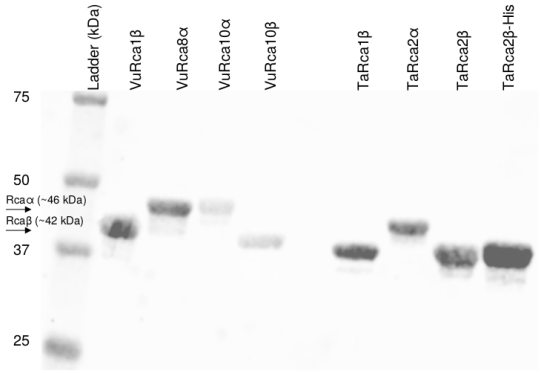
F *ABK286-1-B01-M l pan-Rca*



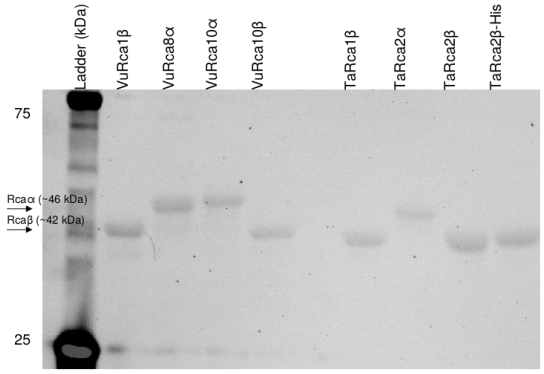
G *ABK302-1-A02-M l pan-Rca*



H *ABK302-1-A02-M l pan-Rca*



I *ABK302-1-E07-M l pan-Rca*



J *ABK302-1-E07-M l pan-Rca*

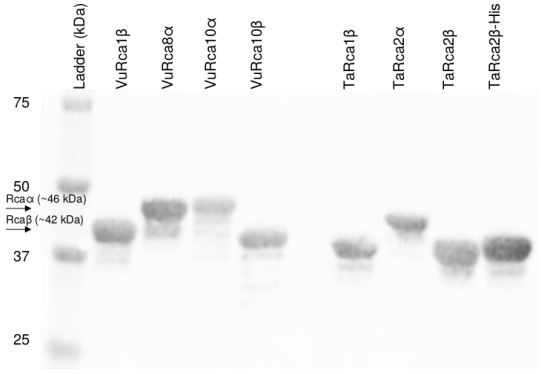
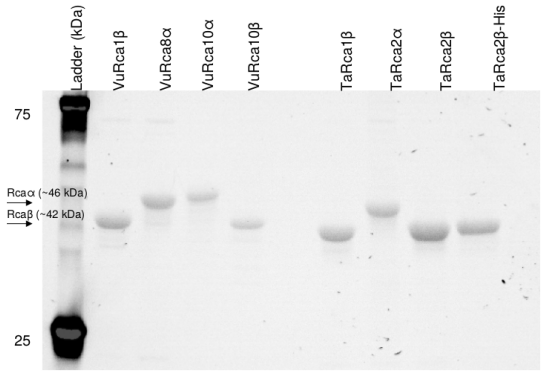
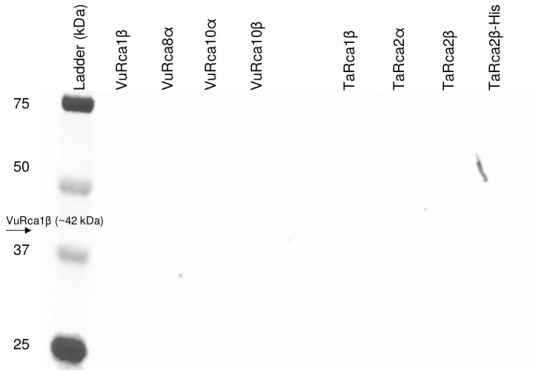


Figure 5.2: *Gels and blots used for validation of pan-Rca Abcalis antibodies.* ***A, C, E, G, I*** *Recombinant Rca from cowpea and wheat were separated by SDS-PAGE and visualised with a Bio-Rad GelDoc Go.* ***B, D, F, H, J*** *The relevant gel was used for immunoblotting with the noted antibody. 2.5 µg of purified protein was loaded in each gel lane.*

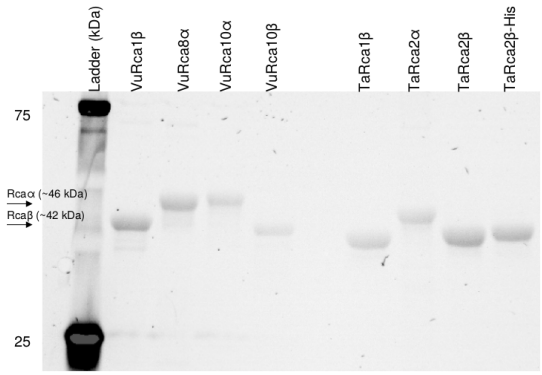
A *ABK296-1-A04-M | VuRca1β*



B *ABK296-1-A04-M | VuRca1β*

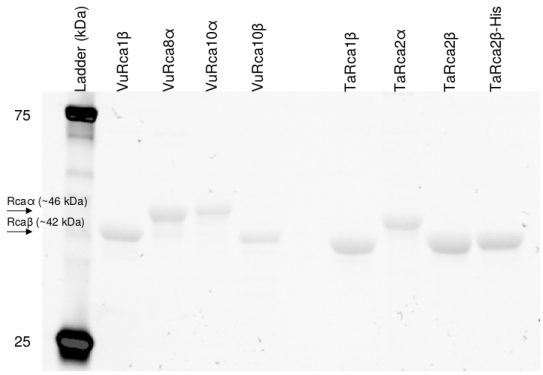


C *ABK298-1-C05-M | VuRca8α*

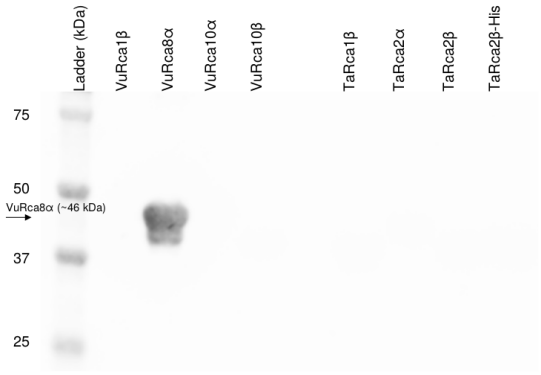


D *ABK298-1-C05-M | VuRca8α*

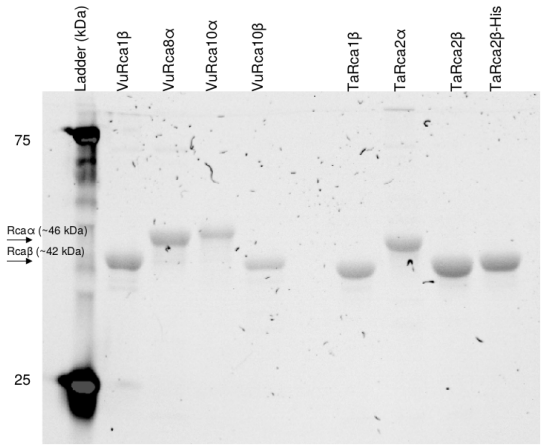
E *ABK298-1-E07-M | VuRca8α*



F *ABK298-1-E07-M | VuRca8α*



G *ABK300-1-B02 I VuRca10*



H *ABK300-1-B02 I VuRca10*

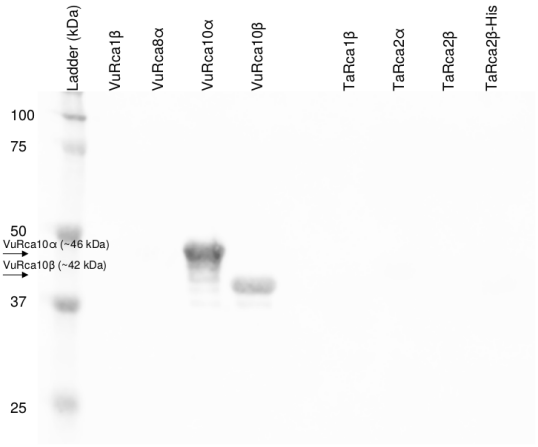


Figure 5.3: *Gels and blots used for validation of isoform-specific Abcalis antibodies.* ***A, C, E, G*** *Recombinant Rca from cowpea and wheat were separated by SDS-PAGE and visualised with a Bio-Rad GelDoc Go.* ***B, D, F, H*** *The relevant gel was used for immunoblotting with the noted antibody. 2.5 µg of purified protein was loaded in each gel lane.*

Table 5.2: *Specificity of antibody binding. Validation results for each antibody produced against pan-Rca or isoforms of cowpea Rca. Antibody ID, target sequence, targeted protein and success or failure of antibody to bind its intended target are shown.*

Antibody ID	Target sequence	Target	Binds target?	Notes
ABK286-1-D03-M	IRDGRMEKFYWAPTREDR	Pan-Rca	True	
ABK286-1-D12-M	IRDGRMEKFYWAPTREDR	Pan-Rca	True	
ABK286-1-E07-M	IRDGRMEKFYWAPTREDR	Pan-Rca	False	Weakly binds VuRca8α
ABK286-1-G12-M	IRDGRMEKFYWAPTREDR	Pan-Rca	False	
ABK286-1-B01-M	IRDGRMEKFYWAPTREDR	Pan-Rca	True	Binds only weakly to VuRca1β
ABK302-1-A02-M	GNAGEPAKLIRQRYREAAD	Pan-Rca	True	
ABK302-1-A04-M	GNAGEPAKLIRQRYREAAD	Pan-Rca	False	
ABK302-1-E01-M	GNAGEPAKLIRQRYREAAD	Pan-Rca	False	
ABK302-1-E07-M	GNAGEPAKLIRQRYREAAD	Pan-Rca	True	
ABK296-1-A04-M	RTDGVPEEEVTKL	VuRca1	False	
ABK298-1-C05-M	KRGAFYGKAAQQINVP	VuRca8	True	
ABK298-1-E07-M	KRGAFYGKAAQQINVP	VuRca8	True	
ABK296-2-B01-M	RTDGIPKEDIIKL	VuRca10	False	
ABK296-2-G01-M	RTDGIPKEDIIKL	VuRca10	False	
ABK300-1-B02-M	KTGNFYGQAAQQVHVP	VuRca10	True	
ABK300-1-E01-M	KTGNFYGQAAQQVHVP	VuRca10	False	

In total, out of nine pan-Rca antibodies, five bind to all tested isoforms of Rca. Of these, one binds only weakly to VuRca1β making it a less viable anti-Rca antibody than the others. Two different sequences were targeted for pan-Rca, each of which produced viable clones.

Only a single clone was supplied against VuRca1β, which did not bind to any isoform of cowpea or wheat Rca.

Two clones were produced against VuRca8α, both of which bind specifically to this isoform and produce similar results.

Two sites were targeted for VuRca10 isoforms. Out of four clones targeting these two different regions, only one clone binds these specifically. This clone binds to both VuRca10α and VuRca10β, and no other tested Rca isoforms.

In summary, at least one antibody clone is successful for each target except VuRca1β. The contribution of VuRca1β to the Rca pool can be estimated using the difference between signal from pan-Rca and VuRca10β.

5.3.3 Antibody Binding Profile is Not Linear Across All Concentrations

The linear range of binding is an important consideration when using antibodies for quantification. Antibodies have a sigmoidal binding profile, meaning that using a linear regression model to quantify amounts outside of the linear range will result in an over, or under, estimation of target amount. To assess the linear range of antibody binding, 0.5-3.0 µg of target protein for successful antibodies was run on SDS-PAGE and transferred to western blots. The signal of each band was quantified, and used for linear regression analysis. For pan-Rca, an antibody against each of the binding targets was used. Figure 5.4 shows the quantification of ABK298-1-E07-M binding to VuRca8α.

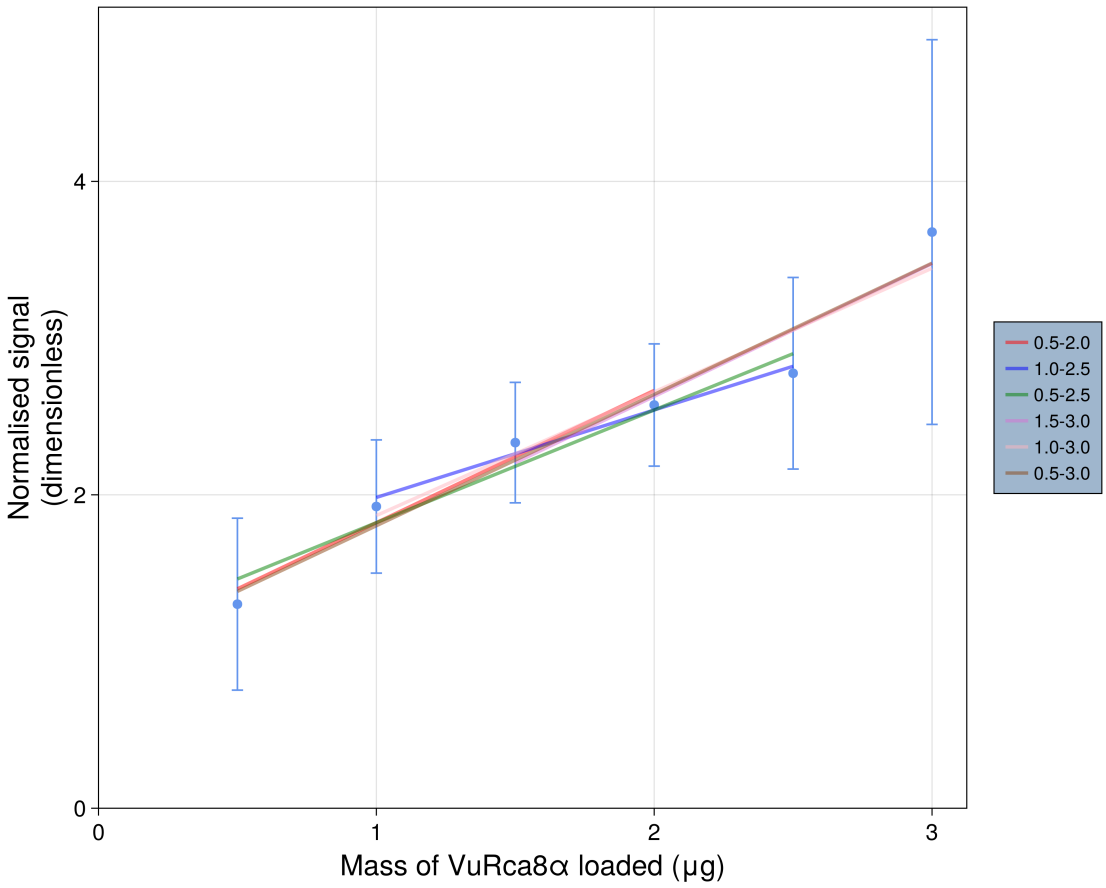


Figure 5.4: *Example quantification for antibody linear range analysis. 0.5-3.0 µg of purified VuRca8α was separated by SDS-PAGE and transferred to western blot. The blot was probed with VuRca8α-specific Abcalis antibody (ABK298-1-E07-M). Signal was normalised by fold increase over the minimum value quantified on each blot. Coloured lines indicate linear regression models applied to subsets of the data. Regression models include all data points spanned by the representative line. Error bars indicate mean ± standard deviation of n = 4 technical replicates.*

Several regression models applied to data subsets are similar, with a higher gradient for models including larger protein amounts. In addition to visual assessment, model statistics such as the coefficient of determination (R^2) and Akaike Information Criterion (AIC) can be used to assess model quality. Many models show a low R^2 due to the inherent variance in western blotting. These metrics are shown for every tested antibody in Table 5.3.

(a) *ABK300-1-B02-M | VuRca10α*

Range of protein loaded (µg)	R^2	AIC
0.5-2.0	0.81	8.91
1.0-2.5	0.79	6.48
0.5-2.5	0.89	7.74
1.5-3.0	0.82	8.5
1.0-3.0	0.89	7.31
0.5-3.0	0.94	7.15

(b) *ABK300-1-B02-M | VuRca10β*

Range of protein loaded (µg)	R^2	AIC
0.5-2.0	0.86	12.63
1.0-2.5	0.87	15.52
0.5-2.5	0.91	16.62
1.5-3.0	0.86	23.53
1.0-3.0	0.89	27.22
0.5-3.0	0.9	32.49

(c) *ABK298-1-E07-M | VuRca8α*

Range of protein loaded (µg)	R^2	AIC
0.5-2.0	0.59	21.65
1.0-2.5	0.37	22.38
0.5-2.5	0.58	29.38
1.5-3.0	0.34	38.37
1.0-3.0	0.45	43.57
0.5-3.0	0.59	49.53

(d) *ABK286-1-D12-M | Pan-Rca*

Range of protein loaded (µg)	R^2	AIC
0.5-2.0	0.73	28.6
1.0-2.5	0.52	44.0
0.5-2.5	0.66	50.51
1.5-3.0	0.39	64.17
1.0-3.0	0.5	74.69
0.5-3.0	0.59	84.85

(e) *ABK302-1-E07-M | Pan-Rca*

Range of protein loaded (µg)	R^2	AIC
0.5-2.0	0.64	29.33
1.0-2.5	0.47	37.33
0.5-2.5	0.67	41.01
1.5-3.0	0.36	63.35
1.0-3.0	0.43	74.3
0.5-3.0	0.53	83.8

Table 5.3: *Model quality statistics for antibody binding profiles. Linear models were applied to subsets of data quantifying western blot band intensity across a range of protein loaded, for different antibodies and their targets. For Pan-Rca antibodies, VuRca1β was quantified. n = 2 reps (ABK300-1-B02-M); n = 4 reps (others). Text coloured green indicates this model is the best performing, judged by the relevant metric. AIC: Akaike Information Criterion (lower is better); R^2: coefficient of determination (higher is better)*

As shown in Table 5.3, models can be of differing quality over different ranges. For example, ABK302-1-E07-M has increasing AIC when including more protein loaded, implying that 0.5-2.0 µg of protein is the best tested range for quantification of target. These model statistics can help in selecting an appropriate range of protein mass to load for quantification of Rca.

5.3.4 Cowpea Rca Can Be Identified in Leaf Extract

In addition to recombinant protein, it is desirable that antibodies will detect protein in leaf extracts. Figure 5.5 shows western blots with cowpea leaf extract, demonstrating that Rca can be identified.

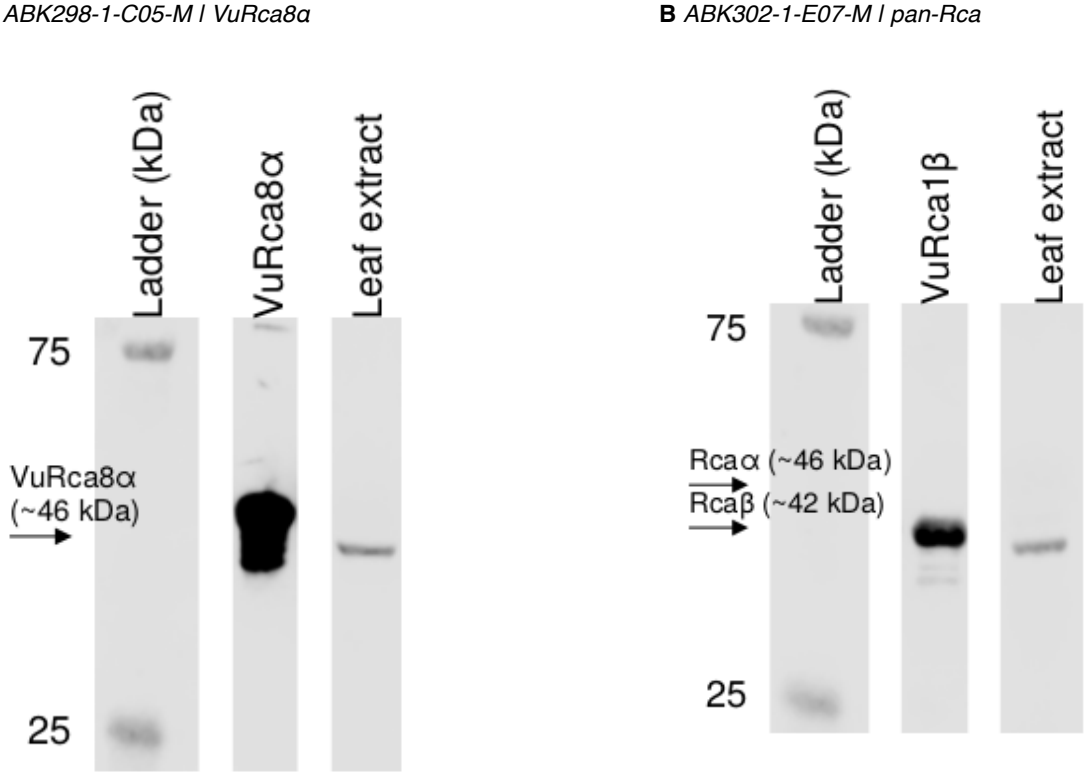


Figure 5.5: *Detection of cowpea Rca in leaf extract. 1 µg of recombinant Rca, and 4 µL of leaf extract from cowpea and were separated by SDS-PAGE. The proteins were transferred to western blot and visualised with the indicated Abcalis antibody.*

In Figure 5.5B, only one band can be seen, presumably due to the higher expression of VuRca1β compared to other VuRca isoforms. The bands run lower in leaf extracts than recombinant proteins, likely due to differences in buffer chemistry and total amount of protein loaded. 700 nm visualisation was omitted due to the high background in leaf extract at this wavelength.

5.4 Discussion

In this chapter, I describe the design process and validation of anti-cowpea Rca and anti-pan-Rca antibodies. These antibodies have been carefully validated for isoform specificity, using heterologously expressed Rca. All supplied clones were ELISA positive (data not shown) but some failed to bind their target in western blot. This may be due conformational differences in the denatured antigen used in western blotting, and the short peptide fragments produced for ELISA. An alternative hypothesis is the chemical differences between buffers used in these assays. All antibodies

were validated with 10 nM antibody concentration, and using 4% milk as a blocking reagent. It is advised that care be taken if changing these parameters as they can affect non-specific binding of antibodies (Pillai-Kastoori et al., 2020a). The linear range of selected antibodies has been tested with supplied model statistics. When using antibodies for quantification, it is important to quantify proteins only within the linear range of antibody binding. For pan-Rca, only VuRca1β binding was tested, the intended target should be used for quantification of other proteins.

With these antibodies, all cowpea Rca proteins except for VuRca1β and VuRca10β can be distinguished from each other on either the basis of size or of antibody specificity. As VuRca1β can be detected by pan-Rca antibody, the relative contribution of VuRca1β and VuRca10β can be estimated by use of both pan-Rca and VuRca10-specific antibodies. The ability to selectively identify and quantify individual isoforms of cowpea Rca enables questions such as:

- Does the protein abundance of cowpea Rca change between different cowpea accessions?
- Does the protein abundance of specific cowpea Rca isoforms change under conditions such as heat stress?
- Does the protein abundance of specific cowpea Rca isoforms change throughout the plant's life cycle (senescence)?
- How does the protein expression of cowpea Rca correlate with mRNA expression?

While these are interesting avenues of research in their own right, antibodies have uses beyond detection and quantification. For example, immunoprecipitation can be used to selectively purify proteins which antibodies bind to. This could be used to purify individual isoforms of native Rca directly from plants.

The choice to go recombinant with these antibodies was driven by the improved reproducibility, and the ethical concern of animal sacrifice in traditional antibody production. Despite the advantages afforded by recombinant antibodies, adoption remains low. Gray et al. (2020) identify that this is not due to intrinsic biochemical properties of recombinant antibodies. Instead, they point to social and economic reasons for the continued use of animal-derived antibodies, such as lack of knowledge of alternatives, and a propensity to stick with pre-existing methods. Likewise, Groff (2020) stress the need for recombinant antibodies to fit in with pre-existing protocols. Reformatting of recombinant antibodies simplifies this as they can be made in the same format as antibodies from whichever animal is needed. For example, the antibodies in this chapter are in the format of mouse IgG, allowing anti-mouse IgG secondary antibodies to be used.

VuRca8α and VuRca1β can be detected by western blot of cowpea leaf extract. Neither VuRca10 isoform was able to be detected in leaf extract (data not shown), presumably due to their low abundance. Concentration of leaf extract sample may enable detection of VuRca10 isoforms by western blot. Polyclonal antibodies hold the advantage of binding to multiple sites of their target. This allows for signal amplification and the detection of lower abundance targets. Mixing

of individual monoclonal antibodies can mimic the benefits of polyclonal antibodies and may allow for the detection of the low abundance VuRca10 isoforms.

The antibodies described herewith are very valuable tools for research into cowpea Rca. A reliable method for detection and quantification of individual isoforms of Rca enables new research questions to be answered which were not previously feasible. mRNA quantification can be used as a proxy for protein abundance, but protein and mRNA abundance rarely correlate 1:1 (Liu et al., 2016; Perdomo et al., 2021). Thanks to recombinant antibody production technology, these can be produced as needed, without worry of batch-to-batch variability. We hope this proves a valuable resource for future research, and provides an example to the photosynthesis research community on the benefits of animal-free antibodies.

5.5 Acknowledgements

The author acknowledges the help of Kilian Zilkens (Abcalis) in design and production of the antibodies used in this chapter. Non-tagged wheat Rca protein was purified by Dr. Dawn Worrall.

Chapter 5 Bibliography

Baker, M. (2015). Reproducibility crisis: Blame it on the antibodies. *Nature*, *521*(7552), 274–276. https://doi.org/10.1038/521274a

Bradbury, A. R. M., Trinklein, N. D., Thie, H., Wilkinson, I. C., Tandon, A. K., Anderson, S., Bladen, C. L., Jones, B., Aldred, S. F., Bestagno, M., Burrone, O., Maynard, J., Ferrara, F., Trimmer, J. S., Görnemann, J., Glanville, J., Wolf, P., Frenzel, A., Wong, J., Koh, X. Y., Eng, H.-Y., Lane, D., Lefranc, M.-P., Clark, M., & Dübel, S. (2018). When monoclonal antibodies are not monospecific: Hybridomas frequently express additional functional variable regions. *mAbs*, *10*(4), 539–546. https://doi.org/10.1080/19420862.2018.1445456

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2), 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Chi, X., Li, Y., & Qiu, X. (2020). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: Mechanism and regulation. *Immunology*, *160*(3), 233–247. https://doi.org/10.1111/imm.13176

Eaglesham, J. B., Garcia, A., & Berkmen, M. (2021). Production of antibodies in SHuffle Escherichia coli strains. In *Methods in Enzymology* (pp. 105–144). Elsevier. https://doi.org/10.1016/bs.mie.2021.06.040

Edfors, F., Hober, A., Linderbäck, K., Maddalo, G., Azimi, A., Sivertsson, Å., Tegel, H., Hober, S., Szigyarto, C. A.-K., Fagerberg, L., von Feilitzen, K., Oksvold, P., Lindskog, C., Forsström, B., & Uhlen, M. (2018). Enhanced validation of antibodies for research applications. *Nature Communications*, *9*(1), 4130. https://doi.org/10.1038/s41467-018-06642-y

Feige, M. J., Hendershot, L. M., & Buchner, J. (2010). How antibodies fold. *Trends in Biochemical Sciences*, *35*(4), 189–198. https://doi.org/10.1016/j.tibs.2009.11.005

Goodman, S. (2018). The antibody horror show: An introductory guide for the perplexed. *New Biotechnology*, *45*, 9–13. https://doi.org/10.1016/j.nbt.2018.01.006

Gray, A., Bradbury, A. R. M., Knappik, A., Plückthun, A., Borrebaeck, C. A. K., & Dübel, S. (2020). Animal-free alternatives and the antibody iceberg. *Nature Biotechnology*, *38*(11), 1234–1239. https://doi.org/10.1038/s41587-020-0687-9

Groff, K. (2020). Increasing the use of animal-free recombinant antibodies. *ALTEX*. https://doi.org/10.14573/altex.2001071

Hattori, T., & Koide, S. (2018). Next-generation antibodies for post-translational modifications. *Current Opinion in Structural Biology*, *51*, 141–148. https://doi.org/10.1016/j.sbi.2018.04.006

Hebbes, T. R., Turner, C. H., Thorne, A. W., & Crane-Robinson, C. (1989). A “minimal epitope” anti-protein antibody that recognises a single modified amino acid. *Molecular Immunology*, *26*(9), 865–873. https://doi.org/10.1016/0161-5890(89)90143-0

Lipman, N. S., Jackson, L. R., Trudel, L. J., & Weis-Garcia, F. (2005). Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources. *ILAR Journal*, *46*(3), 258–268. https://doi.org/10.1093/ilar.46.3.258

Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, *165*(3), 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Perdomo, J. A., Buchner, P., & Carmo-Silva, E. (2021). The relative abundance of wheat Rubisco activase isoforms is post-transcriptionally regulated. *Photosynthesis Research*, *148*(1), 47–56. https://doi.org/10.1007/s11120-021-00830-6

Pillai-Kastoori, L., Heaton, S., Shiflett, S. D., Roberts, A. C., Solache, A., & Schutz-Geschwender, A. R. (2020a). Antibody validation for Western blot: By the user, for the user. *Journal of Biological Chemistry*, *295*(4), 926–939. https://doi.org/10.1016/S0021-9258(17)49905-4

Robinson, M.-P., Ke, N., Lobstein, J., Peterson, C., Szkodny, A., Mansell, T. J., Tuckey, C., Riggs, P. D., Colussi, P. A., Noren, C. J., Taron, C. H., DeLisa, M. P., & Berkmen, M. (2015). Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nature Communications*, *6*(1), 8072. https://doi.org/10.1038/ncomms9072

Roth, K. D. R., Wenzel, E. V., Ruschig, M., Steinke, S., Langreder, N., Heine, P. A., Schneider, K.-T., Ballmann, R., Fühner, V., Kuhn, P., Schirrmann, T., Frenzel, A., Dübel, S., Schubert, M., Moreira, G. M. S. G., Bertoglio, F., Russo, G., & Hust, M. (2021). Developing Recombinant Antibodies by Phage Display Against Infectious Diseases and Toxins for Diagnostics and Therapy. *Frontiers in Cellular and Infection Microbiology*, *11*, 697876. https://doi.org/10.3389/fcimb.2021.697876

Taussig, M. J., Fonseca, C., & Trimmer, J. S. (2018). Antibody validation: A view from the mountains. *New Biotechnology*, *45*, 1–8. https://doi.org/10.1016/j.nbt.2018.08.002

QID: Q57091218

Taylor, T., Denson, J.-P., & Esposito, D. (2017). Optimizing Expression and Solubility of Proteins in *E. coli* Using Modified Media and Induction Parameters. In N. A. Burgess-Brown (Ed.), *Heterologous Gene Expression in E.coli* (pp. 65–82). Springer New York. https://doi.org/10.1007/978-1-4939-6887-9_5

Uhlen, M., Bandrowski, A., Carr, S., Edwards, A., Ellenberg, J., Lundberg, E., Rimm, D. L., Rodriguez, H., Hiltke, T., Snyder, M., & Yamamoto, T. (2016). A proposal for validation of antibodies. *Nature Methods*, *13*(10), 823–827. https://doi.org/10.1038/nmeth.3995

Voskuil, J. L., Bandrowski, A., Begley, C. G., Bradbury, A. R., Chalmers, A. D., Gomes, A. V., Hardcastle, T., Lund-Johansen, F., Plückthun, A., Roncador, G., Solache, A., Taussig, M. J., Trimmer, J. S., Williams, C., & Goodman, S. L. (2020). The Antibody Society's antibody validation webinar series. *mAbs*, *12*(1), 1794421. https://doi.org/10.1080/19420862.2020.1794421

Zaroff, S., & Tan, G. (2019). Hybridoma technology: The preferred method for monoclonal antibody generation for *in vivo* applications. *BioTechniques*, *67*(3), 90–92. https://doi.org/10.2144/btn-2019-0054

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L., Guan, Y., Dai, Z., & Yu, G. (2022). Ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics*, *23*(4), bbac222. https://doi.org/10.1093/bib/bbac222

Chapter 6

General Discussion

Overall, this project has four key outputs. An analysis of Rca sequences and climate data for diverse species has identified three species which may contain thermotolerant Rca - *Bituminaria bituminosa*, *Gleditsia triacanthos* & *Lupinus polyphyllus*. Optimisation of recombinant Rca expression and purification methods has improved Rca purity and increased yield many-fold. Characterisation of the temperature response of cowpea Rca isoforms has identified VuRca10α and VuRca10β as being more thermotolerant than VuRca1β and VuRca8α. Lastly, a collection of recombinant antibodies targeting specific isoforms of cowpea Rca, and non-specific Rca isoforms, were produced and validated, providing useful tools for further Rca research.

Rca is a thermolabile enzyme required for photosynthesis. It can limit rates of photosynthesis under conditions such as heat stress (Salvucci and Crafts-Brandner, 2004a) or fluctuating light (Mott and Woodrow, 2000), and as such is increasingly investigated as a target for crop improvement. Many, but not all plants express multiple isoforms of Rca. *Poaceae* contain a tandem duplication, and these isoforms differ in sequence and expression patterns. It is common for *Fabaceae* to have 3 or more Rca isoforms (Nagarajan and Gill, 2018). Having multiple isoforms of Rca allows for adaptation to environment, such as upregulation for a more thermostable isoform under heat stress (Degen et al., 2021; Ristic et al., 2009). A hypothetical thermostable Rca which is active across a broad temperature range is enticing as a crop improvement concept (Qu et al., 2023). More thermostable isoforms from wild relatives have been suggested to safeguard crop species from rising temperatures, such as the heat adapted *Oryza australiensis* for domesticated rice (*Oryza sativa*) (Scafaro et al., 2016). This concept is the guiding principle behind the bioinformatic analysis in Chapter 2. Comparison to known residues affecting thermostability, and analysis of the breadth of temperatures and climates in which investigated plants grow, identified candidate plants which may have an Rca isoform with a broad temperature optimum. As the species of interest in this study is a *Fabaceae* species, other *Fabaceae* were investigated in this analysis, as it is more likely these effects of identified residues will be shared in plants of more similar genetic backgrounds. In this chapter, three *Fabaceae* species were identified for their diversity of growing conditions - *Gleditsia triacanthos*, *Lupinus polyphyllus*, & *Bituminaria bituminosa*. Of the Rca in

those species, five amino acid residue sites were identified as being potentially related to temperature response. The following AA substitutions are shown in at least one of these three species, but not represented in VuRca: 220L, 220M, 241M, 399H, 399G, 422S, 435N, 435E. Only two of these residues shows diversity between VuRca isoforms. Compared to VuRca10, VuRca8α has amino acid polymorphism K399Q; and VuRca8α & VuRca1β have polymorphism K422N. These substitutions may contribute to the thermal stability of VuRca10 isoforms, and could be further explored by targeted mutagenesis to the amino acid polymorphisms identified above.

To characterise the catalytic properties of cowpea Rca, milligram quantity of pure Rca is required. When purification was first attempted, it became quickly evident that current methods would not reliably fulfil these requirements. Heterologous protein expression and purification is a large and rapidly developing field of study (Rosano et al., 2019), with rigorous optimisation studies being concluded for many proteins. However, to the author's knowledge, there is no thorough publication detailing optimisation of expression and purification of heterologously expressed Rca. Moreover, the full scope of factors influencing heterologous protein expression is often not reported in Rca literature. This makes reproduction inconsistent, as these unreported factors can affect final yield, as well as protein quality and purity.

In the course of expression and purification optimisation, several improvements were made. It was found that later than standard induction by IPTG, a change in expression medium, and 24 °C expression temperature greatly improves Rca expression. Regarding IMAC purification, a more stringent wash produced a more pure product with minimal effect on protein yield, and it was found that 300 mM imidazole was sufficient for protein elution. The optimised expression and purification protocol derived from this work enable purification of sufficient Rca for later experiments.

This kind of one-factor-at-a-time optimisation study can be time consuming and costly. Design of experiments is increasingly being turned to for optimisation of recombinant expression (Uhoraningoga et al., 2018). These multifactor designs allow for larger range of factors to be tested in a smaller time, and in a more reproducible manner. Further expression studies for Rca may wish to use design of experiments to facilitate more robust experiments with quicker turnaround.

Chapter 4 makes use of the improved Rca expression protocol to characterise the individual isoforms of cowpea Rca. Curiously, the investigated cowpea Rca isoforms did not seem to have an optimal concentration for per molar ATPase activity. This contrasts with data for Rca in tobacco (Keown et al., 2013) and spinach (Wang et al., 1993) in which ~1 µM of Rca was found to produce the highest per molar activity. This may imply that cowpea Rca ATPase activity is independent of oligomerisation state in this concentration range. An alternative hypothesis is that cowpea Rca does not change oligomerisation state under these conditions.

VuRca8 has a short splice isoform which is likely not a genuine activase. The function of this extra-short protein is not known, and to the author's knowledge has not been reported in any other genomes. Possible functions of this splice variant could include modulation of Rca activity by interactions with other isoforms, or as an ATPase with other functions. Rca is a member of the AAA+ protein family which is responsible for very diverse functions (Khan et al., 2022). It is

possible this splice variant has diverged into an entirely different purpose.

VuRca10α and VuRca10β proved to be more thermostable than VuRca1β and VuRca8α, and have a broader T_{80}, the range of temperatures across which activity is >80% of its maximum. This applies to both the ATPase and Rubisco reactivation activities, although ATPase activity has a higher temperature optimum. The chromosome 10 isoforms are likely an adaptive measure to high temperatures, as they are upregulated by temperature stress, and could be a target of interest for crop improvement. Upregulation of these isoforms may help in keeping photosynthesis active at higher temperatures.

The more thermostable isoforms - VuRca10α and VuRca10β - were challenging to purify, frequently having issues with solubility. These isoforms are also known to be expressed in very low concentrations *in planta*, demonstrated by the fact they cannot be detected in leaf extract by western blot. Proteins *in vivo* are known to often live "on the edge of solubility" (Vecchi et al., 2020) meaning that small changes to concentration, structure or milieu could drive a protein to insolubility. Taken together, this could imply that the solubility issues were a result of trying to express VuRca10 isoforms beyond their solubility limits. Genetic engineering efforts aiming to improve thermotolerance of cowpea via upregulation of VuRca10 isoforms may wish to focus on improving protein solubility. Tools exist to predict amino acid substitutions which improve protein solubility which may aid in this (Raimondi et al., 2020).

Despite their diversity in catalytic properties, Rca isoforms tend to be similar in structure and sequence. This makes disambiguation of similarly-sized Rca proteins challenging. One potential solution to this is the use of isoform-specific antibodies. Antibodies are vital tools in molecular biology research - with widespread uses including detection, quantification and isolation of their antigen. The fact that antibodies can be developed against a large array of molecules from proteins, to lipids and nucleic acids, makes them incredibly versatile. Despite this, antibodies are a serious source of irreproducible research. Academic literature is plagued with reports of antibodies not binding their intended target, or binding non-specifically. In order to produce trustworthy results, antibodies must be carefully validated. A survey (Freedman et al., 2016) found that almost 1/3 of early career researchers do not validate their antibodies. There have been efforts to provide guidelines for what comprises sufficient antibody validation (Edfors et al., 2018; Taussig et al., 2018; Uhlen et al., 2016), but given the many diverse uses of antibodies it is not possible to make a strict set of rules for this. Furthermore, antibodies were previously some of the least identifiable reagents used in research, although modern standards for reporting are changing this (Menke et al., 2020), in part due to the initiative to report RRIDs (Research Resource Identifiers) (RINL Resource Identification Initiative et al., 2016).

As an alternative to IMAC purification, antibodies against cowpea Rca could be used for Rca purification from leaf extract by immunoaffinity chromatography (Moser and Hage, 2010). This technique uses antibodies as the stationary phase in liquid chromatography to enable selective purification of the antibody's target. The primary advantage of this would be that the purified Rca would be native including all posttranslational modifications, unlike Rca expressed in *E. coli*. This

could enable research such as testing whether Rca activity changes as a result of factors such as environmental stressors or senescence. Additionally, immunoaffinity chromatography is a rapid technique, and faster purification may aid in the solubility issues that frequently arise in purification of VuRca10 isoforms.

Further to the use of antibodies for Rca purification, co-Immunoprecipitation (co-IP) is a technique used to identify and purify interactors of a protein, alongside the protein of interest (Lin and Lai, 2017). This could be used to identify Rca interactors and test their impact on the protein function. For example, Plant Natriuretic Peptide has been identified as an interactor (Turek et al., 2020) but the functional relevance of this is not currently known.

For validation of pan-Rca and isoform-specific cowpea Rca antibodies, pure, recombinant proteins were used. This can prove a lack of cross-reactivity between Rca isoforms. At least one antibody clone against each Rca target except for VuRca1β was successful. These valuable new tools for distinguishing between similarly-sized cowpea Rca isoforms enable new research questions to be answered. In addition to those mentioned above, some potential research applications for isoform-specific Rca antibodies are discussed in Chapter 5.4.

To summarise, this work provides insight into the diversity of Rca from cowpea. The production and validation of isoform-specific antibodies against cowpea Rca provides an invaluable and trustworthy tool for further research into this. The knowledge and tools presented herein lay important foundations for safeguarding cowpea against changing climates.

6.1 Discussion Bibliography

Degen, G. E., Orr, D. J., & Carmo-Silva, E. (2021). Heat-induced changes in the abundance of wheat Rubisco activase isoforms. *New Phytologist*, *229*(3), 1298–1311. https://doi.org/10.1111/nph.16937

Edfors, F., Hober, A., Linderbäck, K., Maddalo, G., Azimi, A., Sivertsson, Å., Tegel, H., Hober, S., Szigyarto, C. A.-K., Fagerberg, L., von Feilitzen, K., Oksvold, P., Lindskog, C., Forsström, B., & Uhlen, M. (2018). Enhanced validation of antibodies for research applications. *Nature Communications*, *9*(1), 4130. https://doi.org/10.1038/s41467-018-06642-y

Freedman, L. P., Gibson, M. C., Bradbury, A. R., Buchberg, A. M., Davis, D., Dolled-Filhart, M. P., Lund-Johansen, F., & Rimm, D. L. (2016). [Letter to the Editor] The need for improved education and training in research antibody usage and validation practices. *BioTechniques*, *61*(1), 16–18. https://doi.org/10.2144/000114431

IPCC. (2018). Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan,

R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.

Keown, J. R., Griffin, M. D., Mertens, H. D., & Pearce, F. G. (2013). Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Are Required for Biological Activity. *Journal of Biological Chemistry*, *288*(28), 20607–20615. https://doi.org/10.1074/jbc.M113.466383

Khan, Y. A., White, K. I., & Brunger, A. T. (2022). The AAA+ superfamily: A review of the structural and mechanistic principles of these molecular machines. *Critical Reviews in Biochemistry and Molecular Biology*, *57*(2), 156–187. https://doi.org/10.1080/10409238.2021.1979460

Lin, J.-S., & Lai, E.-M. (2017). Protein–Protein Interactions: Co-Immunoprecipitation. In L. Journet & E. Cascales (Eds.), *Bacterial Protein Secretion Systems* (pp. 211–219). Springer New York. https://doi.org/10.1007/978-1-4939-7033-9_17

Menke, J., Roelandse, M., Ozyurt, B., Martone, M., & Bandrowski, A. (2020). The Rigor and Transparency Index Quality Metric for Assessing Biological and Medical Science Methods. *iScience*, *23*(11), 101698. https://doi.org/10.1016/j.isci.2020.101698

Moser, A. C., & Hage, D. S. (2010). Immunoaffinity chromatography: An introduction to applications and recent developments. *Bioanalysis*, *2*(4), 769–790. https://doi.org/10.4155/bio.10.31

Mott, K. A., & Woodrow, I. E. (2000). Modelling the role of Rubisco activase in limiting Non-steady-state photosynthesis. *Journal of Experimental Botany*, *51*(suppl_1), 399–406. https://doi.org/10.1093/jexbot/51.suppl_1.399

Nagarajan, R., & Gill, K. S. (2018). Evolution of Rubisco activase gene in plants. *Plant Molecular Biology*, *96*(1), 69–87. https://doi.org/10.1007/s11103-017-0680-y

Qu, Y., Mueller-Cajar, O., & Yamori, W. (2023). Improving plant heat tolerance through modification of Rubisco activase in C3 plants to secure crop yield and food security in a future warming world (E. Carmo-Silva, Ed.). *Journal of Experimental Botany*, *74*(2), 591–599. https://doi.org/10.1093/jxb/erac340

Raimondi, D., Orlando, G., Fariselli, P., & Moreau, Y. (2020). Insight into the protein solubility driving forces with neural attention (E. Alexov, Ed.). *PLOS Computational Biology*, *16*(4), e1007722. https://doi.org/10.1371/journal.pcbi.1007722

RINL Resource Identification Initiative, Bandrowski, A., Brush, M., Grethe, J. S., Haendel, M. A., Kennedy, D. N., Hill, S., Hof, P. R., Martone, M. E., Pols, M., Tan, S. S., Washington, N., Zudilova-Seinstra, E., & Vasilevsky, N. (2016). The Resource Identification Initiative: A Cultural Shift in Publishing. *Neuroinformatics*, *14*(2), 169–182. https://doi.org/10.1007/s12021-015-9284-3

Ristic, Z., Momčilović, I., Bukovnik, U., Prasad, P. V. V., Fu, J., DeRidder, B. P., Elthon, T. E., & Mladenov, N. (2009). Rubisco activase and wheat productivity under heat-stress conditions. *Journal of Experimental Botany*, *60*(14), 4003–4014. https://doi.org/10.1093/jxb/erp241

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in *Escherichia coli* : A 5-year update. *Protein Science*, *28*(8), 1412–1422. https://doi.org/10.1002/pro.3668

Salvucci, M. E., & Crafts-Brandner, S. J. (2004a). Inhibition of photosynthesis by heat stress: The activation state of Rubisco as a limiting factor in photosynthesis. *Physiologia Plantarum*, *120*(2), 179–186. https://doi.org/10.1111/j.0031-9317.2004.0173.x

Scafaro, A., Gallé, A., Rie, J. V., Carmo-Silva, E., Salvucci, M. E., & Atwell, B. J. (2016). Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytologist*, *211*(3), 899–911. https://doi.org/10.1111/nph.13963

Taussig, M. J., Fonseca, C., & Trimmer, J. S. (2018). Antibody validation: A view from the mountains. *New Biotechnology*, *45*, 1–8. https://doi.org/10.1016/j.nbt.2018.08.002
QID: Q57091218

Turek, I., Gehring, C., & Irving, H. (2020). Arabidopsis Plant Natriuretic Peptide Is a Novel Interactor of Rubisco Activase. *Life*, *11*(1), 21. https://doi.org/10.3390/life11010021

Uhlen, M., Bandrowski, A., Carr, S., Edwards, A., Ellenberg, J., Lundberg, E., Rimm, D. L., Rodriguez, H., Hiltke, T., Snyder, M., & Yamamoto, T. (2016). A proposal for validation of antibodies. *Nature Methods*, *13*(10), 823–827. https://doi.org/10.1038/nmeth.3995

Uhoraningoga, A., Kinsella, G., Henehan, G., & Ryan, B. (2018). The Goldilocks Approach: A Review of Employing Design of Experiments in Prokaryotic Recombinant Protein Production. *Bioengineering*, *5*(4), 89. https://doi.org/10.3390/bioengineering5040089

Vecchi, G., Sormanni, P., Mannini, B., Vandelli, A., Tartaglia, G. G., Dobson, C. M., Hartl, F. U., & Vendruscolo, M. (2020). Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proceedings of the National Academy of Sciences*, *117*(2), 1015–1020. https://doi.org/10.1073/pnas.1910444117

Wang, Z. Y., Ramage, R. T., & Portis, A. R. (1993). Mg2+ and ATP or adenosine 5′-[γ-thio]-triphosphate (ATPγS) enhances intrinsic fluorescence and induces aggregation which increases the activity of spinach Rubisco activase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, *1202*(1), 47–55. https://doi.org/10.1016/0167-4838(93)90061-U

Chapter 7

Acknowledgements

Many thanks to my supervisors Elizabete Carmo-Silva and Marjorie Lundgren for their help, support and advice throughout the project, and for always giving me opportunities to pursue my interests. Thanks to Dawn, Peter, Joana, Alejandro, and everyone else in the photosynthesis group for being such a pleasure to work with and making such a welcoming environment. I would also like to thank all of my friends and family, especially my partner Ella, for their continued support and motivation. Finally, I would like to thank everyone who provided their valuable feedback on my research and my thesis, for helping to make it as good as it could be. All these wonderful people provided so much both in and outside of my research life that to properly acknowledge them would double the length of this thesis.