**PROJECT TITLE Bioprospecting plant proteomes for more efficient photosynthesis enzymes**

**AIMS AND BACKGROUND**

The goal of this project is to develop an entirely new way to screen hundreds of plants for enzyme variants that have atypical, and therefore potentially useful, kinetic characteristics. The resulting proteomics methods will be able to quickly screen any species of land plant, which will accelerate the development of new plant varieties across agricultural industries. In addition to innovative methodology, this project will also develop the new theory that amounts of plant enzymes relative to each other are very highly conserved across species.

**The fundamental question of this project is: In conserved metabolic pathways, does abnormal isoform abundance indicate atypical enzyme characteristics?**

The two-step approach, reflected in the project aims below, will first quantify patterns of photosynthesis enzyme abundance and, second, characterise in greater detail the outlier genotypes from those patterns. Although the methods could be applied to many different plant proteins, this project will focus on discovering atypical variants of two key photosynthesis enzymes, Rubisco and Rubisco activase.

**Aim 1: Determine the genetic and environmental drivers of variation in photosynthesis enzyme abundance across native Australian plants.**

**Aim 2: Test the hypothesis that unusual isoform abundance relative to metabolically linked enzymes are predictors of atypical enzyme activities.**

The photosynthesis proteins in 700 leaf samples from different genotypes of plantation-grown and wild *Eucalyptus camaldulensis* (river red gum) will be measured by mass spectrometry and compared against each other and over 1000 samples of other wild eucalypts and Australian native plants. The Rubisco and Rubisco activase from outlier genotypes will then be quantified and kinetically characterised at the isoform level.

**Background**

Finding new ways to improve crops is critically important because, at current rates, global food demand will exceed supply by 2050, and crop yields have stagnated1,2. Improving the efficiency of photosynthesis has high potential to increase crop yields because the leaf enzymes involved in carbon fixation are inefficient and past crop development has improved yields by other mechanisms1,3–5.

Rubisco, the enzyme that fixes CO2 into energy storing metabolites, is enzymatically inefficient because it evolved under ancient atmospheric conditions with higher CO2 and lower O2 than today5,6. Under current conditions its desirable carboxylation reaction competes with an oxygenation reaction that wastes up to >30% fixed carbon3. Rubisco is a widely recognised target for improving crop productivity because increases in Rubisco carboxylation activity have the potential to increase crop yields 15-30%2–4,7,8.

The catalytic properties of Rubisco are diverse across photosynthetic organisms and exploiting that diversity is one of the main approaches of photosynthesis improvement4,6,9–12; *e.g.*, by recombinantly expressing a Rubisco with high specificity for carboxylation over oxygenation. Evergreen plants native to hot and dry environments are promising sources of Rubisco variants with desirable characteristics6,12, suggesting that Australia’s native diversity is likely a rich source of favorable Rubisco variants.

*In vivo* Rubisco catalytic properties depend both on the Rubisco variant and on a second enzyme, Rubisco activase, that interacts with Rubisco to maintain its active state. Rubisco activase is a target for photosynthesis improvement, particularly for increasing crop heat tolerance, because its activity is critical to carbon fixation and it is highly temperature sensitive13–17. Recombinant thermostable Rubisco activase increases *Arabidopsis* growth at high temperatures18. Faster growth at high temperatures by Australian wild rice versus cultivated rice has been attributed to a thermotolerant Rubisco activase variant in wild rice19, demonstrating the potential for discovering important photosynthesis enzyme variants among Australian native plants.

This proposal addresses the need for a faster way to identify potentially useful enzyme variants from among hundreds of genotypes. Rubisco catalytic properties have been measured in less than 0.025% of plants6, and none of the largest kinetics studies have exceeded 100 variants6,8,10,12,20,21, most likely because *in vitro* assays are laborious. Far fewer Rubisco activase variants have been characterised. Also, activity *in vitro* might not reflect *in vivo* activity8,22–24.

The proposed two-step screening strategy will utilize absolute protein quantification by combining two proteomics methods that have been extensively used in medical research, SWATH25,26 and QconCAT27, but have had limited use in plant research28,29. In step 1, photosynthesis enzymes will be quantified across genotypes by focusing on highly conserved peptides, thereby avoiding interference from unknown genetic variation. Conserved enzyme abundance ratios across 1700 genotypes, and how environmental variables affect them, will be established in order to detect outliers suggestive of abnormal enzyme activity (Aim 1). In step 2, the specific isoforms of Rubisco and Rubisco activase from outlier genotypes will be quantified using unique peptides derived from RNA sequence data. Then the catalytic properties of selected isoforms will be compared to isoform abundance data to test the hypothesis of Aim 2.

**PROJECT QUALITY AND INNOVATION**

This project will innovate in three ways:

1. *It will produce new knowledge about the influences of genetics and environment on the absolute amounts of key photosynthesis enzymes and their isoforms.* Recently developed, rapid, absolute protein quantification methods will be used across 120 genotypes of the important plantation timber species *Eucalyptus camaldulensis*, 31 additional eucalypt species, and 70 species of *Acacia* and Proteaceae.
2. *It will quantify the correlations between* in vivo *enzyme amounts and* in vitro *enzyme kinetics parameters.* A subset of Rubisco and Rubisco activase isoforms will be selected based on their abundance relative to other metabolically linked enzymes, and their catalytic parameters measured *in vitro* to establish links between *in vivo* isoformamounts and their kinetic properties.
3. *It will establish new proteomics methods, applicable across all vascular plants, for rapidly identifying enzyme isoforms potentially beneficial to crop plant improvement from among hundreds of genotypes.*

Crop scientists are limited in their ability to screen large populations for beneficial enzyme isoforms because existing approaches are time consuming or they imprecisely resolve consequential genetic differences from phylogenetic noise. This project will address those problems by using new proteomics methods I have developed to rapidly find phenotypically consequential isoform variants among large numbers of samples.

The ability to screen large numbers of genotypes by absolute enzyme quantification will complement and improve upon genomic and transcriptomic approaches because physiology is a product of enzyme kinetics and amounts, which are, at best, imprecisely reflected in sequencing data. My approach aims to identify important genetic variants much faster than reverse or forward genetics, or enzyme kinetics assays, based on the assumption that *amounts* of photosynthesis enzyme isoforms reflect their activity. It uses proteomics to bridge sequence data and physiology: enzyme sequence links to genotype, enzyme amount links to phenotype.

**Theoretical framework**

Two key assumptions underpin this project:

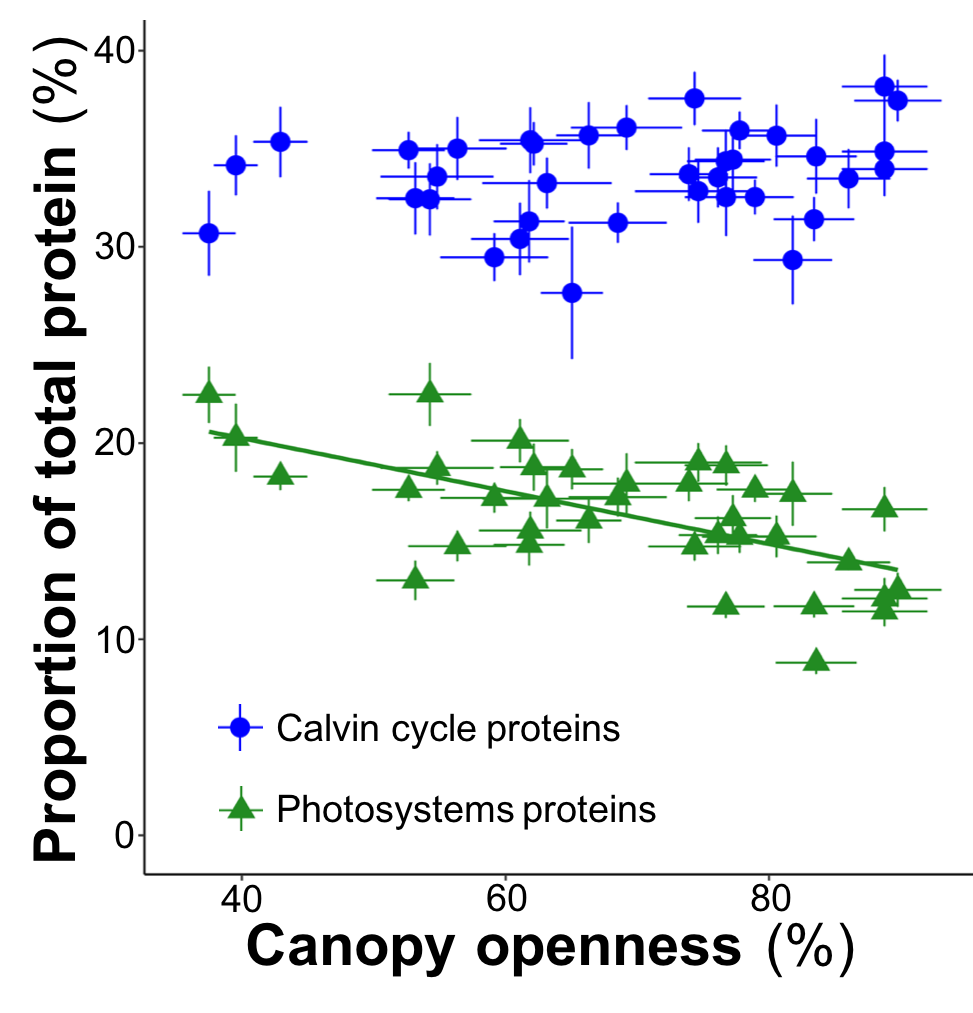
1. *Amounts of photosynthesis enzymes are coordinated so that there are no bottlenecks under average conditions of photosynthesis.* Metabolic flux is governed by collections of enzymes and no one enzyme amount limits that flux30,31, including Rubisco32. This assumption is supported by large-scale physiology and modelling experiments that demonstrate a coordination between electron transport by the light-dependent reactions (*J*max) and carbon fixing by Rubisco (*V*cmax) so that neither is limiting33–36. The fluxes of ATP and NADPH from the light-dependent reactions occur at rates coordinated with their consumption in the Calvin cycle, thus maximising carbon assimilation per total investment in photosynthesis proteins35. At times the fluxes might be out of balance (*e.g.*, low light, temporary drought), but on average they are coordinated. Photosynthesis protein amounts are not adjusted to environmental conditions instantaneously, but instead reflect environmental conditions for approximately the preceding month33,34.
2. *Amounts of Calvin cycle enzymes relative to each other are conserved, but vary predictably with environmental conditions, leaf traits, and whole plant traits.* Large-scale cross-species conservation of photosynthesis protein amounts relative to each other has not been previously studied, most likely because of technological limitations37,38. However, amounts of enzymes common to all life are more conserved than amounts of corresponding transcripts across diverse taxa (fly, nematode, yeast, primate, bacteria, rice), demonstrating that protein amounts are highly constrained by evolutionary pressure39–41. Conserved enzyme activity levels differ across *Arabidopsis* genotypes, but those differences covary among related enzymes37. Amounts of photosynthesis metabolites are conserved across vascular plants and vary predictably with physiological measurements, with some interspecific differences42. Leaf biochemistry responds similarly to changing environmental conditions across plant species, but with varying amplitude38. These observations suggest that photosynthesis enzyme amounts relative to each should be conserved across vascular plants, but also vary around those conserved relationships for genetic and environmental reasons.

Combining the assumptions above provides the basis for the experimental approach through the following conclusions:

1. It should be possible to find trends in amounts of Calvin cycle enzymes across multiple species that vary predictably with environmental conditions and plant traits.
2. Photosynthesis isoforms with atypical kinetic properties inherently contribute differently to the overall flux of photosynthesis metabolites than average isoforms.
3. In order to keep the flux of metabolites coordinated, amounts of kinetically abnormal isoforms, or their metabolic neighbors, must fall outside conserved trends.

*Approach summary*: Conserved trends in photosynthesis enzyme abundance across different levels of genetic diversity will be characterised by regression models of enzyme amounts over ranges of environmental variables. Isoform outliers from the trends will be identified and characterized in kinetics experiments.

Figure 1 contains preliminary data from my current Discovery Project that demonstrates a predictable trend38 in photosynthesis proteins over a gradient of light availability in 32 eucalypt species sampled across Eastern Australia: proportional protein investment in light capture decreases with increasing light. The lack of concomitant decline in Calvin cycle proteins supports the assumption that metabolic flux is coordinated among the enzymes of photosynthesis.



**Figure 1.** Protein amounts in photosystems, as proportions of total leaf protein, decreases with increasing light (R2 = 0.39; p < 10-12). Calvin cycle proteins were not correlated with light environment (p = 0.14). Points are species x site means, 9 measurements per point (3 plants x 3 leaf ages, bars = S.E).

Photosystem proteins are not universally expressed at maximal levels for two reasons: 1) A high flux from the photosystems that exceeds the capacity of the Calvin cycle leads to photoinhibition, which is energetically costly34; 2) Allocation is optimised to make the most of a scarce supply of nitrogen among nitrogen-rich proteins35,36 and protein turnover is a major metabolic expense43. The second reason is applicable to all leaf proteins and supports the coordination assumption above.

Figure 1 also demonstrates that, although there is a significant trend, there is still substantial scatter around that trend, which is the case for most of the protein-environment relationships in my Discovery Project data. Adding additional predictors (*e.g.*, precipitation, temperature, leaf mass per area) in multiple regression models often explains protein variation from 50% to 70%. But, that leaves 30% to 50% unexplained variation, much of it interspecific, which is a major motivator for the proposed project.

It is practically a circular argument to say that interspecific protein amount variation is caused by genetic differences between species, but that is the case. Those genetic differences sometimes encode species-specific isoforms with potentially atypical kinetics characteristics that could explain interspecific variation in protein abundance. Partitioning the variation in abundance of key photosynthesis enzymes among genetic and environmental predictors is Aim 1 of this proposal. Aim 2 tests the idea that outliers in abundance have atypical kinetic properties.

**Experimental approach**

The proposed project will determine the genetic versus environmental contributions to variation in enzyme amounts of key photosynthesis proteins, down to the isoform level for Rubisco and Rubisco activase. To do so it will make comparisons across:

1. 120 genotypes of *E. camaldulensis* subsp *camaldulensis* in a common garden plantation of 422 genotypes at the Hawkesbury Institute for the Environment (HIE, Richmond, NSW). The collection is part of a $4m Science and Industry Endowment Fund (SIEF) collaboration among ANU, HIE, and CSIRO; the lead CI is Prof Graham Farquhar. The genotypes were sourced from across the natural range of *E. camaldulensis* in Southeastern Australia at 26 provenances. There are 4 families (mothers) per provenance, 5 seedlings per mother, and several clones per seedling. Physiology and RNAseq data will be provided by Prof David Tissue and Dr Paul Rymer (HIE).
2. *E. camaldulensis* mother trees at 18 of the 26 provenances for the HIE garden, sampled from across Southeastern Australia.
3. 32 species of eucalypts (*Eucalyptus*, *Corymbia*, *Angophora*) from my current Discovery Project sampled from Eastern Australia across two large environmental gradients: mean annual temperature from 5 °C to 27 °C; mean annual precipitation from 200 mm to 3600 mm. All the samples have already been collected and analysed by the same proteomics methods that will be used in this proposed project.
4. 27 species of *Acacia* and 43 species of Proteaceae from my current Discovery Project. The *Acacia* samples have already been analysed; the Proteaceae will have been analysed by May 2017.

The two Aims of this proposal both contain two hypotheses. The hypotheses for Aim 1 are:

*Hypothesis 1A:* *Rubisco and Rubisco activase isoform sequence diversity does not vary substantially across* E. camaldulensis *provenances.* This hypothesis is based on previous genomic analysis of the same population, which found that most genetic diversity existed within individuals; i.e. they are highly heterozygous44. There is high gene flow across the population and little genetic structure across provenances. Hypothesis 1B is based on a contrasting result from a physiology experiment with genotypes from the same population.

*Hypothesis 1B: Rubisco and Rubisco activase isoforms vary in abundance across genotypes and environments.* Glasshouse grown *E. camaldulensis* from the HIE collection have demonstrated significant physiological differences across 14 genotypes from 6 provenances45. The differences in physiology, in spite of nearly insignificant genetic trends across the population44, suggest that protein isoforms are differentially expressed across the genotypes.

Understanding isoform abundance variation across those scales will create fundamental knowledge of how phenotypic variation arises across species and communities. It will also inform future experimental designs by answering the question: *How much does protein variation across genotypes in controlled environments reflect variation in the wild?* In combination with enzyme kinetics measurements it will answer the question: *Can genetically driven isoform variation among wild plants be separated from confounding environmental effects?* This second question relates to the hypotheses of Aim 2:

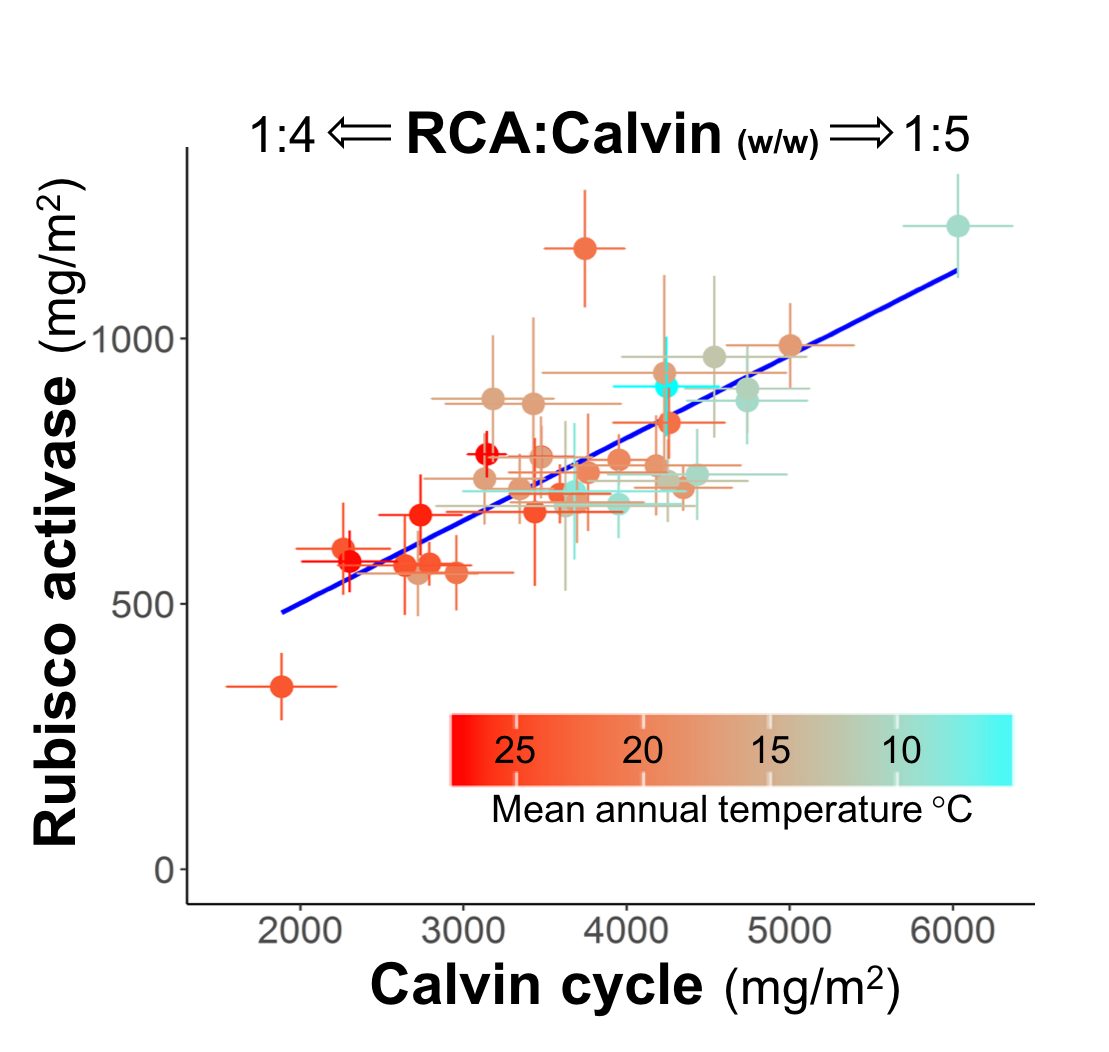
*Hypothesis 2A: Amounts of photosynthesis enzymes relative to neighboring enzymes are highly conserved, but vary predictably across environmental gradients.*

*Hypothesis 2B: Outliers of trends from Hypothesis 2A predict atypical isoform enzyme characteristics.*

Figure 2 demonstrates how amounts of photosynthesis enzymes can be conserved relative to each other, how they can vary predictably across environmental gradients, and how outliers can be detected. In Figure 2, Rubisco activase and Calvin cycle proteins are strongly correlated. Because temperature significantly explains more variation in Rubisco activase than Calvin cycle protein amounts alone, a multiple regression model is more appropriate for detecting outliers than the model in Figure 2, but Figure 2 is shown with one predictor so that it can be visualised easily. The proposed project will use multiple regression models and mixed models to identify trends and outliers.

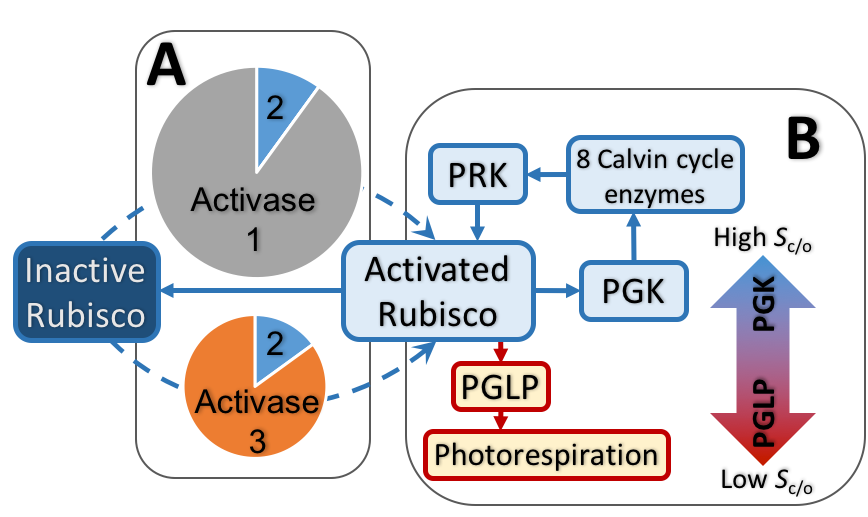
Figure 3 outlines the strategy for identifying isoforms with atypical activity. The strategy differs for Rubisco and Rubisco activase—Rubisco activase kinetics will be inferred from *its* amount relative to neighboring enzymes; Rubisco specificity will be inferred from amounts of neighboring enzymes, for example, PGLP and PGK in Figure 3.

In preliminary results, PGLP and PGK are strongly correlated (R2 = 0.72, not shown) and the ratio of PGLP:PGK ranges from 1:10 at high temperatures to 1:15 at low temperatures, which is consistent with published observations of decreasing Rubisco specificity and increasing photorespiration with increasing temperature8. Preliminary multiple regression models have shown that temperature, precipitation, and light environment are all significant predictors of PGLP:PGK. Once outliers for Rubisco and activase are identified from isoform quantification, then they will be used in kinetics experiments to test Hypothesis 2B and link enzyme kinetics with isoform amounts.



**Figure 2.** Per leaf area amounts of Rubisco activase (RCA) vary predictably with amounts of total Calvin cycle enzymes (R2 = 0.62) in wild eucalypts from broad temperature and precipitation gradients. Ratios vary from 1:4 at low protein amounts to 1:5 at high protein amounts. There is also a trend in protein amounts with temperature, which corresponds with a strong trend of decreasing total protein with increasing temperature (R2 = 0.50, not shown).

There are two clear outliers from the RCA:Calvin cycle trend, which suggests those genotypes potentially have RCA isoforms that differ kinetically from average eucalypt RCA. Points are means (bars = S.E.) for 36 species x site observations, 32 eucalypt species (4 species sampled at 2 sites), 9 samples each, 3 plants x 3 leaf ages.



**Figure 3.** Enzyme isoform activity will be inferred from protein abundance in two ways: A) by directly measuring the amount of the enzyme of interest relative to its metabolic neighbors (Rubisco activase); B) by indirectly inferring activity from neighboring enzymes in contrasting metabolic pathways (Rubisco oxygenation versus carboxylation activity).

**A:** Pie charts represent different hypothetical activase compositions in two different species. The top species has two isoforms, 1 & 2, the bottom species 2 & 3. A comparison of total activase amounts, relative to neighboring enzymes, would predict that the bottom species has a more efficient isoform. Isoform quantification would identify Activase 3 as the most likely explanation for the difference in total activase amount and, therefore, likely a more efficient isoform than isoform 1.

**B:** The efficiency of Rubisco is a product of several kinetic characteristics and activation states, which confound direct inferences based on Rubisco amount. Therefore, this study will make indirect inferences about Rubisco specificity (Sc/o) for carboxylation reactions over oxygenation reactions. Ratios of neighboring enzymes from the subsequent Calvin cycle versus photorespiration pathways will be used as indicators of Rubisco specificity. **Shaded arrow:** Genotypes with higher than average ratios of PGLP (phosphoglycolate phosphatase) to PGK (phosphoglycerate kinase) would be predicted to have lower Rubisco specificities and vice versa.

**Analytical methods**

*Protein extraction and mass spectrometry sample preparation.* Leaf protein will be phenol extracted46, with modifications: use of ovalbumin as an internal standard; leaf powder solvent washes before phenol extraction; phenol extraction at 80 °C, followed by two re-extractions. Precipitated protein will be dissolved in urea-SDS buffer, alkylated, assayed (FluoroProfile kit, Sigma) and 50 μg re-extracted by modified methanol/chloroform extraction47, then Lys-C and trypsin digested in a Rapigest (Waters) buffer.

*Mass spectrometry*. Samples will be analysed on a Sciex TripleTOF mass spectrometer using a SWATH method25 with a 60 min acetonitrile gradient. A reference ion library will be created by the same method, but in data dependent acquisition (DDA) mode for representative samples of each genotype. The DDA results will be matched, allowing for amino acid mutations (ProteinPilot, Sciex), to protein sequences from *E. grandis*48, chloroplast49, and mitochondrial sequences (Uniprot). Functional annotations will be made by Mercator50.

*RNAseq*. Predicted protein sequences for enzymes of interest will be added from RNAseq data provided by Dr Paul Rymer and from RNAseq data for 12 additional eucalypt species I have studied previously. 50 mg leaf samples, already on hand, will be used for RNA extraction51; cDNA library creation and Illumina NextSeq sequencing will be done by the ANU Biomolecular Resource Facility.

*Absolute protein quantification*. A synthetic protein (QconCAT) already on hand contains isotope labelled peptides conserved across all vascular plants (57 peptides from 34 proteins). It will be added during protein digestions to provide internal standards27,29.

SWATH data will be exported to R and proteins with QconCAT standards will be quantified by comparing sample peptide areas to the QconCAT peptide areas. Amounts of proteins not represented in the QconCAT will be estimated from SWATH data by comparison to ovalbumin using a top 2/top 2 peptide/ion method to minimise the effects of different ionisation efficiencies across proteins52.

*Isoform quantification*. A second QconCAT, manufactured by PolyQuant, will contain ~60 peptides that differentiate Rubisco and Rubisco activase isoforms among the genotypes in this project. The ~700 *E. camaldulensis* samples will be analysed by SWATH before the creation of the isoform QconCAT, but the SWATH data will be re-interrogated28,53 to calculate isoform abundance based on QconCAT calibration curves created from a subset of representative samples. In that way isoform abundance will be calculated retrospectively without having to re-run all 700 *E. camaldulensis* samples and other eucalypt samples for which I already have SWATH data.

*In vitro measurements of Rubisco catalytic parameters* (*V*cmax, *V*omax, *K*c, *K*o, *k*catc, *k*cato, *S*c/o). Rubisco will be extracted from leaves of 18 species/genotypes using the *Eucalyptus*-specific method of Sharwood *et al.*54. The following will be measured at 25 °C, 35 °C, and 45 °C by 14CO2 fixation assays by varying [CO2] and [O2]21,54: *V*cmax and *V*omax, the substrate saturated turnover rates for CO2 and O­2, respectively; *K*c and *K*o, Michaelis constants (*K*m) for CO2 and O­2.

Rubisco active sites will be measured by [14C]carboxyarabinitol-P2 binding55. The substrate turnover numbers, *k*catc and *k*cato will be calculated as either *V*cmax or *V*omax divided by the number of Rubisco active sites, respectively.

The CO2/O2 specificity (*S*c/o) of purified Rubisco54,55 will be determined at 25 °C, 35 °C, and 45 °C by measuring radio-labelled carboxylation and oxygenation products of [1-14C]RuBP by HPLC and scintillation counting56.

Rubisco integrity in the enzyme extracts will be analysed by SDS-PAGE. Rubisco isoform compositions in the enzyme extracts will be quantified by SWATH including isotope labelled internal peptide standards.

*Rubisco activase (RCA) isoform activities.* Twenty Rubisco activase isoforms will be expressed in *E. coli* using commercially synthesised genes (GeneArt). The rates at which they activate Rubisco will be quantified at four temperatures. His-tagged RCA-ubiquitin fusion proteins will be purified by nickel affinity chromatography and the ubiquitin removed enzymatically57–59. Purified RCA isoforms will be quantified by protein assays and by SWATH.

RCA activity and temperature response will be determined by *in vitro* Rubisco activation across several temperatures60,61. Purified Rubisco from the eucalypt Rubisco *S*c/o measurements and from *Arabidopsis* will be inactivated by pre-incubation with RuBP (forming Rubisco ER). The rate at which RCA activates Rubisco ER will be determined radiometrically by measuring 14CO2 incorporation into acid stable products over time60.

*Statistical analysis.* Relationships among isoform abundance, enzyme kinetics, environmental conditions, and genotypes will be evaluated in R using linear regression, multiple linear regression, and linear mixed models.

**DECRA CANDIDATE**

I will be based at the ARC Centre of Excellence for Translational Photosynthesis, working full time. I have allocated 0.8 FTE to this project because I expect to receive additional funding to extend my proteomics methods to crop research at ANU. I have experience raising funds in both applied and basic research, >$1m since my PhD (2012) in total. I will work closely with Prof John Evans, A/Prof Spencer Whitney, DECRA fellow Dr Rob Sharwood, and Dr Adam Carroll (Manager of the ANU Joint Mass Spectrometry Facility). All provided input into this proposal; Sharwood and Carroll have helped with my current Discovery Project.

I have experience in all aspects of the project except RNA extraction, which is a well-established method and expertise is readily available at ANU. My experience in the following areas is outlined in Parts C7 and C9 of this proposal: quantitative plant proteomics, enzyme purification and kinetics, recombinant protein expression, ecological experimental design and field work, applied research outcomes, fund raising, and project management.

**FEASIBILITY**

I am confident the four hypotheses of this project will be answered conclusively because the experimental design, which makes comparisons across 1700 genotypes, is extremely statistically powerful. The proteomics methods have already detected trends and outliers across eucalypt species as illustrated in Figure 1, specifically with Rubisco activase as illustrated in Figure 2, and with photorespiration versus Calvin cycle enzymes (not shown).

The key uncertainty is captured by the fundamental question of this proposal—whether or not isoform abundance indicates atypical enzyme activity. Using enzyme activities as proxies for protein amounts is commonplace in plant research37,38 and estimates of protein amounts based on total leaf nitrogen and photosynthesis measurements underpin models from the scale of the cell to the Earth’s surface32–36. At its simplest, this proposal is about making those assumptions in the other direction because the technology now exists to accurately measure many proteins simultaneously, which is much faster than enzymatic assays and adds the important capability of isoform discrimination. Importantly, I do not aim to predict enzyme characteristics quantitatively from protein amounts—which is likely not feasible in the short-term. Rather, my intention is to use protein amounts as signals for detecting atypical enzyme activity, potentially matched to qualitative predictions from protein abundance data.

Meeting project goals within the budget and timeline (below) is feasible because I base them on costs and time requirements of my current project, which has already used the same methods to analyse more samples than will be analysed in the proposed project. Time requirements for enzyme assays are based on estimates by Sharwood and Whitney, leading experts in the area who will assist me with assays.



**Research resources**

*Plant material*. All plant material will be collected from either the Hawkesbury Institute for the Environment SIEF garden of *E. camaldulensis* genotypes, or the field. The SIEF project was initially funded from 2014-2018, but the garden will be maintained through 2020, supported by a NSW Environmental Trust Grant (CIs: Sharwood, Tissue, Farquhar). The greatest risk to that site is fire, but it is split into multiple blocks with genotypes randomized across blocks, so it is unlikely all genotypes could be lost simultaneously. There are multiple options for field sampling—if fire affects a few sites, then other sites can be used.

*Mass spectrometry*. ANU and CSIRO have a reciprocal agreement for mass spectrometry resources so that I will be able to use instruments at either facility for the same price. Proteomics mass spectrometers include three highly capable Thermo instruments and one Sciex TripleTOF that will be used for this project. If the TripleTOF were to become unavailable indefinitely, then the University of Sydney has three TripleTOFs within driving distance from ANU that I currently use.

**Research environment**

The existing environment for photosynthesis research at ANU is world-class—it is difficult to imagine a better environment for this project. The project does not duplicate existing research at ANU, or anywhere else that I know of, and it complements existing research at ANU very well. Specifically, the ARC Centre of Excellence for Translational Photosynthesis at ANU was created to find new ways to improve photosynthesis in crop plants, which provides an immediate pathway of adoption for my methods. I will also be a member of the ANU Research School of Biology, which has world leaders in the field of photosynthesis research from the past 40+ years, as outlined in Part D2, the statement of support from the School.

**BENEFIT**

*Output 1: Quantitative knowledge of how genetic versus environmental variables affect photosynthesis enzyme amounts and variation.* These types of data currently exist only for only a small number of species. Also, the measurements are largely based on proxies for protein amounts, not the actual proteins, which introduces substantial inaccuracies38. Therefore, Output 1, as a high-impact publication, will be original in its methodology, its accuracy, and its scope.

*Benefit of Output 1:* Photosynthesis models from the scale of single cells, to leaves, to the entire photosynthetic capacity of Earth rely on estimates of nitrogen allocation among different functions of photosynthesis based knowingly on oversimplified and overgeneralised treatment of the individual enzymes32–35. This output will make possible more accurate, systems biology-based models on all those scales by filling in an entire missing category of key information—actual measured protein amounts.

*Output 2: Two demonstrations, using Rubisco and Rubisco activase, that enzyme and isoform amounts can be used as indicators for enzyme activity in cross-genotype comparisons.*

*Benefit of Output 2:* The findings specific to Rubisco and Rubisco activase will be beneficial for crop development because they increase knowledge of isoform variation and activity for two of the most important enzyme targets for photosynthesis improvement. The theory behind the output is significant because it could be applied to screens at different scales and with different methods than this project. The output will benefit metabolic engineering approaches for improving crop plants and it could also be applied to industrial microbes.

*Output 3: Analytical and computational methods will be published in a high-impact methods journal.*

*Benefit of Output 3:* Absolute protein quantification is becoming increasingly common in plant research, but not across multiple species or on hundreds of samples at a time because of technological limitations that my methods overcome. My existing methods are already unique because the internal standard (QconCAT) can be used with any species of vascular plant, the mass spectrometry method is very fast, and the cost per sample is low (1/3 the cost of whole transcriptome sequencing). Extending the methods to identifying potentially beneficial isoforms will dramatically increase the adoption of my methods because they will provide a fast and informative new capability to crop researchers. Crop development will benefit through faster discovery of new beneficial isoforms.

**Benefit to Australia**

This project addresses the National Science and Research Priority Food because it aims to create a new technology and new theoretical framework that will accelerate the development of improved crops that are higher yielding under emerging conditions. It does so by leveraging existing research capabilities in a native plantation timber species, *E. camaldulensis*, the red river gum, a national icon that is economically and environmentally important.

The benefit of Output 1, the facilitation of more accurate vegetation models, fits with current Australian research strengths in the Science and Research Priority Environmental Change.

**Cost-effectiveness and value**

ANU’s cash contribution is 36% of the requested funding from ARC. This project leverages substantial previous investments by the ARC, SIEF, ANU, and CSIRO:

1. *The ARC Centre of Excellence in Translational Photosynthesis*, administered by ANU. Methods and results will also be applicable to research at the ARC CoE in Plant Energy Biology, which includes ANU scientists.
2. *The SIEF Forests for the Future project* (lead CI Farquhar, $4m over 5 years) that will provide common garden-grown *E. camaldulensis* genotypes, with additional support from the NSW Environmental Trust.
3. *My current ARC Discovery Project and previous SIEF fellowship*. Most of the analytical methods for the proposed project were developed during my current Discovery Project and past SIEF fellowship. The proposed project will leverage data from the Discovery Project by re-analysing it.
4. *ANU and CSIRO research infrastructure in mass spectrometry and bioinformatics*, including the new ANU-CSIRO Centre for Genomics, Metabolomics, and Bioinformatics, which aims to facilitate the integration of gene discovery with crop deployment and environmental management.

**COMMUNICATION OF RESULTS**

Aim 1 will produce basic knowledge about plants across 102 species and most of Earth’s vegetated climate space, which are unprecedented scales for plant protein data. Results will be of interest to plant physiologists, ecologists, and climate scientists. For those reasons, I will target high-impact generalist journals such as *Nature* and *PNAS*.

Aim 2 will be targeted at leading plant journals such as *Nature Plants*, *Plant Physiology*, and *Plant Cell and Environment*. The theoretical framework behind this project will be applicable to other fields and I will promote it through a commentary article, such as the minireview format of *Molecular and Cellular Proteomics*.

I will promote my work through seminars in Canberra and Sydney. Since 2014 I have organised the Sydney Plant Ecophysiology group meetings, a seminar series across Sydney universities, and I will promote my research at organisations connected with that group. In year 3 of the project I will present its results at the meeting of the American Society for Mass Spectrometry and I will arrange to present it at one or more research institutions during the same trip.

**MANAGEMENT OF DATA**

All data will be synced with local drives using purchased Google Drive storage. All lab and field notebooks will be scanned regularly and stored in Google Drive. Mass spectrometry data will be uploaded to ProteomeXchange62,63, which will make it available through several online resources such as PRIDE, Peptide Atlas, and GPMDB. Calculated protein amounts, geographical and environmental data will be available as electronic supplementary material associated with articles. RNA sequencing data generated from this project will be uploaded to the NCBI Sequence Read Archive.

**REFERENCES**

**1** Ort, D.R. *et al* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* 112, 8529–8536 **2** Long, S.P. *et al* (2015) Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell* 161, 56–66 **3** Long, S.P. *et al* (2006) Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330 **4** Zhu, X.-G. *et al* (2010) Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* 61, 235–261 **5** Evans, J.R. (2013) Improving Photosynthesis. *Plant Physiol.* 162, 1780–1793 **6** Galmés, J. *et al* (2014) Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ.* 37, 1989–2001 **7** Parry, M.A.J. *et al* (2013) Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* 64, 717–730 **8** Orr, D.J. *et al* (2016) Surveying Rubisco Diversity and Temperature Response to Improve Crop Photosynthetic Efficiency. *Plant Physiol.* 172, 707–717 **9** Andersson, I. (2008) Catalysis and regulation in Rubisco. *J. Exp. Bot.* 59, 1555–1568 **10** Prins, A. *et al* (2016) Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *J. Exp. Bot.* 67, 1827–1838 **11** Sharwood, R.E. (2017) Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytol.* 213, 494–510 **12** Galmés, J. *et al* (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ.* 28, 571–579 **13** Yamori, W. *et al* (2012) Rubisco activase is a key regulator of non-steady-state photosynthesis at any leaf temperature and, to a lesser extent, of steady-state photosynthesis at high temperature. *Plant J.* 71, 871–880 **14** Carmo-Silva, E. *et al* (2015) Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ.* 38, 1817–1832 **15** Salvucci, M.E. and Crafts-Brandner, S.J. (2004) Relationship between the Heat Tolerance of Photosynthesis and the Thermal Stability of Rubisco Activase in Plants from Contrasting Thermal Environments. *Plant Physiol.* 134, 1460–1470 **16** Sharkey, T.D. *et al* (2001) Increased heat sensitivity of photosynthesis in tobacco plants with reduced Rubisco activase. *Photosynth. Res.* 67, 147–156 **17** Yamori, W. *et al* (2016) Strategies for Optimizing Photosynthesis with Biotechnology to Improve Crop Yield. In *Handbook of Photosynthesis, Third Edition* 0 vols.pp. 741–759, CRC Press **18** Kurek, I. *et al* (2007) Enhanced Thermostability of Arabidopsis Rubisco Activase Improves Photosynthesis and Growth Rates under Moderate Heat Stress. *Plant Cell* 19, 3230–3241 **19** Scafaro, A.P. *et al* (2016) Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. *New Phytol.* 211, 899–911 **20** Galmés, J. *et al* (2014) Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus Limonium (Plumbaginaceae). *New Phytol.* 203, 989–999 **21** Sharwood, R.E. *et al* (2016) Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C3 photosynthesis. *Nat. Plants* 2, 16186 **22** Galmés, J. *et al* (2015) Temperature responses of the Rubisco maximum carboxylase activity across domains of life: phylogenetic signals, trade-offs, and importance for carbon gain. *Photosynth. Res.* 123, 183–201 **23** Tcherkez, G. (2013) Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters. *Plant Cell Environ.* 36, 1586–1596 **24** Tcherkez, G.G.B. *et al* (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci.* 103, 7246–7251 **25** Gillet, L.C. *et al* (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 11, O111.016717 **26** Liu, Y. *et al* (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Rev. Mol. Diagn.* 13, 811–825 **27** Pratt, J.M. *et al* (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* 1, 1029–1043 **28** Hu, J. *et al* (2015) Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Front. Plant Sci.* 6, **29** Pertl-Obermeyer, H. *et al* (2016) Quantitation of Vacuolar Sugar Transporter Abundance Changes Using QconCAT Synthtetic Peptides. *Front. Plant Sci.* 7, **30** Morandini, P. (2013) Control limits for accumulation of plant metabolites: brute force is no substitute for understanding. *Plant Biotechnol. J.* 11, 253–267 **31** Stitt, M. *et al* (2010) Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth. *Plant Physiol.* 152, 428–444 **32** Farquhar, G.D. *et al* (2001) Models of Photosynthesis. *Plant Physiol.* 125, 42–45 **33** Maire, V. *et al* (2012) The Coordination of Leaf Photosynthesis Links C and N Fluxes in C3 Plant Species. *PLOS ONE* 7, e38345 **34** Walker, A.P. *et al* (2014) The relationship of leaf photosynthetic traits – Vcmax and Jmax – to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. *Ecol. Evol.* 4, 3218–3235 **35** Quebbeman, J.A. and Ramirez, J.A. (2016) Optimal allocation of leaf-level nitrogen: Implications for covariation of Vcmax and Jmax and photosynthetic downregulation. *J. Geophys. Res. Biogeosciences* 121, 2016JG003473 **36** Dong, N. *et al* (2017) Leaf nitrogen from first principles: field evidence for adaptive variation with climate. *Biogeosciences* 14, 481–495 **37** Sulpice, R. *et al* (2010) Network Analysis of Enzyme Activities and Metabolite Levels and Their Relationship to Biomass in a Large Panel of Arabidopsis Accessions. *Plant Cell* 22, 2872–2893 **38** Schöttler, M.A. and Tóth, S.Z. (2014) Photosynthetic complex stoichiometry dynamics in higher plants: environmental acclimation and photosynthetic flux control. *Front. Plant Sci.* 5, **39** Vogel, C. (2013) Protein Expression Under Pressure. *Science* 342, 1052–1053 **40** 40J.M. *et al* (2010) Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* 10, 4209–4212 **41** Khan, Z. *et al* (2013) Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. *Science* 342, 1100–1104 **42** Gago, J. *et al* (2016) Relationships of Leaf Net Photosynthesis, Stomatal Conductance, and Mesophyll Conductance to Primary Metabolism: A Multispecies Meta-Analysis Approach. *Plant Physiol.* 171, 265–279 **43** Penning de Vries, F.W. (1975) The Cost of Maintenance Processes in Plant Cells. *Ann. Bot.* 39, 77–92 **44** Dillon, S. *et al* (2015) Genetic diversity of Eucalyptus camaldulensis Dehnh. following population decline in response to drought and altered hydrological regime. *Austral Ecol.* 40, 558–572 **45** Blackman, C.J. *et al* (2016) Leaf photosynthetic, economics and hydraulic traits are decoupled among genotypes of a widespread species of eucalypt grown under ambient and elevated CO2. *Funct. Ecol.* 30, 1491–1500 **46** Chapman, B. *et al* (2013) Plant Proteogenomics: From Protein Extraction to Improved Gene Predictions. In *Proteomics for Biomarker Discovery* (Zhou, M. and Veenstra, T., eds), pp. 267–294, Humana Press **47** Wessel, D. and Flügge, U.I. (1984) *Anal. Biochem.* 138, 141–143 **48** Myburg, A.A. *et al* (2014) *Nature* 510, 356–362 **49** Bayly, M.J. *et al* (2013) *Mol. Phylogenet. Evol.* 69, 704–716 **50** Lohse, M. *et al* (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* 37, 1250–1258 **51** Chang, S. *et al* (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Report.* 11, 113–116 **52** Ludwig, C. *et al* (2012) Estimation of Absolute Protein Quantities of Unlabeled Samples by Selected Reaction Monitoring Mass Spectrometry. *Mol. Cell. Proteomics* 11, M111.013987 **53** Liu, Y. *et al* (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *PROTEOMICS* 13, 1247–1256 **54** Sharwood, R.E. *et al* (2017) Linking photosynthesis and leaf N allocation under future elevated CO2 and climate warming in Eucalyptus globulus. *J. Exp. Bot.* DOI: 10.1093/jxb/erw484 **55** Sharwood, R.E. *et al* (2008) The Catalytic Properties of Hybrid Rubisco Comprising Tobacco Small and Sunflower Large Subunits Mirror the Kinetically Equivalent Source Rubiscos and Can Support Tobacco Growth. *Plant Physiol.* 146, 83–96 **56** Kane, H.J. *et al* (1994) An Improved Method for Measuring the CO2/O2 Specificity of Ribulosebisphosphate Carboxylase-Oxygenase. *Funct. Plant Biol.* 21, 449–461 **57** Baker, R.T. *et al* (2005) Using deubiquitylating enzymes as research tools. *Methods Enzymol.* 398, 540–554 **58** Hendrickson, L. *et al* (2008) The effects of Rubisco activase on C4 photosynthesis and metabolism at high temperature. *J. Exp. Bot.* 59, 1789–1798 **59** Wachter, R.M. *et al* (2013) Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco–Rubisco activase interaction. *Photosynth. Res.* 117, 557–566 **60** Barta, C. *et al* (2011) Rubisco Activase Activity Assays. In *Photosynthesis Research Protocols* (Carpentier, R., ed), pp. 375–382, Humana Press **61** Carmo-Silva, A.E. and Salvucci, M.E. (2011) The activity of Rubisco’s molecular chaperone, Rubisco activase, in leaf extracts. *Photosynth. Res.* 108, 143–155 **62** Vizcaíno, J.A. *et al* (2014) *Nat. Biotechnol.* 32, 223–226 **63** Deutsch, E.W. *et al* (2017) *Nucleic Acids Res.* 45, D1100–D1106