

Comparative analysis of cell free and *in planta* protein synthesis systems

OpenPlant Progress Report

Primary Team Contact

Susan Duncan, Organisms and Ecosystems Department, Earlham Institute, Norwich Construct design, protein synthesis and analysis. susan.duncan@earlham.ac.uk

Team

Laura-Jayne Gardiner: Wheat bioinformatic expertise: Hall Lab, Earlham Institute laura-jayne.gardiner@earlham.ac.uk

Quentin Dudley: Cell-free protein synthesis expertise: Patron Lab, Earlham Institute quentin.dudley@earlham.ac.uk

Philippa Borrill: Wheat senescence transcription factor expertise: John Innes Centre philippa.borrill@jic.ac.uk

Pallavi Singh: Photosynthesis transcription factor expertise: Hibberd Lab, University of Cambridge ps753@cam.ac.uk

Our Proposal

Wheat is a primary world food crop. It's production needs to increase to meet demand and understanding gene regulatory networks will be critical for designing crops for the future. Although we have good understanding of key networks in the model plant Arabidopsis, the extent to which they have been conserved in wheat remains an open question. In addition, we have very little understanding of how gene regulatory networks operate across complex polyploid genomes. Phylogenetic analyses have revealed many core network transcription factor (TF) homologues, but knowledge of sequence specific binding sites and downstream targets is required to determine the extent to which they can be considered functional orthologues. DNA Affinity Purification Sequencing (DAP-Seq) is a high throughput method that can provide both of these key pieces of information (O'Malley et al., 2016). It uses epitope tagged TFs to pull out associated DNA sequences to reveal binding sites. A recently published DAP-seq dataset for Arabidopsis transcription factors has paved the way for similar studies to be carried out in other plant species and Susan is leading a project in the Hall lab to set this up for wheat. We are requesting Open Plant funding to extend optimisation of the protein synthesis

section of this technique with the aim of improving the ~30% DAP-Seq success rate reported for Arabidopsis TFs. This will allow us to extend the scope of the project and carry out a comprehensive test of three protein synthesis methods. We propose comparing cell free and *in planta* synthesis systems for 42 wheat TFs which include homologues of key Arabidopsis TFs involved in circadian rhythms and photosynthesis as well as a characterized TF that is known to be involved in wheat senescence.

Benefits and Outcomes

We envisaged four major outcomes from this project:

Objective 1: Synthesis of 42 freely available DNA modules for wheat transcription factors

Progress

Construct Design

After obtaining CDS sequences for our 42 candidate wheat TFs, we took a three-step sequence optimisation approach. Firstly, we used an automated pipeline (provided by TWIST Bioscience) to generate sequences that would maximise chances of DNA synthesis success. Secondly we applied a manual codon optimisation approach to ensure that our DNA modules would be suitable for unbiased protein synthesis method assessment. Thirdly, we removed all Bsa I, Bpi I, BsmB I and Sap I sites. Finally we carried out BLAST alignments to confirm that we had preserved TF amino acid sequences during these domestication and codon optimization steps.

Next we modified our proposed Golden Gate level 0 module design. We added 4 flanking sequences: Strep tag, linker sequence, *E. coli* promoter and terminator. This enabled us to order plasmids that could immediately be tested using Quentin's "in-house" *E. coli* Cell Free Protein Synthesis (CFPS) system. We also added BsaI sites at either end of each CDS to ensure that all constructs could function as Level 0 CDS modules and be suitable for downstream Golden Gate reactions (see Figure 1). Our decision to invest in the synthesis of these small flanking sequences greatly accelerated the progress of our project. It effectively removed the assembly and sequencing steps that would have been required before any attempt at protein synthesis could be made. In addition, it enabled us to test all our constructs using the *E. coli* CFPS system - our most cost effective protein synthesis option.

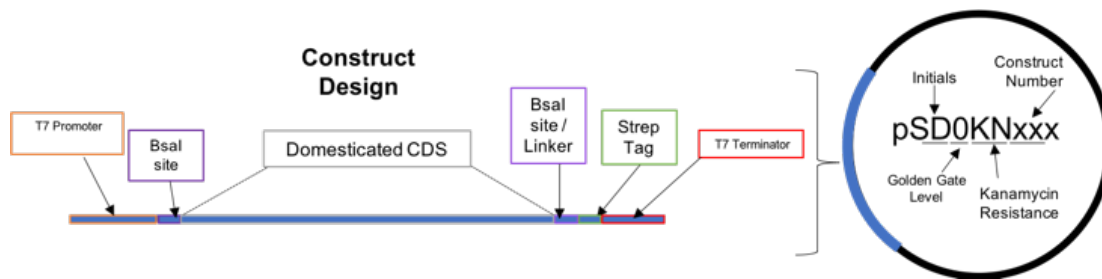


Figure 1. Diagram showing the design strategy used for all constructs.

Construct Synthesis

40 of our 42 construct were successfully synthesised. From these we were able to carry out successful transformations, extract DNA and generate glycerol stocks for 36.

Ongoing

- Transformation is being repeated for the 4 outstanding plasmids.

Objective 2: Valuable comparative data for *in vitro* cell free and *in planta* protein synthesis methods (potential open access publication).

Progress

Initial CFPS Tests

We tested all 36 constructs using the *E. coli* CFPS synthesis system. After optimising western blotting conditions we were able to show that 11 of these constructs had generated tagged protein (of the expected size) in the soluble lysate fractions.

Progress made toward synthesis comparison experiments

Golden Gate DigLig reactions have been completed for all constructs to enable comparative analysis of the wheat germ and *N. benth* expression systems.

Ongoing

- Any of the four outstanding constructs that transform successfully following a second attempt, will be put through our existing CFPS, western pipeline.

- There is evidence that *E.coli* CFPS temperature reduction can increase concentrations of soluble protein. Assessments are ongoing for constructs that have been re-tested using 16 °C CFPS conditions.
- Once CFPS success has been determined for all available constructs, a subset of TFs will be tested in both the wheat germ and transient *N. benthamiana* systems. It is anticipated that results from these experiments will be written up as a short paper and submitted for open access publication.

Objective 3: An optimised protein synthesis pipeline that will underpin an ongoing large DAP-Seq project

Progress

Bead purification efficiency tests for synthesised Strep-tagged proteins

Affinity tagged TFs need to be effectively purified from lysate for use in DAP-Seq experiments. As there are no published studies where Strep affinity tags have been used for this purpose, we set up experiments to determine bead purification efficiency. Results from an SDS PAGE experiment enabled calculations to be made for our first set of DNA affinity tests.

Wheat DNA library construction

A total of 8 amplified KAPA libraries have been constructed. All of these passed quality control tests and are suitable for DAP-Seq experiments. Before investing in whole genome sequencing, some of these libraries are being used for method optimisation and DAP-qPCR experiments. This economic approach enables TF enrichment to be identified at a subset of putative binding sites.

Sequence specific binding of synthesised proteins

A DAP-qPCR experiment has been completed for two proteins synthesised using the *E. coli* CFPS system. We compared sequence specific binding for a control strep-tagged GFP and a wheat circadian gene homologue (LUX). Our results indicate that LUX enrichment occurs at two of our candidate binding sites and GFP does not preferentially bind any of the 10 regions we tested.

Ongoing work

- After confirmation of our preliminary DAP-qPCR, we will proceed with a full DAP-Seq run to identify LUX binding sites across the wheat genome.

Objective 4: New collaborations that would promote knowledge exchange and capitalise on expertise across EI, JIC and University of Cambridge.

- We have successfully synthesised constructs and have succeeded in synthesising at least one TF of interest for each of our collaborators.
- Thanks to OpenPlant funding, we now have an economical in-house *E. coli* CFPS system based at the Earlham Institute. This has already been used extensively and now has a proven success rate for wheat TF synthesis.
- This project has created ongoing, beneficial collaborations between our four labs: Hall Lab (EI), Patron lab (EI), Hibberd Lab (Uni. Cambridge) and Uauy Lab (John Innes).
- It is anticipated that data generated for this project will be used to support future collaborative BBSRC grant applications.

Summary

OpenPlant funding has provided us with a strong foundation from which to pursue our goal - to generate the first genome-wide TF binding site dataset for wheat.