Cell-free proteins synthesis as a resource for generating plant proteins

Primary contact:
Quentin Dudley, *Nicola Patron group*, Earlham Institute (EI), quentin.dudley@earlham.ac.uk will provide the expertise and background in cell-free protein synthesis to purchase appropriate reagents and validate CFPS performance. He will also coordinate the overall effort and oversee the proposed workshops.

Team:
Susan Duncan, *Anthony Hall group,* Earlham Institute**,** susan.duncan@earlham.ac.uk has experience in plant biology and interest in improving the DAP-seq method for identifying transcription factor binding sites. She will provide plant biology expertise, DNA sequences and material, along with help to characterize the primary set of test proteins
Nicolas Larus-Stone*,* *Pietro Lio' group*, Department of Computer Science, University of Cambridge (UC), nl363@cam.ac.uk has experience in modelling of dynamic biological processes. He will provide computational expertise in developing the cost model of various CFPS platforms and provide scientific guidance for additional CFPS applications.

Summary
Cell-free protein synthesis (CFPS) has emerged as a powerful technique for on-demand, *in vitro* protein production which reduces labour and increases experimental throughput. However, cell-free systems can be laborious and expensive to set up and there is a shortage of publicly available data comparing different CFPS systems, particularly regarding plant proteins which can be difficult to express. This proposal will provide a resource for researchers to prototype their experimental plans without setting up the system themselves. We propose comparing an in-house generated *E. coli* S30 crude lysate system with a commercial wheat germ platform to quantify their ability to synthesize transcription factors and other plant proteins. This will provide data that can be used to build a simple model to predict which CFPS platform is best suited to a researcher's needs. Additionally, two workshops (at Norwich and Cambridge) will communicate these results and provide additional expression data by crowdsourcing DNA assembly to workshop participants.

Proposal
The purpose of this project is to set up a cell-free protein synthesis (CFPS) resource based at Norwich BioScience Institutes in collaboration with University of Cambridge. While there is growing interest in cell-free systems as a means for accelerating biological research[1,2], there is a higher barrier in terms of cost (£) and knowledge to setting up CFPS. Indeed, there is a shortage of publicly available information comparing the cost, yield, and flexibility of different CFPS systems. The limited data available often uses standardized non-plant reporter proteins such as green fluorescent protein (GFP) or luciferase which complicates comparison for plant biologists since plant proteins may require unique additives or an optimized folding environment for robust, soluble expression (Supplementary Figure 1). This proposal aims to fill this knowledge gap by comparing a low-cost, in-house PANOx-SP *E. coli* S30 CFPS platform[3] with a commercially available wheat germ[4] CFPS kit from Promega. The reaction conditions for PANOx-SP are publicly available (unlike many commercial *E. coli* kits), *E. coli* lysate can be easily generated with low-cost methods such as sonication[5], and the per reaction cost is far lower than commercial alternatives (£0.04/µg protein, Table 1). While the wheat germ CFPS kit is known to express many plant proteins (£2.42/µg protein), this study will provide useful data comparing the yield versus cost of these two CFPS platforms in effort to minimize overall expense to the researcher.

Table 1. Cost of prominet CFPS platforms

Manufacturer	Source	Kit Description	Product #	# of 50 µL rxns	Cost	Expected yield (µg/mL)	Cost per rxn	Cost per µg protein
n/a	*E. coli*	S30 PANOx SP	n/a	5700	£ 2,807.60	500-1000	£ 0.49	£ 0.02
Promega	*E. coli*	S30 T7 High-Yield Protein Expression System	L1115	24	£ 268.00	500	£11.17	£ 0.45
Promega	Wheat Germ	TnT® SP6 High-Yield Wheat Germ Protein Exp. Sys.	L3260	40	£ 484.00	100	£12.10	£ 2.42
NEB	*E. coli*	PURExpress® In Vitro Protein Synthesis Kit	E6800L	50	£ 2,059.00	10-200	£41.18	£ 4.12

As a first step to characterizing (and eventually improving) the ability of CFPS to make plant derived proteins, we will first attempt to synthesize transcription factors (TFs) from wheat (*Triticum aestivum*). TFs are a key component of the plant circadian rhythm and response to environmental factors[6]. Recent efforts to characterize TF binding sites in *Arabidopsis thaliana* using DNA affinity purification sequencing (DAP-seq) used wheat germ CFPS to generate hundreds of TFs whose DNA binding site(s) were subsequently determined[7,8]. To test our system (and enable DAP-seq at EI), we will synthesize 12 known wheat TFs including several for which DAP-seq data is not available (such as MADs box or helix-loop-helix TFs)[8]. These sequences will be tested side-by-side in *E. coli* and wheat germ CFPS and compared qualitatively using Western blots of C-terminal strep tags[9] and semi-quantitatively using GFP-fusion proteins (expertise provided by **Susan Duncan**).

To lower the uncertainty for researchers considering using CFPS, we will use the TF data and published literature to develop a computational model incorporating a publicly available web interface in which users can select from a list of parameters including protein characteristics, scale of experiment, and other information to evaluate which expression system best meets their need (expertise provided by **Nicolas Larus-Stone**).

To train new users in CFPS, we will host two hands-on workshops (one each in Norwich and Cambridge). To broaden the range of proteins characterized in our model, we will crowdsource DNA assembly by encouraging participants to clone any genes of interest into MoClo-compatible[10,11] CFPS vectors (provide by this funding) and bring these to the workshop. This project will cover costs of CFPS reagents and expression measurement while participants will fund DNA assembly. These results will inform the cost model with a broader range of proteins known to be of interest to the plant biology community.

If successful, this project will enable further exploration to improve *E. coli* as a plant production platform. A few simple experiments include using different *E. coli* source stains (Rosetta™(DE3)) to minimize the codon-optimization requirements, optimize temperature gradients to improve soluble yields (Supplementary Figure 1), add nanodisk membrane mimics to express membrane-bound proteins[12], and add purified plant chaperons and heat shock proteins. Further cost reductions could be made by substituting non-phosphorylated energy substrates for phosphoenolpyruvate (PEP) which is the most expensive part of the PANOx-SP system; however, activating these systems requires specialized lysate preparation which is why our initial effort will use PEP.

Outcomes

1) CFPS expression vectors pJL1 and pF3 WG (BYDV) Flexi® will be modified to be compatible with Type IIS-based restriction cloning (Golden Gate, MoClo, Loop, etc) [10,11]. Standardized backbones will be available through Addgene (if possible).
2) Comparison of cell-free protein synthesis of ~12 transcription factors from wheat (with and without fusion to GFP) using both *E. coli* S30 and wheat germ (Promega) kits. (**Susan Duncan, EI**)
3) Publication of a computational model incorporating a publicly available web interface for comparing and predicting cost of plant protein CFPS (**Nicolas Larus-Stone, UC**).
4) Two hand-on workshops (at Norwich and Cambridge) will allow attendees to run and quantify a protein with the option of providing a DNA template of their own choosing.

Benefits to OpenPlant community

If successful, this project will improve a fundamental tool for plant science and plant synthetic biology. It aligns with the OpenPlant theme of Training and Knowledge Exchange as it funds reagents for a resource that can enable future collaboration and trains new users via a hands-on workshop. The computational model will be open-source and easily accessible to non-computation researchers using a web interface. This project also aligns with the theme of Cell-Free Synthetic Biology by working to lower the cost of expressing plant proteins using CFPS and potentially avoiding the need to purchase expensive commercial kits. This project also represents a new interdisciplinary interaction between individuals at the University of Cambridge and Earlham Institute that leverages unique skills sets of cell-free systems (**Q. Dudley**), plant biology (**S. Duncan**), and computation modelling (**N. Larus-Stone**).

> ***Potential beneficiaries.*** Funding of this OpenPlant application will support the up-front costs of testing the feasibility of future collaborations. The following individuals have already expressed interest in workshops and cell-free systems: **Yao-Min Cai (EI)** yaomin.cai@earlham.ac.uk is building synthetic plant promoter and could use CFPS-derived TFs determine which TFs are binding *in planta.* **Gonzalo Mendoza (UC)** gim23@cam.ac.uk is engineering eukaryotic riboswitches to control transgene expression and would use CFPS as a low-cost and high-throughput testing method. **Nicholas Larus-Stone (UC)** nl363@cam.ac.uk is adapting differential equations and flux balance analysis to predict protein expression in *E. coli* CFPS and could use this platform (or wheat germ) to validate models for spatially complex circuits[13]. **Jim Haseloff (UC)** jh295@cam.ac.uk has initiated efforts to enable a *E. coli* lysate-based CFPS capability at Cambridge for several applications including diagnostics, detection of environmental pollutants, and as an education tool. Dr. Haseloff has graciously offered to share resources, particularly the equipment and biomass needed for lysate generation.

Sponsor for the research and cost centre:

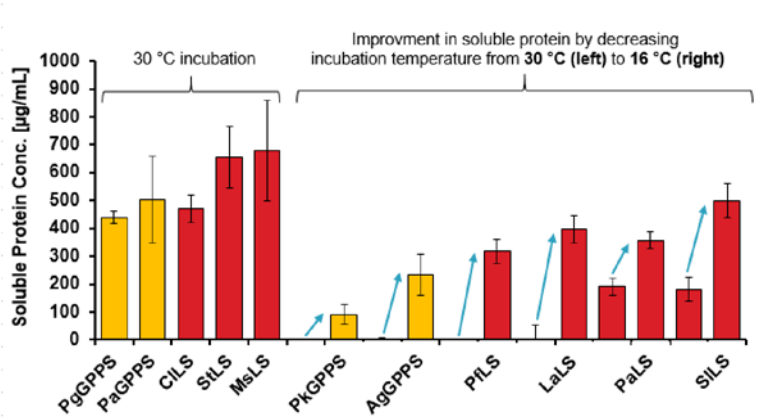
Nicola Patron, Earlham Institute, nicola.patron@earlham.ac.uk will support the project and sponsor the cost-code for any allocated funds. *I confirm that I have the full support of the sponsor listed above and that they can be added to the OpenPlant Fund mailing list to receive project updates (to which they can unsubscribe at any time).*

Budget

Cell-free protein synthesis reagents		£2,807.60
Energy source (PEP)*	£450.00	
T7 RNA Polymerase	£272.00	
Cofactors	£543.00	
NTPs	£522.30	
Amino Acids	£723.40	
Other	£296.90	
E. coli S30 Lysate	*no cost, collaboration with Haseloff group*	
Promega TnT® SP6 High-Yield Wheat Germ Protein Expression System (L3260)**		£484.00
PCR + DNA Assembly Reagents (NEB Gibson Assembly® Master Mix E2611S)		£300.00
DNA Purification Reagents (QIAGEN Plasmid Plus Midi Kit 12943)		£212.00
	TOTAL:	**£3,803.60**

*limiting reagent, enough for 5,000 reactions (50 µL volume)
**permits 40 reactions (50 µL volume)

Supplementary Figure 1. *Lowering incubation temperature improves yield of select plant-derived enzymes in* E. coli *CFPS*



References

1 Dudley, Q. M., Karim, A. S. & Jewett, M. C. Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnology Journal* **10**, 69-82 (2015).

2 Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.* **30**, 1185-1194 (2012).

3 Jewett, M. C. & Swartz, J. R. Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* **86**, 19-26 (2004).

4 Harbers, M. Wheat germ systems for cell-free protein expression. *FEBS letters* **588**, 2762-2773 (2014).

5 Kwon, Y.-C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports* **5** (2015).

6 Greenham, K. & McClung, C. R. Integrating circadian dynamics with physiological processes in plants. *Nature Reviews Genetics* **16**, 598-610 (2015).

7 O'Malley, R. C. *et al.* Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280-1292 (2016).

8 Bartlett, A. *et al.* Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature protocols* **12**, 1659-1672 (2017).

9 Schmidt, T. G. *et al.* Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression Purif.* **92**, 54-61 (2013).

10 Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Syn. Biol.* **3**, 839-843 (2014).

11 Patron, N. J. *et al.* Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist* **208**, 13-19 (2015).

12 Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS letters* **584**, 1721-1727 (2010).

13 Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326-330 (2010).