SYNTHETIC BIOLOGY

Synthetic Biology, 2021, **6(1)**, 1–13
DOI: https://doi.org/10.1093/synbio/ysab029
Advance access publication date: 11 September 2021
Research Article

Biofoundry-assisted expression and characterization of plant proteins

Quentin M. Dudley, Yao-Min Cai, Kalyani Kallam, Hubert Debreyne, Jose A. Carrasco Lopez, and Nicola J. Patron*

Engineering Biology, Earlham Institute, Norwich Research Park, Norwich, Norfolk UK
*Corresponding author: E-mail: nicola.patron@earlham.ac.uk

Abstract

Many goals in synthetic biology, including the elucidation and refactoring of biosynthetic pathways and the engineering of regulatory circuits and networks, require knowledge of protein function. In plants, the prevalence of large gene families means it can be particularly challenging to link specific functions to individual proteins. However, protein characterization has remained a technical bottleneck, often requiring significant effort to optimize expression and purification protocols. To leverage the ability of biofoundries to accelerate design–built–test–learn cycles, we present a workflow for automated DNA assembly and cell-free expression of plant proteins that accelerates optimization and enables rapid screening of enzyme activity. First, we developed a phytobrick-compatible Golden Gate DNA assembly toolbox containing plasmid acceptors for cell-free expression using *Escherichia coli* or wheat germ lysates as well as a set of N- and C-terminal tag parts for detection, purification and improved expression/folding. We next optimized automated assembly of miniaturized cell-free reactions using an acoustic liquid handling platform and then compared tag configurations to identify those that increase expression. We additionally developed a luciferase-based system for rapid quantification that requires a minimal 11–amino acid tag and demonstrate facile removal of tags following synthesis. Finally, we show that several functional assays can be performed with cell-free protein synthesis reactions without the need for protein purification. Together, the combination of automated assembly of DNA parts and cell-free expression reactions should significantly increase the throughput of experiments to test and understand plant protein function and enable the direct reuse of DNA parts in downstream plant engineering workflows.

Key words: biofoundry; cell-free protein synthesis; golden gate DNA assembly; plant biotechnology; automation

1. Introduction

Plant synthetic biology endeavors to apply principles of abstraction, modularity and standardization to engineer plants for useful purposes (1, 2). In support of this goal, the plant community adopted a common syntax, commonly known as the 'phytobrick' standard, that defines the features of DNA parts (3, 4). For expression in plants, phytobricks can be assembled into synthetic genetic circuits using a number of plasmid toolkits including MoClo (5), Loop (6), GoldenBraid (7) and Mobius (8). These systems use Type IIS restriction endonucleases to direct one-pot digestion–ligation assembly reactions, known as Golden Gate (9), that are easily parallelized and miniaturized for automation (10, 11). Parallel DNA assembly has become an enabling technology for biofoundries, which specialize in automating the design–build–test–learn (DBTL) cycle that underpins synthetic biology (12–14).

Plants have long been exploited as sources of bioactive and high-value natural products (15), and recently, applications of model-informed synthetic biology approaches have led to sophisticated engineering of crop traits including biomass and responses to the environment (16–18). There is also growing interest and investment in the use of plants, particularly *Nicotiana benthamiana*, as photosynthetic platforms for the production of recombinant proteins and small molecules for industry and medicine (19–21), including rapid-response vaccines (22–25). For many applications, methods to rapidly characterize protein functions are essential. This particularly applies to understanding the specific functions of members of large protein families such as decorating enzymes and transcription factors (TFs). However, for plant proteins, this is frequently a bottleneck, with researchers often investing considerable time and effort to identify permissive constructs and conditions to obtain useful yields before lengthy protocols to purify recombinant proteins for *in vitro* assays.

Cell-free protein synthesis (CFPS) is an established tool for rapid *in vitro* protein production that combines a DNA template, energy source, amino acids, nucleotide triphosphates (NTPs) and excess cofactors, along with a crude lysate containing the translational machinery (26, 27). The source lysate from *Escherichia coli* is highly active, easy to manipulate and can produce a broad range of proteins (28) including enzymes (29, 30), antibodies (31), glycoproteins (32, 33), as well as proteins containing non-canonical

Submitted: 10 March 2021; **Received (in revised form):** 25 August 2021; **Accepted:** 9 September 2021

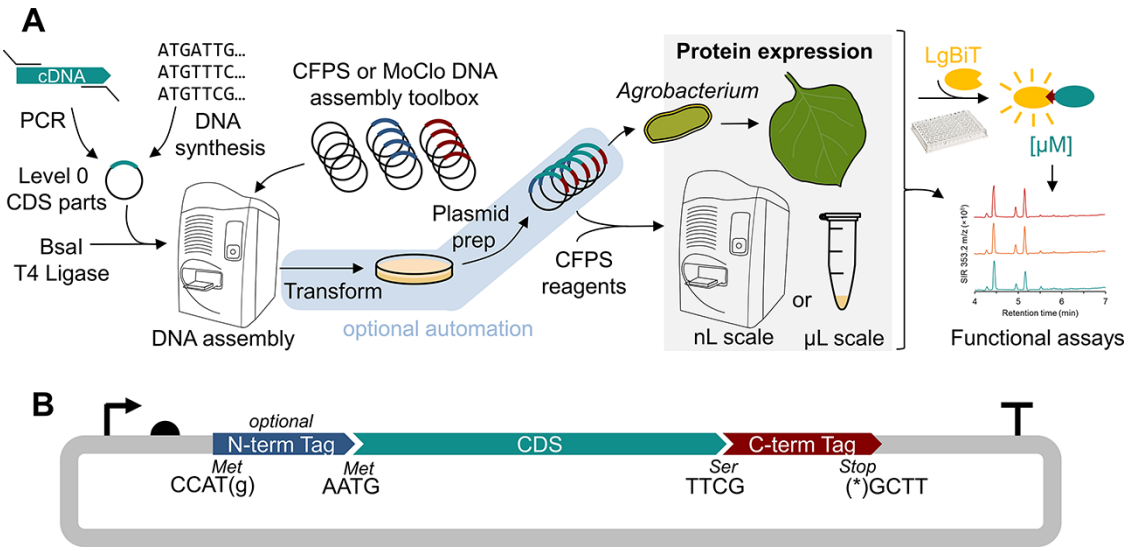

Figure 1. A workflow for biofoundry-assisted DNA assembly and cell-free protein synthesis (CFPS) of plant proteins. (A) Level 0 DNA parts (phytobricks) encoding the protein of interest can be assembled into functional expression plasmids for CFPS or expression *in planta*. Acoustic liquid handling enables the screening of libraries of protein variants to determine optimal expression configurations using the HiBiTLgBiT luminescence. (B) The CFPS cloning toolbox consists of acceptor plasmids that assemble with a CDS and C-terminal tag along with an optional N-terminal tag. The cloning overhangs are compatible with existing plant DNA assembly standards.

amino acids (34–36). It is also possible to use plant-based lysates from wheat germ (37), Arabidopsis (38) and BY-2 tobacco cells (39). Cell-free expression can minimize 'build' times in the DBTL cycle (40) and has been used to prototype metabolic pathways (29, 30, 41, 42) and characterize several natural product biosynthesis pathways (43, 44) including cyanobacterial alkaloids (45) and antibiotic peptides (46). Furthermore, CFPS is amenable to automation and miniaturization (47). The use of automation platforms that utilize acoustic energy to transfer reagents has been shown to reduce operator error and variability in CFPS reactions (48), facilitate active learning-guided optimization of reaction conditions (49) and generally increase the throughput of experiments (50–55).

In this work, we have developed molecular tools and automated biofoundry workflows to enable cell-free expression of plant proteins (Figure 1A). We first developed a phytobrick-compatible plasmid toolkit including (i) plasmid acceptors containing regulatory elements for T7-driven *E. coli* CFPS and the commercial TNT SP6-Coupled Wheat Germ Expression System (Promega), (ii) affinity tags for purification or detection and (iii) a suite of tags for improving the yields of soluble protein. We then use the toolkit to optimize the expression of a range of plant proteins including enzymes and TFs. Finally, we demonstrate the functional activity of the cell-free expressed proteins. Together, our tools and workflows enable the rapid optimization of expression conditions and allow immediate progression to characterization assays, significantly increasing the scale and throughput of experimentation.

2. Materials and methods

2.1 DNA assembly

Acceptor plasmids for CFPS (see Table 1) were designed as 'terminal acceptors' and are not designed for subsequent assembly into plasmids containing multiple transcriptional units; each contains promoter and terminator sequences specific to the expression system being utilized. Acceptors for *E. coli* CFPS were based on pJL1 (Addgene #69496) and some include an N-terminal Expression Tag (NET) that encodes Met–Glu–Lys–Lys–Ile (MEKKI) shown to enhance the expression of some proteins (29). Acceptors for wheat germ expression were based on pEU-Gm23 (Addgene #53738), which uses the pEU backbone (56) containing the E01 translational enhancer (57).

New level 0 DNA parts (i.e. N-terminal tags, coding sequences (CDSs), C-terminal tags) were either synthesized (Twist Bioscience, San Francisco, CA) or amplified by PCR with overhangs containing BpiI (BbsI) recognition sites and assembled into pUAP1 (Addgene #63674) to create parts compatible with the phytobrick standard (3). We included a number of N-terminal tags previously shown to improve translation, solubility and folding (58, 59): the S-tag sequence pancreatic ribonuclease A (KETAAAKFERQHMDS) typically used for quantitation/purification but anecdotally suggested to improve protein solubility (60); a small ubiquitin-related modifier (SUMO) sequence known to increase soluble expression (61) derived from pDest-Sumo (Addgene #106980) (62); a TrxA sequence from *E. coli* and derived from pET32a-TRXtag (Addgene #11516) (63); a glutathione-S-transferase (GST) (64) and thrombin sequence derived from pGEX-2T (GE Healthcare, Chicago, IL), and a maltose-binding protein (MBP) (65) and Factor Xa sequence derived from pMAL-c5X vector (New England Biolabs, Ipswich, MA) with a V313A mutation to be consistent with the native *E. coli* sequence. We also included N-terminal and C-terminal HiBiT tags (Promega, Madison, WI) for quantification using luminescence, a green fluorescence protein (GFP) derived from pJL1 (Addgene #69496) and a C-terminal twin-strep-tag (66). Finally, we included a CDS part encoding the tobacco etch virus (TEV) protease containing a S219V mutation reported to have less autolysis (67).

All DNA assembly reactions were performed in 20 µl (manual) or 2 µl (automated) reaction volumes as previously described (10) and were verified by sequencing. All plasmids used in this study are described in the supplementary information (Supplementary Table S1) along with nucleotide sequences for L0 DNA parts encoding plant proteins (Supplementary Table S2).

Table 1. A plasmid toolkit for cell-free protein expression compatible with the phytobrick assembly standard. Each fusion site lists the top strand only based on the orientation shown in Figure 1B

Plasmid code	Addgene number	Purpose	Description	Fusion sites	
pEPQD0KN0024	162281	Acceptor, *E. coli* CFPS	T7 promoter, RBS, N-term exp. tag, T7 terminator, kanR, pUC ori	AATG	GCTT
pEPQD0KN0244	162282	Acceptor, *E. coli* CFPS	T7 promoter, RBS, N-term exp. tag, T7 terminator, kanR, pUC ori	CCAT	GCTT
pEPQD0KN0025	162283	Acceptor, *E. coli* CFPS	T7 promoter, RBS, T7 terminator, kanR, pUC ori	AATG	GCTT
pEPQD0KN0245	162284	Acceptor, *E. coli* CFPS	T7 promoter, RBS, T7 terminator, kanR, pUC ori	CCAT	GCTT
pEPQD0CB0026	162285	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, folA terminator, ampR, pUC ori	AATG	GCTT
pEPQD0CB0246	162286	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, folA terminator, ampR, pUC ori	CCAT	GCTT
pEPQD0KN0282	162287	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, T7 terminator, kanR, pUC ori	AATG	GCTT
pEPQD0KN0283	162288	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, T7 terminator, kanR, pUC ori	CCAT	GCTT
pEPQD0KN0284	162289	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, folA terminator, kanR, pUC ori	AATG	GCTT
pEPQD0KN0285	162290	Acceptor, TNT SP6 WG	SP6 promoter, EO1 enhancer, folA terminator, kanR, pUC ori	CCAT	GCTT
pEPMY1CB0001	162317	Acceptor, *E. coli*	T7 promoter, lac operator, RBS, T7 terminator, ampR, pUC ori	CCAT	GCTT
pEPYC0CM0258	162291	N-terminal tag	HiBiT	CCAT	AATG
pEPQD0CM0541	162292	N-terminal tag	GST-[thrombin cleavage site]	CCAT	AATG
pEPQD0CM0542	162293	N-terminal tag	HiBiT-GST-[thrombin cleavage site]	CCAT	AATG
pEPQD0CM0543	162294	N-terminal tag	GST-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0544	162295	N-terminal tag	MBP-[Factor Xa cleavage site]	CCAT	AATG
pEPQD0CM0545	162296	N-terminal tag	HiBiT-MBP-[Factor Xa cleavage site]	CCAT	AATG
pEPQD0CM0546	162297	N-terminal tag	MBP-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0547	162298	N-terminal tag	TrxA-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0548	162299	N-terminal tag	HiBiT-TrxA-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0549	162300	N-terminal tag	SUMO-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0550	162301	N-terminal tag	HiBiT-SUMO-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0281	162302	N-terminal tag	[S-tag]-[TEV cleavage site]	CCAT	AATG
pEPQD0CM0551	162303	N-terminal tag	HiBiT-[S-tag]-[TEV cleavage site]	CCAT	AATG
pEPMY0SP0002	162304	N-terminal tag	6xHis tag-[HRV 3C cleavage site]	CCAT	AATG
pEPQD0CM0296	162305	CDS	superfolder GFP, no stop codon	AATG	TTCG
pEPQD0CM0539	162306	CDS	TEV protease (S219V), no stop codon	AATG	TTCG
pEPYC0CM0134	162307	C-terminal tag	HiBiT-stop, frame1 (NNT-TCG)	TTCG	GCTT
pEPYC0CM0257	162308	C-terminal tag	HiBiT-stop, frame2 (TTC-G)	TTCG	GCTT
pEPQD0CM0027	162309	C-terminal tag	9–amino acid linker, superfolder GFP, strep tag, stop codon	TTCG	GCTT
pEPQD0CM0028	162310	C-terminal tag	9–amino acid linker, strep tag, stop codon	TTCG	GCTT
pEPQD0CM0029	162311	C-terminal tag	strep tag, stop codon	TTCG	GCTT
pEPQD0CM0030	162312	C-terminal tag	stop codon	TTCG	GCTT
pEPQDKN0248	162313	Reporter, *E. coli* CFPS	T7prom-RBS-sfGFP-HiBiT	n/a	n/a
pEPQDKN0332	162314	Reporter, TNT SP6 WG	SP6prom-EO1-sfGFP-HiBiT	n/a	n/a
pEPQDKN0329	162315	Protease, *E. coli* CFPS	T7prom-RBS-NET-TEVprotease-HiBiT	n/a	n/a
pEPQDKN0729	162316	Protease, *E. coli* CFPS	T7prom-RBS-NET-TEVprotease	n/a	n/a

2.2 Cell-free protein synthesis

All *E. coli* CFPS reactions used a modified PANOx-SP formula (PEP, amino acids, NAD^+, oxalic acid, spermidine, putrescine) (68, 69) as previously described (29) with minor modifications. Lysate was generated from BL21(DE3) *E. coli* cells and used S30 buffer (10 mM Tris acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol (DTT)) for lysate preparation. Ammonium glutamate was not available and not included in the CFPS reaction. Optimal magnesium concentration was determined to be 8 mM based on the expression of the plasmid pJL1-sfGFP (Addgene #69496) encoding the superfolder green fluorescent protein (sfGFP).

Reactions were assembled in a final volume of 2 μl using a Labcyte Echo® 550 (Beckman Coulter, Brea, CA) via a two-step strategy. Initially, 30–50 μl of each plasmid (or water) was distributed into an Echo® Qualified 384-Well Polypropylene Source Microplate

(384PP, P-05525). Subsequently, 65 μl of a master mix containing all remaining CFPS reagents (with lysate added just before distributing) was aliquoted to fresh wells of the same source plate. Next, the Labcyte Echo® Plate Reformat software v1.6.4 was used to direct the distribution of 1735 nl to each well of a FrameStar® 384-Well PCR Plate (4ti-0384/C; 4titude, Wotton, UK) (i.e. the destination plate) using 384PP_AQ_SP2 as a sample plate–type setting. Each source well can support 20–21 reactions. Then, the Labcyte Echo® Cherry Pick software v1.6.4 (guided by a .csv file) used the 384PP_AQ_BP2 sample plate–type setting to direct the distribution of 265 nl containing 26.6 ng plasmid (equivalent to a final concentration of 13.3 ng/μl) with remaining volume water to all destination wells. The destination plate was then centrifuged briefly at room temperature, covered with adhesive aluminum foil and incubated at 30°C for 20 h (lid temperature 37°C) in a Mastercycler pro384 vapo.protect thermocycler (Eppendorf, Hamberg, Germany). Scaled-up (15 μl) CFPS reactions were performed in 1.5 ml DNA LoBind® microcentrifuge tubes (0030108051; Eppendorf). To quantify soluble/total protein, reactions were centrifuged for 10 min at 4°C at 21 000 × *g* to pellet insoluble protein.

All wheat germ cell-free expressions used the TNT® SP6 High-Yield Wheat Germ Protein Expression System L3260/L3261 (Promega) according to manufacturers' instructions. Plasmids were added at a final concentration of 80 ng/μl, and assembled reactions were incubated at 25°C for 20 h.

2.3 Protein quantification

To compare relative protein concentration of proteins containing GFP, 2 μl of CFPS reaction was diluted with 198 μl of 10 mM Tris-HCl pH 7.5 in a black 96-well plate (655076 PS medium binding; Greiner Bio-One Vilvoorde, Belgium). After a double orbital shake for 10 s at 300 rpm, fluorescence was measured (excitation 470 nm, emission 515 nm) using a CLARIOstar plate reader (BMG Labtech Gmbh, Ortenberg, Germany).

Absolute protein concentration of HiBiT-tagged proteins was measured using the Nano-Glo® HiBiT Extracellular Detection System N2420 (Promega). CFPS reactions were diluted 10^4–10^6-fold using 1× PLB + PI buffer (generated by mixing 2 ml of 5× Passive Lysis Buffer E1941 (Promega) with 8 ml water plus one cOmplete™ Mini EDTA-free 11836170001 (Roche, Basel, Switzerland) protease inhibitor tablet). A standard curve of HiBiT Control Protein N3010 (Promega) was diluted in PLB + PI at concentrations ranging from 0.01 to 2 nM. Luminescence reactions were mixed in a white 96-well plate (Greiner Bio-One 655075 PS medium binding) by combined 40 μl of diluted protein (CFPS reaction or standard protein) with 10 μl of luminescence reagent master mix (containing 10 μl of LgBiT protein, 20 μl of 50× substrate and 970 μl of HiBiT buffer). Luminescence values were recorded in a CLARIOstar plate reader every 2.33 min for an hour to ensure the signal is stable over time with the value at 18 min typically used for sample comparison. CFPS reactions assembled on the Labcyte Echo® 550 were diluted 10^2–10^5-fold using 10 mM Tris-HCl pH 7.5 and frozen at –20°C; 4 μl was later thawed and added to 36 μl of PLB + PI along with 10 μl luminescence reagent master mix for measurement.

2.4 TEV protease cleavage of N-terminal tags and detection by Western blot

Plasmids pEPQDKN0329 (with C-terminal HiBiT tag) and pEPQDKN0729 (no C-terminal tag) were expressed using standard CFPS conditions and pEPQDKN0329 amount quantified via HiBiT. CFPS reactions of pEPQDKN0729 were mixed with CFPS reactions containing 2.5 μg protein of pEPQDKN0313 (MBP-sfGFP), pEPQDKN0310 (GST-sfGFP), pEPQDKN0314 (TrxA-sfGFP), pEPQDKN0316 (SUMO-sfGFP) and pEPQDKN0318 (S-tag-sfGFP) at a w/w ratio of 1:130. Cleavage reactions were incubated at 30°C for 16 h, diluted 1:200 in 10 mM Tris-HCl pH 7.5 and 9 μl loaded (along with 1 μl 1 M DTT and 10 μl 2× Laemmli Buffer) into an Any kD™ Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-rad). For Western blot, proteins were transferred to a polyvinylidene difluoride (PVDF) membrane, incubated with α-GFP-HRP (Santa Cruz Biotech GFP, sc-9996 HRP) and imaged using SuperSignal West Pico chemiluminescent substrate (ThermoFisher, Waltham, MA).

2.5 Purification of His-tagged UGT73 C5

Plasmid pEPQDCB0093 was transformed into BL21(DE3) *E. coli* cells. A starter culture of 50 ml of lysogeny broth (LB) media containing 100 μg/ml carbenicillin was inoculated with 2 ml of saturated overnight culture. Cells were grown at 37°C (250 rpm) until $OD_{600} \sim 1.5$–3 and used to inoculate 1000 ml of 2YT media containing 100 μg/ml carbenicillin to a calculated initial OD_{600} of 0.037. Cells were further grown at 37°C (250 rpm) until $OD_{600} = 0.6$ and transferred to a 18°C shaking incubator and allowed to cool. After 1 h at 18°C (250 rpm), 0.5 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression, and cells were incubated at 18°C (250 rpm) overnight. Cells were pelleted by centrifugation at 8000 × *g* for 15 min at 4°C and resuspended in 1× phosphate buffered saline (PBS). Cells were centrifuged again at 4000 × *g* for 15 min at 4°C, supernatant poured off and flash frozen on liquid nitrogen for storage at –80°C. After thawing, cells were resuspended in 50 ml Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% (v/v) glycerol, 20 mM imidazole) supplemented with 10 mg lysozyme (Sigma, 62 971) and one tablet of cOmplete™ EDTA-free protease inhibitor (Roche, 11873580001). Cells were lysed by a single run through a Cell Disruption System CF1 (Constant Systems Limited, Daventry, UK) cell disruptor at 26 kpsi and centrifuged at 40 000 × *g* for 30 min at 4°C to remove debris. Protein was purified using an AKTA Pure HPLC (GE Healthcare) fitted with a HisTrap HP 5 ml column (GE Healthcare) equilibrated with Buffer A. Samples were step-eluted using Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% (v/v) glycerol, 500 mM imidazole). Eluted protein was further purified on a HiLoadTM 16/600 SuperdexTM 200 column (GE Healthcare) and eluted with Buffer A4 (20 mM HEPES, 150 mM NaCl, pH 7.5). Protein was concentrated using a Vivaspin 500 column (Sartorius, Goettingen, Germany) column following the manufacturer's instructions.

2.6 Enzymatic production of glycosides and detection by liquid chromatography–mass spectrometry

In vitro reactions were performed using both purified enzymes and CFPS reactions enriched in the glycosyltransferase. For purified enzymes, the 100-μl reaction contained 100 mM Tris-HCl (pH 8.0), 1 mM UDP-glucose, 0.5 mM substrate (geraniol or *cis-trans*-nepetalactol) suspended in methanol and 1 μM purified AtUGT73C5. Reactions were incubated at 30°C for 1 h. For CFPS-derived enzymes, the 30-μl reaction contained 100 mM Tris-HCl (pH 8.0), 2 mM UDP-glucose, 2 mM substrate (geraniol or *cis-trans*-nepetalactol) suspended in methanol and 15 μl of CFPS reaction. Geraniol (163333) and UDP-glucose (94335) were purchased from Sigma (St. Louis, MO) and *cis-trans*-nepetalactol from Santa Cruz Biotech (sc-506178). Reactions were quenched using 1× volume of methanol and vortexed for 20 s. Reactions were centrifuged at 21 000 × g for 5 min at room temperature, and 50 μl of supernatant was filtered through a 0.22-μm nylon Corning® Costar® Spin-X® filter CLS8169 (Sigma).

For detection of geraniol and nepetalactol glycosides, 2 μl of quenched *in vitro* reaction was injected and separated on a Waters UPLC with an Acquity BEH C18, 1.7-μm (2.1 × 50 mm) column (40°C) at a flow rate of 0.6 ml/min. Mobile phase A was 0.1% formic acid and mobile phase B was acetonitrile. A linear gradient from 5% B to 60% B in 5.5 min and 60–100% B in 0.5 min was applied for compound separation followed by 100% B for 1 min. The flow was returned to 5% B for 2.5 min to re-equilibrate prior to the next injection. Eluting compounds were subjected to positive electrospray ionization (ESI) and analyzed on a Waters Xevo TQ-s (QqQ) using optimized source conditions: cone voltage 30 eV, capillary voltage 3.0 kV, source temperature 150°C, desolvation temperature 450°C, cone gas 150 l/h and desolvation gas 800 l/h. Multiple reaction monitoring (MRM) transitions monitored for analysis included geraniol-glucoside-H (317.2 *m/z*), geraniol glucoside-Na (339.2 *m/z*), geraniol glucoside-NH_4 (334.2 *m/z*), nepetalactol-glucoside-H (331.2 *m/z*), nepetalactol-glucoside-Na (353.2 *m/z*) and nepetalactol-glucoside-NH_4 (348.2 *m/z*).

2.7 Enzymatic production of chrysanthemol and detection by gas chromatography–mass spectrometry

CFPS reactions expressing pEPKK1KN0203 encode chrysanthemol diphosphate synthase (CcCPPase) from *Chrysanthemum cinerariaefolium* (P0C565.2) (70). Enzymatic reactions, based on refs. (70, 71), consisted of 13 μl of CFPS reaction (containing 3.30 ± 0.22 μM soluble protein) along with added 35 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.6, 5 mM $MgCl_2$, 0.5 mM DTT and 2 mM dimethylallyl diphosphate (DMAPP) to a final volume of 100 μl. The reaction was incubated overnight at 30°C and then heated at 95°C for 2 min. Glycine (500 mM pH 10.5), 5 mM $MnCl_2$ and 20 units of calf alkaline phosphatase (NEB) were added to the cooled solution and incubated at 37°C for 1 h. Approximately 0.1 g of NaCl was added, and the terpenoids were extracted by addition of 500 μl tert-butyl methyl ether. Compounds were analyzed using an HP 6890 gas chromatograph with a 5973 MSD (Hewlett Packard/Agilent, Santa Clara, CA) using 1 μl injection, a Zebron™ ZB-5HT Inferno™ capillary column (30 m × 0.25 mm × 0.10 μm + 5 m Guardian) with split vent and helium carrier gas at a constant flow of 1.0 ml/min. The inlet temperature was 200°C, and the initial column temperature was held at 40°C for 0.5 min, increased at 25°C/min to 70°C, increased at 3°C/min to 120°C and finally increased at 50°C/min to 200°C, which was maintained for 3 more minutes. Mass spectra of relevant peaks were compared with NIST database standards to identify chrysanthemol and lavandulol.

2.8 Electrophoretic mobility shift assay

CPFS reactions expressing AtTGA2 (pEPQDKN0742) were buffer exchanged with protein dilution buffer (20 mM Tris-HCl, 50 mM KCl, pH 7.5) using an Amicon Ultra-0.5 centrifugal filter (UFC-501024; Merck, Kenilworth, NJ). DNA probes were synthesized as oligos (TGA TFBS 5′-gacccctattgcagctatttcacCTGACGTAAGG GATGACGCACAggccatcacgcagta, Random control 5′-gacccctat tgcagctatttcacacataccaacgcttagcgcaatggccatcacgcagta). A primer labeled with Alexa-488 (5′-Alexa488-tactgcgtgatggcc) was designed to anneal to the 3′ end of the probe oligos. A full-length, double-stranded Alexa-488-labeled probe was produced using DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210), following the manufacturer's protocol. About 500 pmol protein and 5.5 pmol probe were mixed with electrophoretic mobility shift assay (EMSA) reaction mixture (20 mM Tris-HCl pH 8.0, 50 mM KCl, 2 mM DTT, 1 mM EDTA, 500 ng poly(dI-dC), 5% glycerol, 0.05% IGEPAL CA630, 0.1 mg/ml BSA). The mixture was incubated at room temperature for 1 h. Bound and unbound probes were separate on a 6% TBE acrylamide gel (Invitrogen, EC6265BOX) and visualized using a ChemiDocTM Touch imaging system (Bio-Rad, Hercules, CA).

3. Results

3.1 A modular type IIs cloning toolbox for cell-free protein synthesis

To enable a pipeline for cell-free expression of plant proteins, we first built a suite of plasmid vectors amenable to miniaturization and automation and able to utilize DNA parts in the phytobrick standard (Figure 1A). We built plasmid toolkits for two cell-free systems: the *E. coli*–based PANOx-SP (29, 68, 69) (driven by a T7 promoter) and the Promega TNT® SP6 High-Yield Wheat Germ Protein Expression System (driven by the SP6 promoter) (37). Using high copy number plasmid backbones, we constructed a suite of 10 different acceptor plasmids with minimal regulatory sequences and strong terminators for both systems, each capable of utilizing Level 0 phytobrick N-tag, CDS and C-tag parts (Figure 1B, Table 1). All plasmids have been deposited in the Addgene plasmid repository.

To expand the utility of the DNA assembly toolbox, we made a library of parts encoding N-terminal and C-terminal tags for purification, detection and improved expression. Facile detection of proteins is important for high-throughput CFPS applications, and although sfGFP was optimized to reduce interference with the folding of its fusion partners, bulky fluorescent protein tags are not always tolerated. To accurately measure protein expression without a GFP fusion or using expensive radiolabeled amino acids, we adapted the HiBiT system sold by Promega. This requires only a minimal 11–amino acid tag to be genetically encoded. Following protein expression, an 18-kDa engineered polypeptide with high affinity for the tag is added, resulting in a complex with luciferase activity proportional to the level of HiBiT-tagged protein. To assess the ability of HiBiT to quantify protein in CFPS reactions, we measured the *E. coli* cell-free expression of sfGFP fused with a C-terminal HiBiT tag (pEPQDKN0248) to be 41.8 ± 3.0 μM (1140 ± 84 μg/ml) (Figure 2A, second bar from top), which is consistent with other PANOx-SP systems using BL21(DE3) (69, 72). The HiBiT tag works for quantification on both the N- and C-terminus of a target protein; however, we found that inclusion on the N-terminus reduced cell-free protein yield when measuring relative GFP fluorescence (Figure 2A).

Several fusion tags have been shown to improve translation, solubility and folding of recombinant proteins; however, it is rarely obvious which expression tag is optimal for a given protein. We therefore included a number of different tags in the toolkit. We initially assessed the functionality of S-tag, SUMO, thioredoxin (TRX), GST and MBP tags by assembling them with the sfGFP CDS and doing *E. coli* CFPS reactions (Figure 2A). We found that many N-terminal tags appeared to express well, although the larger tags (such as MBP) reduced expression. As expected, they did not increase the expression of sfGFP, for which the CFPS reaction conditions have been optimized (34, 69).

As many downstream protein applications require the removal of tags that may interfere with, for example, enzyme functionality, we developed a low-cost method for tag cleavage. First, we expressed the TEV protease S219V using CFPS. Subsequently, this reaction was mixed with the reaction containing the target protein containing the cleavage site (Supplementary Figure

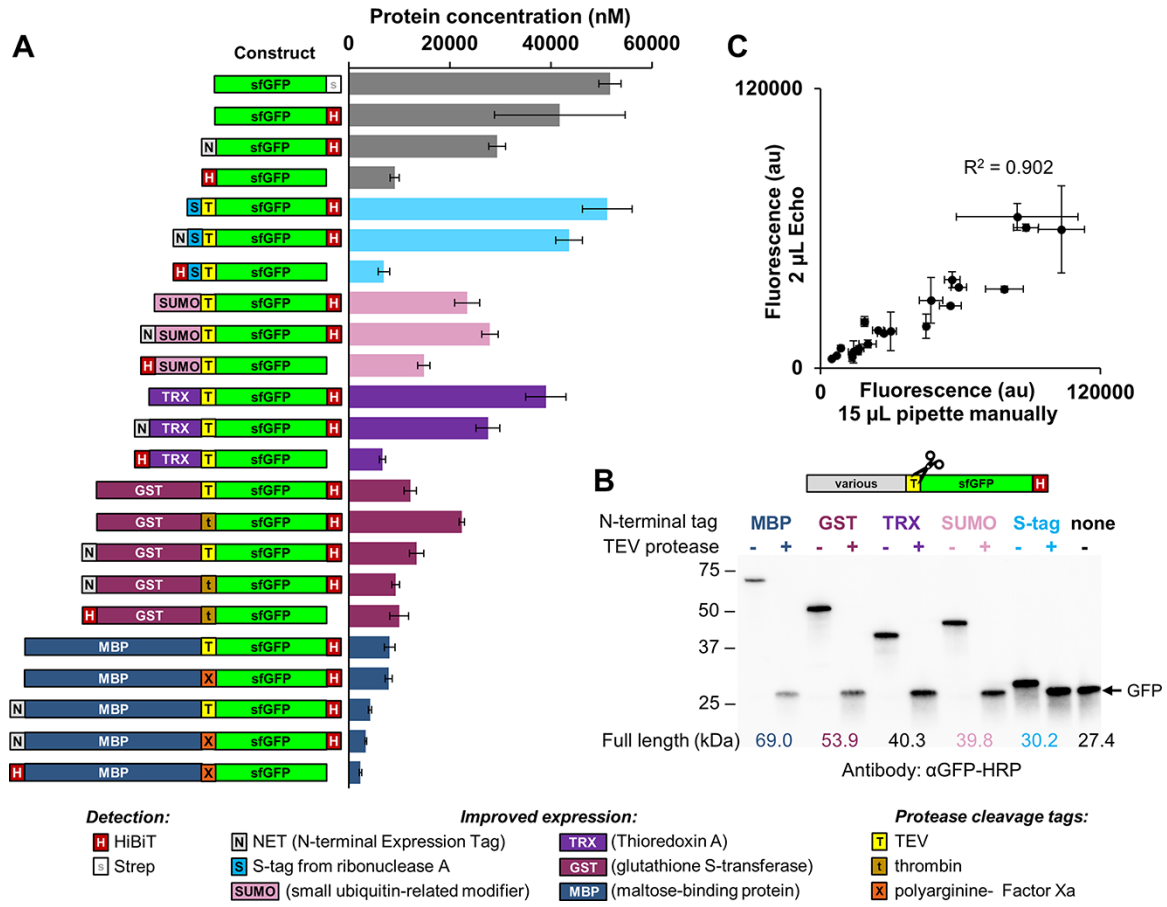

Figure 2. (A) Cell-free expression of sfGFP fused to a variety of N- and C-terminal tags. CFPS reactions were run at the 15-µl scale and incubated for 20 h at 30°C. Relative protein expression was measured by GFP fluorescence and normalized to pEPQDKN0248 quantified by HiBiT. (B) Removal of N-terminal tags by mixing a CFPS reaction expressing TEV protease (1:130 w/w ratio of TEV protease to substrate protein) and incubating at 16 h at 30°C. Protein visualized by western blot with an α-GFP-HRP antibody. (C) Expression of GFP variants from panel A with reactions assembled manually using handheld pipettes (15-µl reactions) or automated using an Echo 550 acoustic liquid handler (2-µl reactions). Values in panels A and C represent averages ($n = 3$), and error bars represent 1 SD.

S1). Mixing the two reactions overnight at a 1:130 w/w ratio of protease:target (~1:2 v/v ratio of cell-free reactions) showed efficient cleavage when assayed by western blot using an anti-GFP antibody (Figure 2B) and did not affect enzyme activity (Supplementary Figure S1C).

3.2 Sequential improvement of automated cell-free protein synthesis

To enable the progression of high-throughput combinatorial experiments, for example, to select tags that enable the best yields for individual proteins, we established an automated workflow for low-volume CFPS. Our aim was to achieve consistent results across replicates and to ensure that the yields obtained from low-volume reactions correlated with those from large-scale reactions, thus demonstrating that conditions established using biofoundries are transferable. To maximize consistency, we pre-mixed all cell-free reaction components into a master mix, which was distributed to all wells of a 384-well plate to which the plasmid template was subsequently added. We first found that calibration of plate type settings was required to enable nanovolumes of master mix to be accurately transferred by the acoustic droplet technology of the Labcyte Echo (Supplementary Figure S2A and B). We then optimized the number of destination wells to which the reaction mix was transferred from each source well to reduce the number of failed reactions (Supplementary Figure S2C). Finally, to reduce the time taken for 384 CFPS reactions to be assembled while maximizing the consistency and yield of protein between reactions, we compared four different reaction volumes (Supplementary Figure S2D) as well software protocols that monitor reagent transfer (Supplementary Figure S2E).

In summary, we found that a single specific plate setting (384PP_AQ_SP2) was essential for consistent distribution and that expression levels obtained from 2000-nl reactions were the most consistent and produced the largest amount of protein. Expression of GFP fusion protein was found to be consistent across all wells of a 384-well plate, and the yields obtained correlated well ($R^2 = 0.902$) with assembly by a manual pipette (15 µl) (Figure 2C). This suggests that automated reaction assembly is an appropriate method for quantifying and comparing expression from different DNA templates.

3.3 Selecting optimal configurations for expression of plant proteins

To test our DNA and cell-free assembly workflow on non-model CDSs, we first selected a range of plant UDP glycosyltransferases (UGTs). UGTs are a superfamily of enzymes that catalyze the addition of glycosyl groups, including many plant specialized

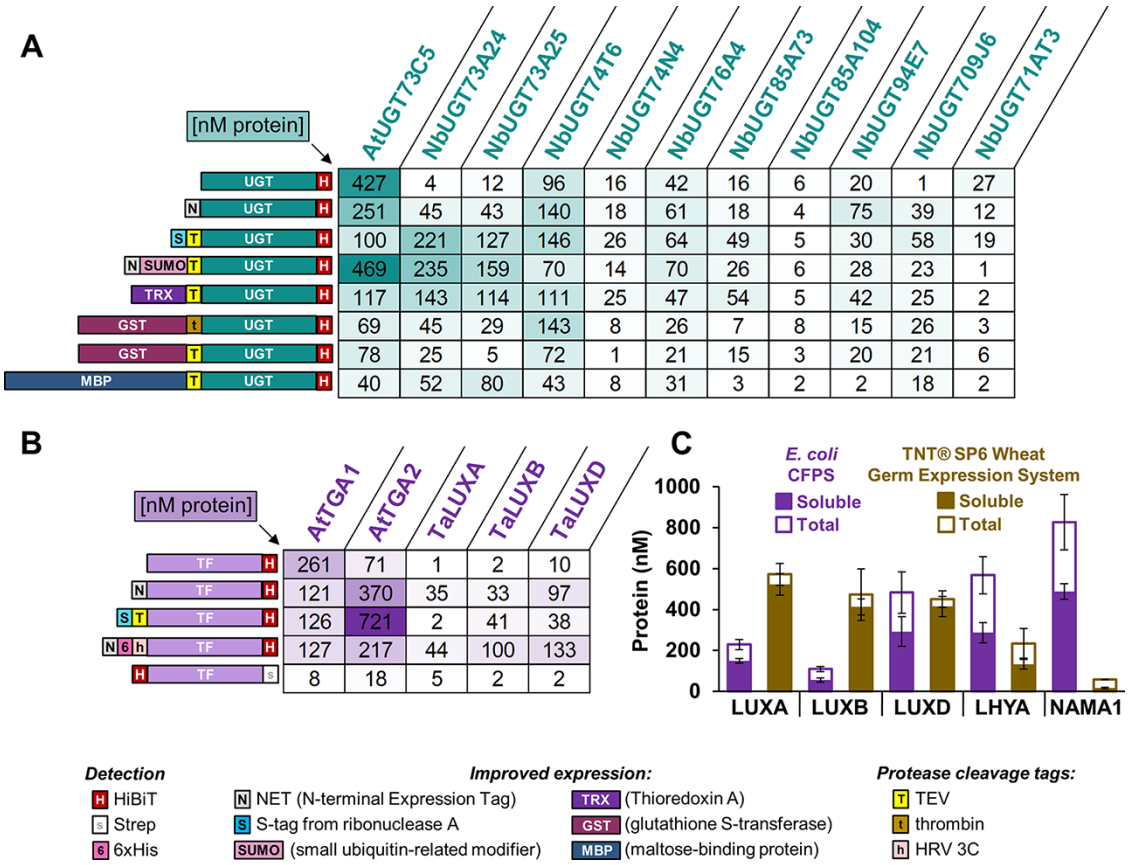

Figure 3. (A–B) HiBiT quantification of plant proteins with various N-terminal tags with cell-free protein synthesis reactions assembled using acoustic liquid handling platforms. (A) Expression of 11 plant UGTs, each with eight different N-terminal tags. (B) Expression of five plant transcription factors, each with five different N-terminal tags. (C) Comparison of *E. coli* CFPS (with NET tag) and the TNT® SP6 High-Yield Wheat Germ Protein Expression System (Promega) for expressing wheat transcription factors at the 15-μl scale. Values represent averages ($n = 3$), and error bars represent 1 SD.

metabolites. Although UGTs can easily be identified from their primary sequence, functional assays are typically required to identify which family members are active on a specific substrate. We included UGT73C5 from *Arabidopsis thaliana*, together with 10 uncharacterized UGTs from *N. benthamiana*. We constructed eight different versions of 11 plant UGTs (88 total plasmids), assembled 2000-nl cell-free reactions and measured total protein level using the HiBiT assay (Figure 3A). In general, we observed that SUMO and S-tags improved the expression of most UGTs; however, several did not express well in any context and may require further optimization or an alternate expression strategy. The GST fusion produced high protein levels for only a handful of UGTs (UGT74T6). Finally, we manually assembled 15-μl reactions for a dozen different UGT-encoding plasmids and again obtained a strong correlation with the low-volume automated reaction (Supplementary Figure S3).

After demonstrating that biofoundry protocols can be used to select optimal construct configurations, we proceeded to test the expression of different classes of proteins. We first tested additional enzymes. CFPS has previously been demonstrated for monoterpene synthases that catalyze a regular head-to-tail 1–4 linkage, for example, geranyl diphosphate synthase or farnesyl diphosphate synthase (29). We successfully used CFPS to express chrysanthemol diphosphate synthase (CcCPPase) from *Chrysanthemum cinerariaefolium* (Supplementary Figure S4), which catalyzes an irregular, non-head-to-tail 1–2 linkage between molecules of DMAPP and also hydrolyzes the diphosphate moiety to produce chrysanthemol that contains a cyclopropane ring and is a precursor to pyrethrins, a widely used class of plant pesticides (73).

We then tested the expression of plant TF proteins, which are of great interest to plant synthetic biologists aiming to engineer complex gene regulatory networks known to control important agricultural traits or to link the synthetic circuits to endogenous processes. Most plants contain large TF families, and many of their cognate DNA sequences remain unknown. We selected two TGA-family TFs from *Arabidopsis thaliana* (TGA) and three TFs from *Triticum aestivum* (bread wheat). All five TFs were expressed using *E. coli* CFPS with five different N-terminal configurations (Figure 3B). The NET and NET-6xHis tags were the most consistent across multiple TFs, but the S-tag proved remarkably helpful in expressing TGA2. To further assess the cell-free expression of TFs from wheat, we wanted to compare *E. coli* CFPS with expression using TNT SP6 High-Yield Wheat Germ Protein Expression System (Promega). As a first test, we first assembled GFP into three different plasmid acceptors containing different combinations of terminator and antibiotic resistance sequences and found that folA terminator from the pEU vector outperformed the T7 terminator (Supplementary Figure S5). We then cloned the three LUX protein CDSs (along with LHYA and NAMA1) into the optimal wheat germ acceptor and measured soluble and total protein expression using the HiBiT assay. While both systems produced

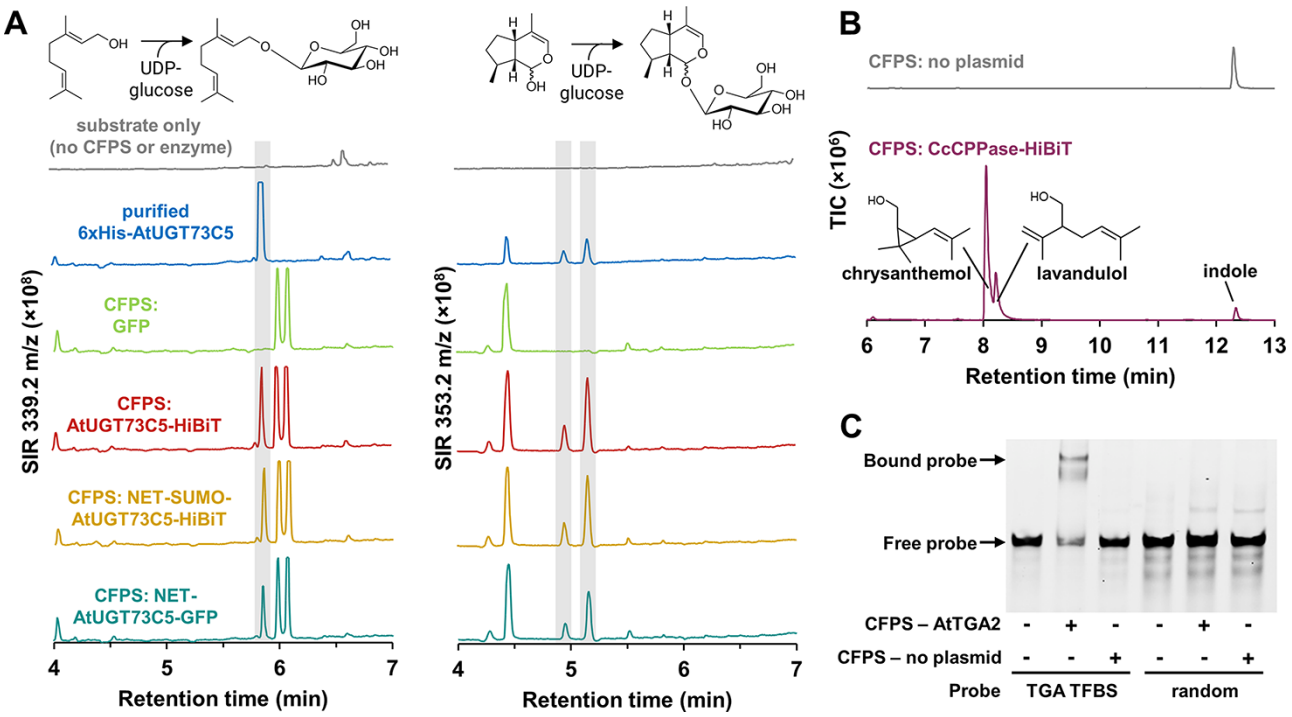

Figure 4. Plant proteins expressing using *E. coli* cell-free protein synthesis are functionally active. (A) UDPglycosyltransferase UGT73C5 from *Arabidopsis thaliana* can glycosylate geraniol or nepetalactol when purified (blue trace) or when expressed using CFPS (red, orange, teal trace) and (B) chrysanthemol diphosphate synthase (CcCPPase) from *Chrysanthemum cinerariaefolium* converts substrate DMAPP to chrysanthemol and lavandulol. (C) Cell-free-expressed AtTGA2 binds a 60-bp DNA probe containing its cognate-binding site but does not impede the mobility of a probe with randomized sequence.

folded protein, the wheat germ kit generally produced higher levels of soluble protein with the exception of NAMA1, only expressed in the *E. coli* system (Figure 3C).

3.4 Cell-free expressed proteins are functionally active

Having identified conditions from which we could obtain a good yield of several proteins, we wanted to test if the proteins obtained were functionally active and whether assays could be conducted without extensive purification protocols. To test UGT enzymatic activity, we expressed UGT73C5 in 15-μl reactions and incubated the reactions with UDP-glucose along with geraniol or *cis-trans*-nepetalactol. Reactions were quenched at 1 h (Supplementary Figure S6) and then analyzed by liquid chromatography–mass spectrometry (LC–MS), which monitored possible glucoside-substrate adducts (Supplementary Figures S7 and S8). Cell-free reactions expressing UGT73C5 in three different tag configurations produced glucosides of both substrates with peaks matching the retention time of glucosides produced by purified UGT73C5 and not present in the GFP negative control (Figure 4A). The cell-free reactions, including GFP negative control, contain additional peaks likely derived from unspecified compounds in the *E. coli* lysate. We were also able to detect the synthesis of chrysanthemol by gas chromatography–mass spectrometry (GC–MS) after incubating cell-free reactions expressing CcCPPase with DMAPP (Figure 4B), obtaining the expected fragmentation pattern (Supplementary Figure S9).

To test the binding ability of the cell-free expressed protein to DNA, CFPS-derived AtTGA2 was incubated with a DNA sequence probe containing a known binding site from the cauliflower mosaic virus (CaMV) 35S promoter (74) and DNA–protein interactions visualized by EMSA. Cell-free expressed TGA2 successfully bound to the probe containing its binding site but not the random control sequence (Figure 4C), suggesting that the DNA-binding domain of the cell-free expressed TGA2 was correctly folded and functional.

4. Discussion

Biofoundries provide a powerful combination of automation platforms and synthetic biology approaches, including the application of engineering principles and computational modeling and analysis to aid the design and evaluation of large datasets. Consequently, they can significantly increase the scale and throughput for experiments testing a given biological problem or question (12–14). Building on earlier works in which we established automation-compliant DNA assembly tools and protocols for the assembly of constructs for engineering plants systems (3, 5, 10), we have developed a toolbox for CFPS compatible with DNA parts used by the plant community (Figure 1, Table 1). This enabled us to obtain useful yields of both enzymes and regulatory proteins, progressing directly to characterization experiments. The resulting toolbox contains 37 plasmids, including acceptors for T7-driven *E. coli* CFPS as well as commercial wheat germ protein expression. While other Golden Gate assembly systems could potentially be used for *E. coli* cell-free expression, they are customized for specific applications (75–77) and are incompatible with DNA parts in the Phytobrick standard. While, for some proteins, optimal expression levels might be obtained by system-specific codon-optimization, optimal expression is

generally unnecessary for rapid characterization and is outweighed by the ability to reuse a large number of parts in prototyping experiments.

Over the past few decades, CFPS has been successfully used to express and prototype bacterial proteins including DNA regulatory elements (50, 77–79), metabolic sensors (80, 81) and ribosomal peptides (82). It has also been applied to metabolic pathways from a range of organisms, mainly to prototype biosynthesis for production in microbial chassis (29, 30, 41, 45, 83, 84). More recently, the advantages of this technology have been applied to enabling the expression of proteins that can be challenging to express, such as cytotoxic proteins and glycoproteins (26, 28, 32, 33, 85, 86). Automated liquid handling has the general advantage of increasing throughput, reducing reaction volumes and improving repeatability. These advantages have already been applied to assembling the components of CFPS reactions (47–54). In this study, we coupled automated nanoscale DNA assembly workflows to CFPS to progress high-throughput experiments that enable the selection of optimal configurations for the expression of multiple plant proteins. We show how software settings, reaction volumes and assembly time are important for consistent cell-free expression and optimize multiple parameters to obtain consistent, repeatable yields from low-volume reactions (Figure 2, Supplementary Figure S2). Our workflows were aided by the use of the high-affinity, 11–amino acid HiBiT tag (87), which is able to complex with a 18-kDa engineered NanoLuc polypeptide (88) and luminesce proportional to the level of HiBiT-tagged protein. We found that this worked well as a C-terminal tag but reduced yields when fused to the N-terminus of sfGFP (Figure 2), perhaps because the HiBiT amino acid sequence is not optimal for early translational elongation (89). Although sfGFP has been optimized to improve folding (90), the minimal size of the HiBiT tag may be less likely to inhibit protein function and accessibility. However, alternatives may be required for proteins in which the C-terminus needs to be folded inside the protein or must be freely available for activity or post-translational modifications. To assist this, we demonstrate that tags can be removed after synthesis by simply incubating with a parallel reaction expressing TEV protease (Figure 2C).

We then applied the twin capabilities of automated high-throughput DNA assembly and low-volume CFPS to rapidly select the tag configurations that resulted in useful levels of expression of different plant proteins. Although we did not expect to find a single configuration that was generally applicable to multiple types of proteins, there was surprisingly little consistency between even closely related proteins. Tags that significantly improved the yields of some family members resulted in little or no expression of others (Figure 3A). This demonstrates the utility of automated high-throughput and low-volume screens to select optimal configurations.

Cell-free expression has yet to be widely applied in plant science but has had an impact in the development of next-generation sequencing techniques such as DNA affinity purification sequencing. In this technique, wheat germ-based cell-free expression systems have been used to produce affinity tagged TFs that are used to capture genomic DNA enabling the identification of TF-binding sites (91, 92). However, when applied to genome-scale collections of TFs, insufficient expression was obtained for several hundred proteins (91). We expressed TGA TFs, known to regulate the expression of defense-related genes (93–97) as well as the CaMV 35S promoter (74, 98–102). We also expressed wheat homologs of TFs that regulate circadian rhythms in Arabidopsis including homologs of LUX ARRHYTHMO (LUX) from each of the three wheat sub-genomes (103). The *E. coli* lysate system provides substantial cost benefits (20-fold less than wheat germ, see Supplementary Figure S10), and although better yields were obtained with the wheat-germ system for three TFs of wheat origin, the *E. coli* system proved beneficial two additional TFs (LHYA and NAMA1) (Figure 3C). We do not, however, expect success with all protein classes and did not, for example, attempt the synthesis of plant proteins that are widely known to express poorly in non-eukaryotes such as cytochrome P450s and oxidosqualine cyclases, which in their native cells are localized to lipid droplets or embedded into the endoplasmic reticulum (104, 105).

Plants synthesize a diverse array of metabolites that contribute to adaptation to ecological niches, serving as attractants for beneficial organisms and providing defense against biotic and abiotic agents (106). Metabolic profiling has been widely applied to assess this diversity, identifying a number of molecules that have been leveraged for use in industry and medicine. However, the genetic basis of biosynthesis for many molecules remains unknown. With many genomes now sequenced and many more underway, a major challenge is to assign the function to sequence, which is particularly challenging for large enzyme families. Sequence similarity allows the classification of enzymes into large and complex superfamilies, but the highly specific reactions, substrates and products that enzymes catalyze remain difficult to predict. For example, there is interest in characterizing the substrate specificity of plant glycosyltransferases as it has been observed that some metabolites are over-glycosylated by native enzymes present in *N. benthamiana* hindering yield (107–110). A significant advantage of cell-free expression was the ability to progress directly to functional analysis without time-consuming cell-disruption and purification protocols. We were able to express and demonstrate the expected activity of an Arabidopsis UGT that has previously been leveraged for overproduction of triterpene saponins (111–113). This workflow is ideally suited to enzymes from plant or microbial secondary metabolism. However, enzymes present in the source lysate may interfere with metabolic reactions of the expressed enzyme, particularly if a cognate of the enzyme of interest is present in the *E. coli* or wheat germ lysate (114). To alleviate this challenge, users could add purification steps to their cell-free expression protocol or utilize source strains deficient in the enzyme of interest (115). Previous studies have shown that some cell-free expressed enzymes exhibit equivalent activity to enzymes produced in cells and subsequently purified (116); however, kinetic characterization experiments will be compromised if lysate proteins that utilize similar substrates or products to the enzyme of interest are present.

The approaches described in this manuscript can be applied to a range of diverse proteins. High-throughput, low-cost experimental pipelines will be useful both for directly assigning function to novel sequences and also for the creation of large datasets that can train machine learning algorithms to predict the characteristics such as specificity from the primary sequence (117, 118). Further, automated reaction assembly may be a helpful approach for adjusting cell-free reaction conditions unique to a given protein of interest (49, 85).

Supplementary data

Supplementary data are available at SYNBIO Online.

Data availability

Plasmids are available from Addgene (#162281-162317). Nucleotide sequences and corresponding accession numbers are provided in the supplementary data.

Funding

The authors acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC), and the Engineering and Physical Sciences Research Council (EPSRC), part of UK Research and Innovation; this research was funded by BBSRC Core Strategic Programme Grant (Genomes to Food Security) BB/CSP1720/1; National Capability Grant BBS/E/T/000PR9815 (Earlham Biofoundry); BB/P010490/1, an industrial partnership award with Leaf Expression Systems; BB/R021554; BBSRC/EPSRC OpenPlant Synthetic Biology Research Centre Grant (BB/L014130/1). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgments

We thank Susan Duncan and Hannah Rees for the provision of coding sequences for wheat transcription factors. We also thank Mark Youles for contributing help with plasmid construction and Mohamed Omar Kamileen for assisting with LC-MS method development. pJL1 was a gift from Michael Jewett (Addgene plasmid # 69496; http://n2t.net/addgene:69496; RRID:Addgene_69496). pEU-Gm23 was a gift from Gabor Igloi (Addgene plasmid # 53738; http://n2t.net/addgene:53738; RRID:Addgene_53738)

Author contributions

Q.M.D. and N.J.P. conceived the study. Q.M.D and H.D. developed the acceptor parts. Q.M.D. expressed tagged GFP variants, ran Echo optimization reactions and tested UGTs using LC-MS. J.A.C.L assisted with the development of automation protocols. Y.C. expressed TGA transcription factors and performed EMSA. K.K. expressed CPPase and performed GC–MS. Q.M.D. and N.J.P. wrote the manuscript. N.J.P obtained funding and provided supervision.

Conflict of interest statement. None declared.

References

1. Patron,N.J. (2020) Beyond natural: synthetic expansions of botanical form and function. *New Phytol.*, **227**, 295–310.
2. Wurtzel,E.T., Vickers,C.E., Hanson,A.D., Millar,A.H., Cooper,M., Voss-Fels,K.P., Nikel,P.I. and Erb,T.J. (2019) Revolutionizing agriculture with synthetic biology. *Nat. Plants*, **5**, 1207–1210.
3. Patron,N.J., Orzaez,D., Marillonnet,S., Warzecha,H., Matthewman,C., Youles,M., Raitskin,O., Leveau,A., Farré,G., Rogers,C. *et al.* (2015) Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol.*, **208**, 13–19.
4. Vazquez-Vilar,M., Orzaez,D. and Patron,N. (2018) DNA assembly standards: setting the low-level programming code for plant biotechnology. *Plant Sci.*, **273**, 33–41.
5. Engler,C., Youles,M., Gruetzner,R., Ehnert,T.-M., Werner,S., Jones,J.D., Patron,N.J. and Marillonnet,S. (2014) A golden gate modular cloning toolbox for plants. *ACS Syn. Biol.*, **3**, 839–843.
6. Pollak,B., Cerda,A., Delmans,M., Álamos,S., Moyano,T., West,A., Gutiérrez,R.A., Patron,N.J., Federici,F. and Haseloff,J. (2019) Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.*, **222**, 628–640.
7. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palací,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.*, **162**, 1618–1634.
8. Andreou,A.I. and Nakayama,N. (2018) Mobius assembly: a versatile golden-gate framework towards universal DNA assembly. *PLoS One*, **13**, e0189892.
9. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**, e3647.
10. Cai,Y.-M., Lopez,J.A.C. and Patron,N.J. (2020) Phytobricks: manual and automated assembly of constructs for engineering plants. In: Chandran S, George KW (eds). *DNA Cloning and Assembly*. Springer, Cham, Switzerland, pp. 179–199.
11. Kanigowska,P., Shen,Y., Zheng,Y., Rosser,S. and Cai,Y. (2016) Smart DNA fabrication using sound waves: applying acoustic dispensing technologies to synthetic biology. *J. Lab. Autom.*, **21**, 49–56.
12. Chao,R., Mishra,S., Si,T. and Zhao,H. (2017) Engineering biological systems using automated biofoundries. *Metab. Eng.*, **42**, 98–108.
13. Hillson,N., Caddick,M., Cai,Y., Carrasco,J.A., Chang,M.W., Curach,N.C., Bell,D.J., Le Feuvre,R., Friedman,D.C., Fu,X. *et al.* (2019) Building a global alliance of biofoundries. *Nat. Commun.*, **10**, 2040.
14. Holowko,M.B., Frow,E.K., Reid,J.C., Rourke,M. and Vickers,C.E. (2020) Building a biofoundry. *Synth. Biol.*, **6**, ysaa026.
15. Cravens,A., Payne,J. and Smolke,C.D. (2019) Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.*, **10**, 2142.
16. Park,S.-Y., Peterson,F.C., Mosquna,A., Yao,J., Volkman,B.F. and Cutler,S.R. (2015) Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature*, **520**, 545–548.
17. South,P.F., Cavanagh,A.P., Liu,H.W. and Ort,D.R. (2019) Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, **363**, eaat9077.
18. Vaidya,A.S., Helander,J.D., Peterson,F.C., Elzinga,D., Dejonghe,W., Kaundal,A., Park,S.-Y., Xing,Z., Mega,R. and Takeuchi,J. (2019) Dynamic control of plant water use using designed ABA receptor agonists. *Science*, **366**, eaaw8848.
19. Buyel,J.F., Twyman,R.M. and Fischer,R. (2017) Very-large-scale production of antibodies in plants: the biologization of manufacturing. *Biotechnol. Adv.*, **35**, 458–465.
20. Reed,J., Stephenson,M.J., Miettinen,K., Brouwer,B., Leveau,A., Brett,P., Goss,R.J., Goossens,A., O'Connell,M.A. and Osbourn,A. (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.*, **42**, 185–193.
21. Schillberg,S., Raven,N., Spiegel,H., Rasche,S. and Buntru,M. (2019) Critical analysis of the commercial potential of plants for the production of recombinant proteins. *Front. Plant Sci.*, **10**, 720.
22. Rattanapisit,K., Shanmugaraj,B., Manopwisedjaroen,S., Purwono,P.B., Siriwattananon,K., Khorattanakulchai,N., Hanittinan,O., Boonyayothin,W., Thitithanyanont,A., Smith,D.R. *et al.* (2020) Rapid production of SARS-CoV-2 receptor binding domain (RBD) and spike specific monoclonal antibody CR3022 in. *Nicotiana Benthamiana*. *Sci. Rep.*, **10**, 17698.
23. Ward,B.J., Gobeil,P., Seguin,A., Atkins,J., Boulay,I., Charbonneau,P.-Y., Couture,M., D'Aoust,M.-A., Dhaliwall,J.,

Finkle,C. *et al.* (2021) Phase 1 randomized trial of a plant-derived virus-like particle vaccine for COVID-19. *Nat Med.*, **27**, 1071–1078.
24. Pillet,S., Aubin,É., Trépanier,S., Bussière,D., Dargis,M., Poulin,J.-F., Yassine-Diab,B., Ward,B.J. and Landry,N. (2016) A plant-derived quadrivalent virus like particle influenza vaccine induces cross-reactive antibody and T cell response in healthy adults. *Clin. Immunol.*, **168**, 72–87.
25. Marsian,J., Fox,H., Bahar,M.W., Kotecha,A., Fry,E.E., Stuart,D.I., Macadam,A.J., Rowlands,D.J. and Lomonossoff,G.P. (2017) Plant-made polio type 3 stabilized VLPs—a candidate synthetic polio vaccine. *Nat. Commun.*, **8**, 245.
26. Gregorio,N.E., Levine,M.Z. and Oza,J.P. (2019) A user's guide to cell-free protein synthesis. *Methods Protoc.*, **2**, 24.
27. Carlson,E.D., Gan,R., Hodgman,C.E. and Jewett,M.C. (2012) Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.*, **30**, 1185–1194.
28. Silverman,A.D., Karim,A.S. and Jewett,M.C. (2019) Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.*, **21**, 151–170.
29. Dudley,Q.M., Karim,A.S., Nash,C.J. and Jewett,M.C. (2020) *In vitro* prototyping of limonene biosynthesis using cell-free protein synthesis. *Metab. Eng.*, **61**, 251–260.
30. Karim,A.S., Dudley,Q.M., Juminaga,A., Yuan,Y., Crowe,S.A., Heggestad,J.T., Garg,S., Abdella,T., Grubbe,W.S., Rasor,B.J. *et al.* (2020) *In vitro* prototyping and rapid optimization of biosynthetic enzymes for cellular design. *Nat. Chem. Biol.*, **16**, 912–919.
31. Yin,G., Garces,E.D., Yang,J., Zhang,J., Tran,C., Steiner,A.R., Roos,C., Bajad,S., Hudak,S., Penta,K. *et al.* (2012) Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs*, **4**, 217–225.
32. Kightlinger,W., Duncker,K.E., Ramesh,A., Thames,A.H., Natarajan,A., Stark,J.C., Yang,A., Lin,L., Mrksich,M. and DeLisa,M.P. (2019) A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.*, **10**, 5404.
33. Hershewe,J., Kightlinger,W. and Jewett,M.C. (2020) Cell-free systems for accelerating glycoprotein expression and biomanufacturing. *J. Ind. Microbiol. Biotechnol.*, **47**, 977–991.
34. Martin,R.W., Des Soye,B.J., Kwon,Y.-C., Kay,J., Davis,R.G., Thomas,P.M., Majewska,N.I., Chen,C.X., Marcum,R.D., Weiss,M.G. *et al.* (2018) Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat. Commun.*, **9**, 1203.
35. Des Soye,B.J., Gerbasi,V.R., Thomas,P.M., Kelleher,N.L. and Jewett,M.C. (2019) A highly productive, one-pot cell-free protein synthesis platform based on genomically recoded. *Escherichia Coli*. *Cell Chem. Biol.*, **26**, 1743–1754.
36. Oza,J.P., Aerni,H.R., Pirman,N.L., Barber,K.W., Ter Haar,C.M., Rogulina,S., Amrofell,M.B., Isaacs,F.J., Rinehart,J. and Jewett,M.C. (2015) Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nat. Commun.*, **6**, 8168.
37. Harbers,M. (2014) Wheat germ systems for cell-free protein expression. *FEBS Lett.*, **588**, 2762–2773.
38. Murota,K., Hagiwara-Komoda,Y., Komoda,K., Onouchi,H., Ishikawa,M. and Naito,S. (2011) Arabidopsis cell-free extract, ACE, a new in vitro translation system derived from Arabidopsis callus cultures. *Plant Cell Physiol.*, **52**, 1443–1453.
39. Buntru,M., Vogel,S., Stoff,K., Spiegel,H. and Schillberg,S. (2015) A versatile coupled cell-free transcription–translation system based on tobacco BY-2 cell lysates. *Biotechnol. Bioeng.*, **112**, 867–878.
40. Chiocchini,C., Vattem,K., Liss,M., Ludewig,L., Reusch,T., Rastogi,I., Webb,B. and Trefzer,A. (2020) From electronic sequence to purified protein using automated gene synthesis and in vitro transcription/translation. *ACS Syn. Biol.*, **9**, 1714–1724.
41. Kelwick,R., Ricci,L., Chee,S.M., Bell,D., Webb,A.J. and Freemont,P.S. (2018) Cell-free prototyping strategies for enhancing the sustainable production of polyhydroxyalkanoates bioplastics. *Synth. Biol.*, **3**, ysy016.
42. Liu,R., Zhang,Y., Zhai,G., Fu,S., Xia,Y., Hu,B., Cai,X., Zhang,Y., Li,Y., Deng,Z. *et al.* (2020) A cell-free platform based on nisin biosynthesis for discovering novel lanthipeptides and guiding their overproduction in vivo. *Adv. Sci.*, **7**, 2001616.
43. Bogart,J.W., Cabezas,M.D., Vögeli,B., Wong,D.A., Karim,A.S. and Jewett,M.C. (2020) Cell-free exploration of the natural product chemical space. *ChemBioChem.*, **22**, 84–91.
44. Rolf,J., Rosenthal,K. and Lütz,S. (2019) Application of cell-free protein synthesis for faster biocatalyst development. *Catalysts*, **9**, 190.
45. Khatri,Y., Hohlman,R.M., Mendoza,J., Li,S., Lowell,A.N., Asahara,H. and Sherman,D.H. (2020) Multicomponent microscale biosynthesis of unnatural cyanobacterial indole alkaloids. *ACS Syn. Biol.*, **9**, 1349–1360.
46. Fleming,S.R., Himes,P.M., Ghodge,S.V., Goto,Y., Suga,H. and Bowers,A.A. (2020) Exploring the post-translational enzymology of PaaA by mRNA display. *J. Am. Chem. Soc.*, **142**, 5024–5028.
47. Marshall,R., Garamella,J., Noireaux,V. and Pierson,A. (2018) High-throughput microliter-sized cell-free transcription-translation reactions for synthetic biology applications using the Echo® 550 liquid handler. *Labcyte App. Note*, http://noireauxlab.org/html%20pages/docs%20website/publications/Marshall%20et%20al%20-%202018B.pdf.
48. Cole,S.D., Beabout,K., Turner,K.B., Smith,Z.K., Funk,V.L., Harbaugh,S.V., Liem,A.T., Roth,P.A., Geier,B.A. and Emanuel,P.A. (2019) Quantification of interlaboratory cell-free protein synthesis variability. *ACS Syn. Biol.*, **8**, 2080–2091.
49. Borkowski,O., Koch,M., Zettor,A., Pandi,A., Batista,A.C., Soudier,P. and Faulon,J.-L. (2020) Large scale active-learning-guided exploration for in vitro protein production optimization. *Nat. Commun.*, **11**, 1872.
50. Moore,S.J., MacDonald,J.T., Wienecke,S., Ishwarbhai,A., Tsipa,A., Aw,R., Kylilis,N., Bell,D.J., McClymont,D.W., Jensen,K. *et al.* (2018) Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. USA*, **115**, E4340–E4349.
51. Marshall,R. and Noireaux,V. (2019) Quantitative modeling of transcription and translation of an all-*E. coli* cell-free system. *Sci. Rep.*, **9**, 11980.
52. McManus,J.B., Emanuel,P.A., Murray,R.M. and Lux,M.W. (2019) A method for cost-effective and rapid characterization of engineered T7-based transcription factors by cell-free protein synthesis reveals insights into the regulation of T7 RNA polymerase-driven expression. *Arch. Biochem. Biophys.*, **674**, 108045.
53. Kopniczky,M.B., Canavan,C., McClymont,D.W., Crone,M.A., Suckling,L., Goetzmann,B., Siciliano,V., MacDonald,J.T., Jensen,K. and Freemont,P.S. (2020) Cell-free protein synthesis as a prototyping platform for mammalian synthetic biology. *ACS Syn. Biol.*, **9**, 144–156.
54. Swaminathan,A., Poole,W., Hsiao,V. and Murray,R.M. (2019) Fast and flexible simulation and parameter estimation for synthetic biology using bioscrape. *bioRxiv*. 10.1101/121152.

55. Chen,Z., Kibler,R.D., Hunt,A., Busch,F., Pearl,J., Jia,M., VanAernum,Z.L., Wicky,B.I., Dods,G., Liao,H. *et al.* (2020) De novo design of protein logic gates. *Science*, **368**, 78–84.
56. Sawasaki,T., Hasegawa,Y., Tsuchimochi,M., Kasahara,Y. and Endo,Y. (2000) Construction of an efficient expression vector for coupled transcription/translation in a wheat germ cell-free system. *Nucleic Acids Symp. Ser.*, **44**, 9–10.
57. Kamura,N., Sawasaki,T., Kasahara,Y., Takai,K. and Endo,Y. (2005) Selection of 5′-untranslated sequences that enhance initiation of translation in a cell-free protein synthesis system from wheat embryos. *Bioorg. Med. Chem. Lett.*, **15**, 5402–5406.
58. Paraskevopoulou,V. and Falcone,F.H. (2018) Polyionic tags as enhancers of protein solubility in recombinant protein expression. *Microorganisms*, **6**, 47.
59. Costa,S., Almeida,A., Castro,A. and Domingues,L. (2014) Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front. Microbiol.*, **5**, 63.
60. Raines,R.T., McCormick,M., Van Oosbree,T.R. and Mierendorf,R.C. (2000) [23] The S· tag fusion system for protein purification. In: Thorner J, Emr SD, Abelson JN (eds). *Methods Enzymol*, Vol. 326. Elsevier, Amsterdam, pp. 362–376.
61. Malakhov,M.P., Mattern,M.R., Malakhova,O.A., Drinker,M., Weeks,S.D. and Butt,T.R. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics*, **5**, 75–86.
62. Nguyen,A.N., Song,J.-A., Nguyen,M.T., Do,B.H., Kwon,G.G., Park,S.S., Yoo,J., Jang,J., Jin,J., Osborn,M.J. *et al.* (2017) Prokaryotic soluble expression and purification of bioactive human fibroblast growth factor 21 using maltose-binding protein. *Sci. Rep.*, **7**, 16139.
63. Bennett,M.J., Huey-Tubman,K.E., Herr,A.B., West,A.P., Ross,S.A. and Bjorkman,P.J. (2002) A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc. Natl. Acad. Sci. USA*, **99**, 11634–11639.
64. Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, **67**, 31–40.
65. Raran-Kurussi,S. and Waugh,D.S. (2014) Unrelated solubility-enhancing fusion partners MBP and NusA utilize a similar mode of action. *Biotechnol. Bioeng.*, **111**, 2407–2411.
66. Schmidt,T.G., Batz,L., Bonet,L., Carl,U., Holzapfel,G., Kiem,K., Matulewicz,K., Niermeier,D., Schuchardt,I. and Stanar,K. (2013) Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression Purif.*, **92**, 54–61.
67. Kapust,R.B., Tözsér,J., Fox,J.D., Anderson,D.E., Cherry,S., Copeland,T.D. and Waugh,D.S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.*, **14**, 993–1000.
68. Jewett,M.C. and Swartz,J.R. (2004) Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.*, **86**, 19–26.
69. Kwon,Y.-C. and Jewett,M.C. (2015) High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.*, **5**, 8663.
70. Rivera,S.B., Swedlund,B.D., King,G.J., Bell,R.N., Hussey,C.E., Shattuck-Eidens,D.M., Wrobel,W.M., Peiser,G.D. and Poulter,C.D. (2001) Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from Chrysanthemum cinerariaefolium. *Proc. Natl. Acad. Sci. USA*, **98**, 4373–4378.
71. Thulasiram,H.V., Erickson,H.K. and Poulter,C.D. (2007) Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science*, **316**, 73–76.
72. Kim,T.-W., Keum,J.-W., Oh,I.-S., Choi,C.-Y., Park,C.-G. and Kim,D.-M. (2006) Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.*, **126**, 554–561.
73. Yang,T., Gao,L., Hu,H., Stoopen,G., Wang,C. and Jongsma,M.A. (2014) Chrysanthemyl diphosphate synthase operates in planta as a bifunctional enzyme with chrysanthemol synthase activity. *J. Biol. Chem.*, **289**, 36325–36335.
74. Krawczyk,S., Thurow,C., Niggeweg,R. and Gatz,C. (2002) Analysis of the spacing between the two palindromes of activation sequence-1 with respect to binding to different TGA factors and transcriptional activation potential. *Nucleic Acids Res.*, **30**, 775–781.
75. Karim,A.S., Liew,F., Garg,S., Vögeli,B., Rasor,B.J., Gonnot,A., Pavan,M., Juminaga,A., Simpson,S.D., Köpke,M. *et al.* (2020) Modular cell-free expression plasmids to accelerate biological design in cells. *Synth. Biol.*, **5**, ysaa019.
76. Iverson,S.V., Haddock,T.L., Beal,J. and Densmore,D.M. (2016) CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Syn. Biol.*, **5**, 99–103.
77. Moore,S.J., Lai,H.-E., Kelwick,R.J., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) EcoFlex: a multifunctional MoClo kit for *E. coli* synthetic biology. *ACS Syn. Biol.*, **5**, 1059–1069.
78. Takahashi,M.K., Chappell,J., Hayes,C.A., Sun,Z.Z., Kim,J., Singhal,V., Spring,K.J., Al-Khabouri,S., Fall,C.P., Noireaux,V. *et al.* (2015) Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription–translation (TX-TL) systems. *ACS Syn. Biol.*, **4**, 503–515.
79. Guo,S. and Murray,R.M. (2019) Construction of incoherent feed-forward loop circuits in a cell-free system and in cells. *ACS Syn. Biol.*, **8**, 606–610.
80. Voyvodic,P.L., Pandi,A., Koch,M., Conejero,I., Valjent,E., Courtet,P., Renard,E., Faulon,J.-L. and Bonnet,J. (2019) Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *Nat. Commun.*, **10**, 1697.
81. Silverman,A.D., Akova,U., Alam,K.K., Jewett,M.C. and Lucks,J.B. (2020) Design and optimization of a cell-free atrazine biosensor. *ACS Syn. Biol.*, **9**, 671–677.
82. Liu,R., Yu,D., Deng,Z. and Liu,T. (2020) Harnessing *in vitro* platforms for natural product research: *in vitro* driven rational engineering and mining (iDREAM). *Curr. Opin. Biotechnol.*, **69**, 1–9.
83. Jiang,L., Zhao,J., Lian,J. and Xu,Z. (2018) Cell-free protein synthesis enabled rapid prototyping for metabolic engineering and synthetic biology. *Synth. Syst. Biotechnol.*, **3**, 90–96.
84. Koo,J., Schnabel,T., Liong,S., Evitt,N.H. and Swartz,J.R. (2017) High-throughput screening of catalytic H2 production. *Angew. Chem. Int. Ed.*, **56**, 1012–1016.
85. Jin,X. and Hong,S.H. (2018) Cell-free protein synthesis for producing 'difficult-to-express' proteins. *Biochem. Eng. J.*, **138**, 156–164.
86. Salehi,A.S., Smith,M.T., Bennett,A.M., Williams,J.B., Pitt,W.G. and Bundy,B.C. (2016) Cell-free protein synthesis of a cytotoxic cancer therapeutic: onconase production and a just-add-water cell-free system. *Biotechnol. J.*, **11**, 274–281.
87. Oh-hashi,K., Furuta,E., Fujimura,K. and Hirata,Y. (2017) Application of a novel HiBiT peptide tag for monitoring ATF4 protein expression in Neuro2a cells. *Biochem. Biophys. Rep.*, **12**, 40–45.

88. Dixon,A.S., Schwinn,M.K., Hall,M.P., Zimmerman,K., Otto,P., Lubben,T.H., Butler,B.L., Binkowski,B.F., Machleidt,T., Kirkland,T.A. *et al.* (2016) NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.*, **11**, 400–408.
89. Verma,M., Choi,J., Cottrell,K.A., Lavagnino,Z., Thomas,E.N., Pavlovic-Djuranovic,S., Szczesny,P., Piston,D.W., Zaher,H.S. and Puglisi,J.D. (2019) A short translational ramp determines the efficiency of protein synthesis. *Nat. Commun.*, **10**, 5774.
90. Pédelacq,J.-D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.
91. O'Malley,R.C., Huang,S.-S.C., Song,L., Lewsey,M.G., Bartlett,A., Nery,J.R., Galli,M., Gallavotti,A. and Ecker,J.R. (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, **165**, 1280–1292.
92. Bartlett,A., O'Malley,R.C., Huang,S.-S.C., Galli,M., Nery,J.R., Gallavotti,A. and Ecker,J.R. (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.*, **12**, 1659–1672.
93. Kesarwani,M., Yoo,J. and Dong,X. (2007) Genetic interactions of TGA transcription factors in the regulation of pathogenesis-related genes and disease resistance in Arabidopsis. *Plant Physiol.*, **144**, 336–346.
94. Fode,B., Siemsen,T., Thurow,C., Weigel,R. and Gatz,C. (2008) The Arabidopsis GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. *Plant Cell*, **20**, 3122–3135.
95. Gatz,C. (2013) From pioneers to team players: TGA transcription factors provide a molecular link between different stress pathways. *Mol. Plant-Microbe Interact.*, **26**, 151–159.
96. Zander,M., Thurow,C. and Gatz,C. (2014) TGA transcription factors activate the salicylic acid-suppressible branch of the ethylene-induced defense program by regulating ORA59 expression. *Plant Physiol.*, **165**, 1671–1683.
97. Hussain,R.M., Sheikh,A.H., Haider,I., Quareshy,M. and Linthorst,H.J. (2018) Arabidopsis WRKY50 and TGA transcription factors synergistically activate expression of PR1. *Front. Plant Sci.*, **9**, 930.
98. Lam,E. and Lam,Y.K.-P. (1995) Binding site requirements and differential representation of TGA factors in nuclear ASF-1 activity. *Nucleic Acids Res.*, **23**, 3778–3785.
99. Strompen,G., Grüner,R. and Pfitzner,U.M. (1998) An as-1-like motif controls the level of expression of the gene for the pathogenesis-related protein 1a from tobacco. *Plant Mol. Biol.*, **37**, 871–883.
100. Niggeweg,R., Thurow,C., Kegler,C. and Gatz,C. (2000) Tobacco transcription factor TGA2. 2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid-and auxin-inducible expression of as-1-containing target promoters. *J. Biol. Chem.*, **275**, 19897–19905.
101. Cai,Y., Kallam,K., Tidd,H., Gendarini,G., Salzman,A. and Patron,N.J. (2020) Rational design of minimal synthetic promoters for plants. *Nucleic Acids Res.*, **48**, 11845–11856.
102. Jores,T., Tonnies,J., Dorrity,M.W., Cuperus,J.T., Fields,S. and Queitsch,C. (2020) Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell*, **32**, 2120–2131.
103. Nusinow,D.A., Helfer,A., Hamilton,E.E., King,J.J., Imaizumi,T., Schultz,T.F., Farré,E.M. and Kay,S.A. (2011) The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature*, **475**, 398–402.
104. Denisov,I., Shih,A. and Sligar,S. (2012) Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.*, **108**, 150–158.
105. Milla,P., Viola,F., Bosso,S.O., Rocco,F., Cattel,L., Joubert,B., LeClair,R., Matsuda,S. and Balliano,G. (2002) Subcellular localization of oxidosqualene cyclases from *Arabidopsis thaliana*, *Trypanosoma cruzi*, and *Pneumocystis carinii* expressed in yeast. *Lipids*, **37**, 1171–1176.
106. Weng,J.-K., Philippe,R.N. and Noel,J.P. (2012) The rise of chemodiversity in plants. *Science*, **336**, 1667–1670.
107. Miettinen,K., Dong,L., Navrot,N., Schneider,T., Burlat,V., Pollier,J., Woittiez,L., van der Krol,S., Lugan,R., Ilc,T. *et al.* (2014) The seco-iridoid pathway from *Catharanthus Roseus*. *Nat. Commun.*, **5**, 3606.
108. Dong,L., Jongedijk,E., Bouwmeester,H. and van der Krol,A. (2016) Monoterpene biosynthesis potential of plant subcellular compartments. *New Phytol.*, **209**, 679–690.
109. Lau,W. and Sattely,E.S. (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, **349**, 1224–1228.
110. Wang,B., Kashkooli,A.B., Sallets,A., Ting,H.-M., De Ruijter,N.C., Olofsson,L., Brodelius,P., Pottier,M., Boutry,M. and Bouwmeester,H. (2016) Transient production of artemisinin in Nicotiana benthamiana is boosted by a specific lipid transfer protein from A. annua. *Metab. Eng.*, **38**, 159–169.
111. Caputi,L., Lim,E.K. and Bowles,D.J. (2008) Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. *Chemistry*, **14**, 6656–6662.
112. Hu,Y., Xue,J., Min,J., Qin,L., Zhang,J. and Dai,L. (2020) Biocatalytic synthesis of ginsenoside Rh2 using *Arabidopsis thaliana* glucosyltransferase-catalyzed coupled reactions. *J. Biotechnol.*, **309**, 107–112.
113. Bi,H., Wang,S., Zhou,W., Zhuang,Y. and Liu,T. (2019) Producing gram-scale unnatural rosavin analogues from glucose by engineered *Escherichia Coli*. *ACS Syn. Biol.*, **8**, 1931–1940.
114. Dudley,Q.M., Nash,C.J. and Jewett,M.C. (2019) Cell-free biosynthesis of limonene using *Escherichia coli* crude extracts. *Synth. Biol.*, **4**, ysz003.
115. Garcia,D.C., Dinglasan,J.L.N., Shrestha,H., Abraham,P.E., Hettich,R.L. and Doktycz,M.J. (2021) A lysate proteome engineering strategy for enhancing cell-free metabolite production. *Metab. Eng. Commun.*, **12**, e00162.
116. Knapp,K.G. and Swartz,J.R. (2004) Cell-free production of active E. coli thioredoxin reductase and glutathione reductase. *FEBS Lett.*, **559**, 66–70.
117. Mou,Z., Eakes,J., Cooper,C.J., Foster,C.M., Standaert,R.F., Podar,M., Doktycz,M.J. and Parks,J.M. (2021) Machine learning-based prediction of enzyme substrate scope: application to bacterial nitrilases. *Proteins: Struct. Funct. Bioinform.*, **89**, 336–347.
118. Robinson,S.L., Smith,M.D., Richman,J.E., Aukema,K.G. and Wackett,L.P. (2020) Machine learning-based prediction of activity and substrate specificity for OleA enzymes in the thiolase superfamily. *Synth. Biol.*, **5**, ysaa004.