Unlocking applications of cell-free biotechnology through enhanced shelf-life and productivity of *E. coli* extracts

Nicole E. Gregorio[# 1,2], Wesley Y. Kao[# 1,2], Layne C. Williams[# 1,2], Christopher M. Hight[1,2], Pratish Patel[3], Katharine R. Watts[1,2], Javin P. Oza[1,2*]

[1]Chemistry and Biochemistry Department, California Polytechnic State University, 1 Grand Ave., San Luis Obispo, CA 93407
[2]Center for Applications in Biotechnology, California Polytechnic State University, 1 Grand Ave., San Luis Obispo, CA 93407
[3]Department of Finance, Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA 93407

# These authors contributed equally to this work
* Correspondence: joza@calpoly.edu; 805-765-2265

Abstract

Cell-Free Protein Synthesis (CFPS) is a platform biotechnology that enables a breadth of applications. However, field applications remain limited due to the poor shelf-stability of aqueous cell extracts required for CFPS. Lyophilization of *E. coli* extracts improves shelf-life but remains insufficient for extended storage at room temperature. To address this limitation, we mapped the chemical space of ten low-cost additives with four distinct mechanisms of action in a combinatorial manner to identify formulations capable of stabilizing lyophilized cell extract. We report three key findings: 1) unique additive formulations that maintain full productivity of cell extracts stored at 4˚C and 23˚C; 2) additive formulations that enhance extract productivity by nearly 2-fold; 3) a machine learning algorithm that provides predictive capacity for the stabilizing effects of additive formulations that were not tested experimentally. These findings provide a simple and low-cost advance toward making CFPS field-ready and cost-competitive for biomanufacturing.

Keywords: Cell-free protein synthesis, *In vitro* transcription/translation, Lyophilization, Synthetic biology, Machine learning

Cell-Free Protein Synthesis (CFPS) is a biotechnology platform that has supported applications in research and industrial settings for protein expression, drug development, and genetic code expansion among several others.[1] Emerging field applications of CFPS primarily leverage *E. coli* crude lysates for *in vitro* transcription and translation to provide point-of-care diagnostics, enhance therapeutics production, and make synthetic biology education feasible.[2–13] The cell extract remains the most sensitive component of CFPS,

requiring careful storage and handling because of its complex mixture of vital RNA and protein-based cellular machinery including ribosomes, auxiliary translation machinery, and central metabolism.[14–17] Aqueous extract is known to degrade quickly under non-ideal conditions, such as room temperature storage.[16] This stability issue engenders a cold chain and presents the primary bottleneck for implementation of CFPS field applications.

A key advancement toward long-term cell extract stability was lyophilization.[16] This approach is robust and has been shown to support CFPS field applications such as paper-based diagnostics devices, biosensors, on-demand synthesis of industrially and clinically relevant molecules, and classroom applications for synthetic biology education.[2-8,15–21] Additionally, the use of lyophilization has been previously shown to mitigate the risk for biocontamination by cell extracts, improving safety for field applications.[25] However, the enhancement to cell extract stability is observed to be limited, as the productivity of lyophilized cell extracts stored at room temperature diminishes notably within the first 14 days.[16,26] This suggests that while lyophilization is an important aspect of achieving extract stability, it remains insufficient on its own.

Successes in the stabilization of various other biological materials used in medicine for diagnostics and therapeutics suggest that the gap in cell extract stabilization could be filled through continued discovery of additive formulations. Numerous lyoprotectants and thermoprotectants have been used to stabilize purified proteins but have not been thoroughly evaluated for their capacity to stabilize complex mixtures. As such, these

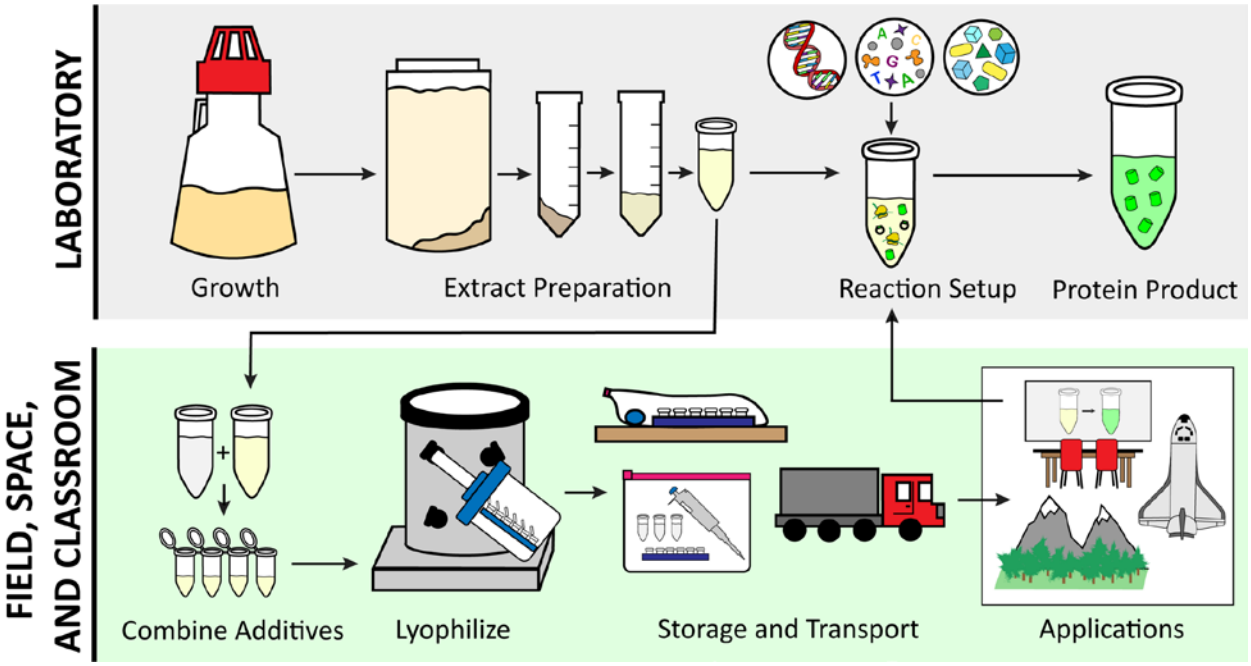

Figure 1. Workflow for traditional cell-free extract production and reaction setup (grey) compared to our modified workflow for improved stability and productivity to support field applications (green). In the modified workflow, additives are combined with cell extract, lyophilized, and stored in airtight bags with desiccant until use.

advances cannot be directly applied to CFPS extracts. An attempt at extract stabilization has also been made by introducing antiplasticized glasses to the extract prior to lyophilization to help retain extract productivity during lyophilization and subsequent storage.[26] While this represents another noteworthy advance, only 60-70% of productivity is retained after two weeks of storage at 25 °C for *E. coli* extracts containing the optimal lyoprotectant formulation. To date, a variety of additives, including sugars, trimethylamine osmolytes, and molecular crowding agents have been used to augment the protein titer of CFPS reactions, but few have been thoroughly evaluated for their ability to stabilize CFPS extract for storage.[27] Furthermore, the field lacks the capacity to predict the performance of various additives toward stabilization of biological materials, necessitating efforts to thoroughly screen putative stabilizers in CFPS extract.

Here we evaluate the impact of ten additives, with four distinct mechanisms of enhancing the shelf-life of lyophilized cell extracts. In addition to establishing the full titration ranges of each additive, we also report the combinatorial effects emerging from binary and ternary co-titrations of different classes of additives on cell-extract stability. These efforts have allowed us to map the additives landscape associated with stabilization of *E. coli* cell extract at -80 °C, 4 °C, and 23 °C (room temperature). We report non-obvious combinatorial formulations of additives that maintain at least 97.7 ± 1.7% of extract productivity, as indicated by their ability to produce the reporter superfolder green fluorescent protein (sfGFP), at all three temperature conditions after two weeks of storage. Unexpectedly, we have also discovered additive formulations that enhance productivity up to 195.4 ± 5.6% for extracts stored at colder temperatures (<4 °C). Lastly, we report a machine learning algorithm emerging from this data that provides predictive capacity for the effects of these additives on extract stabilization at -80 °C, 4 °C, and 23 °C. Together, the advances reported here unite disparate past investigations of additive supplementation in CFPS and provide the ability to predict stabilization from additive combinations that have yet to be experimentally tested. This brings CFPS a step closer to becoming a field-ready biotechnology, representing a transition from a primarily laboratory-based platform to a widely accessible one (Fig. 1).

Results

In order to evaluate the effects of the selected ten additives on *E. coli* cell extract stability, all experiments were conducted by supplementing aqueous extracts with additives prior to lyophilization. Lyophilized extracts were then stored for two weeks at -80 °C, 4 °C, or 23 °C (room temperature) (Fig. S1) alongside aqueous and lyophilized controls with no additives. For all comparisons, the aqueous extract stored at -80 °C for two weeks is referred to as the “benchmark” and all protein yields are reported as a percentage of this internal control to control for batch-to-batch and user-to-user variation in experimentation. This condition is considered the benchmark because it maintains full extract stability for at least a year and is broadly accepted by the research community.[28]

Thus, it sets the standard for evaluating extract treatments herein. The stabilizing effects of lyophilization, single additives, or combinations of additives were evaluated based on the extract's capacity to conduct transcription and translation and generate the reporter protein sfGFP after two weeks of storage. Enhancements to productivity of reporter protein (values over 100% of the benchmark) are distinguished from enhancements to extract stability (values greater than the lyophilized control, but no greater than 100%). While additives are supplemented prior to lyophilization and storage, all concentrations of additives indicate the final concentration in the complete 30 µL CFPS reaction for practical considerations. The mole quantities of each additive supplemented prior to lyophilization are also provided (Table S1). Consistent with recent characterizations of interlaboratory variability in CFPS, we observed ~20% variation in sfGFP yield for the aqueous control stored at -80 °C from user-to-user and from batch-to-batch of reagents, and much lower variation within independent controlled experiments (Table S2).[29]

Lyophilization

Throughout all trials, our results were consistent with previous observations that lyophilization alone can improve cell extract stability. When stored at 4 °C for two weeks, lyophilized extracts retained an average of 79 ± 16% productivity versus the 56.7 ± 8.9% of productivity retained by extracts stored in the aqueous phase. At 23 °C, lyophilized extract retained an average of 35.3 ± 7.3% of the benchmark, while aqueous extract stored at 23 °C was rendered nonfunctional (Fig. 2).[4] Also consistent with previous reports, lyophilization itself did not have a detrimental effect on extract stability, as evidenced by comparable productivity of aqueous and lyophilized controls stored at -80 °C without additives (Fig. 2; Table S2). Importantly, these data provide additional evidence that lyophilization alone remains insufficient to stabilize the full productivity of the cell extracts. As such, additives were identified as beneficial if they improved extract productivity beyond lyophilization alone, with the goal of identifying additives that stabilized the full productivity of extracts when stored at 23 °C.

Single Additive Screening

All ten additives were first screened individually to determine their capacity to stabilize the cell extract on their own, and to determine the top performing additives from each mechanistic category. All data for the single additive screening can be found in Table S3 and Note S2.

Sugars

During lyophilization and subsequent rehydration of cell extracts, sugars are hypothesized to act as lyoprotectants and protein stabilizers by immobilizing proteins in an amorphous glassy matrix, forming hydrogen bond interactions with proteins, and causing preferential hydration of proteins.[30–34] Amongst the sugars available for

stabilizing proteins, nonreducing sugars are favored for protein stabilization due to their ability to interact with proteins without reacting.[33] In other instances, sugars have been proposed to function as crowding agents, thereby increasing the effective concentration of other substances in the reaction and better mimicking the intracellular environment.[35,36] Moreover, it is possible that some sugars, like maltose, may have a metabolic benefit to the CFPS reaction, allowing for the use of alternative metabolic pathways in order to increase productivity of the cell extract.[37] Some sugar supplementation has previously been shown to benefit cell extract stability, however, past work also suggests that sugar choice is non-obvious for most systems and can depend on numerous variables including the unique molecular makeup of the protein or mixture to be stabilized, the sugar to product ratio, the drying technique chosen, and the storage conditions in question, necessitating screening efforts specific to *E. coli*-based CFPS.[38,39] In order to characterize the utility of sugars in stabilizing *E. coli* cell extracts, we selected 5 sugar additives, trehalose, maltose, lactose, raffinose, and sucrose, representing both reducing and non-reducing sugars, and screened them in a defined concentration range from 0 to 150 mM in 30 mM increments.

Of the five sugars, trehalose proved to be the most effective stabilizer. Trehalose has been successfully utilized as a stabilizing agent in a variety of contexts, including stabilization of RNA samples, purified proteins, as well as *Vibrio natriegens* and *E. coli* extract.[39–42] This is in part because of its high glass transition temperature in comparison to other sugars, an important mechanistic property of sugar-based stabilization of proteins.[40] Trehalose was previously shown to stabilize air-dried *E. coli* cell extract at 37 °C, although concentrations under 200 mM in the final reaction were not evaluated.[39] Here, extracts containing 30 mM trehalose stored at 23 °C retained 88 ± 10% of productivity compared to the benchmark, and extracts stored at 4 °C with 30 mM trehalose performed at 107.2 ± 4.8% of the benchmark productivity (Fig 2a). This represents a notable enhancement beyond lyophilization alone, which retained 47.0 ± 2.6% and 78.7 ± 6.5% of benchmark productivity when stored at 23 °C or 4 °C, respectively. Unexpectedly, 30 mM trehalose was also able to boost protein yields in extract stored at -80 °C to 124.2 ± 9.2% of the benchmark (Fig. 2a).

The supplementation of maltose at 30 mM also significantly stabilized the extract for storage at 23 °C and 4 °C and was able to moderately enhance overall productivity at -80 °C (Fig. 2b). Comparatively, lactose, raffinose, and sucrose stabilized cell extract but

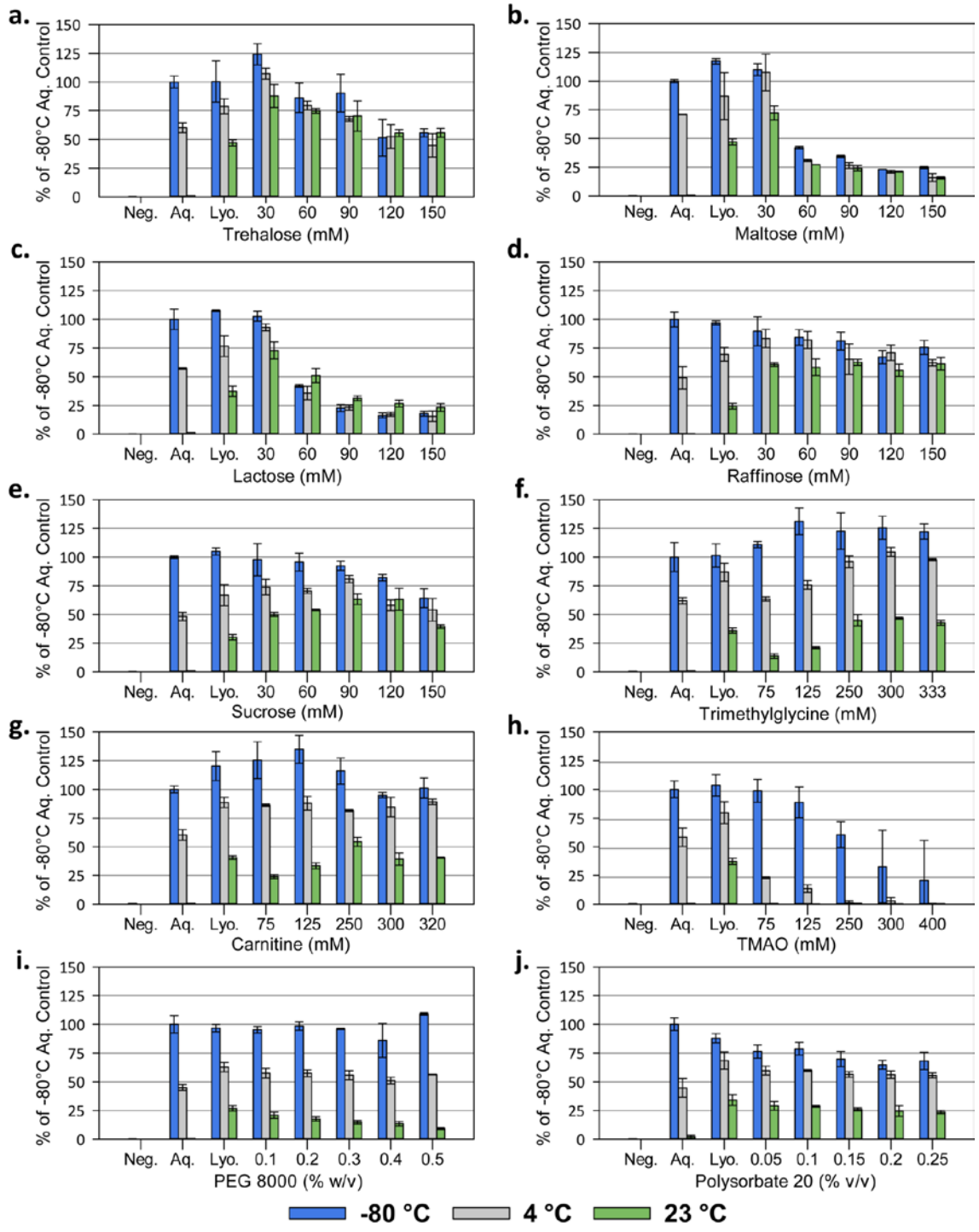

Figure 2. Single additive titrations of ten additives in lyophilized cell extract stored at -80 °C, 4 °C and 23 °C (room temperature) for two weeks. sfGFP reporter protein production was used to quantify reaction productivity. Negative (Neg.) conditions correspond to no added DNA or additives. Aqueous (Aq.) and lyophilized (Lyo.) controls do not contain additives. Concentrations of additives indicate the final concentration in the 30 µL CFPS reaction. All conditions were performed in quadruplicate and a Grubbs' test was used to identify outliers. Error bars represent the standard deviation of a minimum of 3 reactions.

did not demonstrate the capacity to enhance productivity (Fig. 2c-e). Furthermore, none of these sugars were able to maintain the full productivity of the cell extract after storage at 4 °C, whereas maltose and trehalose were (Fig. 2a-e). Lactose and maltose performed similarly at 23 °C, maintaining around 70% of productivity, while raffinose and sucrose maintained only about 60% productivity (Fig. 2b-e). The capacity of trehalose to support 88 ± 10% of productivity retention at 23 °C storage proved to be superior to the other four sugars tested (Fig 2a).

Comparative screening of these sugars resulted in notable trends in concentration dependent stabilization of cell extracts. For trehalose, maltose, and lactose, a 30 mM final concentration in the reaction proved to be the most effective at all temperatures, while raffinose and sucrose exhibited benefits over larger concentration ranges. (Fig. 2a-e). Trehalose, maltose, and lactose also shared similar trends over the chosen titration range, with rapid and significant reductions in productivity as concentration was increased (Fig. 2a-c). These observations are consistent with previous work in which high concentrations of sugars were observed to have an inhibitory effect on protein synthesis, which was proposed to be a result of inorganic acid production rapidly lowering the pH of the reaction.[36,37,39,43] Trends for sucrose varied most drastically with temperature, and higher temperature storage favored higher concentrations of sucrose (Fig. 2e). Raffinose was the only trisaccharide tested and its stabilizing effects were notably concentration independent compared to other sugars tested (Fig. 2d). Interestingly, our data do not support the premise that non-reducing sugars (sucrose, raffinose, and trehalose) should be preferred over reducing sugars (lactose and maltose), as trehalose and maltose provided the most significant stability and productivity improvements, with lactose close behind (Table 1).

Sugars remain prime candidates as extract stabilizers and productivity enhancers. These results provide insights into sugar choice and effective concentration ranges useful for storage conditions that support both field and laboratory applications of CFPS. Overall, 30 mM trehalose emerged as the prime candidate for improved stability and productivity across all temperatures in this category. The machine learning algorithm generated from these data also provides predictive capacity for the stabilizing effects of these sugars at concentrations in between the 30 mM concentration increments evaluated here.

Trimethylamine Osmolytes

Trimethylamine osmolytes are naturally-occurring stabilizing molecules present in a variety of organisms that protect essential proteins against denaturation-inducing environmental stressors without disrupting enzyme activity.[44,45] Given that trimethylamine osmolytes are in contact with a multitude of biomolecules within the cell, they should theoretically stabilize a variety of proteins at one time.[44] This suggests that trimethylamine osmolytes may be beneficial for stabilization of complex enzymatic mixtures, such as cell extract, where the goal is to stabilize a broad range of proteins simultaneously.

Trimethylamine osmolytes have been proposed to function through increasing preferential hydration of proteins by increasing surface tension and the cohesive force of water, due to their strong polarity.[46] Mechanistically, it has been suggested that trimethylamine osmolytes order water molecules, which entropically drives proteins to fold more tightly.[47,48] This can cause tighter binding of ligands, but also may make proteins too rigid to correctly function. Trimethylamine osmolytes can typically be present in large concentrations without affecting protein structure or interactions and are generally unreactive.[49,50] However, they have been shown to inhibit protein expression in CFPS at concentrations above 1 M.[50] In order to characterize the utility of trimethylamine osmolytes in stabilizing *E. coli* cell extracts, trimethylglycine, carnitine, and trimethylamine N-oxide (TMAO) were chosen for analysis.

Reactions containing a final concentration of 300 mM trimethylglycine, a betaine, maintained 46.73 ± 0.76% productivity after 23 °C storage and 104.5 ± 3.9% productivity after 4 °C storage compared to the benchmark. Lyophilized controls only retained 35.9 ± 2.2% and 86.9 ± 7.8% of productivity at the respective temperatures, indicating that trimethylglycine provided some benefits to stability (Fig. 2f). However, at -80 °C storage, 125 to 333 mM trimethylglycine demonstrated a significant productivity boost, slightly above that of trehalose, to a maximum of 131 ± 11% of the benchmark (Fig. 2f). Trimethylglycine has previously exhibited the capacity to increase gene expression in the PURE (Protein synthesis Using Recombinant Elements) system, with greater benefits at lower temperatures for fluorescent proteins, which is consistent with our data at -80 °C.[50]

Unlike trimethylglycine, carnitine had no significant effect on stability at 4 °C storage. However, carnitine provided modest improvements to stability at 23 °C, and substantial increases in productivity at -80 °C storage, similar to trimethylglycine (Fig. 2g). The data trends for both trimethylglycine and carnitine suggest that lower concentrations may

Table 1. Summary of the Effects of Additives on Cell-free Extract Stability and Productivity After Two Weeks of Storage.

	-80 °C			4 °C			23 °C		
Additive	**Optimal Range**	**Highest Productivity**	**ML [#]**	**Optimal Range**	**Highest Productivity**	**ML[#]**	**Optimal Range**	**Highest Productivity**	**ML[#]**
Trehalose	30 mM	124.2 ± 9.2%	128%	30 mM	107.2 ± 4.8%	93%	30 mM	88 ± 10%	87%
Maltose	30 mM	110.1 ± 5.2%	114%	30 mM	108 ± 16%	92%	30 mM	72.3 ± 6.1%	65%
Lactose	30 mM	102.8 ± 4.4%	105%	30 mM	93.1 ± 3.1%	89%	30 mM	72.9 ± 7.4%	67%
Raffinose	30-90 mM	90 ± 13%	98%	30-60 mM	83.4 ± 8.1%	92%	30-90 mM	62.6 ± 2.7%	63%
Sucrose	30-90 mM	98 ± 14%	113%	90 mM	81.0 ± 3.1%	92%	90-120 mM	63.3 ± 4.6%	68%
Trimethylglycine	125-333 mM	131 ± 11%	111%	250-300 mM	104.5 ± 3.9%	83%	250-300 mM	46.73 ± 0.76%	35%
Carnitine	75-250 mM	135 ± 12%	101%	75-320 mM	89.2 ± 2.4%	63%	250 mM	54.3 ± 3.9%	38%
TMAO*	—	—	—	—	—	—	—	—	—
PEG 8000	0.5% w/v	109.0 ± 1.0%	90%	0.1-0.5% w/v	57.4 ± 4.2%	70%	0.1% w/v	20.9 ± 2.8%	40%
Polysorbate 20	0.05-0.1% v/v	78.8 ± 5.5%	90%	0.05-0.1% v/v	59.76 ± 0.88%	70%	0.05% v/v	29.1 ± 3.8%	40%

*TMAO is observed to be detrimental to lyophilized cell extract, no optimal conditions found.

[#] ML: Predicted productivity of lyophilized extract stored for two weeks with the specified additive

function to increase productivity at -80 °C while higher concentrations provide stabilizing effects for improved shelf-life at 4 °C and 23 °C.

Unexpectedly, TMAO was inhibitory to extract productivity at all temperature conditions, with higher concentrations having increasingly detrimental effects (Fig. 2h). Lyophilized extracts containing TMAO were also difficult to resuspend, and a noticeable precipitate formed in the tubes during the CFPS reactions. Previous work with TMAO exhibited enhancements to translational activity but not transcriptional activity, though this was observed using a diluted PURE system, which is likely to behave differently in some ways than a crude lysate-based system.[50] Past work has indicated that TMAO may act through a different mechanism from trimethylglycine, by interacting directly with proteins, rather than ordering water.[51] TMAO is also used to counteract urea inhibition in some organisms.[48] The detrimental effect of TMAO in *E.* coli crude lysate CFPS may be attributed to the fact that in the absence of urea, high TMAO concentrations can lead to excessive stabilization, causing rigidity to the point of non-functionality.[48] Given that TMAO supplementation did not improve productivity or stability of the crude cell extract, we excluded it from combinatorial screening and from the machine learning algorithm.

Miscellaneous Additives

Lastly, we screened large organic molecules, Polyethylene Glycol (PEG) 8000 and Polysorbate 20 (Tween 20), that have previously been utilized to stabilize protein-based products.[31,52] As proteins are about twenty-fold more dilute in CFPS reactions than in *E. coli* cells, PEG is proposed to mimic macromolecular crowding in CFPS reactions, and can also help to stabilize messenger RNA.[53,54] Molecular crowding can significantly increase transcription rates in wheat germ CFPS systems, but can inhibit translation.[54] Unlike sugars, PEG does not stabilize proteins through hydrogen bonding and its mechanism of stabilization may vary based on protein type.[55] Notably, PEG is observed to function as a cryoprotectant, but not as a lyoprotectant.[31] We observed that increasing concentrations of PEG were detrimental to productivity when lyophilized cell extract was stored at 23 °C (Fig. 2i). At 4 °C storage, PEG did not significantly benefit or harm protein output within the titration range of 0.1 to 0.5% w/v, and it showed a minimal improvement to productivity at 0.5% w/v after -80 °C storage (Fig. 2i). Overall, PEG did not appear to substantially improve extract stability or productivity which could be due to competing effects on transcription and translation.

Polysorbates compete with proteins for adsorption at the air-water interface and can bind to hydrophobic regions of proteins to help prevent aggregation, especially in solutions with high concentrations of protein, such as pure protein biologics.[52,56] This mechanism of stabilization could prove beneficial as protein concentration increases exponentially during lyophilization of extract. However, the influence of polysorbates on specific proteins is unpredictable, as they have been shown to both stabilize and destabilize proteins.[52] Storage and handling of polysorbates are also sensitive as

autoxidation at higher temperatures can contribute to destabilization of proteins.[57,58] Additionally, the optimal concentration of polysorbates is highly dependent on the type of stress that the protein undergoes.[57] Our data did not indicate a benefit to productivity or stability for any concentration of polysorbate at any temperature condition (Fig. 2j). In fact, all reactions showed a slight but general decrease in productivity with increased polysorbate. Most notably, polysorbate decreased reaction yields by at least 24% after -80 °C storage, indicating that it may have a baseline inhibitory effect on protein synthesis or that autoxidation had occurred to some extent (Fig. 2j).

Binary Co-titrations

Following the single additive titrations, we co-titrated trehalose and trimethylglycine with the goal of identifying combinatorial effects on stabilization and productivity that could not be achieved using one additive alone (Fig. 3). Stabilization of purified protein solutions during freezing and lyophilization has previously been enhanced by the addition of multiple stabilizing agents that work via complementary mechanisms.[31] Trehalose and trimethylglycine were selected because trehalose out-performed all other sugars and trimethylglycine provided a substantial benefit to stability at 4 °C in comparison to carnitine (Table 1). PEG and polysorbate were not chosen for this analysis as they both had minimal to negative effects on extract productivity. We also refrained from titrating the top two performing additives from a single class, such as trehalose:maltose or trimethylglycine:carnitine, in order to test the hypothesis that combinatorial effects may emerge as a function of combining distinct mechanisms of action. For the trehalose:trimethylglycine co-titration, concentration ranges of 20 to 100 mM trehalose and 66 to 330 mM trimethylglycine in the final reaction were selected. All data can be found in Table S4 and Note S2.

A combination of 66 mM trimethylglycine:60 mM trehalose maintained 78.1 ± 5.1% of benchmark productivity, a substantial increase compared to 29.0 ± 2.1% sustained by the lyophilized control stored at 23 °C for two weeks (Fig. 3c). However, this was not an improvement over stabilization by trehalose itself, which at 30 mM maintained 88 ± 10% productivity. For 4 °C storage, two maxima were observed at 132 mM trimethylglycine:60 mM trehalose and 330 mM trimethylglycine:20 mM trehalose, respectively maintaining 99.6 ± 6.3% and 102 ± 11% productivity compared to the benchmark (Fig. 3b). In comparison, the lyophilized extract stored at 4 °C performed at 60.4 ± 3.2% of the benchmark. Given that 30 mM trehalose alone maintained 107.2 ± 4.8% productivity compared to benchmark, there was no added benefit to combining trimethylglycine and trehalose for 4 °C storage. The binary co-titration stored at -80 °C displayed the most significant productivity boost: hotspots were again observed at 132 mM trimethylglycine:60 mM trehalose and 330 mM trimethylglycine:20 mM trehalose, respectively maintaining 163.5 ± 7.6% and 172 ± 22% productivity compared to the benchmark (Fig. 3a) This was a notable increase compared to the 124.2 ± 9.2% boost

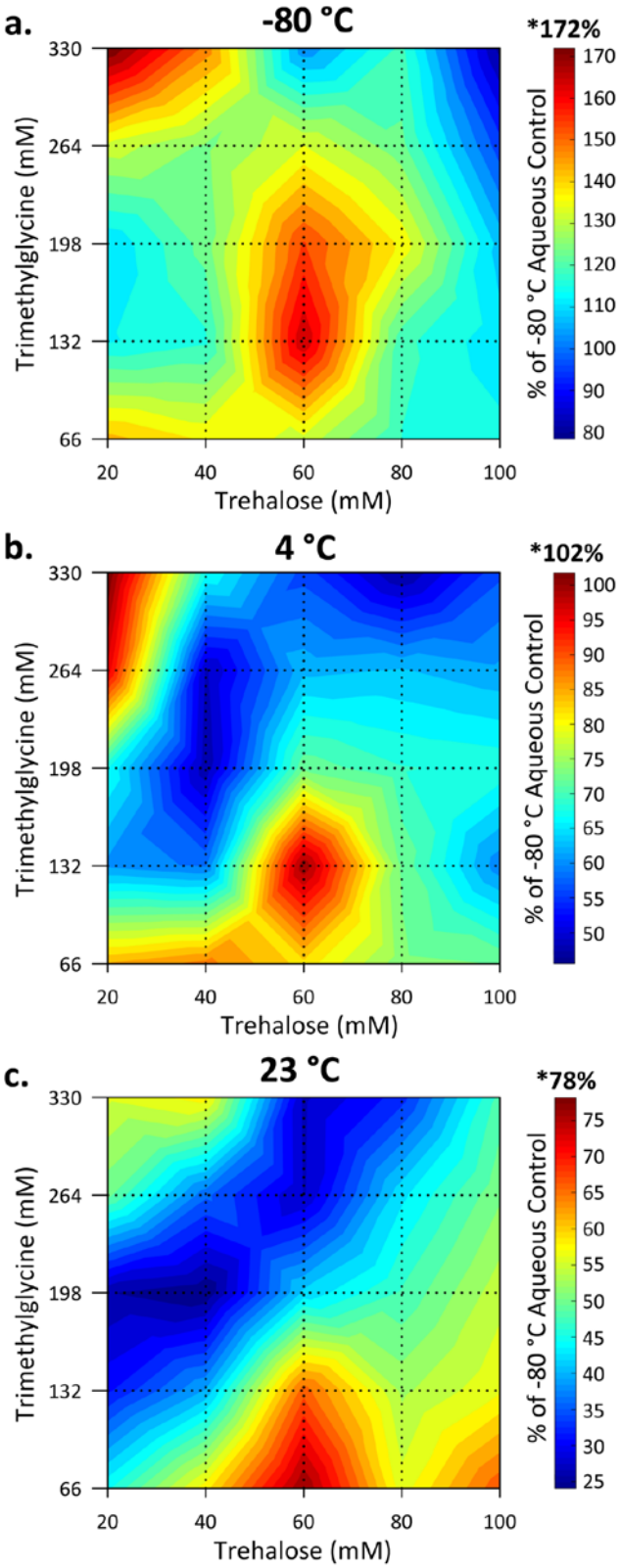

Figure 3. Binary co-titrations of trimethylglycine and with trehalose stored at -80 °C, 4 °C, and 23 °C (room temperature) for two weeks. sfGFP reporter protein production was used to quantify reaction productivity. Concentrations of additives indicate the final concentration in the 30 µL CFPS reaction. All conditions were performed in quadruplicate and a Grubbs' test was used to identify outliers. Each data point represents the average of a minimum of 3 reactions. *Note that color scales represent different values for each temperature condition.

from 30 mM trehalose alone. These data suggest that the combinatorial effects of trehalose:trimethylglycine contribute to the to the enhancement of productivity of lyophilized cell extracts rather than stability during storage at warmer temperatures.

Binary co-titrations enabled the visualization of shifts in productivity “hotspot” as a function of storage temperature (Fig. 3). We observed two shifts as temperature increased: one towards lower concentrations of trimethylglycine with a constant concentration of 60 mM trehalose and one towards lower concentrations of trimethylglycine at 20 mM trehalose.

In sum, the trehalose:trimethylglycine additive combination led to a significant productivity enhancement for -80 °C storage, but was insufficient to produce noteworthy stabilization in higher temperature conditions when compared to trehalose on its own. These findings necessitated continued investigation into additive combinations to identify a mixture of reagents to better stabilize cell extract at field-relevant temperatures.

Ternary Co-titrations

Since stabilization of the cell extract at higher temperatures is desirable for flexibility in field applications and therefore the key objective of this effort, combinations of three additives were surveyed in order to identify additional benefits to storage stability at 4 °C and 23 °C. Following the same reasoning as the binary co-titration, it was hypothesized that a synergistic effect could be achieved with the combined mechanisms of stabilization provided by three different additives. To augment the benefits observed with trehalose and trimethylglycine, PEG was chosen as the third additive. PEG was selected over polysorbate as it was able to maintain extract productivity at -80 °C storage whereas polysorbate displayed a decrease in extract productivity, suggesting that it might inhibit or interfere with the CFPS reaction itself (Fig. 2i and j).

The trehalose:trimethylglycine:PEG ternary co-titrations were conducted within the concentration ranges established in the single and binary titrations. All data are- reported in Table S5 and Note S2. Significant enhancement to productivity was observed for each of the temperature conditions. At 23 °C storage, 97.7 ± 1.7% productivity was maintained at the optimal formulation of 20 mM trehalose:0.4% w/v PEG:0 mM trimethylglycine compared to just 36.7 ± 1.3% in the lyophilized control stored at 23 °C (Fig. 4c). This capacity to maintain extract productivity at 23 °C for two weeks is unprecedented to our knowledge. Additional gains in productivity were observed in the ternary co-titrations at colder storage temperatures. Maxima of 164 ± 13% for 4 °C storage at 20 mM trehalose:0.4% w/v PEG:0 mM trimethylglycine, 195.4 ± 5.6% for -80 °C storage at 20 mM trehalose:0.2% w/v PEG:132 mM trimethylglycine, and 195 ± 17% for -80 °C storage at 20 mM trehalose:0.3% w/v PEG:66 mM trimethylglycine were observed (Figs. 4a and b). The ternary formulations proved superior to the binary and single additive formulations in enhancing *E. coli* cell extract stability and productivity at all temperatures. As observed with the binary co-titrations, ternary co-titration hotspots shifted significantly under different storage temperatures. Here, the shift was towards higher concentrations of trehalose and lower concentrations of trimethylglycine as temperature increased, while there was no clear dependence on PEG (Fig. 4). This generally matched the trends seen

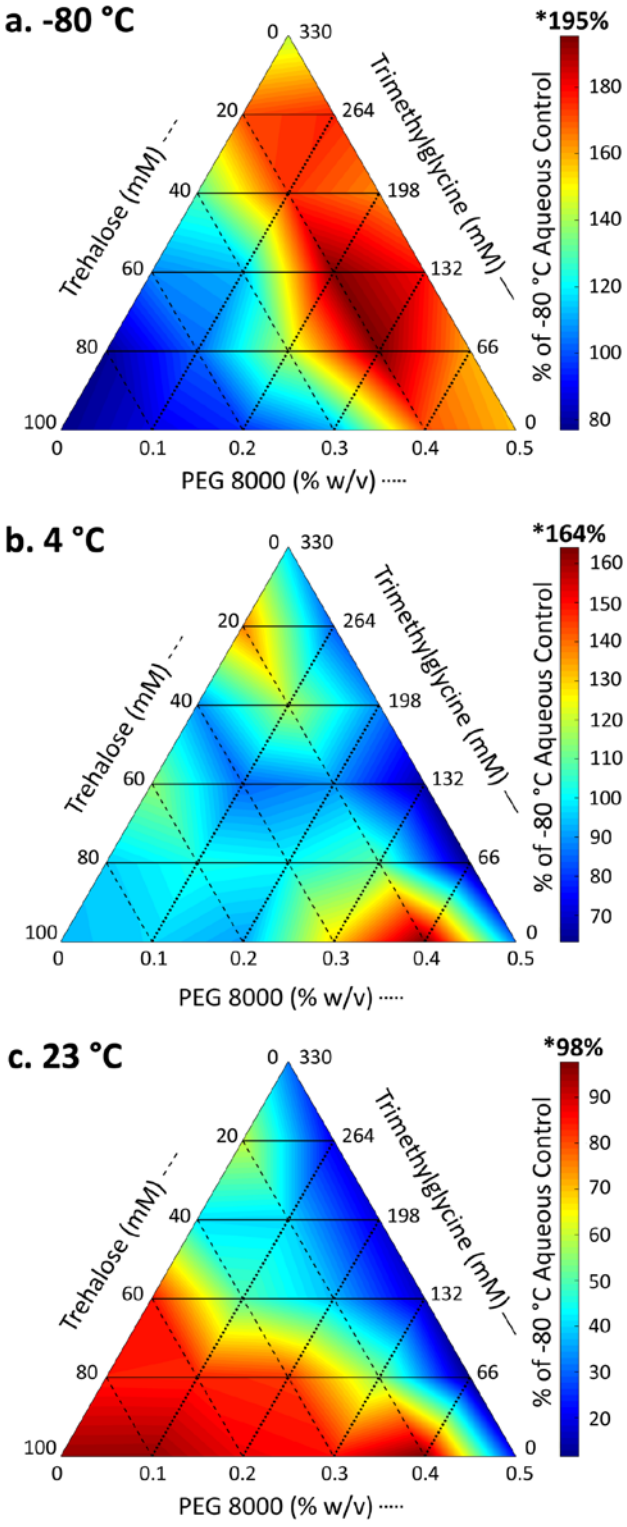

Figure 4. Ternary co-titration of trehalose, trimethylglycine, and polyethylene glycol (PEG) 8000 stored at -80 °C, 4 °C and 23 °C (room temperature) for two weeks. sfGFP reporter protein production was used to quantify reaction productivity. Concentrations of additives indicate the final concentration in the 30 µL CFPS reaction. All conditions were performed in quadruplicate and a Grubbs' test was used to identify outliers. Each data point represents the average of a minimum of 3 reactions. *Note that color scales represent different values for each temperature condition.

in the binary co-titration hotspot migration, but still remained a non-obvious dependence overall.

Lastly, in order to support higher throughput screening efforts, all reactions discussed were performed in PCR tubes. In line with other work done with CFPS, we observed that

reaction vessel size impacts reaction yields; PCR tubes suffer from poor surface area-to-volume ratios, which depress reporter protein titers.[59,60] However, we also demonstrated enhancements to productivity and stability when larger vessels were used, demonstrating that the benefits of the ternary formulations are vessel-independent and provide flexibility for CFPS applications (Fig. S2).

The evaluation of additives in Figures 2-4 was performed on cell extract alone. Energy reagents and substrates contained within our pre-mix solutions A and B and the DNA template were added to lyophilized cell extract to initiate the reaction after the two week storage period. Since many components of the solution A and B mixtures, such as free amino acids, can function as osmolytes, this approach allowed us to reduce such confounding effects. Additionally, this approach provides a modular system to enable a wide variety of CFPS applications. In order to demonstrate that the use of additives can reduce dependency on the cold-chain while maintaining modularity of the system, the CFPS reagents (solutions A, B, and DNA) were combined and lyophilized independently from cell extract containing the best observed additive combination of 20 mM trehalose:0.4% w/v PEG:0 mM trimethylglycine. After storage for 2 weeks at room temperature, the CFPS reagents were rehydrated and added to the supplemented lyophilized cell extract to initiate the reaction. This setup maintained 83.6 ± 8.4% productivity compared to 100 ± 16% productivity observed with the modular system (Fig. S4). Additionally, many applications of cell-free systems would benefit from a one-pot system in which cell extract, substrates, energy reagents, and DNA template are combined prior to lyophilization to enable a 'just add water' reaction at the point-of-use. Therefore, we also evaluated the capacity of the optimal additives formulation to stabilize a one-pot CFPS system for storage at room temperature for two weeks. We observed a productivity of 92.2 ± 8.9% when all additives, energy reagents and substrates, and extract were lyophilized together (Fig. S4). Our results demonstrate that the optimal formulations identified through this work support both modular and one-pot lyophilized systems by retaining nearly full activity of the extract during storage at room temperature for two weeks.

Machine Learning

The need to perform case-to-case optimization of additives for the stabilization of biological materials is necessary because of the lack of predictive capacity for additive performance. Our efforts to map the landscape of cell extract stabilizing additives for CFPS applications has also enabled the opportunity to develop a predictive capacity using machine learning approaches. The resulting algorithm would allow the cell-free systems community to predict the effects of an additive or additive formulation on *E. coli* cell extract stability or productivity when stored at a specified temperature. Toward this goal, we developed four Machine Learning (ML) algorithms: Linear Regression (LR), K-Nearest Neighbor (KNN), Neural Net (NNET) and Gradient Boosting Machines (GBM).

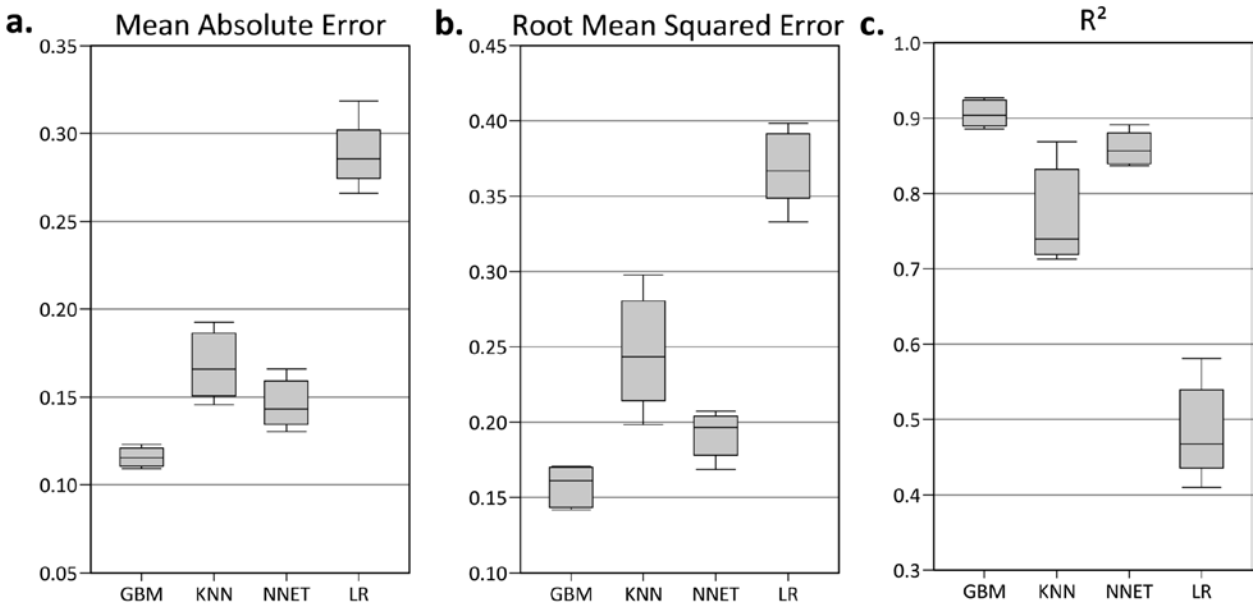

Figure 5. Loss function and overall fit calculated for four machine learning algorithms. The loss is a) calculated using Mean Absolute Error (MAE), b) Root Mean Squared Error (RMSE) and c) the overall fit is measured using R^2. The four algorithms are GBM (Gradient Boosted Model), KNN (K-Nearest Neighbor), NNET (Neural Net) and LR (Linear Regression). The box and whiskers plots represent the minimum value, the 25th percentile, the median, the 75th percentile, and the maximum value.

The algorithms were developed using 70% of our data as the training set, and the remaining 30% of data to validate the four algorithms. The loss function was calculated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for the four algorithms (Fig. 5a and b), where lower loss indicates a better algorithm. We calculated the loss with a five-fold cross validation using the resamples function of the caret package in R. Upon inspection, the GBM algorithm emerged as the top performer while the linear regression benchmark performed the worst in terms of both MAE and RMSE. The outright ranking between KNN and NNET was not possible, but the NNET algorithm appeared to have less variance (Fig. 5a and b). Additionally, the higher error bars around the KNN algorithm may suggest over-fitting relative to other algorithms, which raised the concern that prediction using KNN may be less reliable.

In order to further evaluate and rank the predictive capacity of the four algorithms, we evaluated the overall fit using the R^2, where a higher value indicates a more predictive algorithm (Fig. 5c). Again, the GBM algorithm demonstrated the best predictive capacity. The GBM algorithm managed to achieve an R^2 range of 89% to 93% using the five-fold cross-validation approach. Notably, these R^2 values are close to the theoretical optimum for our dataset established using the average of the experimental replicates. However, the loss calculated using RMSE for GBM was higher than the theoretically smallest value. The R model containing the four algorithms and associated functions are given in the supporting information (Note S1; Supplementary Method 1; Supplementary Equation 1-

4). Based on these analyses, we report the GBM algorithm which enables users to predict the productivity of lyophilized *E. coli* extract as a function of additive identity, concentration, and storage temperature for their applications of cell-free systems.

Discussion

Here we sought to identify low-cost molecules as additives to enhance the stability of *E. coli*-based cell extracts in order to enable field applications of CFPS. The challenge of stabilizing biological materials at ambient temperatures for extended periods of time has been pursued by many, with much success in the fields of industrial and therapeutic proteins. However, the chemical space for putative stabilizers of cell extracts is large, has not been thoroughly mapped, and predictive capacity for stabilization remains non-existent. In order to overcome these, we mapped the additives landscape comprised of 10 additives, falling into four distinct mechanisms of action, individually and in combination.

Of the 10 single additives, sugars emerged as the optimal candidates for extract stabilization, with trehalose outperforming all others. The tested sugars are postulated to work through robust vitrification mechanisms and hydrogen bonding, with potential uses as supplemental energy sources and molecular crowding agents. We observed that the stabilizing effects of different sugars vary, highlighting the lack of predictive capacity for selecting a stabilizing sugar, and underscoring the gaps in our understanding of the molecular basis for biomolecular stabilization. The combinatorial effects of three additives from different classes enhanced the shelf-life of the extract to a greater extent than any one single additive. Remarkably, we have identified a formulation of additives capable of retaining nearly full extract productivity after two weeks of storage at 23 °C compared the benchmark condition of storage at -80 °C (Fig. 4c). This formulation is also capable of maintaining the full productivity of lyophilized extract already containing solutions A, B, and DNA, such that the reaction can be initiated upon addition of water (Fig. S4). Unexpectedly, we observed that in addition to supporting extract stability, unique formulations of additives can nearly double extract productivity when stored at -80 °C (Fig. 4a). These efforts have also elucidated relationships between an additive's optimal concentration and extract storage temperature. Specifically, higher concentrations of trehalose and lower concentrations of trimethylglycine are favored at higher temperatures, while the target range of PEG concentrations broadens at higher temperatures in conjunction with companion additives. The data described in this work have also supported the development of a machine learning algorithm that enables the field to predict and rapidly validate additive formulations that may be suitable for their specific biomanufacturing and field applications of CFPS.

Importantly, the affordability of CFPS remains largely unaltered, increasing the cost per 30 µL reaction by approximately $0.0006. Notably, additives provide a cost benefit to CFPS in cases where yield is increased, for example, the addition of trehalose,

trimethylglycine, and PEG 8000 to extract stored at -80 °C cuts the cost per µg of protein by nearly half. This advance provides an important advance toward making CFPS a cost competitive biomanufacturing platform.

The discoveries made here set the stage for additional in-depth and extensive investigations of additive combinations for CFPS enabled by robotics. While ternary co-titrations are helpful, they are limited in their ability to fully screen combinations of three additives due to the intrinsic nature of the plot's design. As such, it may prove useful to screen unrepresented combinations or to more closely screen areas corresponding to ternary co-titration hotspots for fine-tuned optimization, which can be assisted through predictions from the machine learning algorithm reported herein. Assessment of longer storage times may further illuminate the ruggedness of the cell extract in the presence of stabilizers, and how the hotspots for optimal formulations may migrate with time instead of temperature. Formulations of additives reported herein may also improve yields and stability in other types of cell-free extracts, such as CHO, rabbit reticulocyte, insect, *Vibrio*, and yeast. We anticipate that formulations will need to be tuned to optimize the function of non-model proteins expressed in CFPS. Overall, the formulation of additives identified in this work show an unprecedented ability to maintain or improve productivity of *E. coli*-based crude cell extracts when used in combination with lyophilization, representing a key step making the CFPS a tractable and practical platform for biomanufacturing and field applications.

Methods

Extract Preparation

An overnight culture of BL21* DE3, started from a single colony on a plate, was inoculated into a 2L baffled flask containing 1L of 2x YPTG (5 g NaCl, 16 g Tryptone, 10 g Yeast extract, 7 g KH_2PO_4, 3 g $KHPO_4$, pH 7.2/750 mL solution, 18 g Glucose/250 mL solution) to an optical density of 0.1. Growth at 37 °C and 200 rpm was monitored until an optical density of 0.6, whereupon T7 RNA Polymerase (T7 RNAP) expression was induced by addition of IPTG to a final concentration of 1mM. Once an optical density of 3 was reached, cells were harvested by centrifugation in 1 L bottles at 4 °C, 5,000 xg, for 10 min. Harvested cells were resuspended in 30 mL of S30 buffer (10 mM Tris OAc, pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KOAc, 2 mM DTT) by vortexing, then spun down at 4 °C, 5,000 xg, for 10 min. This buffer exchange was carried out two additional times. Cell pellets were flash frozen and stored at -80 °C or used immediately for extract preparation. Cell pellets were resuspended in 1 mL of S30 buffer per 1 g of cells. 1.4 mL of resuspended cells were aliquoted to a 1.5 mL microfuge tube. The resuspension was sonicated using a Qsonica Q125 Sonicator with a 3.175 mm probe, with the cell resuspension surrounded by an ice water bath. Three pulses of 45 s on and 59 s off, at 50% amplitude were carried out. Immediately after sonication, 4.5 µL of DTT was spiked

into the lysate and the tube was inverted several times to mix. Lysate was centrifuged at 4 °C and 18,000 xg for 10 min. The clear supernatant was pipetted off into a 1.5 mL microfuge tube, which was incubated at 37 °C and 250 rpm for 60 min. After the run-off, the tube was centrifuged at 4 °C and 10,000 xg for 10 min. The resulting supernatant is the cell extract. Extract was supplemented with T7 RNAP to a final concentration of 15.7 μg/mL. The mixture was flash frozen and stored at -80 °C until further use. Detailed protocols for growth, extract preparation, and reaction setup have been previously described.[61]

DNA Purification

pJL1 sfGFP was purified from DH5α cells using an Invitrogen PureLink™ HiPure Plasmid Maxiprep Kit. The final DNA was eluted using warm molecular biology grade water instead of the provided TE buffer, for compatibility with CFPS reactions. A working concentration of 260 ng/μL was used, and the DNA was stored at -20 °C.

Additive Solution Preparation

All additives were dissolved in molecular biology grade water. Sugar solutions were made at a 5.55x final concentration, and all remaining additive solutions were made at a 15x final concentration. All referenced concentrations of additives refer to the final concentration of that additive in the 30 μL CFPS reaction.

Lyophilization

Pre-reaction mixtures were prepared by combining 10.8 μL of T7RNAP supplemented extract and the necessary additive solution(s) (5.4 μL of sugar solutions, 2 μL of other additives) in 200 μL PCR tubes on ice. Each condition tested was performed in quadruplicate. Tubes were briefly centrifuged and gently vortexed to ensure the reagents were well mixed and in a single bead at the bottom of the tube. Tubes were flash frozen with liquid nitrogen on PCR plates. Immediately after, the PCR plates were placed into lyophilization jars and connected to a Labconco 4.5 L Console Freeze Drier System for two hours. Following lyophilization, tubes were partially closed and the PCR plates were placed into sealed, freezer-grade plastic bags containing blue silica gel desiccant beads and stored at the appropriate storage condition (-80 °C, 4 °C, or 23 °C (room temperature)) for two weeks. Room temperature measured via thermostat fluctuated between 21 °C and 25 °C over the storage period, with an average temperature of 23.08 °C (Fig. S1).

CFPS Reactions

Each condition was performed in quadruplicate. In-house pre-mixes containing the necessary energy system and cofactors were used. The final CFPS reaction contained the following concentrations of each reagent: 33.3% v/v cell extract (unsupplemented),

5.7 μg/mL T7 RNAP, 16 ng/μL pJL1 sfGFP DNA plasmid, Solution A (1.2 mM ATP, 0.850 mM GTP, 0.850 mM UTP, 0.850 mM CTP, 31.50 μg/mL Folinic Acid, 170.60 μg/mL tRNA, 0.40 mM Nicotinamide Adenine Dinucleotide (NAD), 0.27 mM Coenzyme A (CoA), 4.00 mM Oxalic Acid, 1.00mM Putrescine, 1.50 mM Spermidine, and 57.33 mM HEPES buffer), Solution B (10 mM $Mg(Glu)_2$, 10 mM $NH_4(Glu)$, 130 mM K(Glu), 2 mM each of the 20 amino acids, and 0.03 M Phosphoenolpyruvate (PEP)), the corresponding concentration of additive(s), and water to a final volume of 30 μL. This reaction volume was selected as a function of unsupplemented 10 μL extract volume used for reproducible lyophilization in PCR tubes. All indicated additive concentrations refer to the final concentration in the complete CFPS reaction, and conversions to the moles of additive in dried extract can be found in Table S1. For lyophilized extracts, the dried pellet was first rehydrated with molecular biology grade water. 80% of the original volume of extract, plus the volume of additives was used for rehydration. All reactions were briefly centrifuged and gently vortexed to ensure the reaction was well-mixed and in a single bead at the bottom of the tube. Reactions were incubated overnight (18 to 24 hrs) at 37 °C, and quantification was completed the next day.

Quantification and Data Analysis

Quantification of protein production took place in half-volume black, opaque bottomed 96-well plate. In each well, 2 μL of CFPS reaction solution was added to 48 μL of 0.05 M HEPES, pH 8. Each individual reaction was quantified in triplicate, giving 12 readings per tested condition. Fluorescence intensity was read with an excitation wavelength of 485 nm and emission wavelength of 510 nm on a Cytation™ 5. A previously established standard curve was used to determine [sfGFP] from the fluorescence measurements (Fig. S3). The [sfGFP] values obtained were averaged for each reaction, resulting in 4 values for each tested condition. Outliers for each condition were identified and removed using a Grubbs' Test for both minimum and maximum values and a 90% confidence interval. Each condition sustained a minimum of N=3 samples and a maximum of N=4 samples. The remaining values were averaged and the standard deviation was determined.

SigmaPlot Binary Co-titration Plotting

Averaged values were plotted in SigmaPlot to obtain heatmaps for binary co-titrations. This was done by selecting Contour plot>Filled Contour Plot>XYZ Triplet and assigning X and Y values to additive titration ranges and Z values to corresponding protein yield. The colormap was defined using the same color values from the Matlab plots. Continuous legend bars were created by running the data through the Matlab ternary plotting script.

Matlab Ternary Co-titration Plotting

Heatmap plots for ternary co-titrations were created using a ternary plot program from MATLAB file exchange and the default "jet" colormap. (https://www.mathworks.com/matlabcentral/fileexchange/7210-ternary-plots)

Machine Learning

Given the concentration of the ten additives $\boldsymbol{x} = \{x_1, x_2, \cdots, x_{10}\}$, we intended to understand the yield y at some temperature T. In machine learning (ML) jargon, elements of the vector $\boldsymbol{x}$ are called predictors or features. The goal of the ML algorithm is to predict y conditional on $\boldsymbol{x}$**.** Mathematically, suppose that the true yield is

$$y = B(T) \times e^{h(\boldsymbol{x};T) + \in}$$

In general, it is easier to predict a scale-independent outcome variable. The function $B(T)$ allows us to achieve scale-independence — it shows the benchmark yield at a certain temperature. That is, the function $B(T)$ is the average lyophilized yield without any additives. The equation says that the yield with additives depends on the benchmark adjusted for two terms. The first term represented by $h(\boldsymbol{x};T)$ shows the effect of additives at respective temperatures. The second term $\in$ shows the experimental error, which encapsulates inconsistencies such as differences in humidity and room temperature, and user-to-user variance. To summarize, we use ML to predict the relative yield $ry \equiv \ln\frac{y}{B(T)}$ where

$$ry = h(\boldsymbol{x};T) + \in$$

The ML algorithm tries to estimate $h(\boldsymbol{x};T)$.

In general, we wanted to compute a "good" prediction of the relative yield ry given additive concentrations $\boldsymbol{x}$. Usually "good" means minimizing some loss function. To compare across the four ML algorithms, we calculated the loss using the Root Mean Squared Error (RMSE). For exposition clarity, we also report the Mean Absolute Error (MAE) and R^2 for the four algorithms as well.

To minimize over-fitting, we followed the best practices. Specifically, the standard consisted of two steps. First, we randomly split the data set in two parts: the training data set and the validation data set. The training data set consists of 70% of the total data and each of the four algorithms are "trained" using it. The validation data set is used to calculate the loss from an out-of-sample forecast. Second, the KNN, NNET and GBM algorithms depend on exogenous parameters. We tune the parameters using five-fold cross-validation which can be summarized by the following procedure:

1. Divide the data into five roughly equal subsets (folds) and label them by s = $1, 2, \cdots, 5$. Start with subset s = 1.
2. Pick a value for the tuning parameter.
3. For each algorithm, fit the model using the k − 1 subsets other than subset s.
4. Predict for subset s and measure the associated loss.
5. Stop if s = k, otherwise increment s by 1 and go to step 2.

In the setup, common sense also dictates some parameters. For example, in the KNN algorithm, one needs to pick the minimum number of neighbors. In the setup, since each control had a maximum of four replicates, we set the minimum neighbors to four. Relatedly, in the GBM algorithm, one chooses maximum interaction depth which controls for interactions between predictors. In the setup, we consider temperature and up to three additives, indicating an interaction depth of four.

Overview of the Algorithms

As explained above, the goal of the four algorithms is to predict the best $h(\boldsymbol{x};T)$. Here we give a brief overview of the four algorithms which were predominantly implemented using the caret package in R.

1. Linear Regression (LR): Mathematically, we restrain the function $h(\boldsymbol{x};T)$ to be linear in its arguments:

$$h(\boldsymbol{x};T) = a_0(T) + x_1 \times \beta_1 + x_2 \times \beta_2 + \cdots + x_{10} \times \beta_{10}$$

   This algorithm is clearly constrained as it does not consider the non-linearity. For example, this algorithm cannot predict the fact that higher sugar concentrations are detrimental to the yield. Nonetheless, the linearity assumption serves as a good benchmark.
2. K-Nearest Neighbor (KNN): The k-nearest neighbor technique is a non-parametric pattern recognition technique that uses the average of the k closest observations in the training set to provide an estimate. Then as expected, small values for k can lead to over-fitting whereas high values of k can under fit. We choose k to equal to 4 as we had 4 replicates, even though we found lower values of k to be optimal from a cross-validation analysis.
3. Neural Network (NNet): This technique is used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. In this sense, neural network is well suited for our setup. NNET models are designed to mimic human brain which comprises interconnected synaptic neurons capable of learning and storing information about their environment. Mathematically, a neuron model comprises three elements: the connecting links characterized by their strength, a linear combiner which combines the weighted input signals and an activation function for limiting the amplitude range of the neuron's output to some finite value. Network types, topologies, and training techniques vary considerably, but a rudimentary explanation of the critical aspects of backpropagation neural networks is contained in Burks et al. (2000) and Rumelhart and McClelland (1986).[62,63]
4. Gradient Boosting Machines (GBM): The generalized boosted regression model algorithm depends on two different algorithms; regression trees and gradient boosting. Regression trees are simple models that fit yield to predictors by partitioning the feature space using a series of partition rules, e.g. binary split, to identify regions in the data having the most consistent responses to predictors. A constant is then fitted

to each region (e.g. mean response for observations in a particular region, in a regression problem). Gradient boosting, on the other hand, combines the regression tree output to produce a more powerful and improved predictive performance. Therefore, the final model would be a combination of several individual regression trees fitted in a forward stage-wise manner.

Abbreviations

CFPS, cell-free protein synthesis; sfGFP, superfolder green fluorescent protein; T7RNAP, T7 RNA polymerase; ML, machine learning; LR, linear regression; KNN, K-nearest neighbor; NNet, neural network; GBM, Gradient Boosting Machines; MAE, mean absolute error; RMSE, root mean squared error.

Supporting Information

The supporting information is available free of charge via the internet at http://pubs.acs.org.

Fluctuation in room temperature during storage; CFPS reaction vessel size comparison; standard curve for quantification of sfGFP; complete and modular lyophilized CFPS reaction stability; additive concentration conversions; post-Grubbs' test data for all experiments; supplementary files and descriptions of each for machine learning; supplementary file and description for pre-Grubbs' test data.

Author Information

* Corresponding author
Email: joza@calpoly.edu

Author Contribution

J.P.O, W.Y.K, N.E.G., and L.C.W. conceived the project, designed the experiments, and interpreted the data. N.E.G, W.Y.K., and L.C.W performed all experiments, collected the data, and generated all figures. N.E.G. managed the project and team. P.A.P carried out machine learning and provided the corresponding written and graphical materials. N.E.G, W.Y.K, L.C.W., and J.P.O. wrote the paper. C.M.H developed the recipes for simplified reagent pre-mixes and performed initial CFPS reactions. K.R.W assisted in manuscript preparation.

Conflicting Interests

The authors declare no competing interests.

Acknowledgment

Authors would also like to acknowledge Andrea Laubscher, Dr. Jennifer VanderKelen, Celine diBernardo, and Kevin Dunham for technical support, Max Z. Levine, Phillip E. J. Smith, and Logan Burrington for helpful discussions. Dr. Michael Jewett for gifting purified T7RNAP, Dr. Gregory Scott and Dr. Ashley McDonald for software and data visualization support, and Dr. Steven Wilkinson, Dr. Chad Immoos, and Dr. Andres Martinez, for helpful discussions. Authors also thank the reviewers of the manuscript for their helpful insights. The research was supported by the Bill and Linda Frost Fund, Center for Applications in Biotechnology's Chevron Biotechnology Applied Research Endowment Grant, Cal Poly Research, Scholarly, and Creative Activities Grant Program (RSCA 2017), and the National Science Foundation (NSF-1708919) grant awarded to K.R.W. and J.P.O.

References

(1) Gregorio, N. E., Levine, M. Z., Oza, J. P. (2019) A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc. 2*, 24.

(2) Huang, A., Nguyen, P. Q., Stark, J. C., Takahashi, M. K., Donghia, N., Ferrante, T., Dy, A. J., Hsu, K. J., Dubner, R. S., Pardee, K., et al. (2018) BioBits™ Explorer: A Modular Synthetic Biology Education Kit. *Sci. Adv. 4*, eaat5105.

(3) Stark, J. C., Huang, A., Nguyen, P. Q., Dubner, R. S., Hsu, K. J., Ferrante, T. C., Anderson, M., Kanapskyte, A., Mucha, Q., Packett, J. S., et al. (2018) BioBits™ Bright: A Fluorescent Synthetic Biology Education Kit. *Sci. Adv. 4*, eaat5107.

(4) Stark, J. C., Huang, A., Hsu, K. J., Dubner, R. S., Forbrook, J., Marshalla, S., Rodriguez, F., Washington, M., Rybnicky, G. A., Nguyen, P. Q., et al. (2019) BioBits Health: Classroom Activities Exploring Engineering, Biology, and Human Health with Fluorescent Readouts. *ACS Synth. Biol. 8*, 1001–1009.

(5) Alam, K. K., Jung, J. K., Verosloff, M. S., Clauer, P. R., Lee, J. W., Capdevila, D. A., Pastén, P. A., Giedroc, D. P., Collins, J. J., Lucks, J. B. (2019) Rapid, Low-Cost Detection of Water Contaminants Using Regulated In Vitro Transcription. *bioRxiv* 619296.

(6) Takahashi, M. K., Tan, X., Dy, A. J., Braff, D., Akana, R. T., Furuta, Y., Donghia, N., Ananthakrishnan, A., Collins, J. J. (2018) A Low-Cost Paper-Based Synthetic Biology Platform for Analyzing Gut Microbiota and Host Biomarkers. *Nat. Commun. 9*, 1–12.

(7) Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., et al. (2016) Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell 165*, 1255–1266.

(8) Salehi, A. S. M., Smith, M. T., Bennett, A. M., Williams, J. B., Pitt, W. G., 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.

(9) Voyvodic, P. L., Pandi, A., Koch, M., Conejero, I., Valjent, E., Courtet, P., Renard, E., Faulon, J.-L., Bonnet, J. (2019) Plug-and-Play Metabolic Transducers Expand the Chemical Detection Space of Cell-Free Biosensors. *Nat. Commun. 10*, 1697.

(10) Silverman, A. D., Akova, U., Alam, K. K., Jewett, M. C., Lucks, J. B. (2019)

Design and Optimization of a Cell-Free Atrazine Biosensor. *bioRxiv* 779827.
(11) Soltani, M., Davis, B. R., Ford, H., Nelson, J. A. D., Bundy, B. C. (2018) Reengineering Cell-Free Protein Synthesis as a Biosensor: Biosensing with Transcription, Translation, and Protein-Folding. *Biochem. Eng. J. 138*, 165–171.
(12) Bundy, B. C., Porter Hunt, J., Jewett, M. C., Swartz, J. R., Wood, D. W., Frey, D. D., Rao, G. (2018) Cell-Free Biomanufacturing. *Curr. Opin. Chem. Eng. 22*, 177–183.
(13) Burgenson, D., Gurramkonda, C., Pilli, M., Ge, X., Andar, A., Kostov, Y., Tolosa, L., Rao, G. (2018) Rapid Recombinant Protein Expression in Cell-Free Extracts from Human Blood. *Sci. Rep. 8*, 1–7.
(14) Hurst, G. B., Asano, K. G., Doktycz, C. J., Consoli, E. J., Doktycz, W. L., Foster, C. M., Morrell-Falvey, J. L., Standaert, R. F., Doktycz, M. J. (2017) Proteomics-Based Tools for Evaluation of Cell-Free Protein Synthesis. *Anal. Chem. 89*, 11443–11451.
(15) Foshag, D., Henrich, E., Hiller, E., Schäfer, M., Kerger, C., Burger-Kentischer, A., Diaz-Moreno, I., García-Mauriño, S. M., Dötsch, V., Rupp, S., et al. (2018) The E. Coli S30 Lysate Proteome: A Prototype for Cell-Free Protein Production. *N. Biotechnol. 40*, 245–260.
(16) Smith, M. T., Berkheimer, S. D., Werner, C. J., Bundy, B. C. (2014) Lyophilized Escherichia Coli-Based Cell-Free Systems for Robust, High-Density, Long-Term Storage. *Biotechniques 186*, 186–193.
(17) Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., Swartz, J. R. (2008) An Integrated Cell-Free Metabolic Platform for Protein Production and Synthetic Biology. *Mol. Syst. Biol. 4*, 220.
(18) Salehi, A. S. M., Shakalli Tang, M. J., Smith, M. T., Hunt, J. M., Law, R. A., Wood, D. W., Bundy, B. C. (2017) Cell-Free Protein Synthesis Approach to Biosensing HTRβ-Specific Endocrine Disruptors. *Anal. Chem. 89*, 3395–3401.
(19) Cameron, D. E., Ferrante, T., Collins, J. J., DaleyKeyser, A., Pardee, K., Yin, P., Green, A. A. (2014) Paper-Based Synthetic Gene Networks. *Cell 159*, 940–954.
(20) Dudley, Q. M., Anderson, K. C., Jewett, M. C. (2016) Cell-Free Mixing of *Escherichia Coli* Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis. *ACS Synth. Biol. 5*, 1578–1588.
(21) McNerney, M. P., Zhang, Y., Steppe, P., Silverman, A. D., Jewett, M. C., Styczynski, M. P. (2019) Point-of-Care Biomarker Quantification Enabled by Sample-Specific Calibration. *Sci. Adv. 5*, eaax4473.
(22) Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., Ferrante, T., McSorley, F. R., Furuta, Y., Vernet, A., et al. (2016) Portable, On-Demand Biomolecular Manufacturing. *Cell 167*, 248-259.e12.
(23) Gräwe, A., Dreyer, A., Vornholt, T., Barteczko, U., Buchholz, L., Drews, G., Ho, U. L., Jackowski, M. E., Kracht, M., Lüders, J., et al. (2019) A Paper-Based, Cell-Free Biosensor System for the Detection of Heavy Metals and Date Rape Drugs. *PLoS One 14*, 1–22.
(24) Benítez-Mateos, A. I., Llarena, I., Sánchez-Iglesias, A., López-Gallego, F. (2018) Expanding One-Pot Cell-Free Protein Synthesis and Immobilization for On-Demand Manufacturing of Biomaterials. *ACS Synth. Biol. 7*, 875–884.
(25) Smith, M. T., Bennett, A. M., Hunt, J. M., Bundy, B. C. (2015) Creating a

Completely “Cell-Free” System for Protein Synthesis. *Biotechnol. Prog. 31*, 1716–1719.
(26) Wilding, K. M., Zhao, E. L., Earl, C. C., Bundy, B. C. (2019) Thermostable Lyoprotectant-Enhanced Cell-Free Protein Synthesis for on-Demand Endotoxin-Free Therapeutic Production. *N. Biotechnol. 53*, 73–80.
(27) Kai, L., Dötsch, V., Kaldenhoff, R., Bernhard, F. (2013) Artificial Environments for the Co-Translational Stabilization of Cell-Free Expressed Proteins. *PLoS One 8*, e56637.
(28) Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., Gold, D. S., Heinsohn, H. G., Murray, C. J. (2011) Microscale to Manufacturing Scale-up of Cell-Free Cytokine Production--a New Approach for Shortening Protein Production Development Timelines. *Biotechnol. Bioeng. 108*, 1570–1578.
(29) 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., Emanuel, P. A., et al. (2019) Quantification of Interlaboratory Cell-Free Protein Synthesis Variability. *ACS Synth. Biol.* acssynbio.9b00178.
(30) Arakawa, T., Timasheff, S. N. (1982) Stabilization of Protein Structure by Sugars. *Biochemistry 21*, 6536–6544.
(31) Carpenter, J. F., Prestrelski, S. J., Arakawa, T. (1993) Separation of Freezing- and Drying-Induced Denaturation of Lyophilized Proteins Using Stress-Specific Stabilization: I. Enzyme Activity and Calorimetric Studies. *Arch. Biochem. Biophys. 303*, 456–464.
(32) Allison, S. D., Chang, B., Randolph, T. W., Carpenter, J. F. (1999) Hydrogen Bonding between Sugar and Protein Is Responsible for Inhibition of Dehydration-Induced Protein Unfolding. *Arch. Biochem. Biophys. 365*, 289–298.
(33) Chatterjee, K., Shalaev, E. Y., Suryanarayanan, R. (2005) Raffinose Crystallization during Freeze-Drying and Its Impact on Recovery of Protein Activity. *Pharm. Res. 22*, 303–309.
(34) Kaushik, J. K., Bhat, R. (2003) Why Is Trehalose an Exceptional Protein Stabilizer? An Analysis of the Thermal Stability of Proteins in the Presence of the Compatible Osmolyte Trehalose. *J. Biol. Chem. 278*, 26458–26465.
(35) Graziano, G. (2012) How Does Sucrose Stabilize the Native State of Globular Proteins? *Int. J. Biol. Macromol. 50*, 230–235.
(36) Bai, L., Guo, X., Zhang, X., Yu, W., Yang, D. (2019) Saccharides Create a Crowding Environment for Gene Expression in Cell-Free Systems. *Langmuir 35*, 5931–5936.
(37) Caschera, F., Noireaux, V. (2014) Synthesis of 2.3 Mg/Ml of Protein with an All Escherichia Coli Cell-Free Transcription-Translation System. *Biochimie 99*, 162–168.
(38) Mensink, M. A., Frijlink, H. W., van der Voort Maarschalk, K., Hinrichs, W. L. J. How Sugars Protect Proteins in the Solid State and during Drying (Review): Mechanisms of Stabilization in Relation to Stress Conditions. *European Journal of Pharmaceutics and Biopharmaceutics*. 2017, pp 288–295.
(39) Karig, D. K., Bessling, S., Thielen, P., Zhang, S., Wolfe, J. (2017) Preservation of Protein Expression Systems at Elevated Temperatures for Portable Therapeutic Production. *J. R. Soc. Interface 14*, 20161039.

(40) Simperler, A., Kornherr, A., Chopra, R., Bonnet, P. A., Jones, W., Motherwell, W. D. S., Zifferer, G. (2006) Glass Transition Temperature of Glucose, Sucrose, and Trehalose: An Experimental and in Silico Study. *J. Phys. Chem. B 110*, 19678–19684.
(41) Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G., Jewett, M. C. (2018) Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from *Vibrio Natriegens*. *ACS Synth. Biol. 7*, 2245–2255.
(42) Jones, K. L., Drane, D., Gowans, E. J. (2007) Long-Term Storage of DNA-Free RNA for Use in Vaccine Studies. *Biotechniques 43*, 675–681.
(43) Kim, H.-C., Kim, T.-W., Kim, D.-M. (2011) Prolonged Production of Proteins in a Cell-Free Protein Synthesis System Using Polymeric Carbohydrates as an Energy Source. *Process Biochem. 46*, 1366–1369.
(44) Santoro, M. M., Liu, Y., Khan, S. M. A., Hou, L. X., Bolen, D. W. (1992) Increased Thermal Stability of Proteins in the Presence of Naturally Occurring Osmolytes. *Biochemistry 31*, 647–654.
(45) Meadows, J. A., Wargo, M. J. (2015) Carnitine in Bacterial Physiology and Metabolism. *Microbiology 161*, 1161–1174.
(46) Arakawa, T., Timasheff, S. N. (1985) The Stabilization of Proteins by Osmolytes. *Biophys. J. 47*, 411–414.
(47) Bruździak, P., Panuszko, A., Stangret, J. (2013) Influence of Osmolytes on Protein and Water Structure: A Step To Understanding the Mechanism of Protein Stabilization. *J. Phys. Chem. B 117*, 11502–11508.
(48) Yancey, P. H., Siebenaller, J. F. (2015) Co-Evolution of Proteins and Solutions: Protein Adaptation versus Cytoprotective Micromolecules and Their Roles in Marine Organisms. *J. Exp. Biol. 218*, 1880–1896.
(49) Rydeen, A. E., Brustad, E. M., Pielak, G. J. (2018) Osmolytes and Protein–Protein Interactions. *J. Am. Chem. Soc. 140*, 7441–7444.
(50) Moriizumi, Y., Tabata, K. V, Miyoshi, D., Noji, H. (2019) Osmolyte-Enhanced Protein Synthesis Activity of a Reconstituted Translation System. *ACS Synth. Biol. 8*, 557–567.
(51) Liao, Y.-T., Manson, A. C., DeLyser, M. R., Noid, W. G., Cremer, P. S. (2017) Trimethylamine N-Oxide Stabilizes Proteins via a Distinct Mechanism Compared with Betaine and Glycine. *Proc. Natl. Acad. Sci. U. S. A. 114*, 2479–2484.
(52) Hoffmann, C., Blume, A., Miller, I., Garidel, P. (2009) Insights into Protein–Polysorbate Interactions Analysed by Means of Isothermal Titration and Differential Scanning Calorimetry. *Eur. Biophys. J. 38*, 557–568.
(53) Jewett, M. C., Swartz, J. R. (2004) Mimicking TheEscherichia Coli Cytoplasmic Environment Activates Long-Lived and Efficient Cell-Free Protein Synthesis. *Biotechnol. Bioeng. 86*, 19–26.
(54) Ge, X., Luo, D., Xu, J. (2011) Cell-Free Protein Expression under Macromolecular Crowding Conditions. *PLoS One 6*, e28707.
(55) Rawat, S., Raman Suri, C., Sahoo, D. K. (2010) Molecular Mechanism of Polyethylene Glycol Mediated Stabilization of Protein. *Biochem. Biophys. Res. Commun. 392*, 561–566.
(56) Agarkhed, M., O'Dell, C., Hsieh, M.-C., Zhang, J., Goldstein, J., Srivastava, A. (2013) Effect of Polysorbate 80 Concentration on Thermal and Photostability of a

Monoclonal Antibody. *AAPS PharmSciTech 14*, 1–9.
(57) Lee, H. J., McAuley, A., Schilke, K. F., McGuire, J. (2011) Molecular Origins of Surfactant-Mediated Stabilization of Protein Drugs. *Adv. Drug Deliv. Rev. 63*, 1160–1171.
(58) Wang, W., Wang, Y. J., Wang, D. Q. (2008) Dual Effects of Tween 80 on Protein Stability. *Int. J. Pharm. 347*, 31–38.
(59) Hong, S. H., Kwon, Y.-C., Martin, R. W., Des Soye, B. J., de Paz, A. M., Swonger, K. N., Ntai, I., Kelleher, N. L., Jewett, M. C. (2015) Improving Cell-Free Protein Synthesis through Genome Engineering of *Escherichia Coli* Lacking Release Factor 1. *ChemBioChem 16*, 844–853.
(60) Voloshin, A. M., Swartz, J. R. (2005) Efficient and Scalable Method for Scaling up Cell Free Protein Synthesis in Batch Mode. *Biotechnol. Bioeng. 91*, 516–521.
(61) Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R., Oza, J. P. (2019) Escherichia Coli-Based Cell-Free Protein Synthesis: Protocols for a Robust, Flexible, and Accessible Platform Technology. *J. Vis. Exp.* e58882.
(62) Rumelhart, D. E., Widrow, B., Lehr, M. A. (1994) The Basic Ideas in Neural Networks. *Commun. ACM 37*, 87–92.
(63) T. F. Burks, T. F., S. A. Shearer, S. A., R. S. Gates, R. S., K. D. Donohue, K. D. (2000) Backpropagation Neural Network Design and Evaluation for Classifying Weed Species Using Color Image Texture. *Trans. ASAE 43*, 1029–1037.

For Table of Contents Use Only

Unlocking applications of cell-free biotechnology through enhanced shelf-life and productivity of *E. coli* extracts

Nicole E. Gregorio[# 1,2], Wesley Y. Kao[# 1,2], Layne C. Williams[# 1,2], Christopher M. Hight[1,2], Pratish A. Patel[3], Katharine R. Watts[1,2], Javin P. Oza[1,2*]

[1]Chemistry and Biochemistry Department, California Polytechnic State University, 1 Grand Ave., San Luis Obispo, CA 93407
[2]Center for Applications in Biotechnology, California Polytechnic State University, 1 Grand Ave., San Luis Obispo, CA 93407
[3]Department of Finance, Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA 93407

# These authors contributed equally to this work