

HTBAA Week 3: Designing a Cell-Free Reaction Composition

Homework Due: October 6, 2026

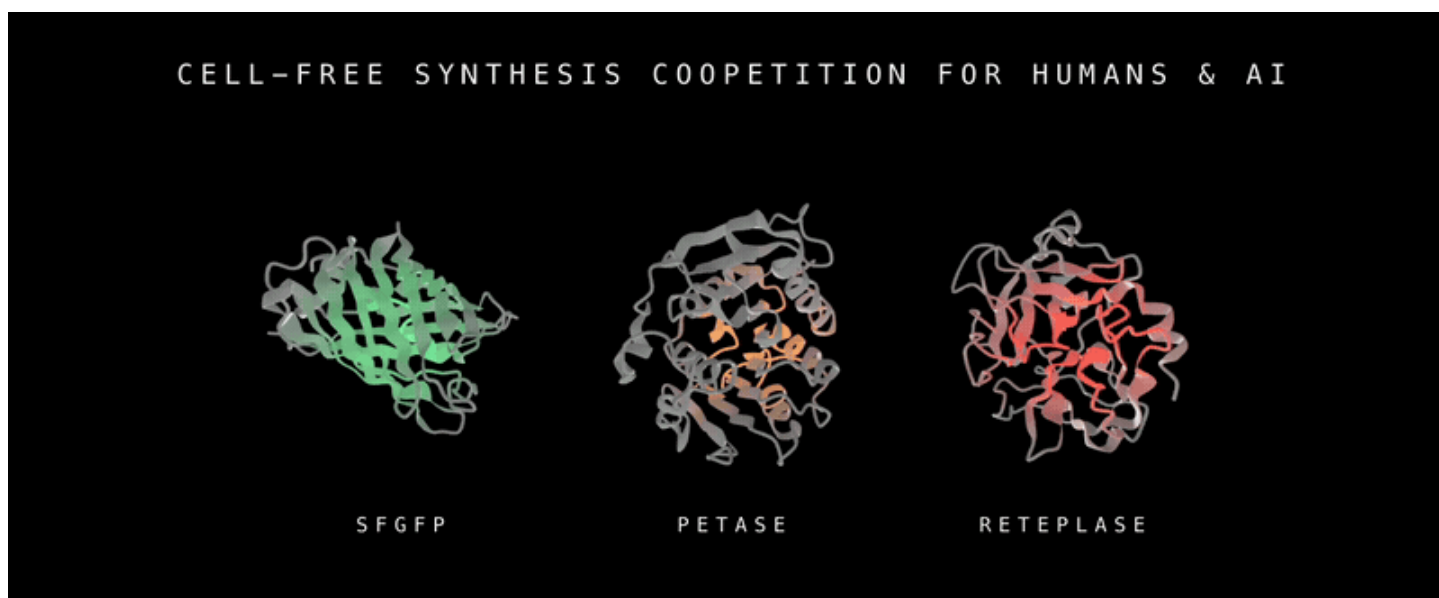
Homework's Associated Website: coopetition.htgaa.org

Coopetition: [coopetition specifications](#)

Homework created by: Ronan Donovan and Michelle Yue

1. Timeline & Overview

- Tue, Sept 29: Reshma Shetty lecture + coopetition introduction
- Thu, Oct 1: Ginkgo field trip + homework assignment introduced by Ronan
- Tue, Oct 6: Kate Adamala lecture + 30-min review of initial cell-free designs
- Thu, Oct 8: Recitation focused on global reaction composition ideation
- Week, Oct 12: Virtual office hours (TBD) to plan 80 unique compositions (no class time)
- Tue, Oct 20: HTBAA submits the official autonomous lab experiment request to Ginkgo



1. INTRODUCTION

A. HOMEWORK OVERVIEW

In this assignment, you will design new reaction compositions for synthesizing three proteins (sfGFP, PETase, and Reteplase) through cell-free protein synthesis (CFPS). In lysate-based CFPS, the protein-synthesis machinery is extracted from cells and used outside of intact living cells. The reaction combines a cell lysate containing the transcription and translation machinery with a DNA template, energy substrates, and other required reaction components to produce proteins in vitro. Because protein synthesis is decoupled from cell growth and viability, CFPS is highly tunable, supports rapid experimental iteration with results often available within hours, and is well suited for automated workflows (Silverman et al. 2020; Rasor et al. 2021).

This homework was designed based on an exciting ongoing project at Ginkgo Bioworks aimed at comparing the performance of human scientists and AI agents in designing cell-free protein synthesis reactions. Teams of researchers and AI agents are independently designing reaction compositions for the same three proteins (sfGFP, PETase, and Reteplase), with the goal of optimizing the cost of producing one gram of functional protein.

A central challenge in reaching this goal is in optimizing the reaction composition in which the CFPS takes place, as this strongly influences both CFPS yield and cost. Cell extract is typically the largest single expense in a CFPS reaction composition, while changes in the concentrations and combinations of other reaction components can alter protein yield and thus cost per gram by one to two orders of magnitude (Olsen et al. 2026; Smith et al. 2026). Because the optimal reaction composition for these proteins has not yet been established, identifying improved formulations requires highly iterative experimental screening.

However, the number of possible combinations of reaction components is too large to evaluate exhaustively. To address this challenge, Ginkgo Bioworks uses a staged screening strategy to efficiently evaluate candidate formulations. Reaction conditions are first tested in 384-well plates at 20 μ L-scale volumes to minimize reagent use, and the best-performing conditions are then re-tested in 96-well plates to assess reproducibility. Final candidates are evaluated in 2 mL tubes. Each stage of screening narrows the candidate reaction composition while testing reproducibility and scale-up viability.



Fig 1. Ginkgo's staged screening strategy. Testing begins in 384-well plates with best CFPS compositions further evaluated in 96-well and 2 mL tubes

B. THE CENTRAL DOGMA IN CFPS

Cells produce proteins through the central dogma of molecular biology:

DNA —transcription→ **mRNA** —translation→ **protein**

During transcription, a sequence of DNA is used as a template to produce messenger RNA (mRNA). During translation, ribosomes read the mRNA sequence in three-nucleotide codons and use this information to assemble amino acids, the building blocks of proteins, into a polypeptide chain. The resulting chain then folds into a specific three-dimensional structure that determines protein function, such as sfGFP producing fluorescence, PETase hydrolyzing PET, and Reteplase dissolving clots.

Cell-free protein synthesis carries out these same processes outside of living cells.

For this project, *E. coli* BL21 Star (DE3) cells are first grown, lysed, and clarified by centrifugation to remove membranes and cellular debris. The resulting lysate, or cell extract, retains functional ribosomes, tRNAs (which transport amino acids to the ribosomes), and enzymes required for transcription and translation. The lysate typically comprises approximately one-quarter of the reaction volume.

A DNA template encoding the desired protein is then combined with the lysate and a chemical mixture containing nucleotide triphosphates (or nucleotide monophosphates) for mRNA synthesis, twenty amino acids

for protein synthesis, an energy system that regenerates ATP, and salts such as magnesium and potassium that are required for proper function of the transcriptional and translational machinery. Once these components are combined, transcription and translation proceed within the reaction.

Importantly for this project, the gene is placed downstream of a T7 promoter and transcribed by T7 RNA polymerase, a fast and robust single-enzyme transcription system derived from a bacteriophage. The polymerase is supplied by the source *E. coli* strain—which is a BL21 Star (DE3) derivative that carries the T7 RNA polymerase gene—and is therefore already present in the lysate. Because transcription is carried out by a single polymerase, the same reaction framework can be used for nearly any gene placed downstream of a T7 promoter, contributing to the flexibility of CFPS programming.

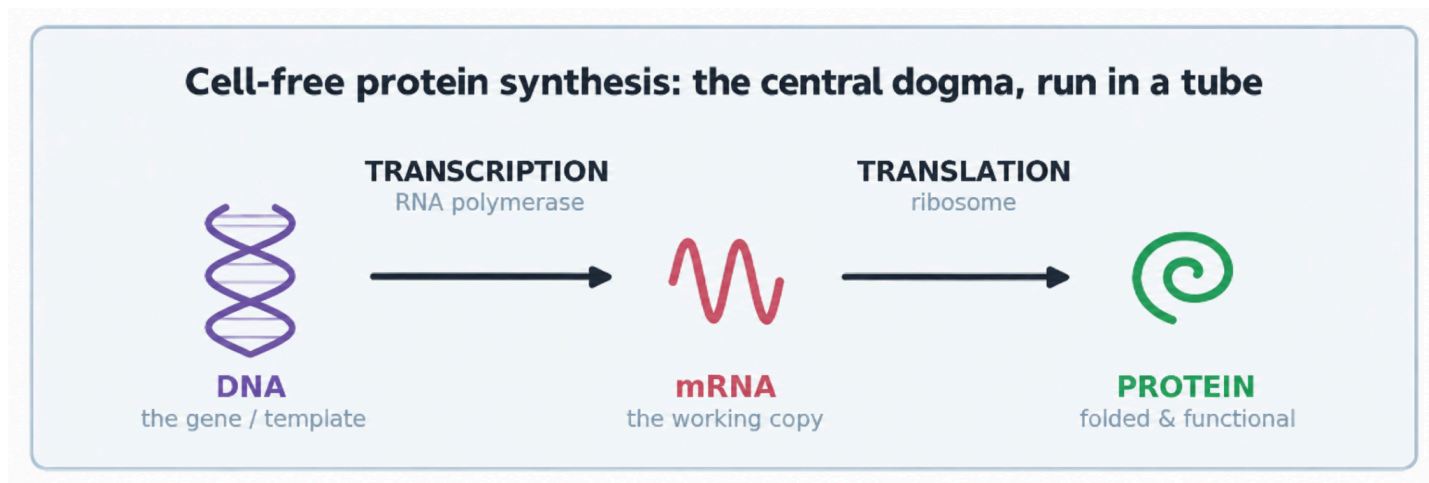


Fig 2. DNA is transcribed into mRNA which is translated by ribosomes into a folded protein, all in a cell free reaction tube.

These features make CFPS well suited for automation, optimization, and bioproduction. Because the reaction does not contain an intact cell membrane and does not require maintenance of living cells, individual components can be added, removed, or adjusted to directly evaluate their effects on protein production. Automated systems are particularly advantageous in that they can prepare and analyze hundreds of reaction compositions in parallel.

2. CFPS EXPERIMENTAL FRAMEWORK

A. OVERVIEW OF PROTEIN TARGETS

Each reaction composition is evaluated using three proteins (i.e., sfGFP, PETase, and Reteplase), which span a range of expression and folding difficulty. **Why these three proteins?**

sfGFP is a widely used fluorescent reporter protein that folds robustly and contains no disulfide bonds; its two cysteines remain as free thiols. It therefore provides a relatively simple baseline for evaluating whether a CFPS reaction is functioning.

PETase is a cutinase from *Thermobifida fusca* (TfCut2) that hydrolyzes polyethylene terephthalate (PET). TfCut2 contains one disulfide bond, making it somewhat more challenging to fold than sfGFP.

Reteplase is a clot-dissolving therapeutic derived from human tissue plasminogen activator (tPA). It contains nine disulfide bonds that must form in the correct pairings for the protein to function, making it the most challenging of the three proteins to produce in an active form.

A reaction composition that performs well for one protein may therefore perform poorly for another. Testing the same composition across proteins with increasing folding difficulty provides a measure of how broadly the

formulation performs. Of note, in one study, a cost-optimized CFPS formulation was tested across twelve additional proteins; six were ultimately detectable by gel electrophoresis and nine were detectable using a more sensitive assay, illustrating how CFPS performance remains partly protein-dependent (Smith et al. 2026).

B. EXPRESSION CHALLENGES OF EACH PROTEIN TARGET

sfGFP: Fluorescent Reporter



Fig 3. Superfolder GFP (PDB 2B3P; UniProt P42212).
The β -barrel structure surrounds the fluorescent chromophore.

Superfolder green fluorescent protein (sfGFP) is ~240 amino acids long and 27 kDa. Its green fluorescence is produced by a chromophore formed autocatalytically from Thr65, Tyr66, and Gly67 through cyclization and dehydration, followed by an oxygen-dependent oxidation step. sfGFP contains two cysteines, but they do not form a disulfide bond. Because sfGFP folds reliably, it is commonly used as a CFPS reference protein; failure to produce fluorescent sfGFP generally indicates that the reaction itself is not functioning properly.

PETase: PET-degrading enzyme

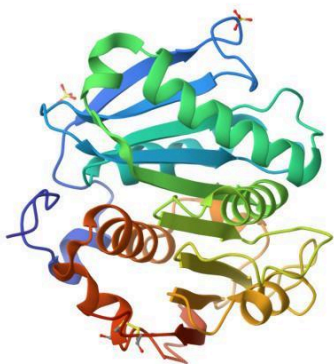


Fig 4. TfCut2, a thermostable PET-degrading cutinase (PDB 4CG1; UniProt Q6A0I4).

The PETase used in this project is TfCut2, a cutinase from *Thermobifida fusca* that is ~260 amino acids and 28 kDa in its active form. TfCut2 hydrolyzes ester bonds in PET, the plastic used in beverage bottles and polyester. It is a serine hydrolase with a catalytic triad composed of Ser170, Asp216, and His248. TfCut2 contains one disulfide bond, making its folding requirements more demanding than those of sfGFP, providing a test of whether a reaction composition can produce an active enzyme rather than only a fluorescent reporter.

Reteplase: therapeutic protease



Fig 5. The catalytic domain of tissue plasminogen activator (PDB 1RTF; tPA UniProt P00750). Reteplase is a truncated form of tPA; yellow sticks indicate disulfide bonds.

Reteplase is a 355-amino-acid, ~39 kDa therapeutic protein derived from human tPA and used to dissolve blood clots. Full-length tPA contains five domains and 17 disulfide bonds, while Reteplase retains the kringle-2 and serine protease domains and contains nine disulfide bonds. These bonds must form in the correct pairings for the protein to adopt its active structure.

It is therefore substantially more difficult to fold than sfGFP or PETase. Industrially, it is produced in *E. coli* as inactive inclusion bodies and refolded in a separate chemical step. A standard reducing CFPS formulation does not support formation of its required disulfide bonds, so active Reteplase requires an oxidizing CFPS formulation.

C. DISULFIDE BONDS: REDUCING AND OXIDIZING CONDITIONS

Disulfide bonds are covalent bonds formed between the sulfur atoms of two cysteine residues. Formation of an S–S bond requires oxidation of two cysteine thiol groups and helps stabilize a protein's three-dimensional structure. Differences in disulfide-bond requirements are therefore a major reason the three target proteins require different CFPS conditions.

Standard CFPS reactions are generally maintained under reducing conditions using DTT (dithiothreitol), which keeps cysteines in their free-thiol state and limits unwanted disulfide-bond formation. This environment is appropriate for sfGFP and remains compatible with TfCut2.

For Reteplase, however, DTT would prevent formation of its nine required disulfide bonds and therefore prevent correct folding. Reteplase instead requires an oxidizing, or disulfide-bond (DSB), formulation with three major modifications:

- 1. Modified lysate:** Lysate is prepared without DTT and treated with iodoacetamide (IAM). IAM irreversibly caps free thiols and inhibits enzymes that would otherwise reduce newly formed disulfide bonds. Excess IAM is quenched with an equal amount of cysteine.
- 2. Redox buffer:** Oxidized glutathione (GSSG) and reduced glutathione (GSH) are added in a controlled ratio to establish an oxidizing environment that favors disulfide-bond formation.
- 3. Folding helper:** DsbC, a disulfide-bond isomerase, is added to rearrange incorrectly paired disulfide bonds. The lysate is also prepared from an *E. coli* strain lacking glutathione reductase (Δ gor), helping preserve the oxidizing environment.

The two reaction regimes can therefore be summarized as follows:

Feature	Reducing: sfGFP, PETase	Oxidizing/DSB: Reteplase
DTT in lysate	Yes	No
Additional lysate treatment	None	IAM alkylation, then cysteine quench
Redox buffer	None	GSSG + GSH
Folding helper	None	DsbC
<i>E. coli</i> strain	BL21 Star (DE3)	BL21 Star (DE3) Δ gor
Best suited for	Proteins with few or no disulfide bonds	Proteins requiring multiple disulfide bonds

C. MEASURING PROTEIN PRODUCTION AND ACTIVITY

Each protein is evaluated using an assay matched to its function. Across all assays, a standard curve converts an experimental signal into an estimated concentration (a proxy for activity), while positive and negative controls determine whether the measurement is reliable. A positive control should produce a strong signal, whereas a negative control (e.g., a reaction lacking DNA), should produce little or no signal.

(a) sfGFP: direct fluorescence

Because sfGFP is intrinsically fluorescent, protein production can be measured directly. The completed reaction is excited near 485 nm, and green fluorescence is measured near 510–520 nm. This is an endpoint assay performed after the reaction is complete.

Greater fluorescence indicates a greater amount of correctly folded, active sfGFP. A standard curve generated using purified sfGFP converts fluorescence into protein concentration. Because only properly folded and active sfGFP fluoresces, the assay measures functional folded protein rather than total protein production.

(b) PETase: substrate hydrolysis at 405 nm

PETase activity is measured using a soluble p-nitrophenyl ester substrate. Hydrolysis releases p-nitrophenol, which is present primarily as the yellow p-nitrophenolate ion at the assay pH and absorbs strongly near 405 nm.

Absorbance at 405 nm is measured repeatedly over time. The initial rate of increase in absorbance reflects the amount of active PETase present: a steeper slope indicates greater enzymatic activity. A standard curve using the colored product converts absorbance into product concentration. Samples are diluted as needed so that measurements remain within the linear range of the assay.

(c) Reteplase: fluorogenic substrate cleavage

Reteplase activity is measured using the fluorogenic peptide substrate IPR-AMC. IPR refers to isoleucine–proline–arginine, while AMC is the fluorophore 7-amino-4-methylcoumarin. While AMC remains attached to the peptide, its fluorescence is quenched. Reteplase cleaves the substrate after arginine, releasing fluorescent AMC. The substrate is added directly to the reaction plate, and fluorescence is measured over time using excitation ~345 nm and emission ~445 nm.

The initial fluorescence rate is proportional to the amount of active Reteplase. Because Reteplase is active only when its nine disulfide bonds are correctly formed, enzymatic activity also provides a functional measure of correct protein folding.

D. Staged screening of reaction compositions

Candidate reaction compositions are evaluated through a staged 384-well → 96-well → 2 mL screening workflow. Each stage reduces the number of candidates while increasing reaction volume.

In the first stage, reaction compositions are screened in 384-well plates using microliter-scale volumes, allowing many conditions to be tested while minimizing reagent use. Each 384-well plate contains 80 reaction compositions, with each composition tested in four replicate wells; the first four columns are reserved for standard curves and controls, while columns 5–24 contain the experimental conditions. Promising compositions are then re-tested in 96-well plates to determine whether their performance is reproducible, and final candidates are evaluated in 2 mL tubes. The study proceeds through three rounds, each using this 384-well → 96-well → 2 mL screening workflow.

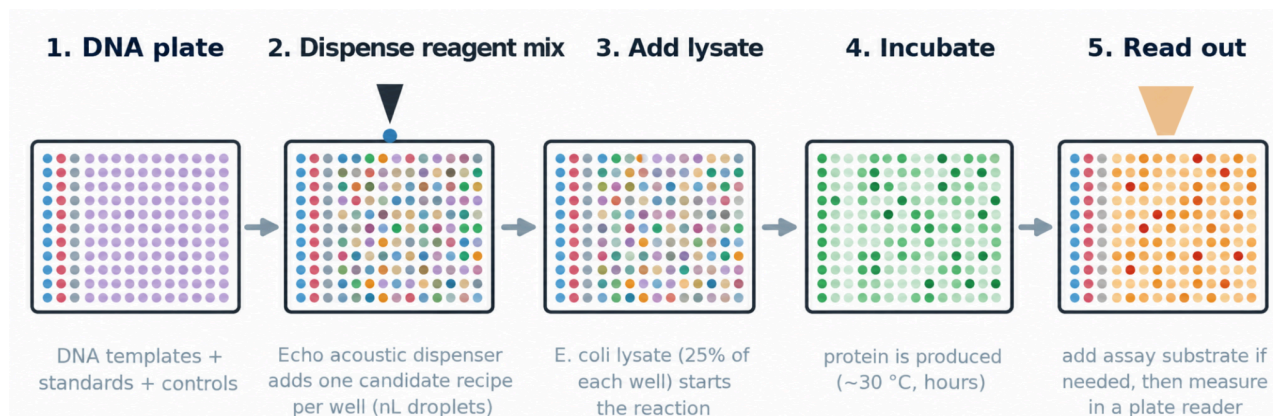


Fig 6. High-throughput CFPS screening workflow.

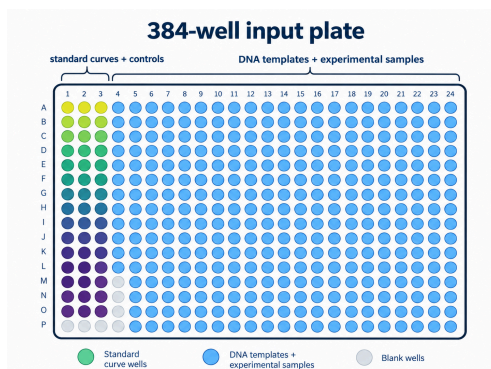


Fig 7. 384-well input plate. The initial 3 columns contain standard curves; column 4 contains positive controls and negative controls, the remaining 320 wells contain DNA templates and experimental samples.

E. Automated experimental platform

High-throughput CFPS screening is performed using Ginkgo's automated Nebula platform. The microplates are transferred between instruments using robotic arms and a MagneMotion magnetic track.

An Echo 525 acoustic dispenser uses sound waves to transfer reagents in droplets as small as 25 nL without pipette tips. A Bravo liquid handler then adds DNA and lysate to initiate the reactions. Plate readers collect

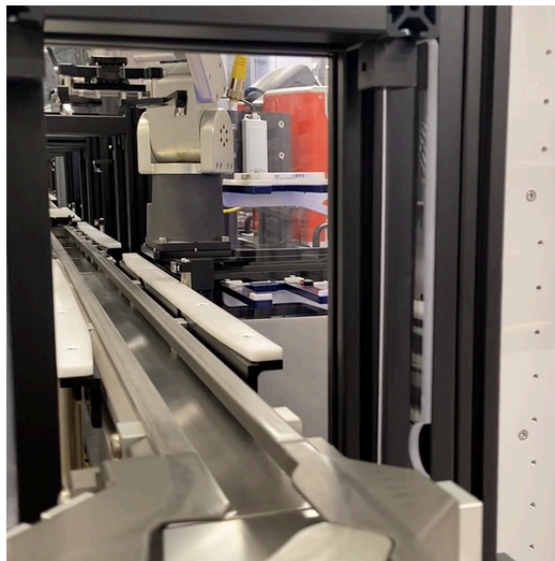
measurements at defined time points, and an Inheco heat block performs the 60°C inactivation step used for PETase assays.

This automation allows thousands of precisely assembled reactions to be prepared and measured on a fixed schedule, significantly for efficiency and reproducibly than what can be done manually.

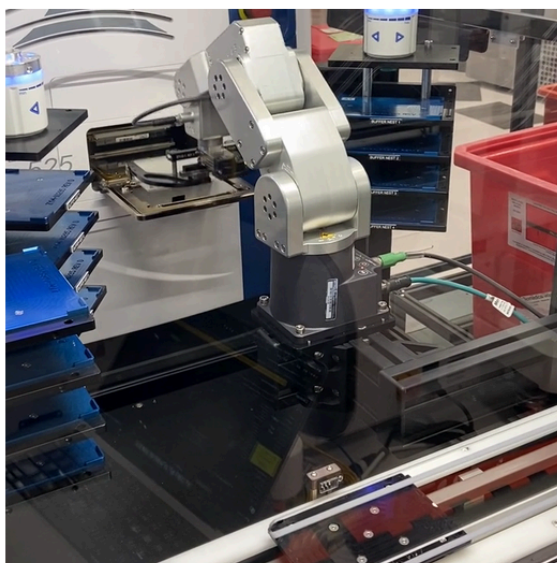
Because each step is programmable, AI systems can also propose new reaction compositions and evaluate them iteratively.



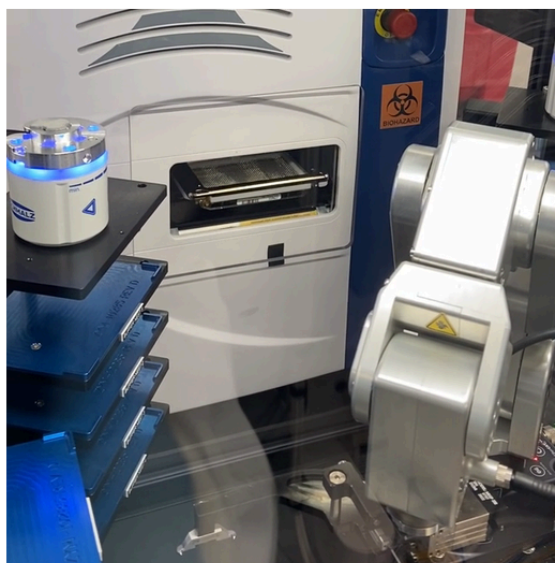
1 · Retrieval from cold storage



2 · Transport on MagneMotion



3 · Plate handling by 6-axis arm



4 · Loading the Echo 525 dispenser

Fig 8. Nebula in Action. (1) A plate is retrieved from cold storage. (2) Plates travel between stations on the MagneMotion magnetic track. (3) A 6-axis robotic arm moves plates on and off instruments. (4) A plate is loaded at the Echo 525 acoustic dispenser, which fires nanoliter reagent droplets with sound instead of pipette tips.

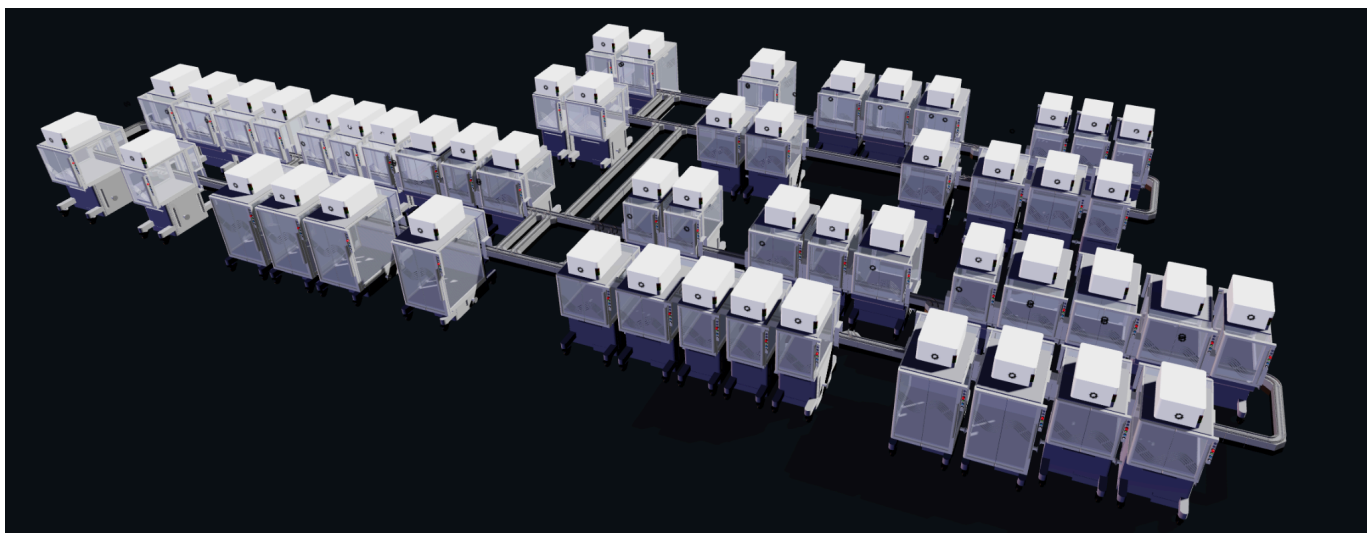


Fig 9. Ginkgo's Nebula System, an automated experiment platform

F. Study design and reaction constraints: the coopetition

The project on which this assignment is based compares human-led and AI-led approaches to CFPS reaction design. Human research teams and OpenAI AI agents independently design reaction compositions for the three target proteins.

Reaction compositions are submitted as JSON files and automatically validated before experiments are performed. In a related Ginkgo study, fewer than 1% of AI-designed plates contained an error that reached the laboratory (Smith et al. 2026). Each reaction has a total volume of 20 μ L and contains both fixed and variable components.

Fixed components:

- 5,000 nL lysate (25% of reaction)
- 2,000 nL DNA template
- 2,000 nL base buffer
- (Total 9,000 nL and are not modified.)

Variable components:

11,000 nL must be selected from the available reagent menu. For any unused volume, nuclease-free water is added to bring the total reaction volume to exactly 20,000 nL. All reagent volumes must be specified in multiples of 25 nL.

The available reagents fall into several functional categories:

Category	Examples	Function
Salts and buffers	potassium glutamate, magnesium glutamate, HEPES-KOH, potassium formate	Control ionic strength and pH
Building blocks	ATP, CTP, GTP, UTP, nucleotide alternatives, amino acids	Supply material for RNA and protein synthesis
Energy regeneration	phosphoenolpyruvate, pyruvate, maltose, maltodextrin, glucose	Regenerate ATP
Cofactors	folinic acid, NAD, coenzyme A, spermidine, cAMP	Support enzymatic reactions
Detergents and solvents	Brij-35, DMSO	Support solubility and stability

Oxygen-related components	catalase, pyruvate oxidase	Influence oxygen availability and metabolism
Disulfide components	DsbC, GSSG, GSH	Promote and correct disulfide-bond formation

Because reaction compositions are evaluated partly by cost, both reagent price and protein yield matter. Lysate costs approximately \$4.00/mL and constitutes 25% of each reaction, making it typically the largest single cost. Several other reagents are also relatively expensive, including CTP, GTP, and UTP at approximately \$28–39/mL, NAD at approximately \$33/mL, and coenzyme A at approximately \$104/mL. Many salts and sugars cost only fractions of a cent, and water is treated as free.

Cost optimization therefore requires reducing expensive components without sacrificing protein yield. In prior work, improved pH buffering, partial replacement of nucleotide triphosphates with less expensive nucleotide monophosphates, and addition of small amounts of spermidine reduced cost per gram of protein produced (Smith et al. 2026). Reagent costs in this assignment follow Olsen et al. (2026).

3. ASSIGNMENT: CFPS REACTION DESIGN FOR SFGFP, PETASE, RETEPLASE

Copy and paste this section (section 3, A-F) into your website! Note, coopetition.htgaa.org is intended to support you as you answer the following questions.

Goal: using the collective intelligence of the HTBAA community and accessible AI systems, design novel CFPS reaction compositions for each of the three target proteins: sfGFP, PETase, and Reteplase to reduce the cost per gram of functional protein. Each design should be compatible with the experimental framework described above and should reflect the distinct expression and folding requirements of that protein. You can learn more about the details of the coopetition [here](#).

A. COMPARE CELL-FREE REACTION COMPOSITIONS (20 minutes)

1. Why does the concentration of potassium glutamate and magnesium glutamate differ between compositions? Which is more sensitive?
2. Explain how a PEP+NTP reaction formulation is different from a Ribose+NMP reaction formulation.
3. Choose two cell-free reaction compositions from the literature, compare and contrast the choice of reagents and the concentration of reagents.

Note: A recommended starting point would be to review a least one published formulation, such as PANOx-SP (Jewett & Swartz 2004), a cost-optimized formulation from Olsen et al. (2026), and/or an AI-designed formulation from Smith et al. (2026).

B. DESIGN ONE REACTION COMPOSITION FOR EACH OF THE THREE PROTEINS (40 minutes)

1. Sign in to coopetition.htgaa.org and publish your first unique 20 μ L CFPS reaction composition for each of the following (don't worry, you can resubmit later):
 - sfGFP
 - PETase

- Reteplase

Note: if you're unable to login with your HTGAA keycloak account, please post in Zulip so that you can participate in the coopetition!

For each composition, specify the reagents you would adjust from the relevant categories, including salts, energy sources, building blocks, cofactors, and redox components.

2. As a HTBAA research team, we will have the opportunity to suggest new reagents to add to the cell-free reagent inventory. Contribute (or upvote) at least one reagent to the "Community Reagents" section of the website and provide a short rationale. Include in your homework documentation how you decided on that specific reagent and provide links to any relevant research.
3. Visit the [cfps-coopetition](#) forum page to ensure you will be able to take part in the discussion. *Bonus: send a note in the channel about a CFPS optimization idea that you have! If you would like clarification on aspects of the project, please message in this dedicated channel.*

C. ASIMOV KERNAL (OPTIONAL)

1. Build T7 compatible plasmids for cell-free expression using the amino acid coding sequences below:

sfGFP

None

```
MSKGEELFTGVVPIILVELDGDVNGHKFSVRGEGEGDATIGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKRHDFFKSAMPEGYVQER
TISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNHNHYITADKQKNGIKANFTVRHNVEDGSVQLADHYQNTPIGDGP
VLLPDNHYLSTQTVLSKDPNEKGTDRDHMLHEYVNAAGITWSHPQFEK
```

PETase

None

```
MEKKIWSHPQFEKGGSGENLYFQSGGASPAGGGANPYERGNPTDALLEASSGPFVSSEENVSRLSASGFGGGTIYYPRENNYGAVAISPQYGTGT
EASIAWLGERIASHGFVVITIDTITTTLDQPDRAEQNLNHNMINRASSTVRSRIDSSRLAVMGHSMGGGGTLRLASQRPDLKAAIPLTPWHLNK
NWSSVTVPPTLIIGADLDTIAPVATHAKPFYNSLPSSISKAYLELDGATHFAPNIPNKIIGKYSVAWLKRFVDNDTRYTQFLCPGPRDGLFGEVEEY
RSTCPFGGSGVSGWRLFKKIS
```

Reteplase

None

```
MEKKIWSHPQFEKGGSGENLYFQSGGASPAGGGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKVYTAQNPSAQUALGLGKHNYCRNPDGDA
KPWCHVLKNRRLTWEYCDVPSCSTCGLRQYSQPQFRIKGLFADIASHPWQAAIFAKHRRSPGERFLCGGILISSCWILSAACHCFQERFPPHHLTV
ILGRTYRVVPGEQEEQKFEVEKYIVHKEFDDDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSERLKEA
HVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDSSGGLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRDNMRP
GGSGVSGWRLFKKIS
```

2. Bonus question: design a T7 helper plasmid that you could add to the cell-free reaction to improve the yield of functional reteplase (general chaperones, disulfide-bond machinery, etc).

4. Use of AI

You are encouraged to use AI and online tools for this assignment to assist your reagent composition development process. As you will present your work during recitation, you must be able to explain and justify the reasoning behind your design.

If you use AI, ensure that you cite the AI client and your prompts used.

5. REFERENCES AND RESOURCES

- Smith, A. A., et al. (2026). Using a GPT-5-driven autonomous lab to optimize the cost and titer of cell-free protein synthesis (Ginkgo Bioworks × OpenAI). *bioRxiv* preprint. DOI: 10.64898/2026.02.05.703998. <https://www.biorxiv.org/content/10.64898/2026.02.05.703998v1>
- Olsen, M. L., Copeland, C. E., Sundberg, C. A., Aw, R., Shaver, Z. M., Rao, G., Swartz, J. R., Karim, A. S., &
- Jewett, M. C. (2026). Design-driven optimization of low-cost reagent formulations for reproducible and high-yielding cell-free gene expression. *Nature Communications*. DOI: 10.1038/s41467-026-69605-8.
- Jewett, M. C., & Swartz, J. R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1), 19–26. DOI: 10.1002/bit.20026.
- Silverman, A. D., Karim, A. S., & Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics*, 21, 151–170. DOI: 10.1038/s41576-019-0186-3.
- Rasor, B. J., Vögeli, B., Landwehr, G. M., Bogart, J. W., Karim, A. S., & Jewett, M. C. (2021). Toward sustainable, cell-free biomanufacturing. *Current Opinion in Biotechnology*, 69, 136–144. DOI: 10.1016/j.copbio.2020.12.012.