

Title Page

Title: Scenario-weighted decision framework for selecting cell-free protein synthesis reactor formats: a PRISMA-informed literature dataset and multi-criteria analysis

Running title: Decision framework for cell-free protein synthesis reactor formats

Authors: Taha Bayar^{1*}

Affiliation: 1. Ankara University, Department of Biology, Ankara, Turkiye

Corresponding author: * Taha Bayar, Email: luminadigitale@gmail.com

Abstract and Keywords

Abstract Cell-free protein synthesis supports rapid prototyping, biomanufacturing, and portable diagnostics, but reactor-format selection is often based on local practice rather than explicit decision criteria. We developed a transparent workflow that combines PRISMA-informed literature screening, unit-standardized quantitative extraction, and scenario-weighted multi-criteria scoring to compare reactor formats across performance and deployment constraints. Fifteen studies met inclusion criteria and were grouped into batch ($n=7$), dialysis or continuous-exchange systems ($n=4$), microfluidic platforms ($n=3$), and hydrogel or compartmentalized systems ($n=1$). Yield values were normalized to ug/mL where possible and analyzed together with reaction time, complexity, scalability, and automation proxies. Batch and dialysis or continuous-exchange formats showed comparable median yields (600.0 and 608.5 ug/mL), while microfluidic systems were faster (median 5.83 h) but lower yielding (median 63.0 ug/mL). Under three locked scenarios (speed-first, yield-first, low-infrastructure-first), batch ranked first across scenarios, whereas microfluidic formats ranked higher only when time was strongly prioritized and lower when infrastructure constraints were emphasized. These results show that no single reactor format is universally optimal; rankings depend on scenario priorities. The framework provides a reproducible decision layer that can be updated as new studies are added.

Keywords: cell-free protein synthesis; reactor format selection; multi-criteria decision analysis; PRISMA; microfluidics

Main Text

1. Introduction

Cell-free protein synthesis has become a practical platform for rapid protein prototyping, membrane-protein production, and deployable diagnostic workflows (1-4). Reactor-format options now include simple batch reactions, dialysis or continuous-exchange setups, and microfluidic or compartmentalized systems (2,5-7). Each format offers a different balance between yield, turnaround time, operational complexity, and infrastructure burden.

Despite this expansion, most cross-format comparisons remain narrative and difficult to operationalize for study planning. Technical reviews and practical guides summarize methods and use-cases, but they do not generally provide a scenario-locked, quantitative framework that translates heterogeneous reports into explicit format rankings (4,8,9).

This study introduces a reproducible decision framework that combines structured evidence extraction and scenario-weighted scoring. The goal is not to replace experimental judgment but to make trade-offs explicit, auditable, and updateable as the literature grows.

2. Materials and Methods

2.1 Study design and reporting approach

This is a literature-only analysis with no new wet-laboratory experiments. Screening logic, extraction rules, and analysis scripts are versioned in the project repository (`protocol/`, `scripts/`, `data_clean/`, `output/`, `figures/`). Reporting flow was organized in a PRISMA-consistent structure (10).

2.2 Eligibility criteria

Inclusion required all of the following: 1. a classifiable cell-free reactor format; 2. at least one absolute quantitative yield or expression value with unit; 3. reaction duration in hours (reported or reliably derivable); 4. sufficient system context to classify format and lysate.

Studies with only relative fold-change outputs and no absolute yield or time were excluded from quantitative synthesis.

2.3 Search and screening

Search-string templates were defined in `protocol/search_strings.md` and screening fields in `protocol/prisma_notes.md`. Final PRISMA counts were: 1. records identified: 22 2. duplicates removed: 2 3. records screened: 20 4. records excluded: 4 5. full-text assessed: 16 6. full-text excluded: 1 7. studies included: 15

The screening flow diagram is provided in `figures/prisma_flow.png`.

2.4 Data extraction and normalization

Data were collected in `data_raw/extraction_template.xlsx`. Required fields were paper identifier, format class, yield value and unit, reaction time, and lysate type. Yield values were normalized to ug/mL where conversions were direct (for example, mg/mL to ug/mL). Non-convertible units were retained in the cleaned table for traceability but excluded from numeric aggregation.

2.5 Criteria and weighting framework

Five criteria were used: 1. yield (higher is better), 2. time-to-result (shorter is better), 3. technical complexity proxy, 4. scalability proxy, 5. automation proxy.

Scenario weights (from `output/table2_mcda_weights.csv`) were pre-locked: 1. speed-first: yield 0.20, time 0.40, complexity 0.15, scalability 0.10, automation 0.15 2. yield-first: yield 0.45, time 0.15, complexity 0.10, scalability 0.15, automation 0.15 3. low-infrastructure-first: yield 0.15, time 0.20, complexity 0.40, scalability 0.10, automation 0.15

2.6 Scoring procedure

Format-level median yield and time were min-max normalized. Time was inverse-scored so lower duration produced higher criterion scores. Scenario totals were calculated as weighted sums across the five criteria.

2.7 Reproducibility

Pipeline order: `scripts/01_cleaning.py` -> `scripts/02_analysis.py` -> `scripts/03_mcda.py` -> `scripts/04_make_figures.py` -> `scripts/05_make_tables.py` -> `scripts/06_prisma_flow.py`.

3. Results

3.1 Dataset overview

Included records spanned four observed format classes: batch (n=7), dialysis or continuous-exchange (n=4), microfluidic (n=3), and hydrogel or compartmentalized (n=1). Publication years in the cleaned dataset ranged from 2014 to 2025.

3.2 Yield-time trade-offs by format

Batch and dialysis or continuous-exchange formats showed similar median yields, but continuous-exchange setups required substantially longer runtimes. Microfluidic systems were faster but had lower median yields in the current evidence window.

Table 1. Format-level summary

format_class	n	yield_median (ug/mL)	yield_IQR (ug/mL)	time_median (h)	time_IQR (h)
batch	7	600.0	320.0-1490.0	10.0	3.0-18.0
dialysis_ecf	4	608.5	192.0-1025.0	36.0	24.0-111.0
hydrogel_compartimentalized	1	197.0	197.0-197.0	4.0	4.0-4.0
microfluidic	3	63.0	50.65-103.65	5.83	3.92-5.92

3.3 Scenario-weighted rankings

Table 2. Scenario scores by format class

format_class	Speed-first	Yield-first	Low-infra-first
batch	0.671884	0.664863	0.710163
dialysis_cecf	0.375000	0.650000	0.450000
hydrogel_compartmentalized	0.574129	0.385541	0.486847
microfluidic	0.527125	0.291422	0.338562

Batch ranked first in all three scenarios in the current dataset. Microfluidic formats ranked second under speed-first priorities but dropped when yield or low-infrastructure constraints were weighted more strongly. Dialysis or continuous-exchange remained competitive in yield-first because of high productivity despite longer time-to-result.

4. Discussion

This analysis shows that reactor-format choice in cell-free protein synthesis is scenario-dependent. A format that performs well in one decision context may not remain optimal when logistical constraints are explicitly included.

The main contribution is methodological integration: PRISMA-informed extraction plus cross-format normalization and scenario-weighted multi-criteria ranking in one reproducible workflow. During targeted checks against available indexed literature, we did not identify a prior study that combines these three components for cell-free reactor-format selection in this exact structure. This should be interpreted as a scoped evidence finding rather than proof of universal absence.

For practical planning, results suggest three use patterns in the current dataset: 1. batch for robust all-around performance, 2. dialysis or continuous-exchange for yield-prioritized objectives, 3. microfluidic for time-sensitive objectives when infrastructure is available.

Limitations include unequal class sample sizes, heterogeneous protein targets and lysates, and proxy-based scoring for complexity, scalability, and automation. One microfluidic record also includes a runtime estimate derived from flow and collection values, which introduces additional uncertainty.

Overall, the framework is designed as a living decision layer. As additional records are added, rankings can be recomputed without changing the pipeline logic.

Material Availability Statement

No new biological materials were generated in this literature-based study. The format taxonomy, scoring definitions, and analysis scripts are available in the project repository. If future versions include newly generated constructs or strains, deposition in an appropriate public repository will be reported in this section.

Data Availability Statement

All data and code underlying this study are available in the project repository and supplementary files. Core files include `data_clean/cfps_dataset_v1.csv`, `data_clean/mcda_scores_v1.csv`, `output/table1_cfps_results.csv`, `output/table2_mcda_weights.csv`, `output/table3_mcda_scores.csv`, and `output/table4_taxonomy_proxies.csv`. Supplementary figures are provided in `figures/`. Public repository URL: <https://github.com/Apeiron01/cfps-decision-matrix>.

Acknowledgements

Generative AI was used for language editing support under direct author supervision. The author reviewed and validated all scientific content, interpretation, and conclusions.

Funding

This research received no external funding.

Conflict of Interest Disclosure

No potential conflict of interest was reported by the authors.

Author Contributions

Taha Bayar: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review and editing.

References

1. Caschera,F. and 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.
2. Stech,M., Brodel,A.K., Quast,R.B., et al. (2014) A continuous-exchange cell-free protein synthesis system based on extracts from cultured insect cells. *PLoS ONE*, 9, e96635.
3. Quast,R.B., Mrusek,D., Hoffmeister,C., et al. (2016) High-yield cell-free synthesis of human EGFR by IRES-mediated protein translation in a continuous exchange cell-free reaction format. *Scientific Reports*, 6, 30399.
4. Silverman,A.D., Karim,A.S. and Jewett,M.C. (2020) Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics*, 21, 151-170.
5. Xiao,X., Li,Y., Wang,M., et al. (2018) Integration of cell-free protein synthesis and purification in one microfluidic chip for on-demand production of recombinant protein. *Biomicrofluidics*, 12, 044106.
6. Murphy,T.W., McCarty,N.S. and Clark,D.S. (2019) On-chip manufacturing of synthetic proteins for point-of-care therapeutics. *Microsystems and Nanoengineering*, 5, 8.

7. Aquino,A.K., Pardo Avila,F., Stark,J.C., et al. (2021) Glycosylation-on-a-Chip: A flow-based microfluidic system for cell-free glycoprotein biosynthesis. *Frontiers in Molecular Biosciences*, 8, 782905.
8. Gregorio,N.E., Levine,M.Z. and Oza,J.P. (2019) A user's guide to cell-free protein synthesis. *Methods and Protocols*, 2, 24.
9. Damiati,S., Mhanna,R., Kodzius,R. and Ehmoser,E.K. (2018) Cell-free approaches in synthetic biology utilizing microfluidics. *Genes*, 9, 144.
10. Page,M.J., McKenzie,J.E., Bossuyt,P.M., et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71.

Figure Legends and Alt Text

Figure 1. Yield distribution by format. Boxplots comparing normalized protein yield (ug/mL) across observed reactor format classes. Alt text: Boxplots show higher median yields for batch and dialysis or continuous-exchange formats, lower yields for microfluidic, and single-point hydrogel data.

Figure 2. Time-to-result distribution by format. Boxplots comparing reaction duration (hours) across observed reactor format classes. Alt text: Boxplots show microfluidic and hydrogel classes with shorter times, and dialysis or continuous-exchange with the longest median runtime.

Figure 3. Scenario-weighted decision matrix heatmap. Heatmap of weighted total scores by format class for speed-first, yield-first, and low-infrastructure-first scenarios. Alt text: Heatmap indicates batch as highest across all scenarios, microfluidic strongest in speed-first relative to its other scenarios, and dialysis or continuous-exchange strongest in yield-first.

Figure 4. Scenario radar comparison of component profiles. Radar chart of average criterion contributions by scenario. Alt text: Radar plot contrasts criterion emphasis across scenarios, with time emphasized in speed-first, yield in yield-first, and complexity in low-infrastructure-first.

Figure 5. PRISMA flow diagram. Flow diagram summarizing study identification, screening, eligibility, and inclusion counts. Alt text: PRISMA diagram shows 22 records identified, 2 duplicates removed, 20 screened, 4 excluded, 16 full texts assessed, 1 full text excluded, and 15 studies included.