

Nucleus Developer Notes

Published Mar 23, 2026

Correspondence to
Scott Riggs
scr@findwhatmatters.ai

Open Access 

Copyright © 2026 Riggs & IGOR. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#), which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Open Source

The materials and processes described in this article are distributed under the terms of the [CERN Open Hardware License Permissive 2.0](#), which enables reusers to make and convey any component or device that apply the information made explicitly available under this license, so long as all notices are retained.

IGOR PPK Optimization: Round 1

Scott Riggs¹ , and IGOR¹

¹Find What Matters

Abstract

To maximize protein yield and synthesis rate in the PURE system, this work employed a multivariate optimization strategy for key components including ATP, GTP, and tRNA. Analysis concluded that optimal performance is not governed by a single factor, but rather a co-dependent relationship between physicochemical viability and kinetic efficiency. The top-ranked hypothesis posits that performance is dictated by the interplay between 'free Mg²⁺' concentration, a function of nucleotide levels, and the stoichiometric ratio of ATP to GTP. This unified model provides a framework for navigating the complex parameter space to identify novel, high-performance operational regimes.

Keywords AI Scientist, PURE, PPK

1. OVERVIEW

This work details the systematic investigation into optimizing the PURE (Protein synthesis Using Recombinant Elements) cell-free system. By integrating new experimental parameters with existing data, a multivariate analysis was conducted to move beyond simple one-factor-at-a-time optimizations. The process culminated in a unified hypothesis that reconciles the system's physicochemical constraints with its kinetic bottlenecks, proposing a co-dependent optimization strategy for achieving maximal protein synthesis.

1.1. Workspace Data

See [workspace/workspace_data.csv](#) for the full exported dataset.

2. RESULTS

The investigation converged on a primary hypothesis that integrates two critical aspects of PURE system function: physicochemical stability and kinetic throughput. This model, termed 'Unified Physicochemical and Kinetic Control,' proposes that peak performance is achieved by co-optimizing the availability of free magnesium ions with the stoichiometric ratio of the primary energy sources, ATP and GTP. The following sections detail the leading hypotheses and the experimental plan designed to validate this unified model.

Discourse Graph Representation

2.1. Detailed Findings

2.2. Executive Summary

The research aimed to maximize the protein expression yield and synthesis rate of the PURE (Protein synthesis Using Recombinant Elements) system. The IGOR process employed a multivariate combinatorial optimization strategy, integrating new features ([ATP], [GTP], [Amino

acid mix], [tRNA]) with existing experimental data. The investigation concluded that optimal PURE system performance is not governed by a single factor but by a complex, co-dependent relationship between physicochemical viability and kinetic efficiency. The top-ranked hypothesis, 'Unified Physicochemical and Kinetic Control,' posits that system performance is dictated by the interplay between the concentration of 'free Mg^{2+} '—a function of total nucleotide levels—and the stoichiometric ratio of ATP to GTP. This unified model provides a robust framework for navigating the high-dimensional parameter space to discover a new, high-performance operational regime, moving beyond simple one-factor-at-a-time optimizations.

2.3. Top-Ranked Hypotheses

2.3.1. 1. Unified Physicochemical and Kinetic Control: Co-optimization of Free Mg^{2+} Availability and Energy Stoichiometry

This hypothesis integrates the foundational physicochemical constraints of the PURE system with its kinetic bottlenecks. It posits that maximal system performance (rate and yield) is not found by optimizing single factors, but by co-optimizing two interdependent layers. The first layer is **physicochemical viability**, governed by the concentration of 'free Mg^{2+} ', which is non-linearly dependent on the total concentration of chelating nucleotides (ATP and GTP). This defines the boundaries of a viable operational space, addressing the failures observed at high total $[Mg^{2+}]$ and $[NTP]$. The second layer is **kinetic efficiency**, which is optimized *within* that viable space by tuning the stoichiometric ratio of ATP to GTP. This 'energetic resonance' balances the ATP-dependent tRNA charging machinery with the GTP-dependent ribosomal translocation machinery. Therefore, peak performance lies in a multi-dimensional sweet spot where sufficient free Mg^{2+} ensures fundamental component function, while the ATP:GTP ratio is optimally matched to prevent kinetic bottlenecks in the translation cycle. **Testable Question:** Can we achieve a >5x improvement in PURE performance by mapping the interaction surface between total nucleotide concentration (Total NTPs from 2-8 mM), the ATP:GTP ratio (from 1:2 to 2:1), and total $[Mg^{2+}]$ (from 6-18 mM)? This multivariate approach aims to identify a globally optimal regime that surpasses performance achievable by optimizing any single variable alone.

Final rationale:

This hypothesis evolves the top-ranked 'Architect 1' by directly incorporating the key strengths of its main competitor, 'Explorer 1'. The debate revealed that while 'Architect 1' correctly identified the fundamental importance of the 'free Mg^{2+} '-NTP relationship, 'Explorer 1' introduced a novel and valuable kinetic argument around the ATP:GTP ratio. This combined hypothesis is superior because it creates a unified model where physicochemical viability is a prerequisite for, and is modulated by, kinetic efficiency. It moves beyond simply preventing system failure (the core of 'Architect 1') to actively tuning for maximal throughput within the viable operational space, thus presenting a more comprehensive and powerful optimization strategy.

Key supporting evidence:

- Architect 1 Rationale: Won debates by addressing the foundational, non-negotiable prerequisite of 'free Mg^{2+} ' availability as a function of total NTP concentration.
- Explorer 1 Rationale: Praised for its novel, multivariate kinetic argument focusing on the stoichiometric ratio of ATP to GTP as a key optimization parameter.
- Li et al., 2014: Finding that Mg^{2+} sensitivity is strongly coupled to ATP/GTP concentrations.

2.3.2. 2. Substrate-Demand Driven Energetic Tuning

This hypothesis refines the concept of 'energetic resonance' by proposing that the optimal stoichiometric ratio of ATP to GTP is not a fixed constant, but is dynamically determined by

the demand from other key substrates. The PURE system's two main energy sinks—ATP for tRNA aminoacylation and GTP for ribosomal translocation—are coupled. The efficiency of the overall system depends on balancing the flux through these two pathways. This hypothesis posits that the 'demand' on the ATP-dependent charging pathway is a direct function of the total concentration of available tRNA and amino acids. Therefore, as tRNA and amino acid concentrations increase, the optimal ATP:GTP ratio required to maximize kinetic rate will shift towards being ATP-rich to prevent the charging machinery from becoming a bottleneck. This entire kinetic optimization is, however, constrained within a viable physicochemical window where total [NTP] and [Mg²⁺] levels do not lead to component precipitation or ribosomal stalling. **Testable Question:** While holding the total [NTP + Mg²⁺] concentrations within a known viable range, does systematically varying the concentration of the tRNA and amino acid pools (e.g., low vs. high levels) cause a predictable and significant shift in the optimal ATP:GTP ratio required to maximize the protein synthesis rate (sigmoid_rate)?

Final rationale:

This hypothesis evolves 'Explorer 1' by addressing the primary critique from its debate against 'Architect 1'—that its kinetic arguments must operate within a fundamentally viable system. It explicitly incorporates the physicochemical constraints (the 'viable window') as a boundary condition. Its primary evolution is to make the 'energetic resonance' concept more concrete and testable by directly linking the optimal ATP:GTP ratio to the 'demand' created by other substrates (tRNA and Amino Acids), a variable that was mentioned but not central to the original hypothesis. This provides a more nuanced, multi-layered interaction model that is mechanistically specific and highly ambitious.

Key supporting evidence:

- Explorer 1 Rationale: Praised for its novel kinetic argument about the ATP:GTP ratio and its coupling to substrate availability.
- Architect 1 Critique of Explorer 1: Highlighted that kinetic optimization can only occur within a system that is fundamentally viable from a physicochemical standpoint.
- Li et al., 2014: Showed that yield improvements are possible through adjustment of translation factors and tRNA, implying that the system is sensitive to the balance of all components, not just energy.

2.4. Proposed Experimental Plan

2.4.1. Objective

To experimentally validate the 'Unified Physicochemical and Kinetic Control' hypothesis by systematically mapping the three-dimensional interaction surface of total nucleotide concentration, ATP:GTP ratio, and Magnesium concentration to identify a globally optimal condition for both PURE system protein synthesis rate and final yield.

2.4.2. Variables & Controls

Variable Type	Parameter	Levels/Range	Notes
Independent	Total Nucleotide Conc. ([ATP] + [GTP])	2 mM, 5 mM, 8 mM	Spans a range from standard to high energy.
Independent	ATP:GTP Molar Ratio	1:2, 1:1, 2:1	Tests GTP-rich, equimolar, and ATP-rich conditions.
Independent	[Magnesium acetate] (mM)	6 mM, 12 mM, 18 mM	Explores the viable range identified in prior experiments.

Variable Type	Parameter	Levels/Range	Notes
Dependent	Final Protein Yield (ng/ μ L)	Continuous	Measured as sigmoid_steady_state.
Dependent	Synthesis Rate (1/h)	Continuous	Measured as sigmoid_rate.
Control	[DNA], [PMix], [Ribosome], [K-glutamate]	Constant	Held at values from a high-performing baseline experiment.
Control	[Amino acid mix], [tRNA]	Constant	Held at the mean of their specified bounds (0.35 mM and 3.125 ug/ μ L, respectively).

2.4.3. Protocol Steps

1. **Design Experiment:** Create a full factorial ($3 \times 3 \times 3 = 27$ conditions) experimental design based on the independent variables. Include triplicate reactions for each condition and baseline controls.
2. **Prepare Stocks:** Prepare concentrated stocks of ATP, GTP, and Magnesium Acetate. Prepare a master mix containing all other constant PURE system components (Protein Mix, Ribosomes, DNA, salts, buffers, Amino Acids, tRNA).
3. **Assemble Reactions:** In a 384-well plate, dispense the master mix into each well. Then, using a liquid handler, add varying volumes of the ATP, GTP, and Magnesium Acetate stocks according to the experimental design. Add nuclease-free water to bring all reactions to a constant final volume.
4. **Incubate & Measure:** Place the plate in a plate reader pre-heated to 37°C. Monitor GFP fluorescence (Excitation: 488 nm, Emission: 507 nm) every 5 minutes for at least 5 hours.
5. **Analyze Data:** Process the time-course fluorescence data for each well. Fit a sigmoid function to each curve to extract the key metrics.

2.4.4. Key Metrics

- **sigmoid_steady_state (ng/ μ L):** The upper plateau of the fitted sigmoid curve, converted to concentration using a standard curve. This quantifies the final protein yield.
- **sigmoid_rate (1/h):** The maximum slope of the fitted sigmoid curve. This quantifies the peak rate of protein synthesis.
- **Response Surface Plots:** Generate 3D surface plots visualizing yield and rate as a function of the three independent variables. These plots will be used to identify the coordinates of the optimal performance region and confirm the hypothesized co-dependence.

2.4.5. Generated Experimental Suggestions

Raw GES representation

Fixed Conditions

All runs shared the following constant parameters:

Parameter	Value	Units
-----------	-------	-------

PEG4K 40%	0	%
RNAse Inhibitor	2000	U/mL
Reaction Volume	10	μL
PPK	~0	μM
PolyP	~0	mM

Note

IsNEB, HasMgAR953, and isplamGFP were effectively zero or constant across all runs and are omitted.

2.4.6. Design Variables

GES Input Parameters

Table 1. *GES Input Parameters*

ID	[DNA] (nM)	[PMix] (mg/mL)	[Ribosome] (μM)	[Mg·OAc] (mM)	[CP] (mM)	[K·Glu] (mM)
GES-1	3.97	1.80	1.80	7.56	21.81	98.41
GES-2	17.14	1.86	4.45	1.00	100.00	40.00
GES-3	6.38	1.71	1.92	6.67	44.69	76.92
GES-4	19.63	1.58	1.74	7.69	98.28	41.53
GES-5	10.48	1.93	1.42	1.00	43.84	40.00
GES-6	3.46	1.81	1.79	8.41	13.38	102.71
GES-7	11.54	1.99	4.50	24.30	100.00	86.17
GES-8	12.93	1.97	1.47	6.51	100.00	56.00
GES-9	4.96	1.89	1.88	6.37	7.91	95.84
GES-10	4.93	1.82	1.63	14.78	52.85	99.84
GES-11	3.74	1.80	1.78	7.65	21.52	97.48
GES-12	20.00	1.98	4.50	25.00	100.00	97.73
GES-13	18.96	2.07	1.23	21.72	100.00	43.10
GES-14	4.68	1.68	1.00	25.00	91.74	178.92
GES-15	2.72	1.88	2.05	21.52	~0	179.52
GES-16	3.71	1.81	1.80	7.90	19.52	99.86
GES-17	4.59	1.86	1.88	7.68	32.87	113.13
GES-18	8.01	1.79	2.05	4.34	23.67	82.66
GES-19	7.12	2.05	4.44	22.25	100.00	67.87
GES-20	16.94	2.21	1.53	23.14	100.00	77.99
GES-21	7.93	1.75	1.19	24.77	99.84	40.00
GES-22	5.42	1.83	1.95	6.80	~0	124.83
GES-23	2.35	1.80	2.04	1.94	41.72	78.16
GES-24	5.24	1.74	1.85	2.36	3.24	75.48
GES-25	2.31	1.88	1.45	4.35	0.64	111.90
GES-26	3.71	1.81	1.80	7.90	19.52	99.86
GES-27	19.32	1.57	4.50	9.87	100.00	40.00
GES-28	1.12	1.77	1.02	6.26	30.12	65.88
GES-29	20.00	1.95	3.68	3.67	~0	56.26
GES-30	3.11	1.35	1.00	21.05	33.90	97.34
GES-31	8.18	1.79	2.00	4.65	30.91	40.00
GES-32	3.18	1.83	1.00	9.13	~0	159.24
GES-33	5.04	1.85	1.96	11.75	4.09	112.19
GES-34	4.55	1.76	1.85	7.52	~0	108.52
GES-35	1.00	2.12	4.50	11.89	100.00	40.00

GES Sigmoid Fit & Success

Table 2. *GES Sigmoid Fit & Success*

ID	Steady State (ng/ μ L)	Rate (1/h)	Time Offset (h)	Success
GES-1	73.30	4.13	0.86	1.000
GES-2	74.17	3.55	1.21	1.000
GES-3	61.68	3.46	1.12	0.975
GES-4	50.51	2.02	1.74	0.827
GES-5	60.45	3.08	1.43	0.994
GES-6	70.81	4.22	0.83	0.998
GES-7	60.91	4.69	0.56	1.000
GES-8	52.12	3.25	1.33	0.950
GES-9	61.18	4.11	0.95	0.999
GES-10	51.96	3.68	1.00	0.984
GES-11	73.40	4.14	0.85	0.997
GES-12	61.69	4.44	0.72	1.000
GES-13	48.70	3.12	1.31	0.998
GES-14	29.44	2.75	1.24	0.985
GES-15	38.61	3.37	1.04	0.999
GES-16	72.97	4.16	0.85	0.998
GES-17	64.82	3.97	0.93	0.988
GES-18	62.62	3.59	1.14	0.993
GES-19	60.11	4.64	0.59	1.000
GES-20	45.55	3.19	1.32	0.998
GES-21	41.89	3.22	1.05	0.999
GES-22	54.84	4.12	0.88	0.998
GES-23	64.90	3.41	1.13	0.993
GES-24	62.61	3.63	1.05	0.999
GES-25	55.25	3.98	0.88	1.000
GES-26	72.97	4.16	0.85	0.998
GES-27	66.69	3.58	1.07	1.000
GES-28	55.33	3.32	1.05	0.999
GES-29	38.97	3.66	1.26	1.000
GES-30	24.95	2.55	1.26	0.998
GES-31	61.90	3.43	1.20	0.997
GES-32	40.47	3.61	0.95	0.999
GES-33	57.83	4.09	0.94	0.998
GES-34	61.13	4.15	0.85	0.999
GES-35	62.67	4.25	0.78	1.000

2.5. References

- Y. Shimizu *et al.* [1], Shimizu Y, 2001, Nature Biotechnology

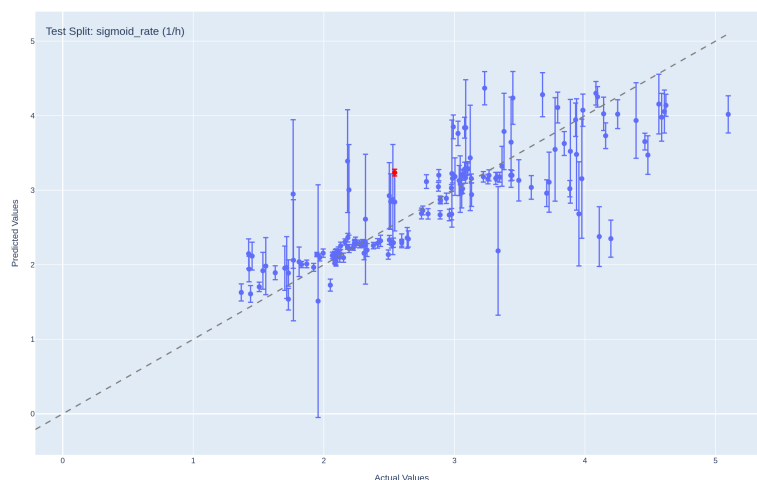


Figure 1. Parity plot showing the predictive performance of the model for protein synthesis rate (*sigmoid_rate*) on a held-out test dataset. The clustering of points along the diagonal indicates the model's ability to generalize to unseen experimental conditions.

- J. Li, L. Gu, J. Aach, and G. M. Church [2], Li J, 2014, PLoS ONE
- E. Rosenberg [3], Robust, and Low-Cost Method To Produce the PURE Cell-Free System, Lavickova B, 2019, ACS Synthetic Biology

2.6. Dead Ends

- **Hypothesis 'Magnesium-ATP/GTP Interaction Governs PURE System Yield':** This hypothesis was consistently rejected in debates. Its core rationale was deemed too remedial and narrow, focusing only on preventing a specific failure mode (precipitation at high Mg^{2+} levels) rather than exploring the broader, multi-dimensional landscape to discover new high-performance regimes. It was judged to be less ambitious and less aligned with the multivariate optimization goal compared to hypotheses that proposed more comprehensive kinetic and physicochemical models.

2.7. Figures

3. CONCLUSIONS AND NEXT STEPS

The analysis revealed that maximizing PURE system performance requires a shift from optimizing individual components to co-optimizing interdependent factors. The top-ranked 'Unified Physicochemical and Kinetic Control' hypothesis provides a robust model where fundamental physicochemical viability, governed by free Mg^{2+} , sets the stage for achieving maximum kinetic efficiency through the precise tuning of the ATP:GTP ratio. The proposed full-factorial experiment is designed to map this multi-dimensional interaction surface, validating the hypothesis and systematically identifying a globally optimal regime for protein synthesis that would be undiscoverable through traditional, linear optimization methods.

This synthesis centers on,.

Grounding evidence includes,.

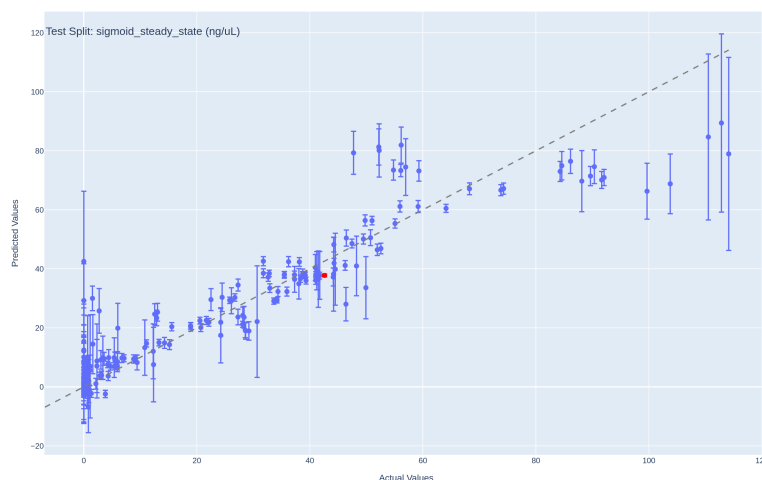


Figure 2. Parity plot showing the predictive performance of the model for final protein yield (*sigmoid_steady_state*) on a held-out test dataset. The correlation between predicted and actual values demonstrates the model's effectiveness in capturing the factors that determine total yield.

REFERENCES

- [1] Y. Shimizu *et al.*, "Cell-free translation reconstituted with purified components," *Nature Biotechnology*, vol. 19, no. 8, pp. 751–755, Aug. 2001, doi: [10.1038/90802](https://doi.org/10.1038/90802).
- [2] J. Li, L. Gu, J. Aach, and G. M. Church, "Improved Cell-Free RNA and Protein Synthesis System," *PLoS ONE*, vol. 9, no. 9, p. e106232, 2014, doi: [10.1371/journal.pone.0106232](https://doi.org/10.1371/journal.pone.0106232).
- [3] E. Rosenberg, "Simple problems, simple solutions," *Society*, vol. 23, no. 4, pp. 24–27, May 1986, doi: [10.1007/bf02701951](https://doi.org/10.1007/bf02701951).

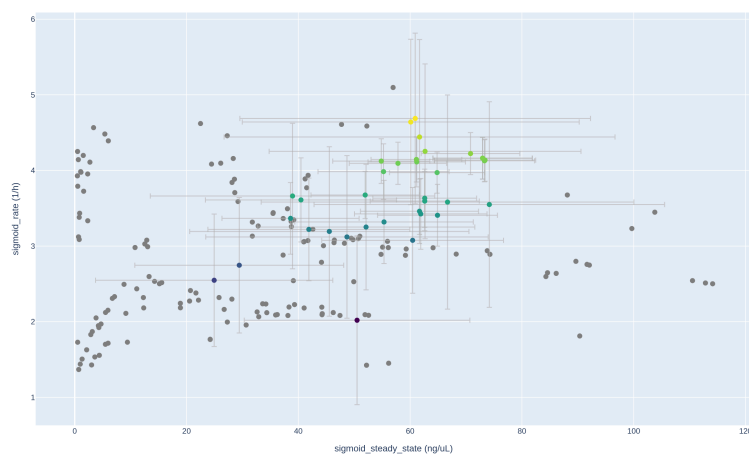


Figure 3. Scatter plot of protein synthesis rate versus final protein yield. This visualization explores the relationship between the two primary optimization targets, indicating whether conditions that improve rate also tend to improve final yield.