

Nucleus Developer Notes

Published Apr 22, 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](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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](https://opendatacommons.org/licenses/by-nc-nd/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 2

Scott Riggs¹ , and IGOR¹

¹Find What Matters

Abstract

We investigate the reduced protein yield in PURE systems assembled by an automated, component-wise liquid handler. Our analysis suggests this is not a failure of final component stoichiometry, but a process-induced limitation rooted in non-equilibrium dynamics. We hypothesize that sequential component addition creates transient concentration spikes, leading to the formation of kinetically-trapped, non-productive states like inorganic precipitates or protein aggregates. A multivariate experiment is designed to test this 'path-dependent yield limitation' by systematically varying both component concentrations and their order of addition to identify an optimized assembly pathway.

Keywords AI Scientist, PURE, PPK

1. OVERVIEW

Automated cell-free protein synthesis platforms like LabCraft promise high-throughput experimentation, but can suffer from performance deficits compared to manual, pre-mixed preparations. This work addresses a persistent reduction in protein yield observed in LabCraft-assembled PURE systems. Through a process of structured debate and hypothesis generation, we moved beyond simple stoichiometric optimization to focus on the assembly process itself. The central hypothesis posits that the sequence of component addition creates path-dependent kinetic traps that limit yield. This note details this hypothesis and presents a multivariate experimental plan to map the interplay between assembly order and key component concentrations, aiming to discover a high-yield protocol.

1.1. Workspace Data

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

2. RESULTS

The investigation converged upon two highly-ranked hypotheses that describe how the automated assembly process itself can limit reaction yield. The core idea is that transient, localized concentrations during sequential mixing create distinct, non-productive states. Below, we outline these hypotheses and the experimental plan designed to distinguish between their proposed failure modes and identify an optimal assembly pathway.

Discourse Graph Representation

2.1. Detailed Findings

2.2. Executive Summary

This research aimed to understand and overcome the reduced protein yield (sigmoid_steady_state) observed in PURE systems assembled via the automated, component-wise LabCraft machine. Through a literature review and a debate-driven hypothesis generation process, the core problem was identified not as a simple issue of final component concentrations, but as a process-induced limitation. The IGOR process converged on the concept of 'path-dependent yield limitation,' where the sequential addition of components creates transient localized concentration spikes, leading to kinetically-trapped, non-productive states (e.g., inorganic precipitates or protein aggregates). The top-ranked hypotheses propose that the specific order of addition of critical components, such as magnesium and the energy regeneration mix, dictates which of these failure modes dominates, thereby setting a ceiling on the final yield. A multivariate experimental plan has been designed to test this by systematically varying both component concentrations and their order of addition, aiming to map this complex interaction and identify an optimized assembly pathway for the LabCraft system.

2.3. Top-Ranked Hypotheses

2.3.1. 1. Path-Dependent Formation of Distinct Kinetically-Trapped States in Automated Assembly

The reduced yield in PURE systems assembled by LabCraft is caused by path-dependent formation of kinetically-trapped, non-productive complexes, a phenomenon that manifests in distinct, experimentally distinguishable forms. Building on the general framework of path-dependency, we propose that the specific class of trap formed is determined by which components experience transient concentration spikes during sequential addition. Specifically, localized spikes of Magnesium and energy components (e.g., Creatine Phosphate, ATP) will favor the formation of inorganic precipitates (e.g., magnesium pyrophosphate), while localized spikes in salts (e.g., Potassium Glutamate) or proteins will favor protein/ribosome aggregation. The final yield is therefore a function of not only the final concentrations but the dominant kinetic trap established by the assembly pathway, providing a more comprehensive model that unifies previous, more specific hypotheses.

Final rationale:

This hypothesis was evolved by combining the winning general framework of 'Explorer 1' (path-dependent, kinetically-trapped states) with the specific, testable mechanism from 'Architect 1' (inorganic precipitation). The debate revealed that 'Explorer 1' was powerful but abstract. This evolution addresses that by proposing multiple, concrete, and distinguishable failure modes (inorganic precipitation vs. protein aggregation) under the same general umbrella. It makes the abstract concept more testable and provides a richer, more nuanced model that has greater explanatory power than either predecessor alone.

Key supporting evidence:

- Rationale from debate (Explorer 1 vs. Architect 1): 'Explorer 1' was praised for its general framework that could encompass multiple failure modes.
- Rationale from debate (Architect 1 vs. Quant 1): 'Architect 1' was praised for its specific, testable mechanism directly addressing the automated vs. manual assembly problem.
- Finding: The method of PURE system assembly, whether from individual components (as with LabCraft) or from pre-mixed solutions, can lead to different performance outcomes, necessitating method-specific optimization. (Bernard-Lapeyre et al., 2026)

2.3.2. 2. Assembly Order Dictates the Dominant Kinetic Trap and Yield Ceiling in LabCraft PURE Systems

The reduced yield in LabCraft-assembled PURE systems is determined by the specific order in which critical components are added, which in turn selects for a dominant, yield-limiting kinetic trap. We hypothesize that the sequence of adding Magnesium Acetate relative to the primary energy source (Creatine Phosphate and ATP/GTP) dictates the failure mode. Specifically, (1) adding concentrated Magnesium to a solution already containing the energy mix will favor inorganic magnesium pyrophosphate precipitation. Conversely, (2) adding the concentrated energy mix to a solution where Magnesium is already present and buffered by other components will mitigate precipitation but may instead favor protein/ribosome aggregation due to localized ionic strength changes. This 'assembly order' hypothesis proposes a direct, controllable experimental variable that determines which non-productive, kinetically-trapped state is formed, thereby setting a path-dependent ceiling on the final protein yield.

Final rationale:

This hypothesis evolves 'Architect 1' by incorporating the more powerful 'kinetically-trapped state' framework from the tournament winner, 'Explorer 1', thus addressing its main criticism of being too narrow. It introduces a significant element of novelty by moving beyond just identifying a problem (precipitation) to proposing a specific, actionable solution rooted in non-equilibrium dynamics: controlling the *order of addition*. This makes the hypothesis not only explanatory but also directly prescriptive for improving the LabCraft process, a key research goal. It elegantly links the abstract concept of 'path dependency' to a simple, testable experimental variable.

Key supporting evidence:

- Rationale from debate (Explorer 1 vs. Architect 1): The winning hypothesis proposed a more general framework of 'kinetically-trapped, non-productive complexes' which is now incorporated.
- Opportunity: The specific mechanistic reasons for reduced yield in PURE systems assembled via automated, component-wise methods (like LabCraft) compared to manual, pre-mixed methods are not fully characterized.
- Analogical support from chemistry where the order of reagent addition is critical in preventing precipitation and controlling reaction pathways.

2.4. Proposed Experimental Plan

2.4.1. Objective

To test the hypothesis that the assembly pathway dictates the formation of distinct, yield-limiting kinetic traps in the LabCraft PURE system. The goal is to identify an assembly order and concentration profile for Magnesium Acetate and Creatine Phosphate that maximizes the final protein yield (*sigmoid_steady_state*) by avoiding these non-productive states.

2.4.2. Multivariate Design Space

This experiment abandons a one-factor-at-a-time approach and instead uses a multivariate combinatorial design. We will simultaneously vary a categorical variable (Assembly Order) and two continuous variables (component concentrations).

1. Categorical Variable: Assembly Order

- **Order_A (Precipitation-Prone):** LabCraft protocol will be programmed to first dispense a master mix of all components *except* Magnesium Acetate. Then, the full volume of concentrated Magnesium Acetate will be added last. This

sequence is hypothesized to create transient, localized Mg^{2+} spikes that favor inorganic precipitation.

- **Order_B (Aggregation-Prone):** LabCraft protocol will be programmed to first dispense a master mix of all components *except* the energy sources (Creatine Phosphate, ATP, GTP). Then, a concentrated mix of these energy components will be added last. This sequence is hypothesized to mitigate inorganic precipitation but may promote protein/ribosome aggregation due to localized ionic strength changes.

2. Continuous Variables: Concentration Matrix

- For *each* of the Assembly Orders defined above, a combinatorial matrix of concentrations will be explored.
- **[Magnesium acetate] (mM):** 5 levels, e.g., {5, 10, 15, 20, 25} mM.
- **[Creatine phosphate] (mM):** 5 levels, e.g., {20, 30, 40, 50, 60} mM.

This design creates a batch of 2 (Orders) x 5 (Mg levels) x 5 (CP levels) = 50 unique experimental conditions, plus controls and replicates. This batch structure will efficiently map the complex interaction landscape between assembly path and component concentrations, directly addressing the epistemic uncertainty around the LabCraft performance gap.

2.4.3. Batch Protocol Steps

1. Prepare concentrated stock solutions for all PURE system components.
2. Program the LabCraft liquid handler with two distinct, multi-step protocols corresponding to Order_A and Order_B.
3. For Order_A, the program will: a. Create an intermediate master mix (IMM-A) of all components (including Creatine Phosphate, ATP, GTP) except Magnesium Acetate. b. Dispense the required volumes of IMM-A for each well. c. Dispense the specified volume of Magnesium Acetate stock solution into each well to achieve the target concentration from the design matrix.
4. For Order_B, the program will: a. Create an intermediate master mix (IMM-B) of all components (including Magnesium Acetate) except the energy sources (CP, ATP, GTP). b. Dispense the required volumes of IMM-B for each well. c. Dispense the specified volume of a concentrated energy mix into each well to achieve the target concentrations.
5. Add the DNA template for deGFP expression to all wells.
6. Seal the plate and immediately begin measurements in the plate reader at 37°C.

2.4.4. Optimization Metrics

1. **Primary Metric (sigmoid_steady_state):** Final fluorescence endpoint will be used to quantify total protein yield. The results will generate two distinct yield landscapes (one for each Assembly Order) as a function of Mg^{2+} and CP concentrations. This directly evaluates the highly coupled interaction between assembly path and stoichiometry.
2. **Secondary Metric (sigmoid_rate (1/h)):** The maximum rate of fluorescence increase will be calculated to determine if the assembly path affects the kinetics of protein synthesis, not just the final yield.
3. **Qualitative Metric (Initial Turbidity):** An initial absorbance reading (e.g., at 600nm) or light scattering measurement will be taken at time zero for all wells. This serves as a direct, physical probe for the formation of precipitates or aggregates hypothesized to be the 'kinetic traps'. A correlation between high initial turbidity and low final yield would provide strong support for the hypothesis.

Feedback Loop to Surrogate Model: The results of this batch experiment—specifically, the `sigmoid_steady_state` values for each of the 50+ conditions—will be fed back into the surrogate model. The `assembly_order` will be encoded as a new categorical feature. The model will then be retrained to learn not only the impact of Mg^{2+} and CP concentrations but also the critical influence of the assembly pathway. This will enable the model to actively optimize the assembly protocol itself, guiding future suggestions toward compositions and assembly sequences that are predicted to avoid kinetic traps and maximize yield, thereby driving the optimization process toward the stated goals.

2.4.5. Generated Experimental Suggestions

Raw GES representation

2.5. References

- Y. Bernard-Lapeyre, C. Cleij, A. Sakai, M.-J. Huguet, and C. Danelon [1], Bernard-Lapeyre et al., 2026, bioRxiv
- J. Li, L. Gu, J. Aach, and G. M. Church [2], Li et al., 2014, PLoS ONE
- J. Li *et al.* [3], Li et al., 2017, Translation
- Y. Shimizu *et al.* [4], Shimizu et al., 2001, Nature Biotechnology
- E. Rosenberg [5], Robust, and Low-Cost Method To Produce the PURE Cell-Free System, Lavickova and Maerkl, 2019, ACS Synthetic Biology

2.6. Dead Ends

- Bimodal Synergy of Magnesium and Creatine Phosphate Drives Protein Yield (ID: Quant 1): This hypothesis was rejected because it focused on rationalizing an existing data pattern rather than providing a novel, mechanistic explanation for the core research problem—the performance deficit of the automated LabCraft assembly process. The winning hypotheses were deemed more impactful as they addressed the *process-induced* limitations, which was central to the research goal.

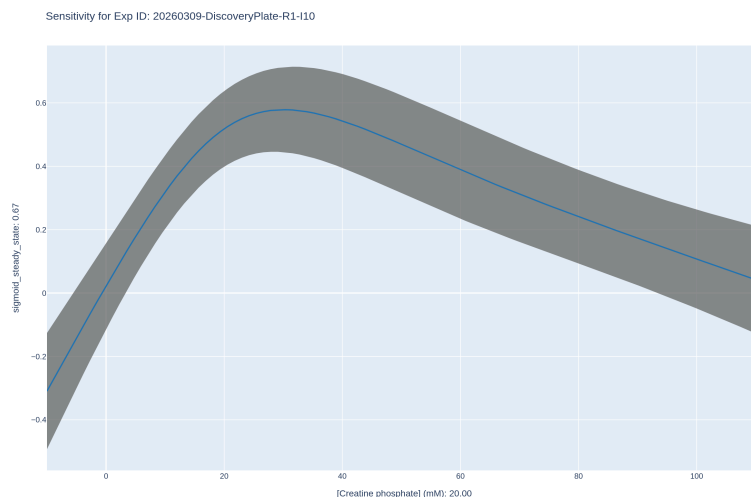


Figure 1. 1D Sensitivity for 20260309-DiscoveryPlate-R1-I10

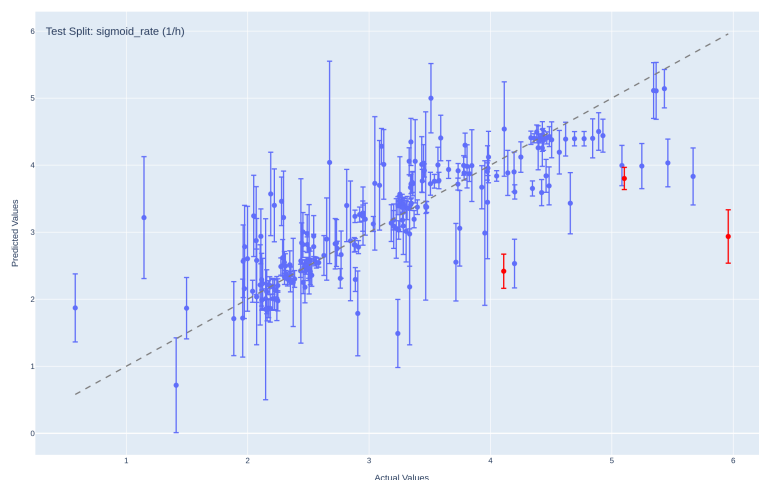


Figure 2. *Test Split: sigmoid_rate (1/h)*

2.7. Figures

3. CONCLUSIONS AND NEXT STEPS

This research has successfully reframed the problem of reduced yield in automated PURE system assembly from a static optimization problem to one of non-equilibrium, process-dependent dynamics. The leading hypotheses concerning path-dependent formation of kinetic traps are both mechanistic and testable. The proposed multivariate experiment is designed to efficiently map the effects of assembly order and concentration, providing a clear path toward not only validating the core hypothesis but also discovering a robust, high-yield protocol for the LabCraft platform. The results will be used to retrain our surrogate model, incorporating assembly sequence as a key variable for future optimization campaigns.

This synthesis centers on,.

Grounding evidence includes,.

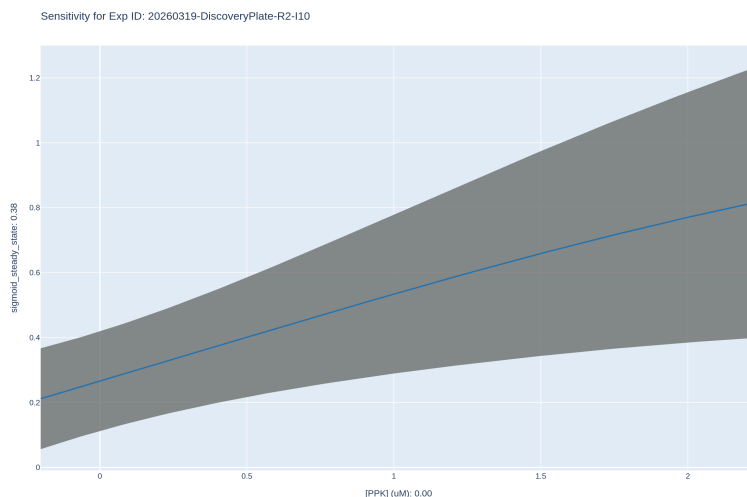


Figure 3. *1D Sensitivity for 20260319-DiscoveryPlate-R2-I10 (2)*

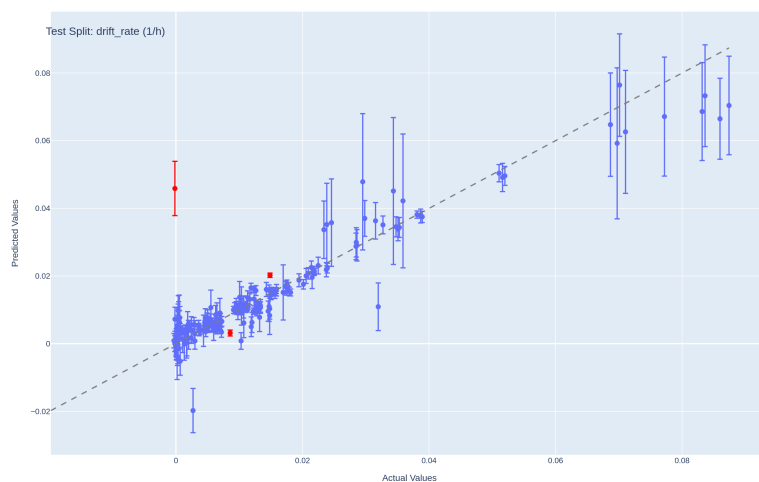


Figure 4. *Test Split: drift_rate (1/h)*

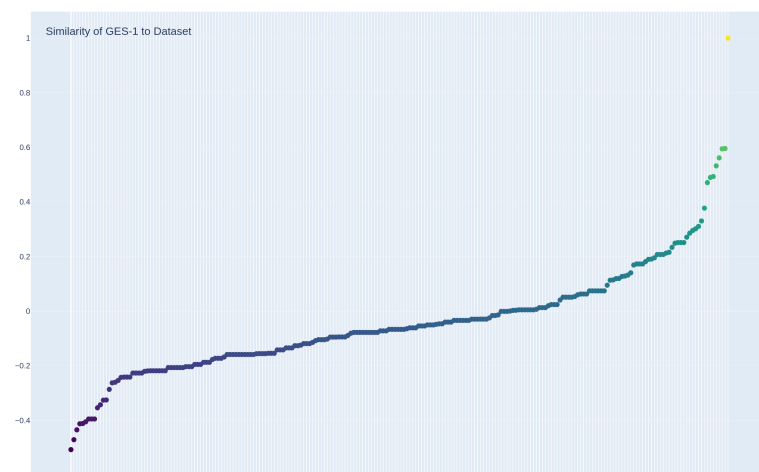


Figure 5. *Similarity of GES-1 to Dataset*

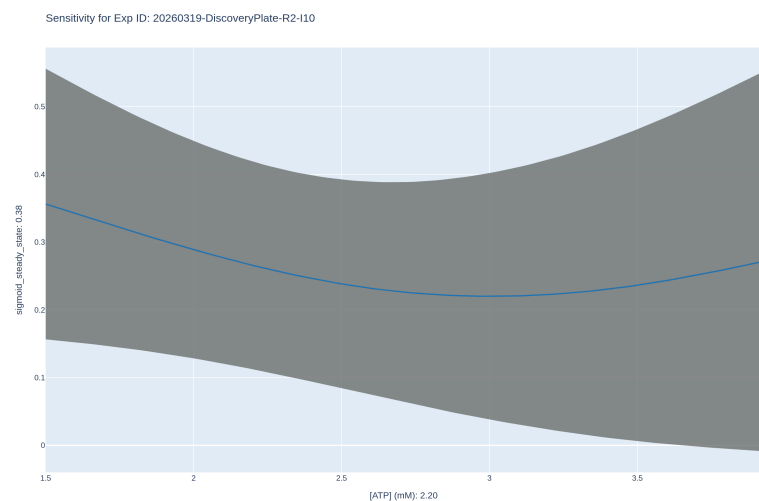


Figure 6. *1D Sensitivity for 20260319-DiscoveryPlate-R2-I10*

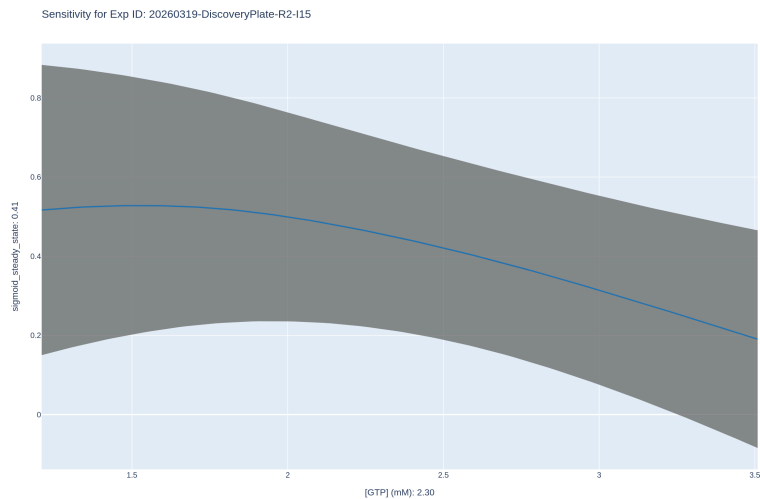


Figure 7. 1D Sensitivity for 20260319-DiscoveryPlate-R2-I15

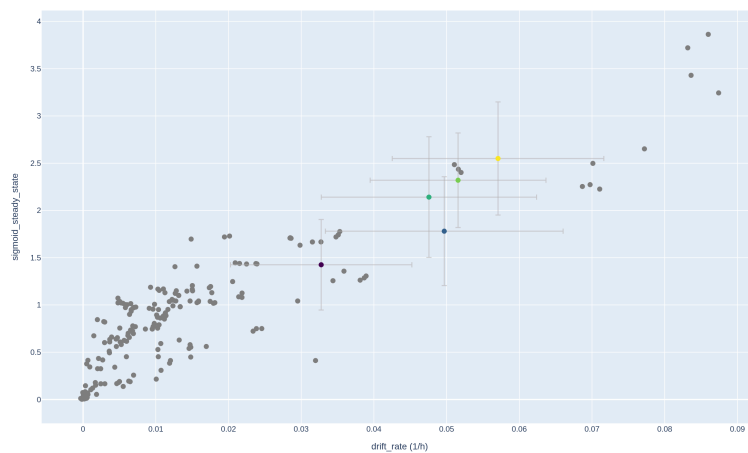


Figure 8. *sigmoid_steady_state* vs *drift_rate* (1/h)

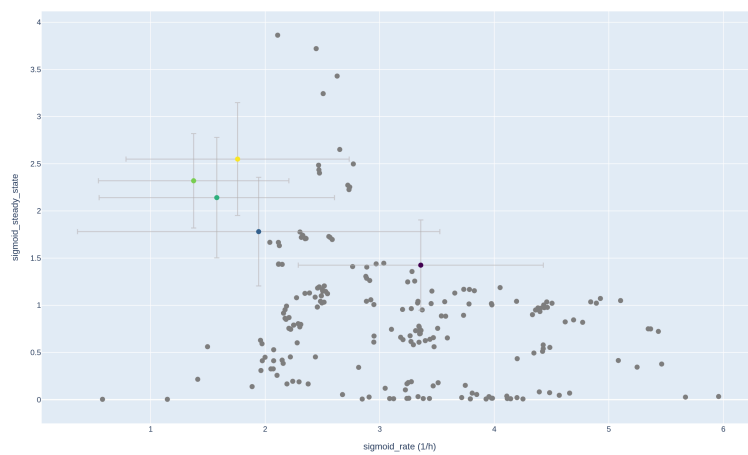


Figure 9. *sigmoid_steady_state* vs *sigmoid_rate* (1/h)

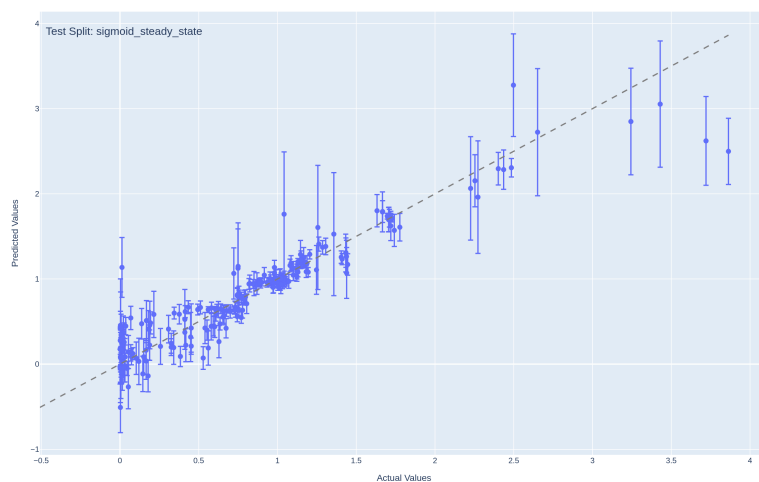


Figure 10. *Test Split: sigmoid_steady_state*

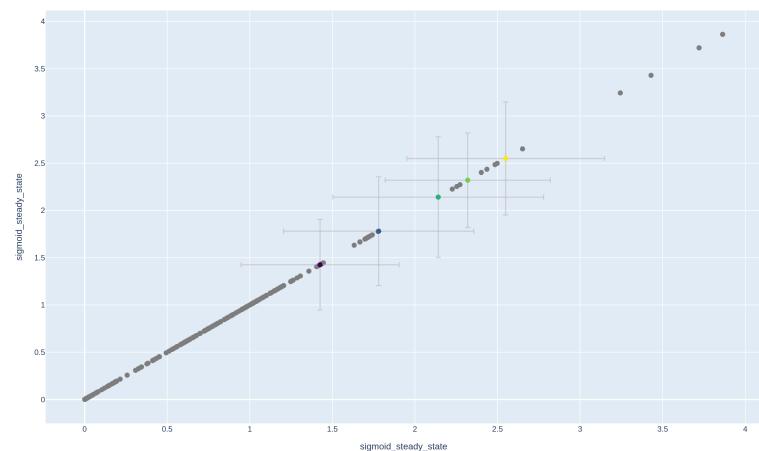


Figure 11. *sigmoid_steady_state vs sigmoid_steady_state*

REFERENCES

- [1] Y. Bernard-Lapeyre, C. Cleij, A. Sakai, M.-J. Huguet, and C. Danelon, "Optimization of PURE system composition using automation and active learning," Mar. 2026, doi: [10.64898/2026.03.23.713685](https://doi.org/10.64898/2026.03.23.713685).
- [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] J. Li *et al.*, "Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system," *Translation*, vol. 5, no. 1, p. e1327006, Jan. 2017, doi: [10.1080/21690731.2017.1327006](https://doi.org/10.1080/21690731.2017.1327006).
- [4] 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).
- [5] 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).