

Nucleus Developer Notes

AI Scientist: Base Module Report

IGOR¹, b.next , and FindWhatMatters 

¹Find What Matters

AbstractKeywords AI

Note

This article was written by a **human author**.

1. OVERVIEW

This report documents iterative optimization by the AI scientist IGOR (Iterative Guide and Orchestrated Research) of the base PURE system to meet deliverable D1 of the ARIA AI-driven Cell-Free Energy Development and Optimization AI scientist project. The report is shared as a Developer Note, encapsulating human-readable information, digital data, and computational analysis. This page summarizes the key experimental data, model predictions, optimized PURE specification, and the project workflow. In addition, the Developer Note contains:

- Raw experimental data and recipes:
 - Data used and generated through iterative rounds is in the `data/` directory of the DevNote.
 - A full curated dataset is available as a CSV at [20260417-iteration-2-data_jl-curated.csv](#)
- Mechanistic interpretation of each round of experiments by IGOR:
 - [IGOR Report: Iteration 1 \(also independently here\)](#)
 - [IGOR Report: Iteration 2 \(also independently here\)](#)

2. WORKFLOW

Experimental data was collected over three rounds:

- First, we curated historical data from human-run experiments on base PURE system and PPK energy module implementation, to pre-seed the IGOR bayesian engine.
- Next, we operated a prototype experiment using a combination of IGOR-informed and manually-created PURE compositions and executed using a combination of automation and hand-running of experiments.
- Finally, we performed a full discovery plate round, designed by IGOR and collected using the automated experiment platform.

These rounds allowed us to prototype, test, and validate our experimental platform, the interfaces between b.next and FWM, and the ability of IGOR to interpret PURE compositional data.

Published Apr 26, 2026

Correspondence to
b.next
anton@bnext.bio

Open Access 

Copyright © 2026 IGOR *et al.*. 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.

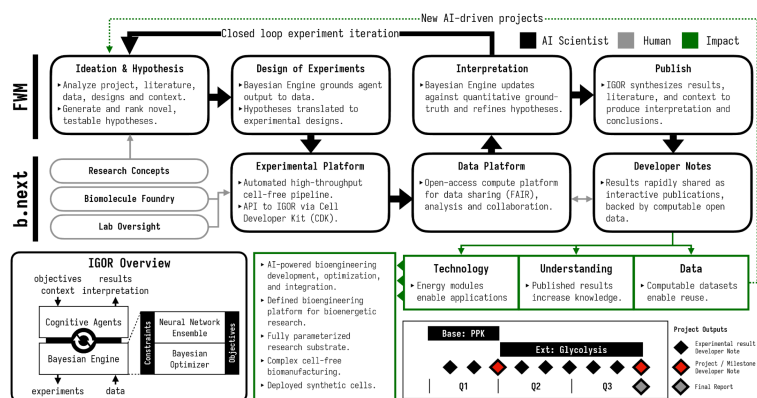


Figure 1. Overview of AI scientist optimization workflow. IGOR designs experiments which are run on the b.next experimental platform. Data from the platform is used to update the IGOR bayesian engine and informs further experiments, and also feeds the orchestrated research component for mechanistic interpretation. Intermediate results are published as Developer Notes.

3. EXPERIMENTAL DATA

The final Discovery Plate round examined 35 different compositions across 7 PURE parameters, producing yields from 0 to 0.62 normalized fluorescence units. The best-performing single composition was I14, with steady-state yield of 0.62 and maximum velocity of 0.31 (Figure 2).

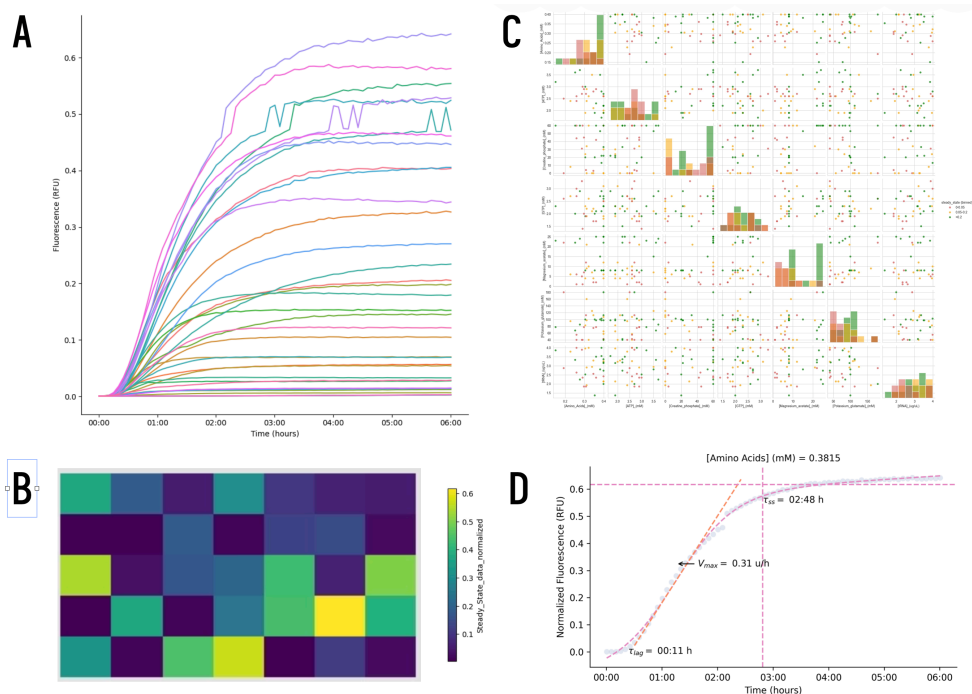


Figure 2. Normalized fluorescence data from Round 1 Discovery Plate. **A** Raw time-series fluorescence traces, one trace per discovery plate well. **B** Wellplate heatmap of steady-state fluorescence, determined by CDK kinetics analysis. **C** Pairwise analysis of discovery plate data, binned by steady-state fluorescence. **D** Kinetic analysis of top-performing composition experiment.

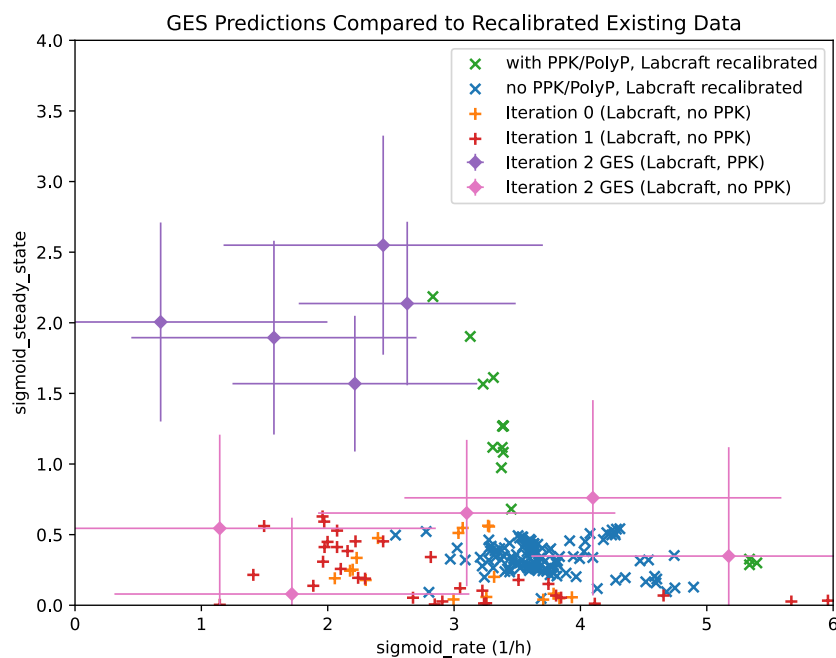


Figure 3. Generated Experimental Suggestions (GESes) predicted by IGOR for subsequent rounds of PURE optimization. The IGOR model was used to adjust data visualization to compensate for differences in experiment assembly and measurement.

4. GES (GENERATED EXPERIMENTAL SUGGESTIONS)

Experimental data produced in the most recent round was used to generate further predictions with IGOR. Including the PPK energy module, IGOR predicts a steady-state yield of 2.55 RFU; excluding the PPK energy module, it predicts a steady-state yield of 0.86 RFU.

5. PURE SPECIFICATIONS

Experiment (-PPK)

Table 1. *Highest-performing PURE composition (direct experiment)*

Property	Value
ID	20260319-DiscoveryPlate-R2-I14
Well	I14
Date	3/19/2026 0:00
Experiment	20260319-DiscoveryPlate-R2
Name	DiscoveryPlate-R2-GES10
Type	Sample
PMix ID	AR-975
NEB in Pmix	FALSE
[PMix] (mg/mL)	1.8
Ribosome ID	AR-905
[Ribosome] (uM)	1.8
SMS ID	220260319-DiscoveryPlate-R2-SMS
tRNA ID	AR-902
[tRNA] (ug/uL)	2.8
DNA ID	AR-1028
[DNA] (ng/uL)	0
PMix Vol (uL)	0
Rxn Volume (uL)	10
[DNA] (nM)	5
[Magnesium acetate] (mM)	15
[Creatine phosphate] (mM)	53
[Potassium glutamate] (mM)	100
[ATP] (mM)	3.5
[GTP] (mM)	2.7
[Amino acid mix] (mM)	0.3815
[HEPES] (mM)	50
[PPK] (uM)	0
[PolyP] (mM)	0
[CTP] (mM)	1
[UTP] (mM)	1
[TCEP] (mM)	1
[Folinic acid] (mM)	0.02
[Spermidine] (mM)	2
[RNase Inhib] (U/mL)	2000
Reader Type	Cytation5
Reader Serial Number	1705168
Gain	ext
Read Type	M
HEPES ID	AR-977
PMix ID	AR-975
Ribosome ID	AR-905
tRNA ID	AR-902
DNA ID	AR-1028

Predicted (-PPK)**Table 2.** *Highest-performing PURE composition (predicted, PPK excluded)*

Property	Value
Experiment	20260319-DiscoveryPlate-R2
NEB in Pmix	False
[PMix] (mg/mL)	1.80
[Ribosome] (uM)	3.60
[tRNA] (ug/uL)	4.46
[DNA] (nM)	9.40
[Magnesium acetate] (mM)	5.00
[Creatine phosphate] (mM)	54.33
[Potassium glutamate] (mM)	98.54
[ATP] (mM)	1.20
[GTP] (mM)	0.92
[Amino acid mix] (mM)	0.18
[PPK] (uM)	-0.00
[PolyP] (mM)	-0.00
[RNase Inhib] (U/mL)	2000.00
Reader Type	Cytation5
Gain	ext
Read Type	M
[PEG] (%)	-0.00
Product	deGFP
Condition	Labcraft
success prediction	0.39
sigmoid_steady_state prediction	0.86
sigmoid_rate (1/h) prediction	5.03
drift_rate (1/h) prediction	0.01
success uncertainty	0.39
sigmoid_steady_state uncertainty	0.46
sigmoid_rate (1/h) uncertainty	1.15
drift_rate (1/h) uncertainty	0.01

Predicted (+PPK)

Table 3. *Highest-performing PURE composition (predicted, PPK included)*

Property	Value
Experiment	20260319-DiscoveryPlate-R2
NEB in Pmix	False
[PMix] (mg/mL)	1.75
[Ribosome] (uM)	2.81
[tRNA] (ug/uL)	4.42
[DNA] (nM)	5.54
[Magnesium acetate] (mM)	5.00
[Creatine phosphate] (mM)	93.68
[Potassium glutamate] (mM)	97.19
[ATP] (mM)	1.67
[GTP] (mM)	1.17
[Amino acid mix] (mM)	0.26
[PPK] (uM)	2.52
[PolyP] (mM)	37.50
[RNase Inhib] (U/mL)	2000.00
Reader Type	Cytation5
Gain	ext
Read Type	M
[PEG] (%)	-0.00
Product	deGFP
Condition	Labcraft
success prediction	0.08
sigmoid_steady_state prediction	2.55
sigmoid_rate (1/h) prediction	2.21
drift_rate (1/h) prediction	0.06
success uncertainty	0.18
sigmoid_steady_state uncertainty	0.75
sigmoid_rate (1/h) uncertainty	1.27
drift_rate (1/h) uncertainty	0.02

6. DISCUSSION

We had several goals for the base module component of the project (months 1–3):

- Increase protein yield of the PURE system by optimizing its composition across its 108 protein, ribonucleic, and small molecule components.
- Prototype, test, and improve IGOR as an AI scientist, the automated experimental platform, and the interfaces between the two.
- Begin prototyping Developer Notes as a means of publishing human- and machine-readable reports on AI progress.

- Test and validate our iterative loop, to set up for further PURE optimization and energy module implementation in the extension (months 4–9) section of the project.

Through the first three months, we validated the ability of IGOR to design compositional PURE experiments, and the ability of the experimental platform to run them. Although we were not able to run as many iterations as we originally anticipated, IGOR's GESes (Generated Experimental Suggestions) in the most recent round potentially span improved performance spaces for the PURE system with and without PPK. In addition, these early rounds led to the validation and curation of a substantial amount of pre-existing PURE performance data (included in the [curated data set](#)), improvements to experimental platform performance, and improvements to IGOR's Bayesian Engine. In particular:

- We implemented scalable handling of categorical input parameters within IGOR to address a clear need to capture batch parameters such as reagent source, experimentalist, and measurement instrument, particularly in our curated historical dataset. Adding these categorical variables substantially changed IGOR's predictions on our most recent data; we hope to validate these improved predictions over subsequent experimental iterations.
- Working with the manufacturer, we increased the capability of our high-throughput nanoliter dispensing system to handle highly-viscous protein samples, increasing the degrees of freedom of the experimental system and allowing us to attain higher protein concentrations within assembled reactions (an expanded experimental range).
- We developed a Developer-Note publishing agent, Sam Cell, as a starting point to the autonomous publication of DevNotes by IGOR as a means of enabling oversight and guidance by humans in the loop, and to open the door to use of the computational DevNotes artifacts for downstream replication and meta-analysis by other AI scientist agents.

Now that IGOR and the experimental platform are running together smoothly we intend to operate at a two-week experimental cadence, a balance between running more iterations and providing time for upgrades, learning, and analysis between iterations. Over the remainder of the project we will:

- Further optimize the PURE base cytosol, expanding the range of parameters simultaneously manipulated
- Validate and optimize the PPK energy module
- Implement and optimize a glycolysis energy module.

In addition, our work so far has identified further opportunities for upgrades and experiments on the system:

- More effective constraints encoding on the AI scientist side, allowing us to better indicate capabilities and limitations of the experimental platform so that IGOR-generated experiments are able to be physically implemented.
- Direct ingest and analysis of raw data by IGOR. At present, IGOR is returned analyzed data from the experimental platform (for example, steady state yield and kinetic parameters). Giving IGOR access to raw time-series information may increase its capability to optimize the system.
- Matched experiments, where IGOR-generated GESes are run simultaneously with human-designed experiments on the experimental platform. This would allow us to compete IGOR head-to-head with an expert scientist, while also allowing us to determine whether supplementing IGOR's input dataset with human experiments increases performance.

- Automated agentic output as Developer Notes. At present, the AI-generated developer notes are an indicative prototype. We look forward to expanding the capability of the agentic Developer Note output, as well as its computation backing, to fully realize the potential of the publications.

7. CONCLUSION

We have now built out the fundamentals of our compositional-biology-based AI scientist system. The experimental platform is increasingly able to manipulate many parameters (proteins, small molecules, etc.) of the cytosolic system, and IGOR is able to interpret data, design experiments grounded in this data, and provide mechanistic hypothesis grounded in both data and literature, published as Developer Notes. Over the next six months of the project we are particularly excited to explore our ability to autonomously implement, integrate, and optimize cytosol modules, and the potential of Developer Notes as a computable, digitally-native means of overseeing, understanding, and disseminating the results of AI-driven science.

8. APPENDIX

The full set of computational data backing this Developer Note is available as a download using the “Computational Archive” link at the top of this page. It may also be browsed in an ephemeral Jupyter environment using the launcher below:

