

Nucleus Developer Notes

Published Jul 25, 2026

Correspondence to
b.next
anton@bnext.bio

Open Access 

Copyright © 2026 b.next & Matters. 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.

Defining AI Scientific Workflows: Using IGOR for Optimization of PURE

b.next , and Find What Matters 

Abstract**Keywords** AI Scientist, Self-Driving Laboratories, Bayesian Optimization,

Synthetic Biology, PURE

Note

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

1. OVERVIEW

Building on Deliverable 1, we continued the co-optimization of the base PURE cell-free system alongside the PPK energy module. In [Round 3](#), this effort achieved a 1.9-fold increase in protein yield over canonical baseline recipes, demonstrating substantial unmapped optimization potential. To capture additional performance gains, we sought to simultaneously tune a broader set of reaction components.

Thus far, we have constrained the number of components being simultaneously manipulated to eight. Fully optimizing PURE, however, requires expanding our parameter space beyond the eight-channel constraint. Exploring this strategy, we prototyped a new, higher risk, compositional approach, wherein we constructed intermediate mixes. These mixes would enable us to vary many more components and then assemble individual reactions by combining these intermediate basis mixtures on deck.

To enable this approach, we developed a batch Bayesian optimization workflow to optimize 11 target components given the constraint of working in an 8-dimensional intermediate working-solution basis ([Round 4, Joseph Lozier, 2026](#)). Natively enforcing shared intra-batch physical constraints directly within the optimization loop addresses an open challenge in bioprocess engineering, one that has historically limited the deployment of Bayesian optimization in high-throughput self-driving laboratories ([Maximilian Siska *et al.*, 2026](#)). Unfortunately, operating within this intermediate basis exposed uncharacterized failure modes across both hardware calibration and chemical formulations.

Throughout this phase, we continued our work on optimization of the base PURE cytosol and of the PPK energy module, in order to focus on building and testing key new capabilities and methods motivated by our results thus far. We expanded the automation platform to handle a wider variety of liquid properties and prototyped a new compositional method. Within the AI scientist, we implemented intra-batch Bayesian optimization and expanded the Orchestrated Research agents with direct data connectors for greater synthesis of organizational knowledge and literature repositories.

Throughout the next phase, we will continue optimization of the PURE cytosol, optimize new energy modules, and further test and expand the IGOR AI scientist.

2. MOTIVATION

An **ARIA-defined** AI scientist aims to execute the entire scientific workflow: ideation, hypothesis generation, experimental design, automated execution, and result interpretation. However, a critical, if not primary, determinant of successful AI deployment is the translation layer between human intent and AI execution. Before novel scientific insights can be generated, relevant experimental context must be rigorously described and characterized, including defining proper feature embeddings, comprehending inherent noise sources, and diagnosing uncontrolled systematics like hardware drift or environmental fluctuations (Martin Seifrid *et al.*, 2022, Gary Tom *et al.*, 2024). Failing to identify and control for these physical realities can result in AI erroneously attributing signal to noise when digesting experimental results, a realization of the classic “garbage in, garbage out” principle.

Autonomous AI deployment magnifies the importance of precise specification of experimental context. Sensitivity to this specification commonly emerges when AI is tasked with navigating high-dimensional spaces (e.g., tuning all variables simultaneously without running replicates (Maximilian Siska *et al.*, 2026)), which pushes experimental setups beyond their standard limits and exposes novel failure modes. Human-driven, highly localized sampling strategies (like Edisonian “one-factor-at-a-time” (OFAT) or fractional factorial experiments) artificially confine experiments to narrow, safe operational windows with better-controlled systematics. While these strategies minimize the probability of experimental failures during runtime, they limit AI’s ability to explore novel regions of input space. The novel experimental conditions that lead to innovative breakthroughs rarely reside within the well-mapped experimental regions (if they did, the scientist would have probably already discovered them); instead, innovation generally requires exploration of as-yet unexplored experimental regions. To effectively and efficiently explore these spaces, AI systems should proactively ingest multimodal information to learn, model, and control for experimental noise and physical limitations. This information can be provided to AI systems in three synergistic ways:

- **textually**, via literature findings and historical insights (and also including textually-encodable information, like graphics, videos, PowerPoint presentations, audio transcriptions of meetings, etc.)(Figure 1);
- **quantitatively**, via the provision of curated data (used, for example, in refining model training, failure mode classification, optimization, and other quantitative treatments) (Figure 2); and
- **interactively**, via active domain-expert-provided, human-in-the-loop feedback (Figure 3).

The Iterative Guide and Orchestrated Research (IGOR) platform, Find What Matters’ AI scientist, can ingest these textual, quantitative, and interactive information sources to produce Generated Experimental Suggestions (GESs): experimental recommendations that maximize the probability of achieving scientific objectives under the model, conditional on being executable in the lab. GESs are obtained via Bayesian optimization (Eric Brochu *et al.*, 2010),

Figure 1. Advanced contextual information ingestion and knowledge corpus synthesis in IGOR (V2).

An AI co-scientist avoids hallucinations, refines hypotheses, avoids experimental failure modes, and generates higher quality insights when it can access and leverage both internal and external repositories of relevant scientific knowledge. This video demonstrates the multi-agentic AI co-scientist platform IGOR (V2) expanding its contextual-grounding capabilities by summarizing, indexing, and organizing multi-source research inputs into a unified knowledge corpus, highlighting: (i) expanded multi-format file support, including Word documents, PDFs, PowerPoint presentations, Excel/CSV files, and plain or markdown text; (ii) dynamic persistent memory storage to capture critical findings derived during expert interactions; (iii) literature integration housing peer-reviewed papers and URLs retrieved via Open Annex searches or direct user uploads; (iv) automated compilation of workspace artifacts into exportable, web-publishable packages; (v) direct connections to internal repositories (e.g., Notion, Google Drive, and Slack); and (vi) retrieval from chat logs to continuously capture domain-expert-provided insights across sessions.

Figure 2. Quantitative data ingestion and multi-objective optimization engine in IGOR (V2). (i) The platform ingests continuous and categorical experimental variables to optimize multiple target outcomes simultaneously, such as PURE yield and reaction kinetics. (ii) The surrogate model can incorporate multi-fidelity learning across historical and varied metrology datasets. (iii) To account for physical process limits, integrated binary and multi-class classifiers flag non-viable experimental regimes, mapping unrunnable failure modes and penalizing risky candidate regions while maximizing target regressor values.

Figure 3. Interactive domain knowledge injection and agentic audit logging in IGOR (V2). In low-data regimes, purely statistical surrogate models can propose unfeasible or physically unrunnable experiments, consuming valuable laboratory resources. To accelerate optimization and prevent unviable execution, IGOR enables domain experts to directly inject qualitative heuristics, pseudo-data, and parameter constraints into the optimization loop. This video demonstrates some of IGOR's interactive information ingestion capabilities, highlighting: (i) pseudo-data classification, where expert intuition (e.g., identifying zero-magnesium formulations as unviable) classifies proposed candidates as "unrunnable" and adapts parameter bounds; and (ii) agentic audit trailing, where IGOR logs every user decision, constraint modification, and literature highlight into persistent memory artifacts for continuous system learning across iterations.

an iterative, model-based strategy for black-box function optimization that is particularly tailored to situations in which the black-box function is difficult to model from first principles, high-dimensional, convolved with noise, and expensive to query (restricting the total budget of allowable function observations). Relative to other optimization methods, Bayesian optimization converges to global optima in significantly fewer iterations in the general setting, primarily due to its model-guided approach to trading off exploration and exploitation and its tendency to sample multivariately over the entire domain of the black-box function (i.e., "the search space", and more specifically the set of valid input features over which it is possible to query the black-box function for its corresponding output value). Successful deployment of Bayesian optimization depends sensitively on the definition of the search space through precise delineation of:

- **Bounds (i.e., per-feature, axis-aligned constraints):** Strict maximum and minimum limits, applied independently to individual input features. An example from this phase includes a minimum value for creatine phosphate (CP) being set to 0 mM (no CP) and the maximum user-defined value of 100 mM.
- **Constraints (i.e., mixed feature, non-axis-aligned constraints):** Mathematical expressions of input features that must also be respected by each valid experiment. An example from this phase includes ensuring that the total combined volume of five reagents remains strictly below 10 μ L.

In addition to delineating the search space that each *individual* experiment must respect, there are also **intra-batch constraints** (Joseph Lozier, 2026): restrictions to the experiments that Bayesian optimization recommends arising from the experimental imperatives to run multiple experiments in parallel at a single iteration. Running batches of experiments at a single iteration reflects the diminishing marginal costs of experiments once initial preparatory lab work has been undertaken (e.g., once reagents can be dispensed and mixed in a single sample well on a well plate, there is relatively less overhead to mixing different ratios of the same reagents in a separate sample well), shared-state conditions across experiments at that iteration (e.g., if reactions are distributed across a thermally-anchored well plate, then all experiments in the batch must respect that single, uniform temperature), and multiplexed hardware limitations (e.g., if reactions use a liquid-dispensing tool to pipette solutions held in a fixed number of cartridges, but those solutions are themselves composed of a significantly larger number of reagents, and experimentation requires efficiently sampling the reagent feature space). Asking for a batch of recommendations requires changes to the Bayesian optimization routine in order to ensure that batch optimization remains computationally tractable (Roman Garnett, 2023) and maintains sufficient diversity across the batch (Jenna Fromer *et al.*, 2025).

Given the algorithmic complexity associated with training mathematically principled AI systems, coupled with recent advancements in large language models (LLMs), much of current research has focused on ostensibly simpler, prompt-driven AI scientist frameworks over Bayesian optimization approaches. While such an implementation lowers the burden to initial deployment, the responses that LLMs generate in this mode are prone to returning unreliable results due to structural biases like hallucinations (Jiawei Gu *et al.*, 2026) and genuine gaps in their ability to mathematically reason (Iman Mirzadeh *et al.*, 2025). Moreover, absent a mathematical model and a mathematical principle to generate more promising candidate designs and to rank-order their viability (as a Bayesian optimization approach accomplishes with a surrogate model and an acquisition function), non-model-based approaches ultimately end up offloading the design refinement and post-filtering tasks to the human user (Joy Datta *et al.*, 2025, Alexis A. Smith *et al.*, 2025). Performing these tasks is highly nontrivial even for an individual with substantial expertise in both AI and the specific scientific domain (Claudio Zeni *et al.*, 2025). The ultimate effect is to decelerate or even preclude novel insight generation / achieving scientific objectives. These drawbacks highlight the advantages of a model-based optimization approach for accelerating scientific innovation, even when accounting for its added implementation challenges.

Instead of being *substitutable* for model-based approaches, agentic / LLM-based components are *complementary* to model-based approaches, and an exemplary AI scientist should integrate both throughout its experimentation and knowledge-distillation processes. This is precisely the design ethos underlying IGOR.

3. AI SCIENTIST WORKFLOW

3.1. Problem Setup: From Deliverable-1 (D1) to Deliverable-2 (D2)

During the first phase of the project, we established our ability to design and run compositional PURE experiments: curating historical data from human-run experiments to pre-seed IGOR's Bayesian engine, running prototype experiments, and collecting the first full discovery plate round designed by IGOR and executed using our automation platform (D1 Report). We deliberately limited IGOR's bounds and GESs to small molecules because their dispensing methods are well characterized and the critical reagents were readily available. This allowed us to prototype, test, and validate the workflows between IGOR and b.next automation pipelines, with the last round optimizing small molecules in the base cytosol recipe: ATP, GTP, amino acids, creatine phosphate, magnesium acetate, potassium glutamate, and tRNAs.

During this second phase, we expanded on this capability to run a larger screen focused on optimizing the base PURE recipe and the PPK energy module. In order to give IGOR the ability to manipulate all component concentrations simultaneously, we developed and attempted a novel experimental process: constructing individual reactions through IGOR-designed "intermediate mixes". This new method pushed the limits of our composition and automation platform, with our first experiment failing during intermediate assembly. However, work to design this method led to substantial technical upgrades to IGOR, as well as new insight into compositional assembly of PURE.

3.2. Wrangling and Curation of Information for Surrogate Modeling

A fundamental aspect of experimentally deploying any AI co-scientist is translating (i.e., "representing" or "embedding") what happens experimentally into a numerical form that can be processed by the AI. This translation layer is context-specific, non-regimented, and iterative, but it broadly consists of two stages: wrangling and curating. During the wrangling stage, all of the relevant experimental context is aggregated, standardized, and cleaned. Domain experts ensure that all important input and output quantities are recorded, that metadata

associated with those input and output quantities are properly specified (e.g., distinguishing input from output quantities; labeling quantities in a uniform fashion—for example, all with ASCII characters, all with units, or all without spaces; specifying data types; etc.), and that the data are free of certain flaws (e.g., indicated numbers accurately parameterize the experiment, instead of referring to intended set points; that null values are handled; that missing, duplication, and transcription errors are corrected; that experimental identifiers are unique; etc.). Curation then transforms the wrangled data into a form suitable for AI ingestion and data modeling. Domain experts downselect wrangled features (e.g., excluding features that are held constant throughout experimentation, and therefore whose causal relationships to outputs cannot be estimated), consolidate dependent feature columns into a minimal basis set of feature columns (e.g., excluding feature columns that are scalar multiples or linear superpositions of other feature columns, since including the whole feature column set can adversely affect the model's ability to disambiguate the relationship between features and outputs during model training), and suitably transform data columns (e.g., one-hot encoding of categorical features, mathematically transforming columns due to numerical conditioning or motivated by scientific insight) to produce a curated dataset with properly indicated metadata. During this milestone, we executed 1) the transformation of raw data sheets into a single **wrangled** comma-separated value (CSV) file, 2) the transformation of the wrangled CSV into an IGOR **ingestible** CSV file, and 3) the transformation of the wrangled CSV file into a **curated** Excel file through manual scripting / editing; however, concurrently, we also developed and integrated an agent-based wrangler / curator into IGOR, making it possible to accomplish the same data transformations wholly within the AI co-scientist platform (Figure 4). While this wrangler / curator capability might seem primarily directed towards user convenience (and, indeed, it does streamline the wrangling / curation process), access to the history of wrangling and curating decisions is helpful to IGOR's agents in acquiring information from the user that it will use later on in the workflow to synthesize detected data patterns and connect them to the broader body of scientific knowledge.

Production of the curated dataset is the main precondition for commencing the Bayesian optimization portion of IGOR's co-scientific workflow. Executing Bayesian optimization requires 1) the precise specification of the domain (**Motivation**), 2) the furnishing of output goals, and 3) the configuration of a suitable objective function (i.e., a function that reduces a set of experimental outcomes into a scalar number that can be used to rank the utility of each experiment in the curated dataset). While defining the output goals is usually clear from context (e.g., maximizing a yield, minimizing the rate of steady-state drift, achieving a target protein concentration, etc.), quantitatively defining the types of acceptable tradeoffs ("How valuable is an increment of yield improvement if it comes with a corresponding penalty to steady-state drift and protein concentration?") and quantitatively restricting the domain based on scientific knowledge / experimental conditions both require a bit more care in order to guide IGOR towards experimental goals.

Figure 4. End-to-end dataset wrangling and curation pipeline in IGOR (V2). The main precondition for commencing Bayesian optimization is the production of a curated dataset; however, the process of executing curation steps (e.g., aggregating data sources, specifying metadata, and downselecting features) is traditionally manual, error-prone, and destructive of important context that can be helpful to a downstream AI scientist. This video demonstrates the multi-agent AI co-scientist platform converting raw, multi-source research inputs into curated, AI-ready datasets, highlighting: (i) workspace initialization and objective definition from uploaded supporting files; (ii) contextual grounding via access to literature and internal information repositories; (iii) context-aware ingestion and combination of tabular data; (iv) autopopulation of input, output, and metadata and user review; (v) automated data cleaning (e.g., rectifying missing values, outliers, and duplicate IDs); and (vi) feature embedding via downselection.

3.3. Bayesian Optimization Strategy for PURE

By default, Bayesian optimization treats all outputs as equally important. While this is the appropriate default (indeed, any other default would necessarily be customized to a particular context), it can result in substandard recommendations as evaluated by a domain expert if (by contrast) certain outputs are valued more highly than others. This mismatch between default and expectation arose during the present milestone (the steady-state fluorescence fit parameter output, proportional to PURE protein yield, was the primary focus of the instant application, of more importance than the secondarily important steepness fit parameter output, characterizing reaction rate), and it had a simple remedy, domain expert evaluation that the optimizer distributed insufficient force towards the more important goals, and manual corrective adjustment of the goal limits in response.

Another Bayesian optimization default regards the input feature domain itself: training and optimization of the surrogate model requires embedding the set of valid input features into a vector space (by default, the real vector space of dimension n , where n is the number of input features), which can then be further subset by bounds, constraints, and additional restrictions derived from the scientific application. The bounds used for this round can be found here ([Bounds](#)), and there were no explicit constraints applied by IGOR while generating GESs; however, an important, application-specific restriction that arose during this milestone regarded the liquid handling robot used to dispense the nanoliter-scale reagents. While the liquid handling robot plate assembly allows for ≈ 100 experiments to be run in near-parallel, all of the experiments had to be made from the same few working solutions (eight working solutions, due to known degradation when reagents incubate at room temperature for extended periods ([Surendra Yadav, 2026](#))). Experimental design thus became a question of choosing the sets of working solutions, which can directly reconstruct and densely sample the raw input feature dimensions ([Figure 5](#)). To the best of our knowledge, using IGOR to optimize in the working solution space represents the **first deployment of an AI co-scientist performing Bayesian optimization in an intermediate basis**. The eight-dimensional working-solution basis was constructed to span the largest volume of promising experiments as determined by the surrogate model, while also containing the control within the interior of its convex hull and avoiding functional failure of protein components using the domain-expert-provided heuristic (working solutions in [concentrations](#) and in [volume fractions](#)) ([Figure 6](#)). Unfortunately, two of the eight working solutions (specifically, working solution 4 and working solution 5) could not be thoroughly mixed due to salt precipitation, thereby inhibiting the ability to test the corresponding GESs.

Both the choice of working solutions and the batch of GESs (re-translated into [raw features](#)) are entry points for interactive user evaluation that significantly affect the scientific optimization trajectory. IGOR did not consider any specific fluid or mixing reagent properties while generating candidate working solution bases beyond the rule of thumb provided; however, such properties could have been incorporated with further information provided. Examples of such additional information include: 1) uploading textual information describing the relevant reagent fluid properties, and allowing IGOR to evaluate the working solutions in light of this textual information; 2) experimenting with the proposed working solutions, observing aberrant fluid behavior, modeling that behavior (e.g., via a new classifier output), and refining the working solution basis in light of that newly modeled behavior; and 3) additional domain expert evaluation, beyond the rule of thumb provided. These actions are actively being discussed as we plan for the next milestone.

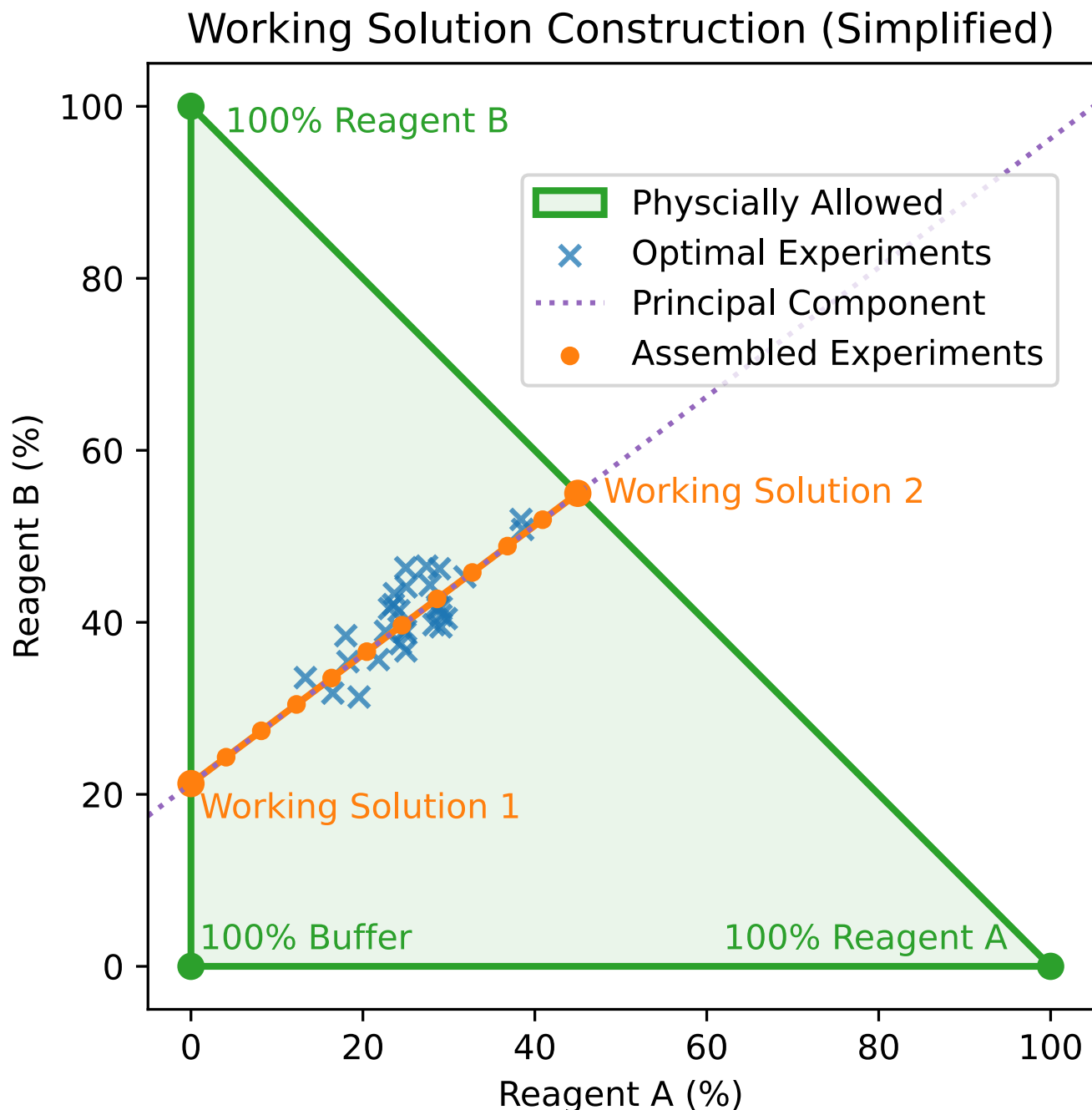
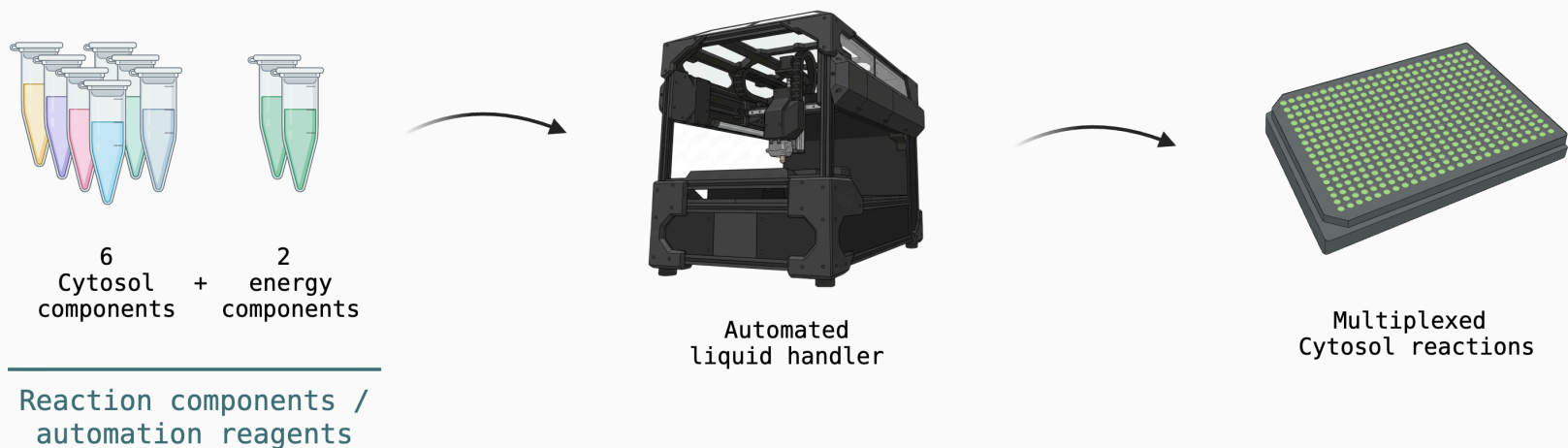


Figure 5. Simplified intermediate basis problem for automated plate assembly Automated droplet printers are powerful tools as they facilitate high-throughput parallel experimentation. The PURE workflow added another operational complexity, as sampling over an optimized space required preparing mixtures from a limited set of working solutions (in our setup, using eight working solutions to vary eleven input parameters). Experimental design thus transforms into selecting optimal working-solution sets that allow the platform to sample the optimized parameter space. For visualization purposes, consider a reduced-dimension formulation of the working-solution selection problem. Experimental formulations are composed of two reagents and a buffer solution, where valid mixtures populate the interior of a triangle (green shaded area) with vertices 100% Buffer, 100% Reagent A, and 100% Reagent B (green points). Enforcing a hardware constraint where the liquid handler dispenses only three working solutions into a fixed reaction volume restricts accessible formulations to a linear segment between by Working Solution 1 and Working Solution 2. Once these endpoints are selected, the platform can densely and cost-effectively sample this linear space. To determine optimal working solutions, candidate recipes are first generated without the working-solution constraint. Bootstrapping an existing dataset, via repeated resampling, re-modeling, and recipe re-optimization yields a distribution of candidate optima (blue crosses) that explicitly accounts for model uncertainty due to limited sampling and experimental

Round 3 workflow: Simple component input for automation robot



Round 4 workflow: Complex design in intermediate basis

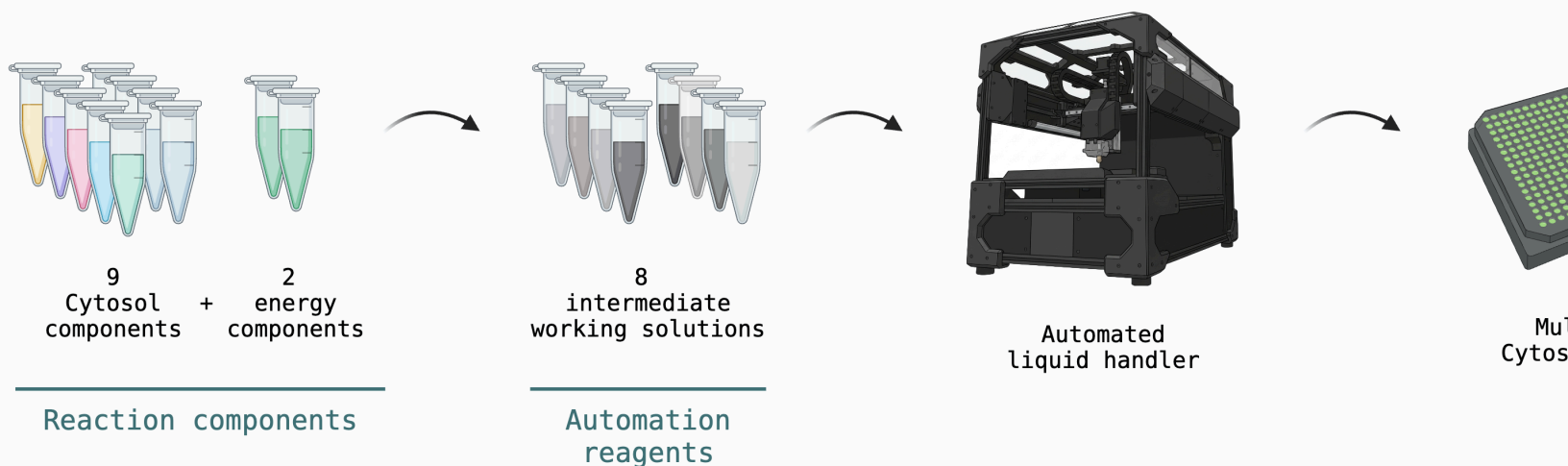


Figure 6. Progression from direct component dispensing to intermediate-basis batch optimization using an automated liquid handler. (Top) Round 3 workflow: Direct component dispensing scheme. A 1:1 mapping is maintained between reaction components (6 cytosol + 2 energy module components) and the 8 automated liquid handling cartridges, directly assembling multiplexed cytosol reactions on-plate. (Bottom) Round 4 workflow: Intermediate-basis optimization scheme engineered to overcome physical liquid-handling limits. To sample an expanded 11-dimensional parameter space (9 cytosol + 2 energy module components) without exceeding the 8 available physical dispensing channels, reaction components projected into an 8-dimensional set of intermediate working solutions. These intermediate mixtures serve as the automation reagents loaded into the liquid handler to construct high-dimensional, multiplexed cytosol reaction plates.

4. EXPERIMENTAL RESULTS

4.1. Round 3: PURE Composition and PPK Energy Module Optimization

Building on our results from the first phase of the project, we expanded our optimization to a wider range of features critical to cytosol optimization. We introduced PPK and PolyP (the PPK energy module) for optimization, in addition to a subset of base cytosol small molecules: tRNA, magnesium, creatine phosphate, ATP, GTP, and amino acids. In order to validate the experimental platform and provide a baseline against less-sophisticated optimization techniques, we included compositions constructed using Latin Hypercube Sampling (LHS), controls assembled by hand and with automation, and metrological standards. Overall, this round included: 29 unique IGOR GESs in duplicate, 20 LHS-designed samples in duplicate, 5 manual controls in triplicate, 3 automated controls, 2 reagent controls + PPK/PolyP controls, and 3 fluorescein standards, for a total of 106 experiments (Figure 7, Figure 8).

During analysis, we discovered discrepancies that revealed issues with liquid dispensing, resulting in differences between nominal design compositions and the effective composition within the reaction. We characterized this discrepancy, reconstructing the true input concentrations for the purposes of analysis. As a result, this round tested approximations of the IGOR- and LHS-designed compositions.

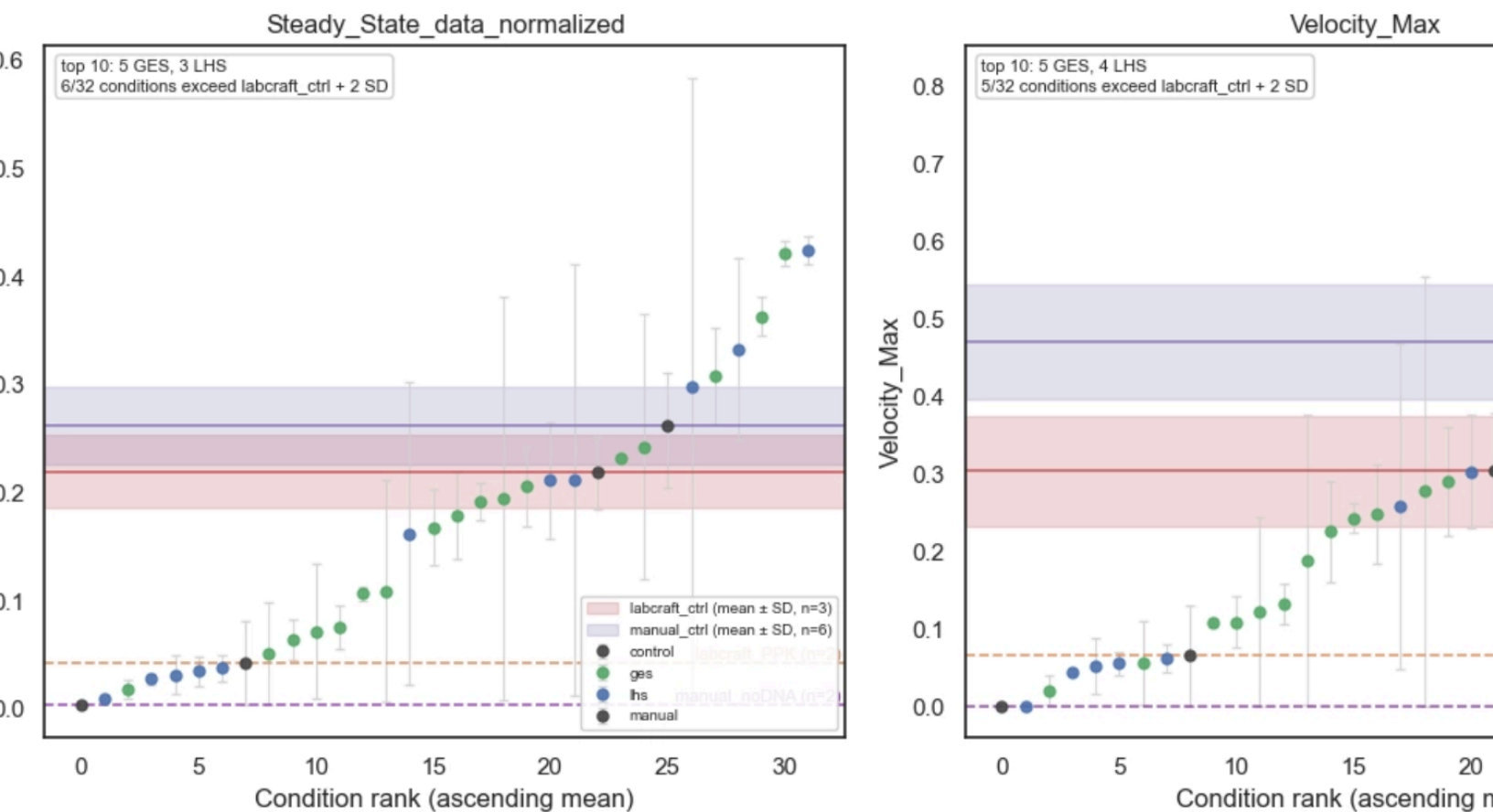


Figure 7. Round 3 experimental results. Steady state (left) and maximum translation rate (right) performance of experimental compositions relative to manually-assembled (blue) and automated (red) controls. Dotted lines indicate negative controls. Five compositions exceed the performance of the controls.

Name	ATP (mM)	CP (mM)	DNA (nM)	GTP (mM)	Mg (mM)	PolyP (mM)	PPK (mM)	tRNA (mM)	Mean
10	3.27	29.60	2.15	3.81	10.49	10.70	3.63	4.13	0.0
02	2.07	48.62	2.10	2.35	72.10	31.27	1.42	3.02	0.0
12	2.07	57.41	2.09	2.34	81.60	31.16	0.24	3.88	0.0
1	2.29	28.99	2.18	6.19	32.14	18.03	2.45	4.55	0.0
13	1.87	30.87	1.89	1.87	23.34	0.00	0.00	2.50	0.0
7	2.72	39.70	2.25	5.82	17.29	33.62	2.03	4.62	0.0
ual Control	2.00	20.00	3.00	2.00	8.00	0.00	0.00	3.50	0.0
21	2.12	41.43	2.14	2.12	50.52	31.97	2.65	3.27	0.0
16	2.03	42.34	2.05	4.70	72.97	0.00	4.16	2.54	0.0
mation Control	1.89	18.85	1.91	1.89	7.54	0.00	0.00	3.16	0.0

Figure 8. Table of Round 3 experimental results. Compositions outperforming the automated positive control. 8 compositions exceeded control performance, with two reaching nearly double that of the control. We expect a manual control to have higher expression due to differences in assembly.

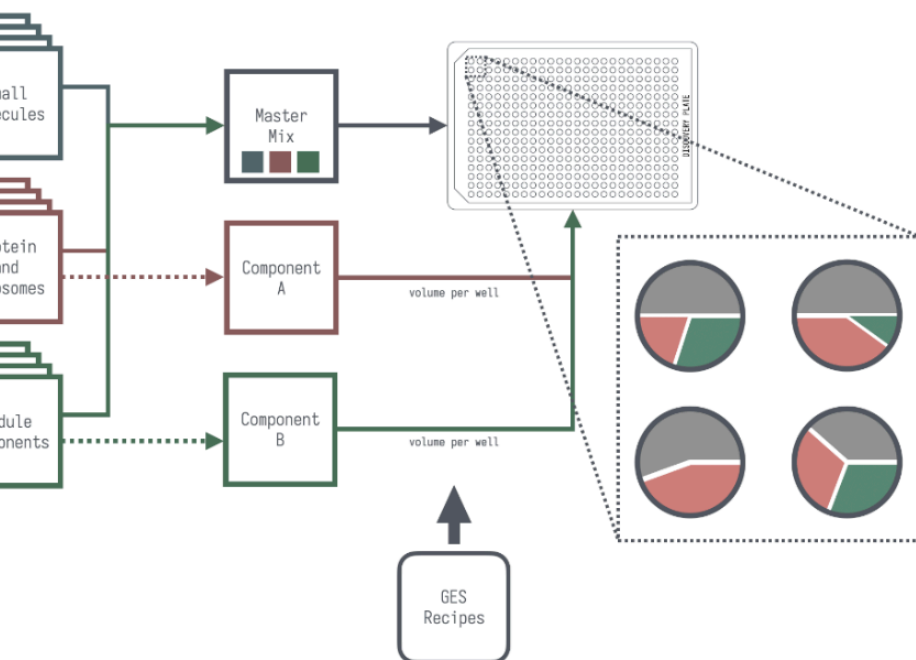
This round identified 8 PURE compositions exceeding the performance of the reference control. These compositions spanned a range of 0 to 3.63 mM PPK and 0 to 31.27 mM PPK, demonstrating performance improvements of both the base PURE cytosol as well as improvements potentially driven by the PPK energy module. Both IGOR-generated GESs and LHS-designed compositions showed improved performance. We anticipate follow-up manual experiments to confirm the performance of the recipes optimized here.

4.2. Round 4: Intermediate Mixtures Basis

Given the success of Round 3, we wanted to push IGOR and our automation capabilities further and expand the range of testable conditions. In particular, we wanted to establish an experimental procedure which could sample the concentrations of many more components of a given reaction simultaneously, to more efficiently and effectively explore compositional space. To do so, we attempted a novel means of experiment composition: constructing intermediate mixtures and then combining them downstream (Figure 9).

The higher volume of the intermediate mixes enables us to combine more components with higher accuracy and across a wider range of components. These intermediates are then as-

A. Direct Composition



B. Intermediate Basis Composition

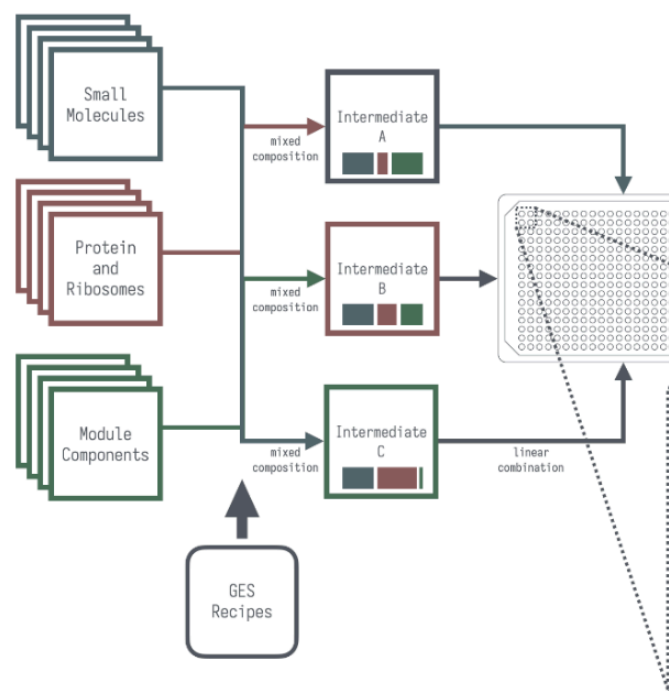


Figure 9. Approaches for assembling PURE compositions. (A). Compositions are directly assembled, loading an experiment plate with a master mix and directly dispensing the cytosol components being tested into the Master Mix using low-volume liquid handling. (B). IGOR designs “intermediate bases” which are assembled in higher volume, upstream of the experiment plate, enabling manipulation of more cytosol components simultaneously. These intermediate bases are then combined in the final experiment plate in order to test a wide range of compositions.

sembled in linear combination using low-volume liquid handling. These linear combinations of the intermediate bases screen a wider range of concentrations while requiring a smaller number of individual handling steps on the automation platform.

In order to test this approach, we built intermediate mixes to optimize similar compositions to those explored in Rounds 2 and 3, while increasing the number of components manipulated. We built eight intermediate mixes spanning 11 unique reagents (potassium, magnesium, ATP, GTP, CP, amino acids, tRNA, polyphosphate, PPK, ribosomes, and protein mix). To retain protein function, every reagent except DNA was required to comprise 10% of its baseline concentration. DNA was titrated in separately at the end to start the reaction.

Unfortunately, this first attempt at intermediate basis composition failed during execution. During assembly of the intermediate basis mixtures, small molecules (in particular ATP) precipitated out of solution due to their relative concentration to other components. Furthermore, the liquid characteristics of these assembled intermediates were substantially different from the calibrated individual components used in prior rounds, leading to inaccurate dispensing. Tighter bounds on intermediate mixture composition, and additional liquid handling calibration, could resolve these issues.

While intermediate basis mixtures provide a powerful way to explore wider ranges of compositions more effectively, we anticipate falling back to our validated direct composition experimental method for subsequent optimization rounds. Building on our results in [Round 3](#), we

will expand our optimization of PURE energy components. We anticipate building towards optimization of the glycolysis pathway by testing additional energy modules. In particular, we are exploring the pyrovate-acetate pathway module as a second step for optimization: the pathway has more components than the PPK energy module, but fewer than glycolysis, and uses more commonly available and less hard to work with energy substrates. We will continue to explore intermediate composition as a separate and lower-priority activity.

5. FURTHER DISCUSSION

The process of offering corrective feedback to the translation layer when the AI co-scientist fails to consider important context, extrapolates too far, or otherwise returns results of tenuous reliability is context-specific, non-regimented, and iterative. Especially at the early stages of experimental deployment, progress is largely characterized by enabling the ingestion and integration of such corrective, domain-specific information. Without the ability to do so, the iterative optimization loop gets stuck suggesting experiments that do little to advance experimental progress and/or cannot even be tested. While seeding the AI co-scientist with information about known experimental constraints and failure modes is the ideal practice for co-scientific deployment, predicting the most relevant obstacles *a priori* and aggregating a corresponding knowledge corpus is highly nontrivial. A more pragmatic approach is to enact the co-scientist workflow in light of known issues and to course correct dynamically (especially through the provision of supplementary data and textual information) when / if additional experimental or co-scientific issues arise. A few of the specific issues that we encountered in this milestone are discussed in [Experimental Results](#).

While some of these issues might have been overcome through subsequent experimental iterations and additional domain expert involvement, others (especially reagent mixing order and working solution mixing) exposed inadequacies of earlier versions of IGOR that then demanded capability development concurrent with this milestone. In particular, we overhauled our agentic workflow to more naturally elicit (and incorporate) domain expert feedback and prompted IGOR with relevant experimental checks provided by information pipelines (i.e., “connectors”) to internal repositories (e.g., Slack, Notion, Google Drive, etc.). Access to natural language documentation stored in b.next’s internal systems directly surfaced the reagent mixing order issue, an experimental factor that at least could have been explicitly presented as a checklist item during the experimental planning phase and, with both more data-summoning and backend development effort, could have been explicitly modeled and synthesized within IGOR’s Bayesian optimization engine. Specifically, this backend development could be directed towards enabling model architecture changes that enforce the importance of the time ordering of the relevant reagent features. More generally, we have continued to reorganize and (as indicated by experimental need) to augment our suite of agents in order to 1) add relevant information that buttresses IGOR’s insights (e.g., via agents leveraging information connector channels), 2) facilitate the integration of important domain context (e.g., via agents aggregating information during data wrangling and curation), and 3) streamline active domain expert involvement in a manner that is natural to the domain expert through modifications to IGOR’s user interface.

6. CONCLUSION

Within the context of autonomous deployment of an AI scientist in a particular scientific domain (e.g., synthetic biology), while there may be overarching scientific objectives (e.g., increasing protein yield and rate), the path to achieving those overarching objectives typically requires overcoming a series of intermediate obstacles (e.g., efficient mixing of novel reagent and calibrated dispensation of nanoliter fluids) that are difficult to anticipate *a priori* and are themselves high-dimensional, stochastic, black-box optimization problems. Both AI- and

non-AI-based approaches can stumble at these intermediate stages, as it is challenging to aggregate relevant experimental data characterizing the sub-problems, retrieve relevant additional information from external resources, and distill both into an actionable plan. Model-based optimization strategies are systematic, principled, and achieve objectives significantly faster than non-model-based approaches, but they require proper configuration and information-seeding so that the AI scientist can understand relevant experimental context and dynamically course-correct in light of newly emergent information.

Our experience deploying IGOR to optimize PURE conformed to this pattern, requiring us to extend IGOR's capabilities to enable batch Bayesian optimization in an intermediate basis and with our overall objectives interrupted by diagnosing various systematic effects arising from this experimental setup. Further changes to IGOR's workflow, such as the ability to retrieve additional relevant experimental context via information connector agents to internal knowledge ecosystems (e.g., Slack, Notion, Google Drive) also appears to be important for refining IGOR's ability to execute the entire scientific workflow (as defined by [ARIA](#)) and for soliciting important feedback from domain experts. With the improvements developed in this phase, IGOR's neural-network-backed, intra-batch-constrained, Bayesian-optimization-driven capabilities are well-poised to overcome the experimental challenges associated with using an AI scientist for high-throughput synthetic biology optimization ([Maximilian Siska et al., 2026](#)).