nature portfolio

Peer Review File

An Autonomous Lab for Data-Driven Homogeneous Catalysis

Corresponding Author: Professor Milad Abolhasani

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
This manuscript presents a well executed and technically sophisticated autonomous catalysis platform for Rh catalyzed hydroformylation, integrating batch high throughput experimentation with hierarchical Bayesian optimization, and the authors should be commended for clearly addressing several points raised in prior reviews by more explicitly differentiating this work from their earlier flow based studies and by introducing a plate constrained decision framework that reflects real experimental limitations . In particular, the transition from continuous variable optimization in flow to mixed discrete and continuous exploration in a batch format, along with the explicit treatment of shared reactor setpoints, represents a sensible and practically relevant evolution of their platform.

However, despite these improvements, the manuscript remains largely an incremental extension of previously published work, with the same reaction system, similar optimization philosophy, and comparable outputs centered on Pareto front mapping and ligand condition relationships, such that the core conceptual advance is limited. The study is further constrained by its reliance on a relatively small predefined ligand set and a substantial initialization dataset, which together position the work closer to structured optimization within a bounded space rather than true autonomous discovery, and the resulting chemical insights remain largely consistent with established steric and electronic trends rather than revealing fundamentally new mechanistic understanding .

Additional limitations include the lack of demonstrated generalization beyond the explored ligand space, the dependence on post hoc descriptor analysis, and the observed drop in performance upon scale up which underscores unresolved transport and mixing effects that are not deconvoluted in the current framework . Overall, while the platform is robust and thoughtfully developed, and the revisions have improved clarity and positioning, the work reads primarily as an engineering refinement of an existing paradigm rather than a step change in autonomous catalysis or catalyst discovery, and this distinction should be more clearly articulated in assessing its suitability for a broad audience.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
Due to a conflict with the original Referee 2, I was asked to step in given my similar expertise in high-throughput experimentation (HTE) and automation for organic synthesis. In the review below, I primarily address the authors' point-by-point responses to reviewer #2, along with an evaluation of the work as a whole.

General comments to the authors
Miller, Abolhasani, and co-workers describe the development of a system for autonomous optimization of high-pressure, high-temperature gas–liquid reactions. Although some aspects of the implementation would benefit from additional clarity, the overall workflow appears to integrate commercial automation hardware and software (e.g., Chemspeed, AutoSuite, ArkSuite) with Python based tools for modeling and batched Bayesian optimization.

The authors use the industrially important Rh-catalyzed hydroformylation of propylene as a case study, demonstrating that the end-to-end workflow can efficiently identify conditions that favor linear selectivity, branched selectivity, or operationally flexible conditions for either outcome. To address practical constraints associated with high-throughput experimentation (mixed continuous and discrete variables; plate-based constraints), the authors developed a batch Bayesian optimization strategy tailored to this setting. The results are presented clearly and provide convincing validation of the methodology; however, the use of batched Bayesian optimization reads more as an extension of established approaches than a fundamentally new scientific advance, which raises the question of whether the work is best suited for Nature Communications or for a more specialized journal. In addition, the scale-up demonstration is limited (10×) and does not address mass-transfer characterization (e.g., kLa) or transport–kinetic deconvolution, both of which would likely be important for industrial translation and for establishing broader generality.
Response to Referee 2: general comments
Based on my reading of the revised manuscript, the authors more clearly articulate the paper's central contribution: constraint-aware Bayesian optimization of hydroformylation under different objective functions. The identification of a single, operationally flexible set of conditions that can be tuned to favor either branched or linear selectivity is intriguing. That said, it would be helpful to understand whether performance could be further improved by incorporating descriptor-informed parameters (e.g., catalyst/ligand or substrate descriptors) into the search space. In addition, relative to other reports of autonomous optimization campaigns, the stated experimental budget (320 initialization experiments followed by up to 360 autonomous runs) seems high. Would a comparable number of experiments be required when adapting the platform to a new reaction system? More generally, a demonstration that the approach can identify broadly applicable conditions across multiple substrates (similar to bandit-optimization) or scale would strengthen the case for generality and impact.

Comment 1
The authors clearly define what is "human-initiated" versus computer-controlled in the design and execution of the experiments. The rationale for the batch size used in each run, and how it was determined, is also communicated clearly. I would still appreciate additional detail in the Supporting Information to aid interpretation of the relevant plots (e.g., more explicit discussion and references). The comparison of constraint-aware Bayesian optimization with random sampling and Latin hypercube sampling provides clear motivation for the selected approach. While I am generally hesitant to suggest additional experiments in a future publication it would be informative to include a benchmark of batch Bayesian optimization against other batch acquisition strategies.

Comment 2
The authors have improved the mechanistic section by making it less speculative and positioning it more clearly as a complement to the data-rich study. The added text and the reworked figure help highlight potentially informative ligand properties for design guidance. However, while the authors note that "extending the framework to descriptor-informed priors or out-of-library generalization is a natural future direction," this leaves an important open question: would the current model have performed better had such descriptors been included from the outset?

Comment 3
The authors do not directly address the original concerns regarding scale-up, but instead revise the manuscript to include a limitations statement. For this approach to support more generalizable scale-up in the future, measurement and understanding of kLa (and, more broadly, mass-transfer effects) would be important to incorporate. This limitation has now been appropriately noted by the authors.

(Remarks on code availability)
I am not familair with "zenodo" but it does appear to have all the relevant information in a structured format. I found the readme to be informative but didn't have time to fully review each part of the code.

I will not the github page at the bottom (https://github.com/AbolhasaniLab/Flex-Cat) was empty. [This could be intentional]

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Reviewer #3

General Comments to the Author:

This manuscript presents a well executed and technically sophisticated autonomous catalysis platform for Rh catalyzed hydroformylation, integrating batch high throughput experimentation with hierarchical Bayesian optimization, and the authors should be commended for clearly addressing several points raised in prior reviews by more explicitly differentiating this work from their earlier flow based studies and by introducing a plate constrained decision framework that reflects real experimental limitations . In particular, the transition from continuous variable optimization in flow to mixed discrete and continuous exploration in a batch format, along with the explicit treatment of shared reactor setpoints, represents a sensible and practically relevant evolution of their platform. However, despite these improvements, the manuscript remains largely an incremental extension of previously published work, with the same reaction system, similar optimization philosophy, and comparable outputs centered on Pareto front mapping and ligand condition relationships, such that the core conceptual advance is limited. The study is further constrained by its reliance on a relatively small predefined ligand set and a substantial initialization dataset, which together position the work closer to structured optimization within a bounded space rather than true autonomous discovery, and the resulting chemical insights remain largely consistent with established steric and electronic trends rather than revealing fundamentally new mechanistic understanding. Additional limitations include the lack of demonstrated generalization beyond the explored ligand space, the dependence on post hoc descriptor analysis, and the observed drop in performance upon scale up which underscores unresolved transport and mixing effects that are not deconvoluted in the current framework. Overall, while the platform is robust and thoughtfully developed, and the revisions have improved clarity and positioning, the work reads primarily as an engineering refinement of an existing paradigm rather than a step change in autonomous catalysis or catalyst discovery, and this distinction should be more clearly articulated in assessing its suitability for a broad audience.

Response: We thank the reviewer for the thoughtful evaluation and for recognizing the technical sophistication of Flex-Cat and the improvements made in distinguishing this work from our prior flow-based studies. We respectfully note that the central contribution of this manuscript is not intended to be the discovery of a fundamentally new hydroformylation mechanism or an unconstrained search across all possible ligand structures. Rather, the key advance is the integration of high-pressure gas–liquid batch automation with a hierarchical, plate-constrained Bayesian optimization framework that explicitly accounts for shared reactor setpoints and mixed discrete/continuous variables. This combination addresses a practical and important limitation in autonomous catalysis and HTE: the fact that many parallel reactor platforms do not allow each well to independently vary temperature, pressure, or gas composition.

To avoid overstatement, we have added a clarification to the revised manuscript that Flex-Cat performs autonomous ligand–condition mapping within a predefined safety-bounded design space, rather than open-ended catalyst discovery across an unlimited ligand universe.

> **Main Text, Page 10, paragraph 5, highlighted in yellow:** "Thus, the autonomy demonstrated here should be interpreted as closed-loop optimization and ligand–condition mapping within a predefined, safety-bounded design space, rather than unconstrained discovery across an open-ended ligand universe."

The reviewer also notes the finite ligand library and the post hoc nature of the descriptor analysis. This is a fair boundary of the present study, but it does not diminish the value of the data-rich autonomous optimization campaign. The ligand set was intentionally selected to span a chemically meaningful phosphorus-ligand family, and descriptors were used only after the campaigns to interpret experimentally observed trends. We have clarified this point in the revised manuscript:

> **Main Text, Page 14, paragraph 2, highlighted in yellow:** "Since ligand descriptors were not prospectively included in the closed-loop decision policy, determining whether descriptor-informed priors would have improved sample efficiency or final performance would require a separate detailed study."

Regarding scale-up, the 20 mL experiments were included as an initial automated protocol-transfer demonstration across an order of magnitude in reaction volume within the same platform infrastructure. A full transport and scale-up study, including kLa measurement and transport–kinetic deconvolution, would require a dedicated investigation, which we have already clarified in the manuscript.

Reviewer #4

Due to a conflict with the original Referee 2, I was asked to step in given my similar expertise in high-throughput experimentation (HTE) and automation for organic synthesis. In the review below, I primarily address the authors' point-by-point responses to reviewer #2, along with an evaluation of the work as a whole.

General Comments to the Authors:

Miller, Abolhasani, and co-workers describe the development of a system for autonomous optimization of high-pressure, high-temperature gas–liquid reactions. Although some aspects of the implementation would benefit from additional clarity, the overall workflow appears to integrate commercial automation hardware and software (e.g., Chemspeed, AutoSuite, ArkSuite) with Python based tools for modeling and batched Bayesian optimization. The authors use the industrially important Rh-catalyzed hydroformylation of propylene as a case study, demonstrating that the end-to-end workflow can efficiently identify conditions that favor linear selectivity, branched selectivity, or operationally flexible conditions for either outcome. To address practical constraints associated with high-throughput experimentation (mixed continuous and discrete variables; plate-based constraints), the authors developed a batch Bayesian optimization strategy tailored to this setting. The results are presented clearly and provide convincing validation of the methodology; however, the use of batched Bayesian optimization reads more as an extension of established approaches than a fundamentally new scientific

advance, which raises the question of whether the work is best suited for Nature Communications or for a more specialized journal. In addition, the scale-up demonstration is limited (10×) and does not address mass-transfer characterization (e.g., kLa) or transport–kinetic deconvolution, both of which would likely be important for industrial translation and for establishing broader generality. Response to Referee 2: general comments. Based on my reading of the revised manuscript, the authors more clearly articulate the paper's central contribution: constraint-aware Bayesian optimization of hydroformylation under different objective functions. The identification of a single, operationally flexible set of conditions that can be tuned to favor either branched or linear selectivity is intriguing. That said, it would be helpful to understand whether performance could be further improved by incorporating descriptor-informed parameters (e.g., catalyst/ligand or substrate descriptors) into the search space. In addition, relative to other reports of autonomous optimization campaigns, the stated experimental budget (320 initialization experiments followed by up to 360 autonomous runs) seems high. Would a comparable number of experiments be required when adapting the platform to a new reaction system? More generally, a demonstration that the approach can identify broadly applicable conditions across multiple substrates (similar to bandit-optimization) or scale would strengthen the case for generality and impact.

Response: We thank the reviewer for evaluating the revised manuscript and for recognizing the improved framing of the central contribution. We respectfully emphasize that the novelty of Flex-Cat is not the use of Bayesian optimization (BO) alone, but its physically constrained implementation in a high-pressure gas–liquid HTE setting where plate-level reactor parameters are shared across wells while ligand identity and formulation variables are selected at the well level. This constraint-aware decision structure is central to the experimental reality of the platform and is distinct from standard independent-condition BO implementations.

The reviewer asked whether the relatively large initialization budget would be required for new reaction systems. The 320-experiment initialization was selected for this specific demonstration because the study combined 16 ligands, multiple shared reactor variables, and no descriptor-informed priors. It should not be interpreted as a universal requirement. We have clarified this in the revised Discussion.

> **Main Text, Page 16, paragraph 2, highlighted in yellow:** "The 320-experiment initialization reflects the combination of a 16-ligand library, plate-coupled reactor variables, and the deliberate exclusion of descriptor-informed priors in this study. The LHS size selected in this study should not be interpreted as a fixed requirement for adapting Flex-Cat to other reaction systems, where smaller libraries, narrower operating windows, transfer learning, or descriptor-informed priors could reduce the initialization burden."

Comment 1: *The authors clearly define what is "human-initiated" versus computer-controlled in the design and execution of the experiments. The rationale for the batch size used in each run, and how it was determined, is also communicated clearly. I would still appreciate additional detail in the Supporting Information to aid interpretation of the relevant plots (e.g., more explicit discussion and references). The comparison of constraint-aware Bayesian optimization with random sampling and Latin hypercube*

sampling provides clear motivation for the selected approach. While I am generally hesitant to suggest additional experiments in a future publication it would be informative to include a benchmark of batch Bayesian optimization against other batch acquisition strategies.

Response: We thank the reviewer for this comment. The Supporting Information (SI) document already benchmarks the selected hierarchical BO strategy against random sampling, LHS, and a linear-objective BO variant under matched virtual budgets. A comprehensive benchmark against all possible batch acquisition functions would be useful but is more appropriate for a separate methodological study. We have added a short statement to clarify this scope.

> **SI, Page 10, paragraph 3, highlighted in yellow:** "A broader benchmark against additional batch acquisition strategies, including qEI/qNEI, Thompson sampling, qUCB, and hypervolume-based multi-objective policies, is outside the scope of the present experimental study but is enabled by the released modular codebase."

Comment 2: *The authors have improved the mechanistic section by making it less speculative and positioning it more clearly as a complement to the data-rich study. The added text and the reworked figure help highlight potentially informative ligand properties for design guidance. However, while the authors note that "extending the framework to descriptor-informed priors or out-of-library generalization is a natural future direction," this leaves an important open question: would the current model have performed better had such descriptors been included from the outset?*

Response: We thank the reviewer for this comment. Descriptors were intentionally not included in the closed-loop policy so that the autonomous campaign was driven by Flex-Cat-generated experimental measurements rather than ligand-prior assumptions. Testing whether descriptor-informed priors would improve the campaign would require a controlled ablation or prospective comparison. We have clarified this point in the revised manuscript:

> **Main Text, Page 14, paragraph 2, highlighted in yellow:** "Since ligand descriptors were not prospectively included in the closed-loop decision policy, determining whether descriptor-informed priors would have improved sample efficiency or final performance would require a separate detailed study."

Comment 3: *The authors do not directly address the original concerns regarding scale-up, but instead revise the manuscript to include a limitations statement. For this approach to support more generalizable scale-up in the future, measurement and understanding of kLa (and, more broadly, mass-transfer effects) would be important to incorporate. This limitation has now been appropriately noted by the authors.*

Response: We thank the reviewer for this comment. The 20 mL experiments were included as an initial automated protocol-transfer demonstration across an order of magnitude in reaction volume within the same platform infrastructure. A full transport and scale-up study, including k_La measurement and

transport–kinetic deconvolution, would require a dedicated investigation, which we have already clarified in the manuscript.

(Remarks on code availability): *I am not familiar with "zenodo" but it does appear to have all the relevant information in a structured format. I found the readme to be informative but didn't have time to fully review each part of the code.*
Response: We thank the reviewer for reviewing Flex-Cat's Zenodo documentation.