Future-Proofing Data for Catalyst Fingerprinting and Better Design Validation

Wesley McNutt, Jacob Jessiman, Kashel Judin, Leo Eggar, Jason Hein
Department of Chemistry, University of British Columbia, Vancouver, Canada

Correspondence to: [Wesley McNutt] wesley.mcnutt@ubc.ca

1. Introduction

The development and implementation of catalytic systems for organic synthesis have revolutionized the way useful molecules are prepared through improved route efficiency or access to novel motifs. Catalyst development has often followed a linear sequence where leveraged mechanistic understanding and iteration is used to design, build, and screen catalysts against a given reaction based on prior knowledge and experience.[1] The potential benefits of *in silico* approaches with the ability to build and screen novel catalysts could greatly expedite this labour-intensive process and save considerable resources for reaction invention and optimization (fig.1a). Furthermore, this process could also be improved by the development and application of artificial intelligence to further enhance the rate and volume of chemical space explored.

Regardless, high-quality datasets of actual reactions are required to invent novel catalysts and validate these designed materials meet the intended/expected outcomes once identified. Unfortunately, these dynamic catalytic systems are often complex and as a result are often simplified to single-point outcomes (like yield, or turnover number (TON)) which limits their fitness (fig. 1b) for deeper analysis. This reduction of catalytic reaction networks to a single time-dependent value at a chosen time greatly simplifies analysis and reduces the burden of data acquisition, but severely reduces the completeness of data collected. Modern advances in reaction monitoring techniques and publications featuring the utility of detailed timecourse data calls for a shift in how reaction optimization and catalyst design is conducted and reported.[2]

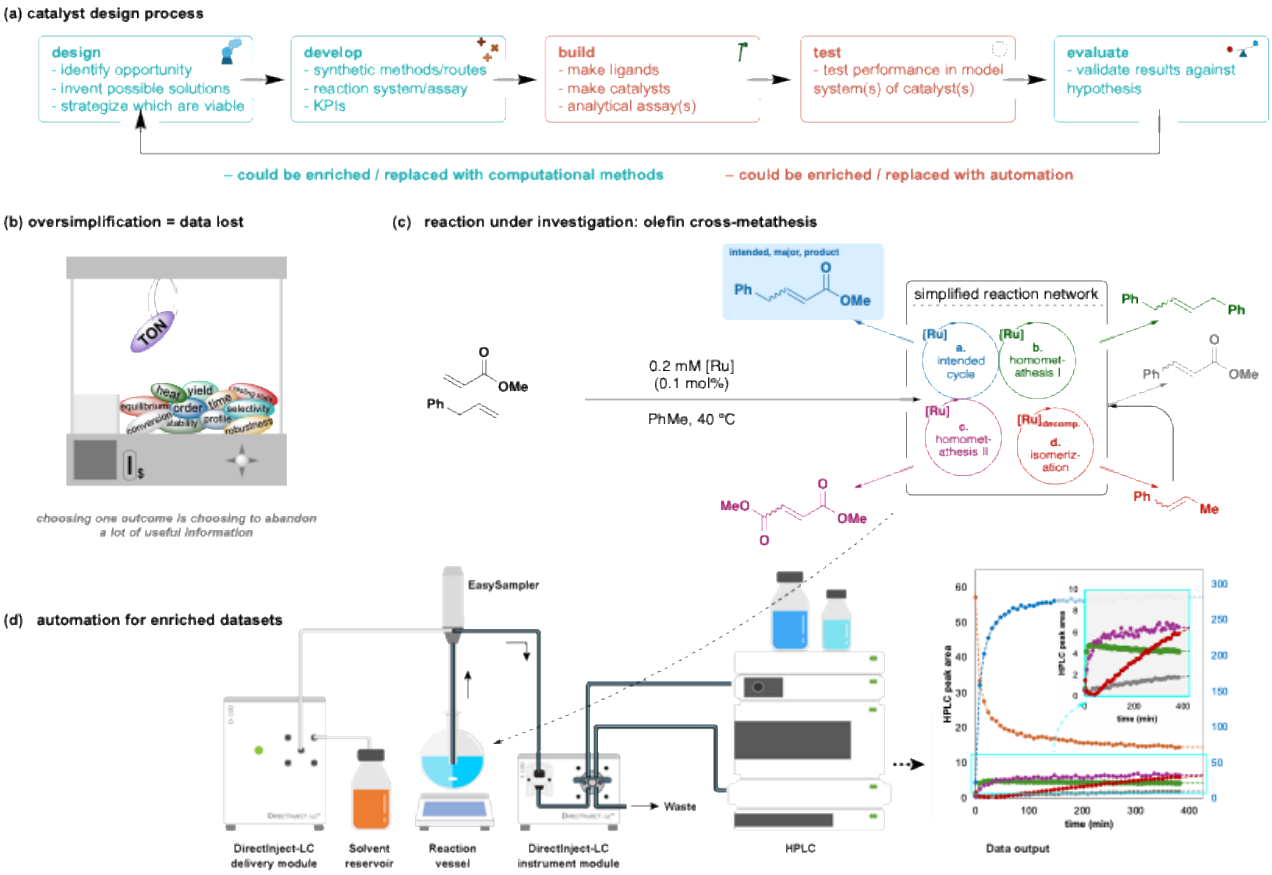

Fig.1: (a) Typical example of catalyst development process. (b) Simplified reaction network for model system in this study and (c) visualization of experimental setup (trend colours correspond to structures in (c)).

3. This Work

Previous work on catalyst development in this context is extensive and a number of excellent reviews describe previous reaction systems.[3] Standardized benchmarking strategy has been proposed, but has had limited uptake by the community at large.[4] With the increasing relevance of computational design methods we hope previous perspectives[5] and our work here can inspire others in the community to move towards standardized benchmarking and an emphasis on improved data collection and reporting beyond monodimensional oversimplifications of chemical reactions.

This initial disclosure seeks to leverage contemporary techniques in automated on-line reaction monitoring to generate more complete temporal datasets that can provide a fingerprint of a given catalysts' impact within a given reaction network (fig.1c–d). With intentional selection of reaction system and experimental conditions we are hoping to standardize benchmarking approaches for Ru-catalysed olefin metathesis catalysts, while also generating more granular datasets that could be useful for validation of *in silico* design of *de novo* catalysts. Aside from the specific utility of this class of catalytic reactions, the general philosophy of standardizing catalyst benchmarking and improving the granularity and quality of datasets should be treated as a necessity in order to better design/validate catalytic processes now, as well as create more future-proof datasets as AI becomes more integral to the process.

Our model system was chosen due to its industrial and academic relevance, as well as the complexity of the reaction network involved which includes an array of possible side-products (fig.2a). The degree and distribution of side-products relates to the catalysts activity within the reaction network as well as serving as an indicator of catalyst decomposition. An example comparison of two similar reaction conditions demonstrates that these chosen conditions behave very differently with respect to product distribution over time, but judicious choice of reaction endpoint and reporting of a single value could suggest that they perform comparably (fig.2b). For example, if conversion or yield (or TON) are taken at different times one may come to the conclusion that: Condition A (filled circles) is better at an early time, that they are the same at *ca* 250 minutes, or given that Condition B (hollow circles) is still progressing that may give a "better" result if the reaction was run for longer. The impurity profile is also different for the two trials and may or may not be significant depending on other factors (*i.e.* toxicity, or cost of removal). Further examples employing this approach will highlight how choosing a single metric can be risky and misleading while also limiting choice of goals to an inflexible scoring strategy.

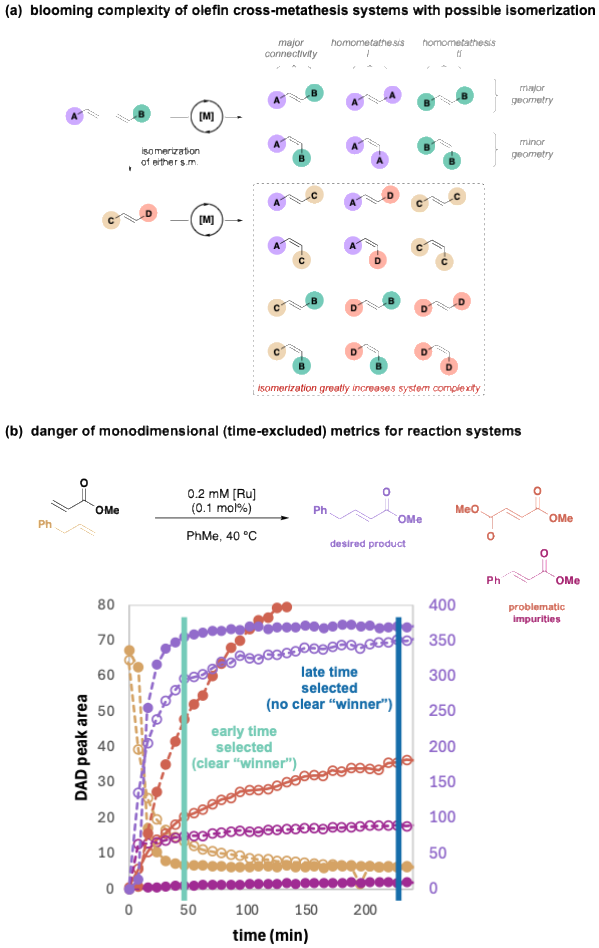

Fig.2: (a) Simple demonstration of product distribution complexity for olefin cross-metathesis and (b) timecourse comparison of two different reaction conditions.

Acknowledgments

Financial support for this work was provided by the Research Council of Norway (Project 331967), Canada Foundation for Innovation (CFI-35833, CFI-44843), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2021-03168, Discovery Accelerator), and The University of British Columbia. This research was undertaken thanks in part to funding provided to the University of Toronto's Acceleration Consortium from the Canada First Research Excellence Fund (CFREF-2022-00042).

References

[1] S. N. Alektiar, Z. K. Wickens, *Chem* **2023**, *9*, 1352–1354.

[2] (a) J. Burés, *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 16084–16087; (b) J. Bures, I. Larrosa, *Nature* **2023**, *613*, 689-695.

[3] (a) L. Monsigny, A. Kajetanowicz, K. Grela, *Chem. Rec.* **2021**, *21*, 3648-3661; (b) A. Brotons-Rufes, S. Posada-Pérez, J. P. Martínez, S. P. Nolan, A. Poater, *Coord. Chem. Rev.* **2025**, *542*.

[4] A. H. T. Ritter, A. G. Wenzel, T. W. Funk, R. H. Grubbs, *Organomet.* **2006**, *25*, 5740–5745.

[5] T. Bligaard, R. M. Bullock, C. T. Campbell, J. G. Chen, B. C. Gates, R. J. Gorte, C. W. Jones, W. D. Jones, J. R. Kitchin, S. L. Scott, *ACS Catalysis* **2016**, *6*, 2590-2602.