REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
29-12-2021	Final Report	17-Feb-2016 - 17-Feb-2021

4. TITLE AND SUBTITLE
Final Report: Automated System for Knowledge-based Continuous Organic Synthesis (ASKCOS)

5a. CONTRACT NUMBER
W911NF-16-2-0023

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Massachusetts Institute of Technology (MIT
77 Massachusetts Avenue
NE18-901
Cambridge, MA 02139 -4307

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
68395-CH-DRP.45

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Klavs Jensen
UU	UU	UU	UU		19b. TELEPHONE NUMBER 617-253-4589

RPPR Final Report
as of 11-Jan-2022

Agency Code: 21XD

Proposal Number: 68395CHDRP **Agreement Number: W911NF-16-2-0023**
INVESTIGATOR(S):

Name: Klavs F. Jensen
Email: kfjensen@mit.edu
Phone Number: 6172534589
Principal: Y

Organization: **Massachusetts Institute of Technology (MIT)**
Address: 77 Massachusetts Avenue, Cambridge, MA 021394307
Country: USA

DUNS Number: 001425594 EIN: 042103594
Report Date: 17-May-2021 Date Received: 29-Dec-2021
Final Report for Period Beginning 17-Feb-2016 and Ending 17-Feb-2021
Title: Automated System for Knowledge-based Continuous Organic Synthesis (ASKCOS)
Begin Performance Period: 17-Feb-2016 **End Performance Period:** 17-Feb-2021
Report Term: 0-Other

Submitted By: Klavs Jensen Email: kfjensen@mit.edu
Phone: (617) 253-4589

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 18 **STEM Participants:** 45

Major Goals: Our goal is to create a fully automated system that will rapidly and efficiently produce any specified organic target molecule – an extraordinary challenge considering past practice. Currently, syntheses are often unreliable and require reoptimization or redevelopment at each phase of scale-up. The challenge is to automate the design of an efficient synthetic route to a target molecule, specified in sufficient detail that an automated reactor system can carry it out.

Our approach builds on recent advances in machine learning, cheminformatics, and computational chemistry to identify promising synthetic routes to arbitrary target molecules. Machine learning can effectively utilize the large corpus of known reactions to infer essential information about new (unreported) reactions such as optimal conditions and likely yields. Such predictive methods are fast and robust. In addition to knowledge extracted from literature reactions, we will apply computational chemistry approaches to provide increasingly accurate predictions of reaction yields and physical properties (e.g., molecular solubility). Such detailed knowledge of each individual reaction will be combined to identify optimal multistep syntheses. These syntheses will then be evaluated and optimized experimentally generating new data to feedback and refine the synthesis planning process.

Combining chemistry and computer science will lead to transformative advances, as already seen in the biological sciences. In part, we will be automating and fully integrating steps currently done by humans, but we will also do tasks impossible even for the most efficient team of human experts. We will teach a computer to use millions of literature syntheses, systematically find commonalities and differences among different reactions, and then use these results to make inferences about new reactions. The computer will identify a small number of promising reactions out of a vast set of candidate reactions, and apply computational chemistry to further guide machine learning predictions. This process is likely to discover more efficient synthetic routes than those chosen by a human organic chemist constrained by limitations of equipment and experience to choose from a relatively small set of possible reactions. Establishment of a useful intersection between machine learning and computational chemistry lies at the core of our proposal.

In order to realize the identified reaction pathway, we aim to build a fully automated, reconfigurable multistep synthesis and purification system based on our extensive prior experience. This will also function as a highly compact continuous manufacturing system that combines both synthesis and final drug production. The use of continuous flow within the system enables efficient heat and mass transfer as well as process intensification.

The Make-it system hardware will be designed for flexibility to accommodate diverse chemical syntheses identified through data and knowledge based computational techniques. For flexibility, ease of operation, and maintenance,

the proposed system will have a number of bays into which we insert “generalized process units” (GPUs). The GPU can be an individual unit operation, such as a reactor, separator, or absorption column, or it can include a higher level assembly containing an integrated set of basic unit operations to carry out a specific synthesis function, e.g. alkylation, hydrogenation, reduction, condensation. An automated fluid rerouting system will address different bays.

The feedback control system will automatically reconfigure itself to match the system configuration. The system will include on-line process analytical techniques (PAT) such as Fourier Transform Infrared (FTIR), Near Infrared (NIR), and Ultraviolet-Visible (UV-Vis) spectroscopic methods along with on-line high performance liquid chromatography / mass spectroscopy (HPLC/MS). Advanced model-based optimal design and control procedures that consider model uncertainties will maximize production rate while minimizing startup and changeover time from synthesis of one target to another.

Accomplishments: Please see uploaded pdf document.

Training Opportunities: Nothing to Report

Results Dissemination: The results of the research has been disseminated through conferences presentations at major national and international conferences, multiple publications in peer reviewed journals (see products),PhD theses, and regular meetings with DARPA and related government organizations, e.g. representatives from ARL.

In addition, we participated in a DARPA organized Industry Day October 3, 2019 where results was presented to major companies.

The machine learning efforts form the basis for the formation of an industrial consortium at MIT, Machine Learning for Pharmaceutical Discovery and Synthesis (mlpds.mit.edu)

RPPR Final Report
as of 11-Jan-2022

Honors and Awards: Klavs Jensen received the Founders Award from the AICHE and was elected to the National Academy of Sciences.

Klavs F. Jensen was elected to the National Academy of Sciences.

Klavs F. Jensen received the inaugural Corning International Prize for Outstanding Work in Continuous-Flow Reactors and Chemistry for a Greener and Safer World.

Klavs F. Jensen was selected as the John Prausnitz American Institute of Chemical Engineers Institute Lecturer.

Richard D. Braatz was elected Fellow of the American Institute of Chemical Engineers (AIChE), road terms, for his research in systems and control theory and its application to classes of chemical processes including pharmaceutical manufacturing.

Richard D. Braatz was selected as the President of the American Automatic Control Council, which is a federation of nine professional societies (AIChE, AIAA, ASCE, ASME, IEEE, ISA, SCS, SIAM, and APS) that manages U.S. control activities including the organization of the annual American Control Conference.

Richard Braatz was elected to the U.S. National Academy of Engineering for “contributions to diagnosis and control of large-scale and molecular processes for materials, microelectronics and pharmaceuticals manufacturing.”

Richard Braatz was selected to receive the 2019 American Institute of Chemical Engineers’ Separations Division Innovation Award for research on the design and control of pharmaceutical crystallization processes. He was invited to give a talk on his results in the Separations Division Plenary Session at the AIChE Annual Meeting in Orlando, Florida,

Richard D. Braatz received the AIChE Separations Division Innovation Award which recognizes “outstanding contributions to scientific, technological, or industrial areas involving separations technologies.”

Frank Gupton was part of a team winning ACS Award for Affordable Green Chemistry for the synthesis of Nevirapine.

Frank Gupton Peter J. Dunn Award for Green Chemistry & Engineering Impact in the Pharmaceutical Industry.

Connor W. Coley was named DARPA Riser and one of ACS Chemical and Engineering News Talented 12. A highly unusual honor for a graduate student.

Wengong Jin received the Ho-ching and Han-ching Fund Award.

Shankul Vartak received the AICHE Separations Division Student Award in Crystallization.

Protocol Activity Status:

Technology Transfer: System for synthesis of precursors to energetic materials completed and transferred to Dr. Jesse Sabatini at ARL.

Dr. Brian Barnes (ARL) use ASKCOS to design precursors to energetic molecules.

AKSCOS enabled the formation of the MIT Consortium Machine Learning in Pharmaceutical Discovery and Synthesis. The consortium aims to develop modern machine learning tools to enhance the work of medicinal and process chemists, including new approaches to synthesis planning and structure-activity/property relationships. It brings together scientists and engineers from 14 major pharmaceutical companies with a cross-disciplinary team of MIT researchers (mlpds.mit.edu)

PARTICIPANTS:

Participant Type: PD/PI

Participant: Klavs F. Jensen
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: Y

Participant Type: Co PD/PI
Participant: Regina Barzilay
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Kyle Bishop
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Richard D Braatz
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: William H Green
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Frank B Gupton
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Tommi S Jaakkola
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Timothy F. Jamison
Person Months Worked: 3.00 **Funding Support:**

Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Allan S Myerson
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: Rachel Beingessner
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: Andrea Adamo
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: John Schreck
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Phalgun Lolur
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yuran Wang
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Robert W Hicklin
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Ashley Longstreet
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Anne C Bedard
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yi Shen
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Naomi Briggs
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Ridade Sayin
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Anastasia Nikolakopoulou
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Matthias E Freiherr von Adrian-Werburg
Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

Participant Type: Graduate Student (research assistant)
Participant: Connor Coley
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Liam P Kelly
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Dale A Thomas
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Leia Dwyer
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Brennan Lee
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Tim Plump
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Jenson Vergese
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)

Participant: Caleb Kong
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Saeed Ahmad
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Swara Fadnis
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yiming Wan
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Eramda Harinath
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Hanyu Gao
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Benson Chen
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yunsie Chung
Person Months Worked: 5.00 **Funding Support:**

Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Wengong Jin
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Phillip Tran
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: John Poitti
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Joshua Fishman
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Suzanne O'Meara
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Michelle Chen
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Joshua Byington
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Other Professional
Participant: Hideki Moriguchi
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Joshua Britton
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Justin Alexander MacDonald Lummiss
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Pedro Garcia Barrantes
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Jonathan Jaworski
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Andre Y Bessette
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Laurel M Heckman
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

Participant Type: Graduate Student (research assistant)
Participant: Mary G Russell
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Lucrece Nicoud
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Bernard F Gupton
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Bernard F Gupton
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Bernard F Gupton
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Perrer Nounagnon Tossa
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Sudha Korwar
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)

Participant: Eric Chris Yu
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Eliseu Ortega De Oliveria
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Nakul Sudhir Telang
Person Months Worked: 11.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Gerard C Mendez
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jennifer Schall
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Shankul Vartak
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Gregory Hammersmith
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Other Professional
Participant: Perrer N Tossa
Person Months Worked: 15.00 **Funding Support:**

Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Somi Amir
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Juekun Wen
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Erin Stryker
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Wai-Chung Fu
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Timothy Monos
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Staff Scientist (doctoral level)
Participant: John Stephens
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jessica Weber
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kelley Danahy
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Erica Flear
Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Aria Fodness
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Tabrez Alam
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kosisochukwu Aroh
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Lorenz Baumgartner
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Natalie Eyke
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Travis Hart
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Christina Kuhnle
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Timothy Kulesza
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yiming Mo
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Anirudh M Nambiar
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Victor Schultz
Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Thomas Struble
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student

RPPR Final Report
as of 11-Jan-2022

Participant: Kabir Ahuja
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Colin Grambow
Person Months Worked: 13.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yi-Pei Li
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Lagnajit Pattanaik
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Duminda Ranasinghe
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Xiang Fu
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Vikas Garg
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yilun Zhou
Person Months Worked: 2.00 **Funding Support:**

Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Jiang Guo
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Victor Quach
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Darsh Shah
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kyle Swanson
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Adam Fisch
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Adam Yale
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Clemence Neuror
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Nakul Telang
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Raymond Dominey
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Erin Stryker
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Benjamin Kurzban
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kumar Nambiar
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Haomiao Zhang
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Geetha Bolla
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Matthias Werberg
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Eranda Puwakkatiya
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Dongying Shen
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yiming Wan
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Paresh Malalur
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jonas Mueller
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yujie Qian
Person Months Worked: 11.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)

Participant: Antonio Santiago Ibanez Lopez
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Justin Yu
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Jason Zhao
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Michael Forsuelo
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kevin Spiekermann
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Florence Vermeire
Person Months Worked: 11.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Shangyuan Tong
Person Months Worked: 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Nithin Bududma
Person Months Worked: 2.00 **Funding Support:**

Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Dimitri Livitz
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Sarah Atassi
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Technician
Participant: Juekun Wen
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Janaka Gamaethige
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Alpana Thorat
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Wilhelm Hutzler
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Wai Chung
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Preston MacQueen
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Grace Ahlqvist
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Chris Breen
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Katie McGeough
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Sarah Mear
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Corshai Williams
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Prajwal Tumkur Mahesh
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

RPPR Final Report
as of 11-Jan-2022

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Bioorganic & Medicinal Chemistry
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.bmc.2017.02.002
Volume: Issue: First Page #:
Date Submitted: 8/25/17 12:00AM Date Published: 2/5/17 3:42AM
Publication Location:
Article Title: Minimizing E-factor in the continuous-flow synthesis of diazepam and atropine
Authors: Anne-Catherine Bédard, Ashley R. Longstreet, Joshua Britton, Yuran Wang, Hideki Moriguchi, Robert W
Keywords: flow chemistry, synthesis, E-factor
Abstract: Minimizing the waste stream associated with the synthesis of active pharmaceutical ingredients (APIs) and commodity chemicals is of high interest within the chemical industry from an economic and environmental perspective. In exploring solutions to this area, we herein report a highly optimized and environmentally conscious continuous-flow synthesis of two APIs identified as essential medicines by the World Health Organization, namely diazepam and atropine. Notably, these approaches significantly reduced the E-factor of previously published routes through the combination of flow chemistry techniques, computational calculations and solvent minimization. The E-factor associated with the synthesis of atropine was reduced by 94-fold (about two orders of magnitude), from 2245 to 24, while the E-factor for the synthesis of diazepam was reduced by 4-fold, from 36 to 9.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: ACS Central Science
Publication Identifier Type: DOI Publication Identifier: 10.1021/acscentsci.7b00064
Volume: 3 Issue: 5 First Page #: 434
Date Submitted: 6/13/17 12:00AM Date Published: 4/18/17 12:00PM
Publication Location: Washington, DC
Article Title: Prediction of organic reaction outcomes using machine learning
Authors: Connor Coley, Regina Barzilay, William Green, Tommi Jaakkol, and Klavs Jensen
Keywords: machine learning, retrosynthesis, organic chemistry, synthesis
Abstract: Computer assistance in synthesis design has existed for over 40 years, yet retrosynthesis planning software has struggled to achieve widespread adoption. One critical challenge in developing high-quality pathway suggestions is that proposed reaction steps often fail when attempted in the laboratory, despite initially seeming viable. The true measure of success for any synthesis program is whether the predicted outcome matches what is observed experimentally. We report a model framework for anticipating reaction outcomes that combines the traditional use of reaction templates with the flexibility in pattern recognition afforded by neural networks. Using 15 000 experimental reaction records from granted United States patents, a model is trained to select the major (recorded) product by ranking a self-generated list of candidates where one candidate is known to be the major product.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.6b00601
Volume: Issue: First Page #:
Date Submitted: 8/25/17 12:00AM Date Published: 7/1/17 4:00AM
Publication Location:
Article Title: Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction
Authors: Connor W. Coley, Regina Barzilay, William H. Green, Tommi S. Jaakkola, Klavs F. Jensen
Keywords: Convolutional Embedding, Physical Property Prediction, Machine Learning, Neural Networks
Abstract: Learning an expressive molecular representation is central to developing quantitative structure–activity and property relationships. Traditional approaches rely on group additivity rules, empirical measurements or parameters, or generation of thousands of descriptors. In this paper, we employ a convolutional neural network for this embedding task by treating molecules as undirected graphs with attributed nodes and edges. Simple atom and bond attributes are used to construct atom-specific feature vectors that take into account the local chemical environment using different neighborhood radii. By working directly with the full molecular graph, there is a greater opportunity for models to identify important features relevant to a prediction task. Unlike other graph-based approaches, our atom featurization preserves molecule-level spatial information that significantly enhances model performance. Our models learn to identify important features of atom clusters for the prediction task.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: ACS Central Science
Publication Identifier Type: DOI Publication Identifier: 10.1021/acscentsci.7b00355
Volume: Issue: First Page #:
Date Submitted: 12/11/17 12:00AM Date Published:
Publication Location:
Article Title: Computer-Assisted Retrosynthesis Based on Molecular Similarity
Authors: Connor Coley, Luke Rogers, William Green, and Klavs Jensen
Keywords: machine learning, organic synthesis, retrosynthesis
Abstract: We demonstrate molecular similarity to be a surprisingly effective metric for proposing and ranking one-step retrosynthetic disconnections based on analogy to precedent reactions. The developed approach is intended to mimic the retrosynthetic strategy defined implicitly by a corpus of known reactions. Using 40,000 reactions from the patent literature as a knowledge base, the recorded reactants are among the top 10 proposed precursors in 88.1% of 5,000 test reactions within ten common reaction classes. Extension of the one-step strategy to multi-step pathway planning is demonstrated and discussed.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Tetrahedron Symposium-in-Print
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.tet.2017.11.068
Volume: Issue: First Page #:
Date Submitted: 12/11/17 12:00AM Date Published:
Publication Location:
Article Title: Selective N-monomethylation of primary anilines with dimethyl carbonate in continuous flow
Authors: Hyowon Seo, Anne-Catherine Bédard, Willie P. Chen, Robert W. Hicklin, Alexander Alabugin, and Timo
Keywords: Monomethylation of anilines, Continuous flow chemistry, Green chemistry, Dimethyl carbonate, in situ Protection-deprotection
Abstract: Selective N-monomethylation of anilines has been achieved under continuous flow conditions using dimethyl carbonate as a green methylating agent in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene. Our methodology takes advantage of the expanded process windows available in the continuous flow regime to safely induce monomethylation in superheated solvents at high pressure. We propose selective N-monomethylation is achieved via an in situ protection-deprotection pathway, which is supported by the observed reactivities of several putative reaction intermediates. The robust and scalable method was applicable to a broad range of primary aniline substrates including ortho-, meta-, and para-substituted anilines, as well as electron-rich and electron-deficient anilines. The synthetic precursor of diazepam, 5-chloro-2-(methylamino)benzophenone, was selectively synthesized under our optimized conditions.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Synlett
Publication Identifier Type: DOI Publication Identifier: 10.1055/s-0039-1690884
Volume: 31 Issue: First Page #: A
Date Submitted: 8/29/20 12:00AM Date Published: 4/9/20 4:00AM
Publication Location: Germany
Article Title: A Continuous Flow Synthesis of Tramadol from Cyclohexanone
Authors: Timothy M. Monos, Jonathan N. Jaworski, John C. Stephens, Timothy F. Jamison
Keywords: continuous processing, flow chemistry, E-factor, Grignard reaction, Mannich bases
Abstract: A multioperation, continuous-flow platform for the synthesis of tramadol, ranging from gram to decagram quantities, is described. The platform is segmented into two halves allowing for a single operator to modulate between preparation of the intermediate by Mannich addition or complete the fully concatenated synthesis. All purification operations are incorporated in-line for the Mannich reaction. ‘Flash’ reactivity between meta-methoxyphenyl magnesium bromide and the Mannich product was controlled with a static helical mixer and tested with a combination of flow and batch-based and factorial evaluations. These efforts culminated in a rapid production rate of tramadol (13.7 g°h–1) sustained over 56 reactor volumes. A comparison of process metrics including E-Factor, production rate, and space-time yield are used to contextualize the developed platform with respect to established engineering and synthetic methods for making tramadol.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: European Journal of Organic Chemistry
Publication Identifier Type: DOI Publication Identifier: 10.1002/ejoc.201701002
Volume: 2017 Issue: 44 First Page #: 6495
Date Submitted: 8/30/18 12:00AM Date Published: 12/1/17 5:00AM
Publication Location: Weinheim, Germany
Article Title: The Application of a Continuous Grignard Reaction in the Preparation of Fluconazole
Authors: Sudha Korwar, Somi Amir, Perrer N. Tosso, Bimbisar K. Desai, Caleb J. Kong, Swara Fadnis, Nakul S.
Keywords: Continuous flow, Fluconazole, Flow, Continuous, Turbo Grignard
Abstract: The application of continuous methods in the synthesis of active pharmaceutical ingredients continues to receive significant attention in the academic as well as the industrial research communities. One of the major advantages of continuous methods is the ability to safely access kinetic synthons as well as highly reactive reagents that are typically unavailable through traditional batch methods. In this work, we report the high-yielding, clean formation of an aryl-turbo Grignard and its selective addition to a highly-enolizable 1,3-dichloroacetone, for the continuous synthesis of a key intermediate for fluconazole, a widely-prescribed anti-fungal agent. In addition, process optimization of the final API was also carried out to arrive at a semi-continuous method to this essential medicine.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Accounts of Chemical Research
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.accounts.8b00087
Volume: 51 Issue: 5 First Page #: 1281
Date Submitted: 8/31/18 12:00AM Date Published: 5/1/18 12:00AM
Publication Location: United States
Article Title: Machine Learning in Computer-Aided Synthesis Planning
Authors: Connor W. Coley, William H. Green, Klavs F. Jensen
Keywords: synthesis planning, route development, reaction prediction
Abstract: Computer-aided synthesis planning (CASP) is focused on the goal of accelerating the process by which chemists decide how to synthesize small molecule compounds. The ideal CASP program would take a molecular structure as input and output a sorted list of detailed reaction schemes that each connect that target to purchasable starting materials via a series of chemically feasible reaction steps. Early work in this field relied on expert-crafted reaction rules and heuristics to describe possible retrosynthetic disconnections and selectivity rules but suffered from incompleteness, infeasible suggestions, and human bias. With the relatively recent availability of large reaction corpora (such as the United States Patent and Trademark Office (USPTO), Reaxys, and SciFinder databases), consisting of millions of tabulated reaction examples, it is now possible to construct and validate purely data-driven approaches to synthesis planning.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.7b00622
Volume: 58 Issue: 2 First Page #: 252
Date Submitted: 8/31/18 12:00AM Date Published: 1/1/18 12:00AM
Publication Location: United States
Article Title: SCScore: Synthetic Complexity Learned from a Reaction Corpus
Authors: Connor W. Coley, Luke Rogers, William H. Green, Klavs F. Jensen
Keywords: molecular complexity, synthesis complexity, machine learning
Abstract: Several definitions of molecular complexity exist to facilitate prioritization of lead compounds, to identify diversity-inducing and complexifying reactions, and to guide retrosynthetic searches. In this work, we focus on synthetic complexity and reformalize its definition to correlate with the expected number of reaction steps required to produce a target molecule, with implicit knowledge about what compounds are reasonable starting materials. We train a neural network model on 12 million reactions from the Reaxys database to impose a pairwise inequality constraint enforcing the premise of this definition: that on average, the products of published chemical reactions should be more synthetically complex than their corresponding reactants. The learned metric (SCScore) exhibits highly desirable nonlinear behavior, particularly in recognizing increases in synthetic complexity throughout a number of linear synthetic routes.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Chemical Science
Publication Identifier Type: DOI Publication Identifier: 10.1039/c8sc04228d
Volume: 10 Issue: First Page #: 370
Date Submitted: 8/30/19 12:00AM Date Published: 11/23/18 5:00AM
Publication Location: Cambridge, UK
Article Title: A graph-convolutional neural network model for the prediction of chemical reactivity
Authors: Connor Coley, Wengong Jin, Luke Rogers, Timothy Jamison, Tommi Jaakkola, William H. Green, Regi
Keywords: neural networks, machine learning, reaction prediction, template free
Abstract: We present a supervised learning approach to predict the products of organic reactions given their reactants, reagents, and solvent(s). The prediction task is factored into two stages comparable to manual expert approaches: considering possible sites of reactivity and evaluating their relative likelihoods. By training on hundreds of thousands of reaction precedents from the patent literature, the neural model makes informed predictions of chemical reactivity. The model predicts the major product correctly over 85% of the time, a significantly higher accuracy than achieved by previous machine learning approaches, and performs on par with expert chemists with years of formal training. We gain additional insight into predictions via the design of the neural model, revealing an understanding of chemistry qualitatively consistent with manual approaches.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: ACS Central Science
Publication Identifier Type: DOI Publication Identifier: 10.1021/acscentsci.8b00357
Volume: 4 Issue: 11 First Page #: 1465
Date Submitted: 8/30/19 12:00AM Date Published:
Publication Location: United States
Article Title: Using Machine Learning to Predict Suitable Conditions for Organic Reactions
Authors: Hanyu Gao, Thomas Struble, Connor Coley, Yuran Wang, William Green, Klavs Jensen
Keywords: machine learning, reaction conditions, reaction prediction
Abstract: Reaction condition recommendation is an essential element for the realization of computer-assisted synthetic planning. Accurate suggestions of reaction conditions are required for experimental validation and can have a significant effect on the success or failure of an attempted transformation. However, de novo condition recommendation remains a challenging and under-explored problem and relies heavily on chemists' knowledge and experience. In this work, we develop a neural network model to predict the chemical context (catalyst(s), solvent(s), reagent(s), as well as temperature) most suitable for any particular organic reaction. Trained on ~10 million examples from Reaxys, the model is able to propose conditions where a close match to the recorded catalyst, solvent and reagent is found within the top-10 predictions 69.6% of the time, with top-10 accuracies for individual species reaching 80%-90%.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: J. Phys. Chem. A
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jpca.8b10789
Volume: 123 Issue: 10 First Page #: 2142
Date Submitted: 8/30/19 12:00AM Date Published: 2/13/19 5:00AM
Publication Location: United States
Article Title: Self-Evolving Machine: A Continuously Improving Model for Molecular Thermochemistry
Authors: Yi-Pei Li, Kehang Han, Colin Grambow, William Green*
Keywords: Machine Learning, Active Learning, Convolutional Neural Network, Molecular Fingerprint, Uncertainty, Bootstrap Aggregation, Dropout, Thermochemistry
Abstract: Because collecting precise and accurate chemistry data is often challenging, chemistry datasets usually only span a small region of the chemical space, which limits the performance and the scope of applicability of data-driven models. To address this issue, we integrated an active learning machine with automatic ab initio calculations to form a self-evolving model that can continuously adapt to new species appointed by the users. In the present work, we demonstrate the self-evolving concept by modeling the formation enthalpies of polycyclic species. By combining a molecular graph convolutional neural network with a dropout training strategy, the model we developed can predict enthalpies for a broad range of polycyclic species and estimate uncertainty in each enthalpy value. For the species which the current model is uncertain about, the automatic ab initio calculations provide accurate enthalpies to improve the performance of the model.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.9b00286
Volume: 59 Issue: 6 First Page #: 2529
Date Submitted: 8/30/19 12:00AM Date Published: 6/1/19 12:00AM
Publication Location: United States
Article Title: RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application
Authors: Connor W. Coley, William H. Green, Klavs F. Jensen
Keywords: RDKit, stereochemistry, SMILES, molecular representation
Abstract: There is a renewed interest in computer-aided synthesis planning, where the vast majority of approaches require the application of retrosynthetic reaction templates. Here we introduce RDChiral, an opensource Python wrapper for RDKit designed to provide consistent handling of stereochemical information in applying retrosynthetic transformations encoded as SMARTS strings. RDChiral is designed to enforce the introduction, destruction, retention, and inversion of chiral tetrahedral centers as well as the cis/trans configuration of double bonds. We also introduce an open-source implementation of a retrosynthetic template extraction algorithm to generate SMARTS patterns from atom-mapped reaction SMILES strings. In this application note, we describe the implementation of these two pieces of code and illustrate their use through many examples.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Science
Publication Identifier Type: DOI Publication Identifier: 10.1126/science.aax1566
Volume: 365 Issue: 6453 First Page #:
Date Submitted: 8/30/19 12:00AM Date Published: 8/1/19 12:00AM
Publication Location: United States
Article Title: A robotic platform for flow synthesis of organic compounds informed by AI planning
Authors: Connor W. Coley, Dale A. Thomas, Justin A. M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen,
Keywords: chemical synthesis, computer aided synthesis, robots, machine learning
Abstract: The synthesis of complex organic molecules requires several stages, from ideation to execution, that require time and effort investment from expert chemists. Here,we report a step toward a paradigm of chemical synthesis that relieves chemists from routine tasks, combining artificial intelligence–driven synthesis planning and a robotically controlled experimental platform. Synthetic routes are proposed through generalization of millions of published chemical reactions and validated in silico to maximize their likelihood of success. Additional implementation details are determined by expert chemists and recorded in reusable recipe files, which are executed by a modular continuous-flow platform that is automatically reconfigured by a robotic arm to set up the required unit operations and carry out the reaction.This strategy for computer-augmented chemical synthesis is demonstrated for 15 drug or drug-like substances.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.9b00237
Volume: 59 Issue: 8 First Page #: 3370
Date Submitted: 8/30/19 12:00AM Date Published: 7/1/19 4:00AM
Publication Location: United States
Article Title: Analyzing Learned Molecular Representations for Property Prediction
Authors: Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez
Keywords: property prediction, molecular representation, machine learning
Abstract: Advancements in neural machinery have led to a wide range of algorithmic solutions for molecular property prediction. Two classes of models in particular have yielded promising results: neural networks applied to computed molecular fingerprints or expert-crafted descriptors and graph convolutional neural networks that construct a learned molecular representation by operating on the graph structure of the molecule. However, recent literature has yet to clearly determine which of these two methods is superior when generalizing to new chemical space. Furthermore, prior research has rarely examined these new models in industry research settings in comparison to existing employed models. In this paper, we benchmark models extensively on 19 public and 16 proprietary industrial data sets spanning a wide variety of chemical end points. In addition, we introduce a graph convolutional model that consistently matches or outperforms models using fixed molecular descriptors.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Reaction Chemistry and Engineering
Publication Identifier Type: DOI Publication Identifier: 10.1039/c9re00348g
Volume: 5 Issue: First Page #: 367
Date Submitted: 8/29/20 12:00AM Date Published: 1/2/20 5:00AM
Publication Location: Cambridge, UK
Article Title: Combining retrosynthesis and mixed-integer optimization for minimizing the chemical inventory needed to realize a WHO essential medicines list
Authors: Hanyu Gao, Connor W. Coley, Thomas J. Struble, Linyan Li, Yujie Qian, William H. Green, Klavs F. Jen
Keywords: machine learning, optimization, minimization of inventory for synthesis
Abstract: The access to essential medicines remains a problem in many low-income countries for logistic and expiration limits, among other factors. Enabling flexible replenishment and easier supply chain management by on demand manufacturing from stored starting materials provides a solution to this challenge. Recent developments in computer-aided chemical synthesis planning have benefited from machine learning in different aspects. In this manuscript, we use those techniques to perform a combined analysis of a WHO essential medicines list to identify synthetic routes that minimize chemical inventory that would be required to synthesize the all the active pharmaceutical ingredients. We use a synthesis planning tool to performretrosynthetic analyses for 99 targets and solve a mixed-integer programming problem to select a combination of pathways that uses the minimal number of chemicals.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Reaction Chemistry and Engineering
Publication Identifier Type: DOI Publication Identifier: 10.1039/d0re00071j
Volume: 5 Issue: First Page #: 896
Date Submitted: 8/29/20 12:00AM Date Published: 3/31/20 4:00AM
Publication Location: Cambridge, UK
Article Title: Multitask Prediction of Site Selectivity in Aromatic C-H Functionalization Reactions
Authors: Thomas J. Struble, Connor W. Coley, and Klavs F. Jensen
Keywords: C-H functionalization, site selectivity, aromatic reactions
Abstract: Aromatic C–H functionalization reactions are an important part of the synthetic chemistry toolbox. Accurate prediction of site selectivity can be crucial for prioritizing target compounds and synthetic routes in both drug discovery and process chemistry. However, selectivity may be highly dependent on subtle electronic and steric features of the substrate. We report a generalizable approach to prediction of site selectivity that is accomplished using a graph-convolutional neural network for the multitask prediction of 123 C–H functionalization tasks. In an 80/10/10 training/validation/testing pseudo-time split of about 58000 aromatic C–H functionalization reactions from the Reaxys database, the model achieves a mean reciprocal rank of 92%. Once trained, inference requires approximately 200 ms per compound to provide quantitative likelihood scores for each task.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: ACS Central Science
Publication Identifier Type: DOI Publication Identifier: 10.1021/acscentsci.9b00055
Volume: 5 Issue: 6 First Page #: 970
Date Submitted: 8/30/19 12:00AM Date Published: 5/1/19 12:00AM
Publication Location:
Article Title: Learning Retrosynthetic Planning through Simulated Experience
Authors: John S. Schreck, Connor W. Coley, Kyle J. M. Bishop
Keywords: synthesis planning, machine learning, reinforcement learning
Abstract: The problem of retrosynthetic planning can be framed as a one-player game, in which the chemist (or a computer program) works backward from a molecular target to simpler starting materials through series of choices regarding which reactions to perform. This game is challenging as the combinatorial space of possible choices is astronomical, and the value of each choice remains uncertain until the synthesis plan is completed and its cost evaluated. Here, we address this search problem using deep reinforcement learning to identify policies that make (near) optimal reaction choices during each step of retrosynthetic planning according to a user-defined cost metric. Using a simulated experience, we train a neural network to estimate the expected synthesis cost or value of any given moleculebased on a representation of its molecular structure. We show that learned policies based on this value network can outperform a heuristic approach that favors symmetric disconnections.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 5-Submitted
Journal: Automatica.
Publication Identifier Type: Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 8/31/19 12:00AM Date Published:
Publication Location:
Article Title: Polynomial chaos-based H2-optimal output-feedback control of systems with probabilistic parametric uncertainties.
Authors: Dongying E. Shen, Yiming Wan, Sergio Lucia, Rolf Findeisen, and Richard D. Braatz
Keywords: output-feedback, rrobust control, parametric uncertainties, polynomial chaos, stochastic systems
Abstract: In this paper, H2 static and dynamic output-feedback control problems are investigated for linear time-invariant uncertain systems. It aims at minimizing the averaged H2 performance in the presence of nonlinear dependence on time-invariant probabilistic parametric uncertainties. By applying the polynomial chaos theory, the control synthesis problem is solved using a high-dimensional expanded system which characterizes stochastic state uncertainty propagation. Compared to existing polynomial chaos-based controls, the proposed approach addresses the simultaneous presence of parametric uncertainties and white noises. The effect of truncation errors due to using ♀finite-degree polynomial chaos expansions is captured by time-varying norm-bounded uncertainties, and is explicitly taken into account by adopting a guaranteed cost control approach. This feature avoids the use of high-degree polynomial chaos expansions to alleviate the destabilizing effects of expansion truncation errors.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 5-Submitted
Journal: Automatica
Publication Identifier Type: Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 8/31/19 12:00AM Date Published:
Publication Location:
Article Title: An Improved Polynomial Chaos Approach to Robust H Static Output-Feedback Control
Authors: Yiming Wan, Dongying E. Shen, Sergio Lucia, Rolf Findeisen, and Richard D. Braatz
Keywords: output-feedback control, uncertain linear time-invariant systems, polynomial chaos
Abstract: This article considers the H1 static outputfeedback control for linear time-invariant uncertain systems with polynomial dependence on probabilistic time-invariant parametric uncertainties. By applying polynomial chaos theory, the control synthesis problem is solved using a highdimensional expanded system which characterizes stochastic state uncertainty propagation. A closed-loop polynomial chaos transformation is proposed to derive the closed-loop expanded system. It explicitly accounts for the closed-loop dynamics and preserves the L2-induced gain, which results in smaller transformation errors compared to published polynomial chaos transformations. The effect of using finite-degree polynomial chaos expansions is first captured by a norm-bounded linear differential inclusion, and then addressed by formulating a robust polynomial chaos based control synthesis problem ...
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Control Engineering Practice.
Publication Identifier Type: Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 12/29/21 12:00AM Date Published:
Publication Location:
Article Title: Plantwide control of a compact modular reconfigurable system for continuous-flow pharmaceutical manufacturing
Authors: Anastasia Nikolakopoulou, Matthias von Andrian, Richard D. Braatz.
Keywords: plantwide control, modular reconfigurable system, dynamic nonlinearities, model predictive control
Abstract: This article considers the design of plantwide control for a portable modular reconfigurable system for the on-demand continuous-flow production of pharmaceuticals. The existing physical system has a regulatory control layer that is designed to control key states such as reactor temperatures at specified setpoint values, but lacks a higher level control system for the control of plantwide objectives such as overall yield and production rate. This article presents the design of a model predictive control-based plantwide control system, whose online optimization enables on-demand changing of plantwide control objectives while satisfying operational constraints. The controller was applied to a simulated plant based on the physical system. Strong dynamic nonlinearities that arise when operating the system near optimality complicate the design of a model predictive control system based on a linear process model...
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Angewandte Chemie International Edition
Publication Identifier Type: DOI Publication Identifier: 10.1002/anie.201909987
Volume: Issue: First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 6/1/20 4:00AM
Publication Location: Germany
Article Title: Autonomous Discovery in the Chemical Sciences Part I: Progress
Authors: Connor W. Coley, Natalie S. Eyke, Klavs F. Jensen
Keywords: automation, chemoinformatics, machine learning, drug discovery, materials science
Abstract: Flow chemistry devices have seen an increase as research tools in recent years. We present a modular system for benchtop continuous flow that allows user flexibility and reliability in the pursuit of challenging chemistry. An oscillating baffle reactor mold-formed in perfluoroalkoxy alkane enables handling reactions forming solids. A LabView control system with help from Omega temperature controllers allows for simple and precise regulation of reaction conditions with real-time monitoring of the temperature and pressure. A new pressurized sample collection vessel allows for inert and pressurized sampling into multiple vials. An array of case studies serves to validate the system by demonstrating the use of interchangeable reactors with in-line heating and cooling, along with the ability to handle highly corrosive and reactive chemical reagents such as white fuming HNO3.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Angewandte Chemie International Edition
Publication Identifier Type: DOI Publication Identifier: 10.1002/anie.201909989
Volume: Issue: First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 6/1/20 8:00AM
Publication Location: Germany
Article Title: Autonomous Discovery in the Chemical Sciences Part II: Outlook
Authors: Connor W. Coley, Natalie S. Eyke, Klavs F. Jensen
Keywords: automation, chemoinformatics, machine learning, drug discovery, materials science
Abstract: Flow chemistry devices have seen an increase as research tools in recent years. We present a modular system for benchtop continuous flow that allows user flexibility and reliability in the pursuit of challenging chemistry. An oscillating baffle reactor mold-formed in perfluoroalkoxy alkane enables handling reactions forming solids. A LabView control system with help from Omega temperature controllers allows for simple and precise regulation of reaction conditions with real-time monitoring of the temperature and pressure. A new pressurized sample collection vessel allows for inert and pressurized sampling into multiple vials. An array of case studies serves to validate the system by demonstrating the use of interchangeable reactors with in-line heating and cooling, along with the ability to handle highly corrosive and reactive chemical reagents such as white fuming HNO3.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Medicinal Chemistry
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jmedchem.9b02120
Volume: 63 Issue: 16 First Page #: 8667
Date Submitted: 8/30/20 12:00AM Date Published: 4/1/20 4:00AM
Publication Location: United States
Article Title: Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis
Authors: Thomas J. Struble, Juan C. Alvarez, Scott P. Brown, Milan Chytil, Justin Cisar, Renee L. DesJarlais, Ola
Keywords: Artificial Intelligence, Machine Learning, Synthesis Planning
Abstract: Artificial intelligence and machine learning have demonstrated their potential role in predictive chemistry and synthetic planning of small molecules; there are at least a few reports of companies employing in silico synthetic planning into their overall approach to accessing target molecules. A data-driven synthesis planning program is one component being developed and evaluated by the Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) consortium, comprising MIT and 13 chemical and pharmaceutical company members. Together, we wrote this perspective to share how we think predictive models can be integrated into medicinal chemistry synthesis workflows, how they are currently used within MLPDS member companies, and the outlook for this field.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.0c00403
Volume: 60 Issue: 7 First Page #: 3398
Date Submitted: 8/30/20 12:00AM Date Published: 6/1/20 4:00AM
Publication Location: United States
Article Title: Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning
Authors: Michael E. Fortunato, Connor W. Coley, Brian C. Barnes, Klavs F. Jensen
Keywords: reaction template recommendation, computer-aided synthesis planning
Abstract: This work presents efforts to augment the performance of data-driven machine learning algorithms for reaction template recommendation used in computer-aided synthesis planning software. Often, machine learning models designed to perform the task of prioritizing reaction templates or molecular transformations are focused on reporting high-accuracy metrics for the one-to-one mapping of product molecules in reaction databases to the template extracted from the recorded reaction. The availabletemplates that get selected for inclusion in these machine learning models have been previously limited to those that appearfrequently in the reaction databases and exclude potentially useful transformations. By augmenting open-access data sets of organic reactions with explicitly calculated template applicability and pretraining a template-relevance neural network on this augmented applicability data set, we report an increase in the template applicability recall.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Organic Process Research & Development
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.oprd.0c00160
Volume: Issue: First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 6/1/20 4:00AM
Publication Location: United States
Article Title: Development of a Versatile Modular Flow Chemistry Benchtop System
Authors: Travis Hart, Victor L. Schultz, Dale Thomas, Tim Kulesza, Klavs F. Jensen
Keywords: continuous flow, small volume continuous, modular, fluid dynamics, organic reaction, nitration, industrial manufacturing, energetic reactions
Abstract: Flow chemistry devices have seen an increase as research tools in recent years. We present a modular system for benchtop continuous flow that allows user flexibility and reliability in the pursuit of challenging chemistry. An oscillating baffle reactor mold-formed in perfluoroalkoxy alkane enables handling reactions forming solids. A LabView control system with help from Omega temperature controllers allows for simple and precise regulation of reaction conditions with real-time monitoring of the temperature and pressure. A new pressurized sample collection vessel allows for inert and pressurized sampling into multiple vials. An array of case studies serves to validate the system by demonstrating the use of interchangeable reactors with in-line heating and cooling, along with the ability to handle highly corrosive and reactive chemical reagents such as white fuming HNO3.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Scientific Data
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41597-020-0460-4
Volume: 7 Issue: 1 First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 5/1/20 4:00AM
Publication Location:
Article Title: Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry
Authors: Colin A. Grambow, Lagnajit Pattanaik, William H. Green
Keywords: geometry optimizations, frequency calculations, eaction SMILES, activation energies, enthalpies
Abstract: Reaction times, activation energies, branching ratios, yields, and many other quantitative attributes are important for precise organic syntheses and generating detailed reaction mechanisms. Often, it would be useful to be able to classify proposed reactions as fast or slow. However, quantitative chemical reaction data, especially for atom-mapped reactions, are difficult to find in existing databases. Therefore, we used automated potential energy surface exploration to generate 12,000 organic reactions involving H, C, N, and O atoms calculated at the ?B97X-D3/def2-TZVP quantum chemistry level. We report the results of geometry optimizations and frequency calculations for reactants, products, and transition states of all reactions. Additionally, we extracted atom-mapped reaction SMILES, activation energies, and enthalpies of reaction. We believe that this data will accelerate progress in automated methods for organic synthesis and reaction mechanism generation.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: The Journal of Physical Chemistry Letters
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jpclett.0c00500
Volume: 11 Issue: 8 First Page #: 2992
Date Submitted: 8/30/20 12:00AM Date Published: 3/1/20 5:00AM
Publication Location: United States
Article Title: Deep Learning of Activation Energies
Authors: Colin A. Grambow, Lagnajit Pattanaik, William H. Green
Keywords: template-free deep learning model, gas-phase quantum chemistry reactions
Abstract: Quantitative predictions of reaction properties, such as activation energy, have been limited due to a lack of available training data. Such predictions would be useful for computer-assisted reaction mechanism generation and organic synthesis planning. We develop a template-free deep learning model to predict the activation energy given reactant and product graphs and train the model on a new, diverse data set of gas-phase quantum chemistry reactions. We demonstrate that our model achieves accurate predictions and agrees with an intuitive understanding of chemical reactivity. With the continued generation of quantitative chemical reaction data and the development of methods that leverage such data, we expect many more methods for reactivity prediction to become available in the near future.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Reaction Chemistry & Engineering
Publication Identifier Type: DOI Publication Identifier: 10.1039/D1RE00350J
Volume: 6 Issue: 12 First Page #: 2306
Date Submitted: 12/29/21 12:00AM Date Published: 9/21/21 4:00AM
Publication Location: Cambridge, UK
Article Title: Design of dynamic trajectories for efficient and data-rich exploration of flow reaction design spaces
Authors: Federico Florit, Anirudh M. K. Nambiar, Christopher P. Breen, Timothy F. Jamison, Klavs F. Jensen
Keywords: data rich experimentation, flow chemistry, transient experiments
Abstract: Batch and continuous reactors both enable exploration of a chemical design space. The former rely on transient experiments, thus experiencing a wide variety of operating conditions over time, whereas the latter are usually operated at steady state and are representative of only one set of conditions. Operating a continuous reactor under dynamic conditions allows more efficient exploration of the underlying reaction space for extraction of kinetics and optimization of performance. We present a methodology to efficiently explore a design space using a tubular flow reactor installed on an automatic platform (equipped with FTIR and HPLC analysis) operated in a transient regime using sinusoidal variations of the parameters. This datadense method proves to be quicker with respect to steady-state operations because of the larger amount of information collected during a single experiment. A computational analysis provides a simple criterion for the design of dynamic experiments in order ...
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Trends in Chemistry
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.trechm.2021.02.005
Volume: 3 Issue: 5 First Page #: 373
Date Submitted: 12/29/21 12:00AM Date Published: 5/1/21 4:00AM
Publication Location:
Article Title: Ready, Set, Flow! Automated Continuous Synthesis and Optimization
Authors: Christopher P. Breen, Anirudh M.K. Nambiar, Timothy F. Jamison, Klavs F. Jensen
Keywords: automation, flow chemistry, review
Abstract: Synthetic chemistry provides access to advanced materials that facilitate innovation in key industries such as medicine, energy, and agriculture. Automation is poised to challenge the traditional process of chemical synthesis and development. Continuous flow chemistry has recently come into maturity and provides a flexible platform amenable to automation. The merger of synthesis and automation promises to democratize access to custom complex small molecules for nonexperts as well as accelerate the development of new synthetic protocols by relieving expert chemists of routine tasks. In this contribution, we discuss recent case studies that present strategies towards realizing automated synthesis with a further focus on works that leverage continuous flow chemistry as an enabling technology.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 56th IEEE Conference on Decision and Control
Date Received: 31-Aug-2018 Conference Date: 13-Dec-2017 Date Published: 13-Dec-2017
Conference Location: Melbourne, Australia
Paper Title: A Piecewise Polynomial Chaos Approach to Stochastic Linear Quadratic
Authors: Yiming Wan, Eranda Harinath, and Richard D. Braatz
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 34 th International Conference on Machine Learning
Date Received: 30-Aug-2018 Conference Date: 06-Aug-2017 Date Published: 06-Aug-2017
Conference Location: Sydney, Australia
Paper Title: Deriving Neural Architectures from Sequence and Graph Kernels
Authors: Tao Lei, Wengong Jin, Regina Barzilay, Tommi Jaakkola
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 31st Conference on Neural Information Processing Systems (NIPS 2017),
Date Received: 11-Dec-2017 Conference Date: 04-Dec-2017 Date Published: 04-Dec-2017
Conference Location: Long Beach, CA, USA.
Paper Title: Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network
Authors: Wengong Jiny Connor Coley, Regina Barzilay, Tommi Jaakkola
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Conference on Machine Learning
Date Received: 31-Aug-2018 Conference Date: 10-Jul-2018 Date Published: 10-Jul-2018
Conference Location: Stockholm, Sweden
Paper Title: Junction Tree Variational Autoencoder for Molecular Graph Generation
Authors: Wengong Jin, Regina Barzilay, Tommi Jaakkola
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: American Control Conference
Date Received: 30-Aug-2019 Conference Date: 10-Jul-2019 Date Published: 10-Jul-2019
Conference Location: Philadelphia, PA
Paper Title: Offset-free Input-Output Formulations of Stochastic Model Predictive Control Based on Polynomial Chaos Theory
Authors: Matthias von Andrian, Richard Braatz
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2020 American Control Conference 2020
Date Received: 31-Aug-2020 Conference Date: 01-Jul-2020 Date Published:
Conference Location: Denver, CO, USA
Paper Title: Stochastic Dynamic Optimization and Model Predictive Control based on Polynomial Chaos Theory and Symbolic Arithmetic
Authors: Matthias von Andrian, Richard D. Braatz
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2020 American Control Conference
Date Received: 31-Aug-2020 Conference Date: 01-Jul-2020 Date Published:
Conference Location: Denver, CO, USA
Paper Title: Fast Model Predictive Control of Startup of a Compact Modular Reconfigurable System for Continuous-Flow Pharmaceutical Manufacturing
Authors: Anastasia Nikolakopoulou, Matthias von Andrian, Richard D. Braatz
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 11-Jan-2022

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Federation of Automatic Control World Congress
Date Received: 31-Aug-2020 Conference Date: 11-Jul-2020 Date Published: 11-Jul-2020
Conference Location: Virtual
Paper Title: Fast stochastic model predictive control of unstable dynamical systems
Authors: Mathias von Andrian, Richard D. Braatz
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International Federation of Automatic Control World Congress
Date Received: 31-Aug-2020 Conference Date: 11-Jul-2020 Date Published: 11-Jul-2020
Conference Location: Virtual
Paper Title: Self-Optimizing Control of a Continuous-Flow Pharmaceutical Manufacturing Plant
Authors: David Perez Pineiro, Anastasia Nikolakopoulou, Johannes Jaschke, Richard D. Braatz
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 3-Accepted
Conference Name: Conference on Decision and Control
Date Received: 31-Aug-2020 Conference Date: 14-Dec-2020 Date Published:
Conference Location: Virtual
Paper Title: Output Feedback Control of Dynamic Artificial Neural Networks Using Linear Matrix Inequalities
Authors: Anastasia Nikolakopoulou, Moo Sun Hong, Richard D. Braatz
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: American Control Conference
Date Received: 31-Aug-2020 Conference Date: 10-Jul-2019 Date Published: 10-Jul-2019
Conference Location: Philadelphia
Paper Title: Plantwide Control of a Compact Modular Reconfigurable System for Continuous-Flow Pharmaceutical Manufacturing
Authors: Anastasia Nikolakopoulou, Matthias von Andrian, Richard D. Braatz
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Klavs F. Jensen
Signature Date: 12/29/21 1:49PM

Accomplishments

Table of Contents

1 Task 1 and 2 Machine Learning

1.1 General improvements to the ASKCOS web-based application

We have focused on making the ASKCOS web-based application more computationally efficient, as well as improving the web interface to make it more user friendly. We reduced the memory footprint from ~40 GB to ~5 GB RAM by reducing duplication of reaction template data in memory and serving machine learning predictions from a centralized API, rather than duplicating models. We continued to migrate tools in the user interface from static, server-side rendered HTML forms to dynamic client-side (browser) applications built with VueJS. These interactive tools consume data from an API, allowing components of the interface to change independently.

1.2 Rare template applicability

For data-driven, template-based retrosynthetic models, it can be difficult to extract full value out of available reaction data due to rare transformations that show up infrequently. With limited examples for these rare reaction templates, machine learning models are implicitly biased towards prioritizing popular reactions which tend to be reactions that minimally decrease synthetic complexity. In order to provide more information during model training, we augmented reaction data with computationally determined, synthetic examples that represent hypothetical reaction precedents for rare reaction templates. This process effectively expands the scope of product molecules for each template so that the model better learns which input features are relevant for each output class (template). A template applicability model was trained on this augmented data set during a pre-training stage, which was followed by fine-tuning on the ground truth, un-augmented reaction data to prioritize known chemical transformations before hypothetical chemical transformations. Without a significant drop in top-k accuracy, the template relevance model pre-trained on template applicability information achieved significantly higher template applicability precision - meaning a high fraction of the top prioritized templates apply to the input target molecule and generate a precursor.

1.3 Predicting the likely products and yield. [Columbia][1]

The ability to predict the likely products and yields of a reacting mixture under specified conditions remains an outstanding challenge. Recently, models trained on literature data have made significant progress in identifying the likely products of unfamiliar reactions; however, additional data is required to refine these predictions and to quantify the role of reaction conditions. In this context, the design of experiments to maximize the relevant information provided by these additional data can greatly reduce the amount of data required to train predictive models. We have implemented a Bayesian design platform to guide optimal experiments based on strong guiding models with unknown parameters. The design of experiments to learn these parameters efficiently remains challenging for nonlinear models that give rise to multimodal distributions. We address this problem using a sequential Monte Carlo (SMC) approach to sample parameter distributions and estimate the expected utility of candidate experiments. This

[1] Headings with no brackets designates work done at MIT

platform has been validated using simulated data and is now being applied to guide parameter estimation in sequences of automated experiments.

1.4 QM descriptors for selectivity prediction

Accurate and rapid evaluation of whether substrates can undergo the desired the transformation is crucial and challenging for both human knowledge and computer predictions. Despite the potential of machine learning in predicting chemical reactivity such as selectivity, popular feature engineering and learning methods are either time-consuming or data-hungry. We introduce a new method that combines machine-learned reaction representation with selected quantum mechanical descriptors to predict regio-selectivity in general substitution reactions. We construct a reactivity descriptor database based on ab initio calculations of 130k organic molecules, and train a multi-task constrained model to calculate demanded descriptors on-the-fly. The proposed platform enhances the inter/extrapolated performance for regio-selectivity predictions and enables learning from small datasets with just hundreds of examples.

Furthermore, the proposed protocol is demonstrated to be generally applicable to a diverse range of chemical spaces. For three general types of substitution reactions (aromatic C-H functionalization, aromatic C-X substitution, and other substitution reactions) curated from a commercial database, the fusion model achieves 91%-97% top-1 accuracy in predicting the major outcome. Using predicted descriptors, the fusion model is end-to-end, and requires approximately only 70ms per reaction to predict the selectivity from reaction SMILES strings.

1.5 Quantitative prediction of reaction kinetics

Accurate knowledge of kinetic parameters is a crucial step towards quantitative modeling of condensed phase systems. For example, estimation of drug shelf life via computational modeling of decomposition and oxidation processes requires extensive knowledge of the relevant reaction networks involved. To this end, quantitative predictions of reaction energetics enable assessment of the relative importance of reaction candidates or even allow for the determination of major product and side-product yields. Due to a lack of chemical reaction data annotated with kinetic information, we devised a workflow to generate more than 30,000 reactions based on automated potential energy surface exploration methods using density functional theory calculations. We used the newly calculated data to develop a model for the prediction of activation energies. To encode the reactions, we adapted a graph convolutional neural network, which preferentially focuses on the atoms involved in the reactive center as these contribute most to the activation energy. The model predicts activation energies with high accuracy (<2 kcal/mol). Much of this work was completed in the previous calendar year, but we published manuscripts detailing both the generation of the dataset and development of the activation energy prediction model.

1.6 Predicting octanol-water partitioning coefficients

A machine learning model is built for the prediction of octanol-water partitioning coefficients. Standard chemprop is used for machine learning with slight modifications to the atom and bond features. Features more important to solvation are added, such as the ability to act as a hydrogen donor/acceptor and the electronegativity of atoms. Two databases are used to train the machine learning model. A quantum chemical database is constructed using COSMOtherm calculations at the BP-TZVPD-FINE level of theory. A second experimental database is compiled from different open-source databases including PHYSPROP, OChem and Drugbank. The quantum chemical database is used to pre-train the model, such that it learns the physical behavior of solutes partitioned between an octanol and water phase. The parameters of this first model for the latent molecular representation are fixed and transferred to a second model that is finetuned on the experimental data. Transfer learning from the quantum chemical data to the experimental data improves the generalizability of the model and hence the performance on unseen elements or molecule types. Overall the model can predict octanol-water partitioning coefficients on a random test split with a logP RMSE of 0.47.

1.7 Predicting temperature dependence of solvation free energy

Solvation free energy (ΔGsolv) can be used to calculate solutes' relative solubilities in differentsolvents and thus is helpful in identifying solvents that minimize risk of precipitation during liquid-liquid extractions for clean-up after synthesis steps. There has been extensive literature work on improving solvation energy prediction, but the existing work mostly focuses on the prediction at 298 K, and temperature dependence is seldom considered. In order to improve ΔGsolv prediction at elevated temperatures, we have developed a simple method that can predict the temperature dependence of ΔGsolv from near room temperature up to the critical temperature of the solvent only based on Δgsolv and ΔHsolv (solvation enthalpy) at 298 K and solvent's temperature dependent density. Our method employs the piece-wise function combining two existing semi-empirical formulas and solves for needed empirical parameters from ΔGsolv and ΔHsolv at 298 K. We have tested our method over 47 solute-solvent systems and compared with a simple linear extrapolation method and an ab initio method using COSMO-RS BP/TZVPD-FINE level. Results indicate that our new temperature dependence method can provide fast and accurate prediction across all temperatures and is comparable or better than the computationally expensive ab initio method. The mean absolute errors of our method and the ab initio COSMO method are 0.38 and 0.50 kcal/mol respectively for 47 solute-solvent systems. Our new method is beneficial for designing good liquid-liquid extractions as one can predict the effect of temperature on extraction efficiency and select optimal solvents based on temperature dependent estimation.

1.8 Predicting solvation free energy and enthalpy at 298 K

In the previous section, we introduced a new method to predict temperature dependence of solvation free energy only based on solvation free energy and enthalpy at 298 K and the solvent's density as a function of temperature. However, ΔGsolv and ΔHsolv at 298 K for arbitrary solute and solvent pairs are usually not readily available. Therefore, we have developed machine learning (ML) models to predict ΔGsolv and ΔHsolv at 298 K from 2D structures of solutes and solvents. We took two different approaches for these machine learning models: (1) direct ML (2) indirect method using solute parameter ML. For the first approach using the direct ML, we trained ML models to directly predict ΔGsolv and ΔHsolv when solute and solvent SMILES strings are given as inputs. To train the models, we collected approximately 20,000 experimental ΔGsolv data from various sources including Minnesota Solvation Database, CompSol, FreeSolv, OChem, Physprop, Drugbank, and literature. These include 6000 solutes in 1400 solvents. We collected 6300 experimental ΔHsolv data from Acree Database and CompSol, which include 1700 solute in 1300 solvents. For the second indirect approach, we used a linear solvation energy relationship to calculate ΔGsolv and ΔHsolv. This relationship requires several parameters, so we trained an ML model to predict these parameters given a solute's SMILES string. Initial results show that the direct ML model gives more accurate predictions for both ΔGsolv and ΔHsolv predictions on all testing sets.

2 Task 3 Chemistry

2.1 Tramadol

A telescoped continuous synthesis of tramadol according to the ASKCOS retrosynthetic suggestion has been achieved during this reporting period (Figure 1). A highly detailed report of this work was also been published in the journal *Synlett* in early 2020 (DOI: 10.1055/s-0039-1690884).

The sequence begins with the Mannich reaction of formaldehyde, dimethylammonium chloride, and cyclohexanone. The reaction takes place under process intensive conditions to afford the desired Mannich product in 92 % yield. This stream was then fed into a purification sequence wherein excess cyclohexanone was removed via organic extraction, followed by neutralization and desiccation of the Mannich product salt. Water content of the dried Mannich product was assessed using inline FTIR, which was then reacted with 3-methoxyphenylmagnesium bromide in a Grignard reaction to afford tramadol.

Identification of Grignard reaction conditions that increased throughput to 13.7 $g \cdot h^{-1}$ and reactor space-time efficiency to 4.2 $g \cdot h^{-1} \cdot mL^{-1}$ compared to previously reported processes was a major achievement of this work. The complexity and number of unit operation needed to execute this process likely prohibit the complete transfer to the automated synthesis platform. However, it may worth perusing the Grignard reaction unit operation of this process on the automated synthesis platform. In this example, the solvent sensitivity of the Grignard reaction can be investigated and optimized in an autonomous fashion.

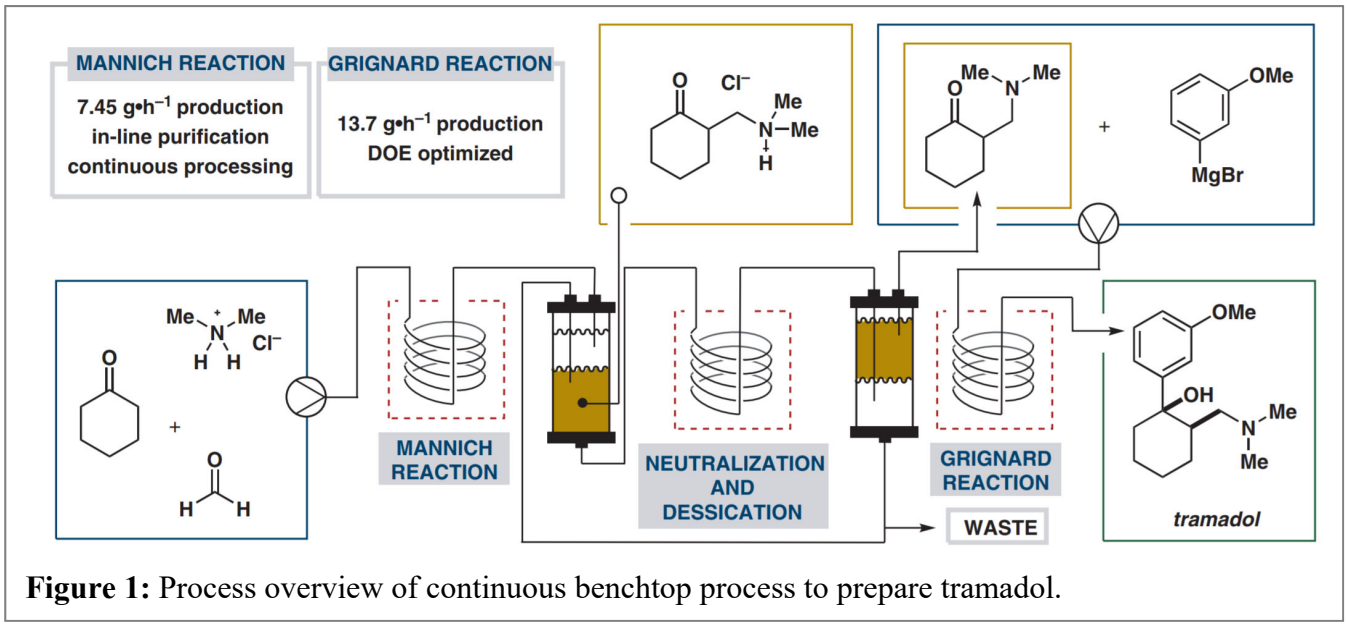


Figure 1: Process overview of continuous benchtop process to prepare tramadol.

2.2 Ribavirin

Further efforts towards the synthesis of ribavirin have been pursued both as a benchtop process and as an automated process on REX (Figure 2). We previously reported that 1,2,4-triazole-3-carbonitrile was a competent glycosyl acceptor in offline batch processes towards ribavirin, but conversion to a flow continuous process yielded inferior results. In order to improve the chances of success for this target we turned to ASKCOS for new retrosynthetic analysis along with forward synthesis condition recommendations (Figure 2A). The second highest ranking suggestion for the first precursor involved aminolysis that can be accomplished in the forward direction by reacting the ester-protected intermediate with ammonia in methanol. ASKCOS terminated the analysis by suggesting an *N*-glycosylation reaction between commercially available tetra-acetylated β-D-ribofuranose and methyl-1,3,5-triazole carboxylate. Forward condition recommendations include an elevated reaction temperature and a Brønstead acid catalyst, such as *p*-toluenesulfonic acid.

A brief series of benchtop experiments was conducted prior to implementation as an automated optimization process on the reaction execution platform (REX) described in Task 4. First, the solubility of all starting materials was tested. It was found that methyl-1,3,5-triazole carboxylate displayed limited solubility, except in amide solvents. We proceeded with NMP as a suitable solvent in benchtop flow experiments. High conversion of starting material could be achieved at 140 °C and a residence time of 12 minutes.

With a fully soluble reactivity profile established we proceeded to transfer further optimization efforts to REX (Figure 2B). The offline experiments led us to define a residence time range of 10-40 minutes, a catalyst loading range of 10-20 mol %, a temperature range of 30-150 °C, and two different pyridinium brønstead acid catalysts as discrete variable options. REX performed 15 experiments over 17 hours of uninterrupted operation. The optimal yielding condition occurred at a temperature of 150 °C, 20 mol%

catalysts loading, 40 minutes of residence time, and with the pyridinium triflate catalyst. While conversion reached about 90%, the inline HPLC yield was only 35%. This is likely due to the sluggish reactivity of the methyl-1,3,5-triazole carboxylate glycosyl acceptor promoting unproductive conversion. This set of experiments serves to validate the automated optimization capabilities of REX, however the limited optimization potential of ribavirin has led us to deprioritize this chemistry in favor of targets with a strong optimization potential.

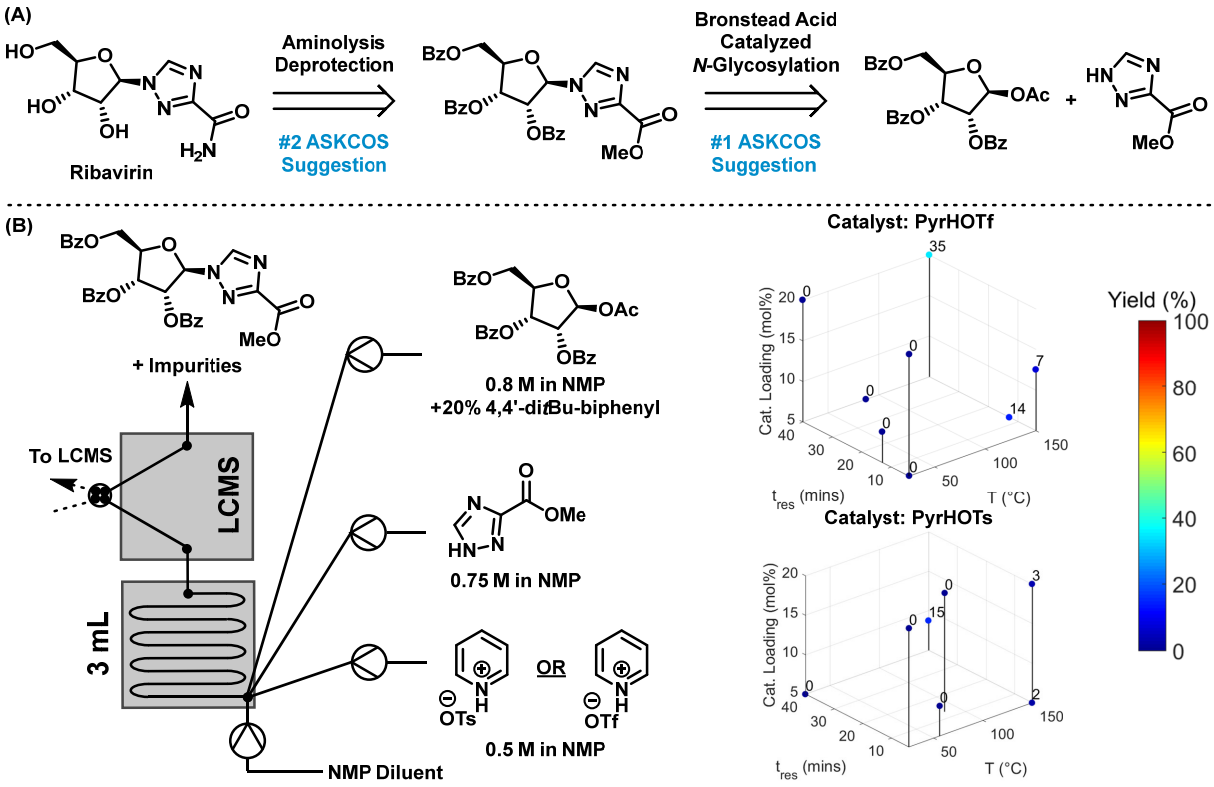


Figure 2: (A) Retrosynthetic analysis for ribavirin according to ASKCOS. (B) Process configuration of REX during the automated optimization campaign and experimental results.

2.3 Camostat

Camostat is a serine protease inhibitor currently in Phase 1 and Phase 2 clinical trials for the treatment of COVID-19. We are currently pursuing this synthetic target in order to showcase the utility of REX and ASKCOS when robust, scalable production of an active pharmaceutical ingredient is needed to support rapidly evolving material demands.

Investigations commenced with a thorough analysis of synthesis plans provided by ASKCOS (Figure 1A). To produce the desired product, an esterification facilitated by either DCC or EDC with 4-guanidinobenzoic acid and alkylated 4-hydroxyphenylacetic acid was proposed. The alkylated 4-hydroxyphenylacetic acid is not commercially available and must be prepared by an S_N2 esterification with 4-hydroxyphenylacetic acid and either bromodimethyl acetamide, or chlorodimethyl acetamide along with a nitrogenous base. Additionally, ASKCOS suggested an amide solvent would be suitable for both steps.

Offline process development involved solubility studies, one-pot batch experiments, and outlining general continuous process parameters. Ultimately, the one-pot batch experiments confirmed that the ASKCOS forward conditions yielded camostat and the reagent and solvent conditions could be used with

only one substitute – $NEt(iPr)_2$ was substituted for NEt_3 due to limited solubility of $NEt_3 \cdot HCl$ in the ASKCOS recommended solvent.

The first step on this process was then transferred to REX for a rigorous reaction condition screening effort (Figure 2B). The process configuration of REX utilized both an inline LCMS sampling module for quantitative analysis as well as an inline FTIR for steady-state monitoring. The optimization scheme included base discrete variable candidates (2,6-lutidine and $NEt(iPr)_2$) as well as halodimethylacetamide discrete variable candidates (bromodimethyl acetamide and chlorodimethyl acetamide). The objective

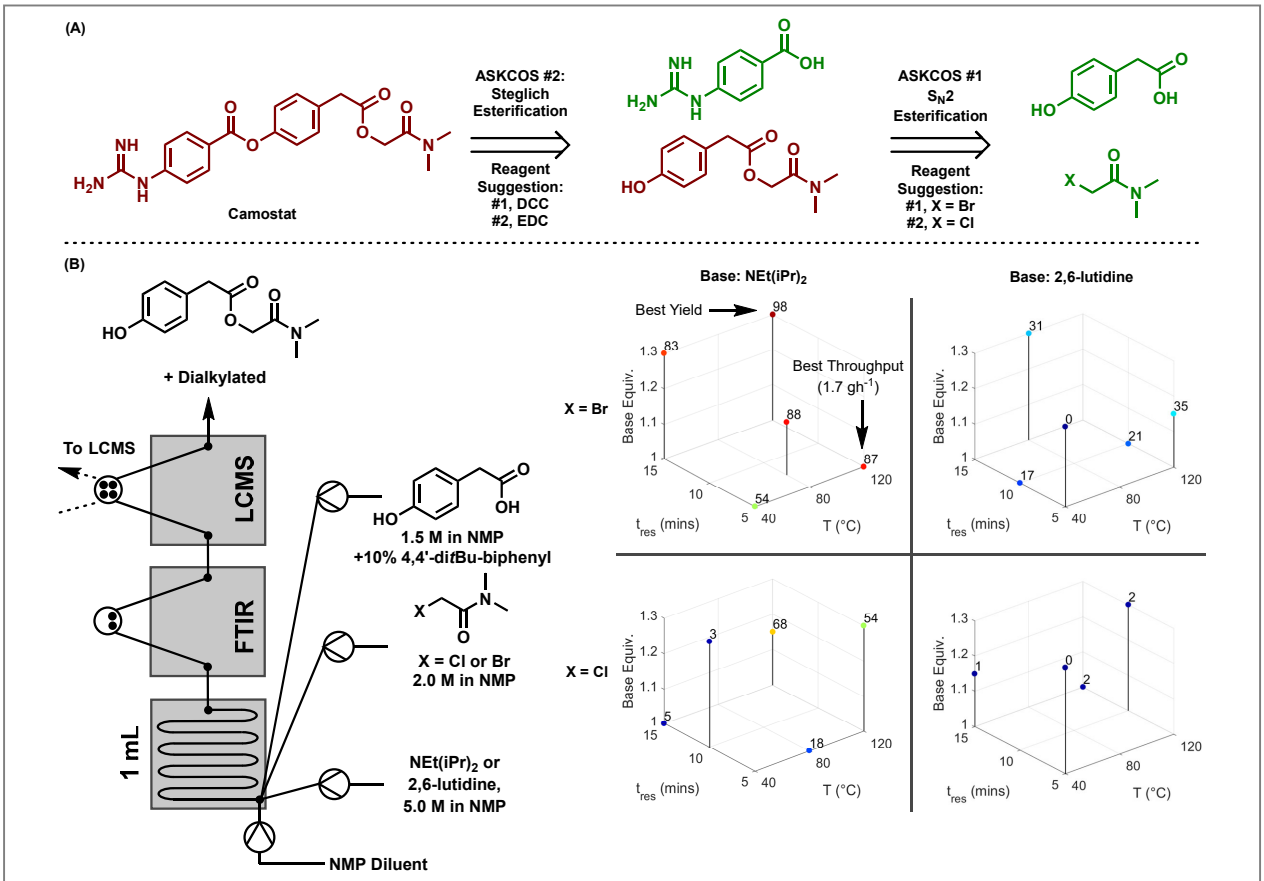


Figure 3: (A) Retrosynthetic analysis for camostat according to ASKCOS. (B) Process configuration of REX during the automated optimization campaign and experimental results.

function was set to optimize for throughput. After performing 19 experiments over the course of 12 hours, REX established the optimization landscape. Interestingly, the highest yielding experiment was not the highest throughput experiment. The combination of $NEt(iPr)_2$ with bromodimethyl acetamide yielded optimal results during this initial optimization campaign and further investigation of this step is underway. Efforts towards completing the two-step synthesis will commence shortly thereafter.

2.4 Arylex

We are currently pursuing the synthesis of arylex, a modern auxin herbicide product, in order to diversify synthetic targets away from active pharmaceutical ingredients, thereby expanding the utility of REX to different sectors of the chemical industry. The core scaffold of arylex has been assembled by Pd-catalyzed C-C coupling and this option was also identified by ASKCOS (Figure 4A). The Suzuki coupling may be performed with or without an *N*-protecting group. It is worth noting that a C-N coupling pathway promoting the formation of undesired products may proceed if a strong base is used. However, Suzuki coupling conditions containing unprotected nitrogen atoms have been developed.

The *N*-protected Suzuki coupling was tested using reaction conditions suggested by ASKCOS (Figure 4B). The recommended conditions proved to be extremely effective, providing *N*-acetly arylex in 96%

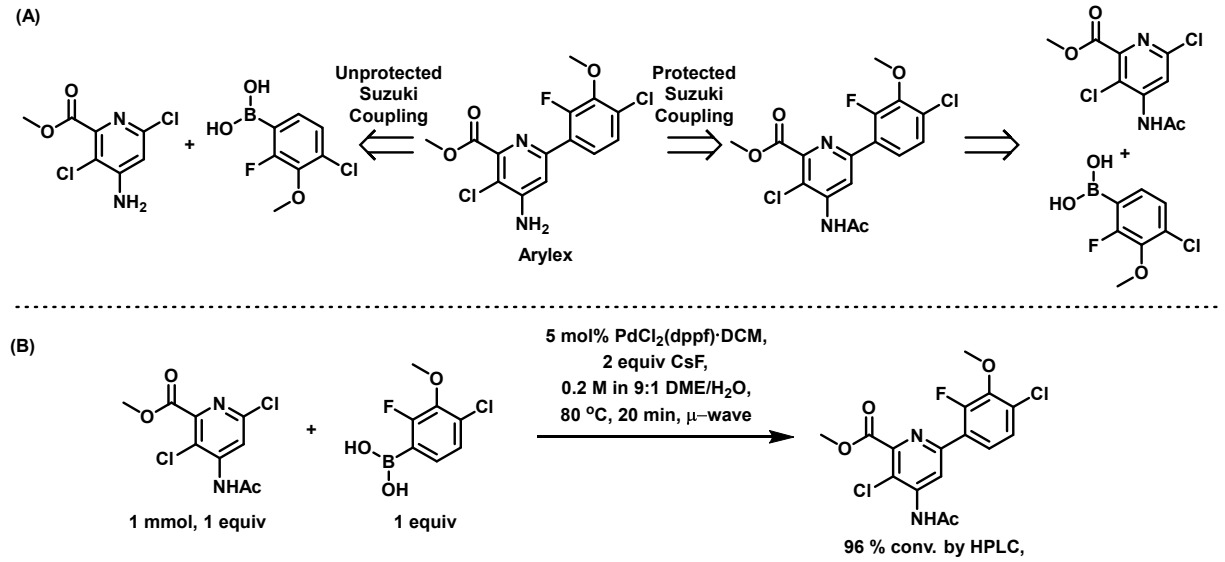


Figure 4: (A) Retrosynthetic analysis for arylex according to ASKCOS. (B) Initial batch data for *N*-protected Suzuki coupling using ASKCOS recommended conditions.

conversion in just 20 minutes of reaction time at °C. Future investigation will focus on the unprotected Suzuki coupling and transfer on the process to REX.

2.5 Tiazofurin [VCU]

Our previous efforts to develop a continuous flow synthetic route for Tiazofurin led to lower yields (69%) as a result of significant by-products due to de-silylation of TMS-thiazole synthon. However, the reaction of 5-amido adduct of TMS-thiazole with halosugar gave the predominant β-conformer with minimal desilylated product. These batch conditions were translated to continuous flow process, gave 70% ee of the desired β-conformer.

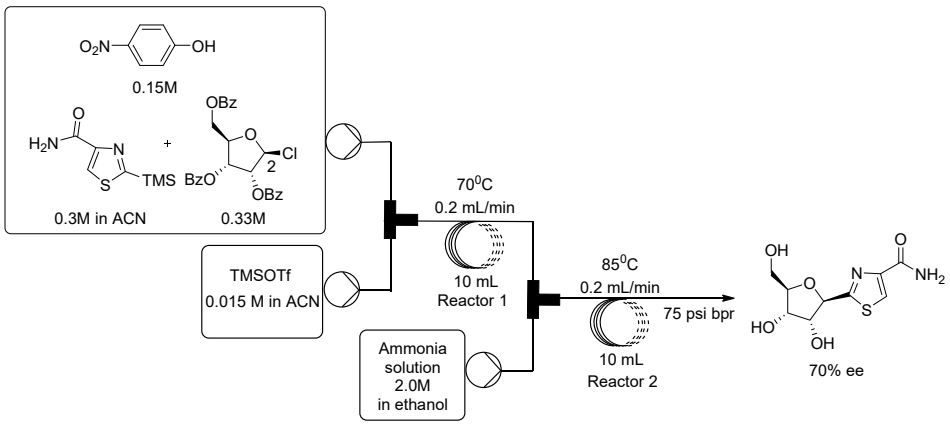


Figure 5: Continuous flow synthetic route to tiazofurin.

2.6 Pregabalin [VCU]

A fully continuous route to pregabalin has been proposed that relies on an asymmetric enantioselective Michael addition using principles of phase transfer catalysis. The scheme begins with condensation of 5-

nitro isoxazole with isolvaleraldehyde to give the highly selective *E*-alkene in good yields. This is further subjected to Michael addition of nitromethane using quinidine asymmetric phase transfer catalyst (PTC), after which the product undergoes Sarti-Fantoni reaction (ring opening of 5-nitro isoxazole) to give enantioselective nitro acid. Currently, the ability to recycle the PTC as well as the reduction of the nitro acid along with isolation and purification of the final pure product is under investigation.

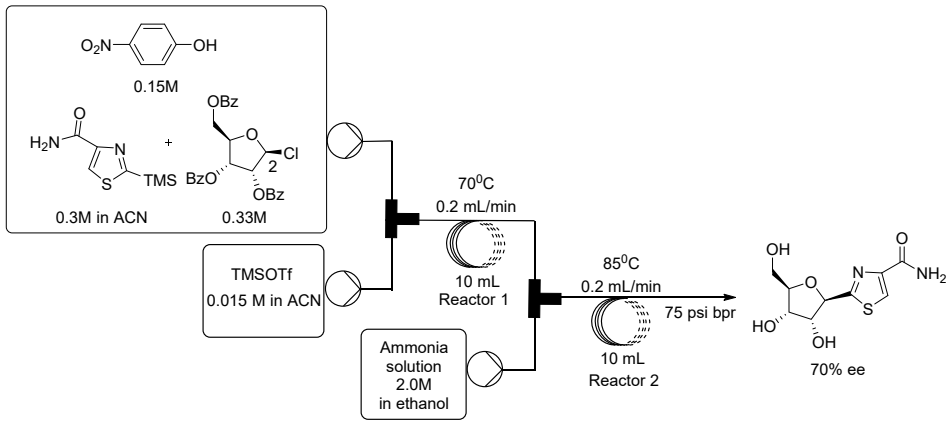


Figure 6: Laboratory scale continuous flow synthesis of pregabalin.

2.7 Summary of Key Outcomes

1) Completed lab scale synthesis of tramadol.
    a. Telescoped flow synthesis of tramadol was achieved.
    b. Improvements to process throughput and reactor space-time efficiency have been improved compared to previous reports.
2) Progress of lab scale synthesis of ribavirin.
    a. Batch conditions superior to continuous flow conditions.
    b. Discrete variable screening campaign performed on REX suggests limited improvement from further investigation.
3) Progress of lab scale synthesis of camostat
    a. Rapid bench-top optimization with minimal deviation from ASKCOS forward condition recommendation.
    b. Discrete variable screening campaign for first step performed on REX with good results.
    c. Discrete variable screening campaign for second step not yet performed.
4) Progress of lab scale synthesis of arylex
    a. Commenced screening of forward synthesis conditions recommended by ASKCOS.
    b. Different reactor configurations present unique discrete variable screening opportunity.
5) Completion of lab-scale batch and continuous flow synthesis of tiazofurin
6) Progress on lab-scale continuous flow synthesis of pregabalin
    a. Individual optimization of the first two steps of alkene formation is complete.
    b. Optimization of Michael reaction using resin catalyst and hydrolysis of the formed Michael adducts is in progress

2.8 Challenges

Minimizing optimization experiments performed by a human operator remains a significant challenge towards achieving greater levels of automation in our chemical development scheme. While improvements to both ASKCOS and the automated synthesis platform have narrowed the gap between

computer generated forward synthesis prediction and automated process execution, offline testing by the expert human operator remains an essential step. Key areas for offline evaluation include assessing reagent solubility, identification of chemical compatibility issues, and performing an overall process safety evaluation.

3 Task 4. Systems: Reaction Execution System

3.1 Summary of important outcomes

For the robotic flow chemistry platform (Figure 7), the following key outcomes were achieved:

1. New XYZθ gantry robot enabling simpler path planning, 2.5x faster assembly, and high repeatability.
2. Improved fluidic connection system with lower dead volume and convergent synthesis capability.
3. Inline reaction monitoring with HPLC-MS and FT-IR modules at any/multiple points within a multistep synthesis.
4. Algorithm for automated optimization of reaction conditions involving both continuous (e.g., temperature, residence time, equivalents) and discrete (e.g., catalyst, base) variables.

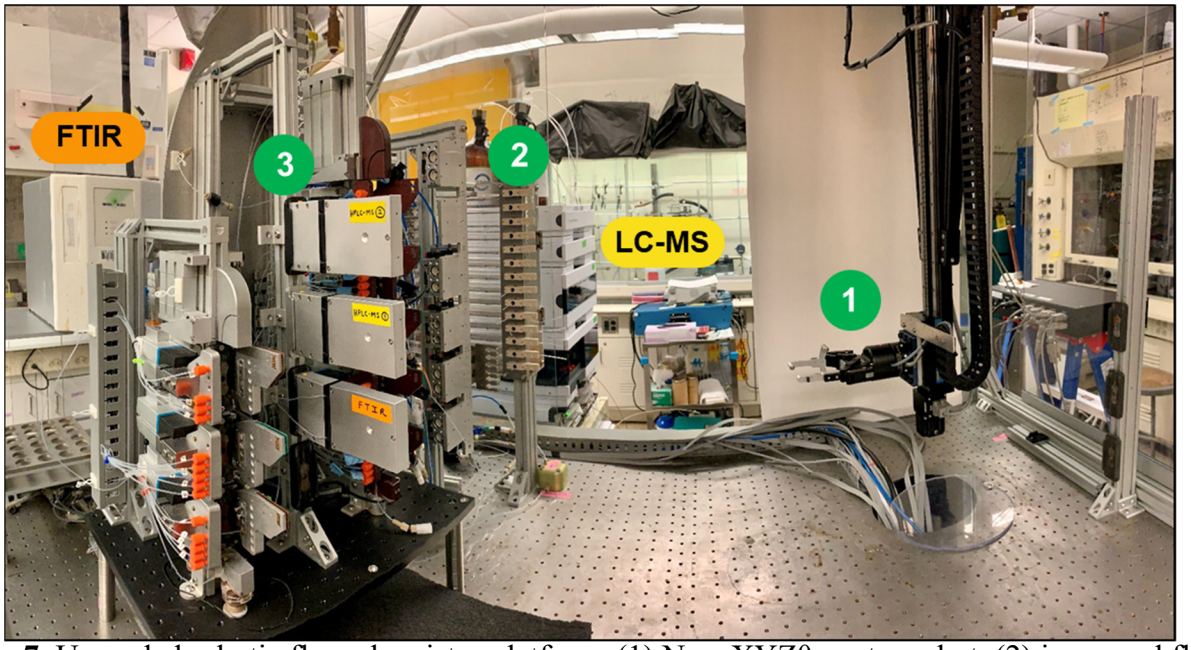


Figure 7. Upgraded robotic flow chemistry platform. (1) New XYZθ gantry robot, (2) improved fluidic connection system, and (3) HPLC-MS and FT-IR modules for inline reaction monitoring.

3.2 New XYZθ Gantry Robot

We faced several challenges with the previous 6-axis robotic arm (model: UR3 by Universal Robots). When reactor modules are placed onto the process bays, a linear motion along a straight path is required to prevent collisions with surrounding hardware (e.g., mechanical components, adjacent reactor modules). Because the UR3 is an articulated robot with interacting rotary joints, performing the linear movements required was a major challenge. As a result, users spent a significant amount of time on path planning to avoid collisions and singularities (a situation where the robot joints become blocked in certain directions). Furthermore, pick-and-place movements also ended up being extremely slow (2 minutes per pick-and-place).

To overcome these challenges, we turned our attention to gantry robots (an assembly of multiple linear axes) which are designed to move in Cartesian XYZ directions. After considering several options, we

purchased a XYZθ gantry robot from Newmark Systems Inc. with the additional rotary stage providing 360° access to all sides of the system enclosure. Once the initial and final positions of reactor modules are recorded, path planning is straightforward in Cartesian coordinates. Each linear axis can move at 10 cm/sec (5 micron unidirectional repeatability) and the rotary axis at 25°/sec (5 arc-sec unidirectional repeatability). This results in a pick-and-place time of around 45 seconds, which is 2.5x faster than the typical time of 2 minutes required by the previous robotic arm. All axes are equipped with homing switches and encoders enabling accurate zeroing of the gantry's position and tracking of absolute position during movement. The existing gripper was attached to the Z-stage of the gantry and the force-torque sensor on its wrist is used as a safety feature to stop the robot if the gripper collides into an object while moving.

3.3 Improved Fluidic Connection System

The old fluidic connection system required 60 inches of tubing to be wrapped inside spring-loaded reels (white reels shown in Figure 8 left image). This resulted in significant dead volume. Moreover, the tensioning force provided by the spiral spring coils inside the reels was not always reliable.

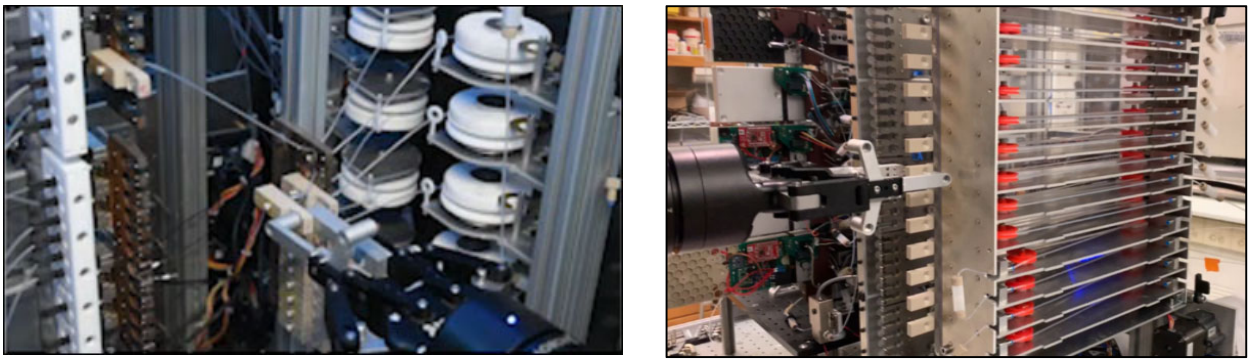


Figure 8. (Left) Previous fluidic connection system with spring-loaded reels. (Right) New system with spring-loaded pulleys with lower dead volume and convergent synthesis capability.

To address these problems, a new design (Figure 8, right image) was implemented in which tubing is wound around pulleys (red components in image) and a constant-force spring acts on the pulley to tension the tubing. This design cut the required tubing length in half (from 60 to 30 inches) and switching from 0.03 to 0.02 inch inner diameter tubing for the inlet lines further reduced the dead volume, resulting in a total 5x reduction (0.7 to 0.15 mL). In addition to providing a reliable tensioning force, the modular pulley design makes it easier for the user to install the tubing and troubleshoot. Furthermore, the new design contains two levels that are connected via a union to enable convergent synthesis by connecting fluid streams between the two process stacks on the system.

3.4 Inline Reaction Monitoring

To analyze the reaction directly on the system instead of manual offline analysis of the collected product, modules for inline reaction monitoring were implemented (Figure 9). These modules (like the reactor modules) can be picked up by the robot from an analytics stack and placed at any location on the process stack providing flexibility. They direct the reaction stream to a nearby sampling valve for injection into an HPLC-MS (high-performance liquid chromatography-mass spectrometry) or to an FTIR (Fourier-transform infrared) spectroscopy flow cell, and then return the stream back to the process stack to enable further reactions. HPLC-MS is a generally applicable, accurate quantitation technique and mass spectra provide confirmation of product formation and valuable insight into the identity of any byproducts formed. FTIR spectroscopy provides reaction data in seconds (compared to HPLC method times of ~10 minutes) but is generally noisier and limited to reactions where IR-active functional groups react.

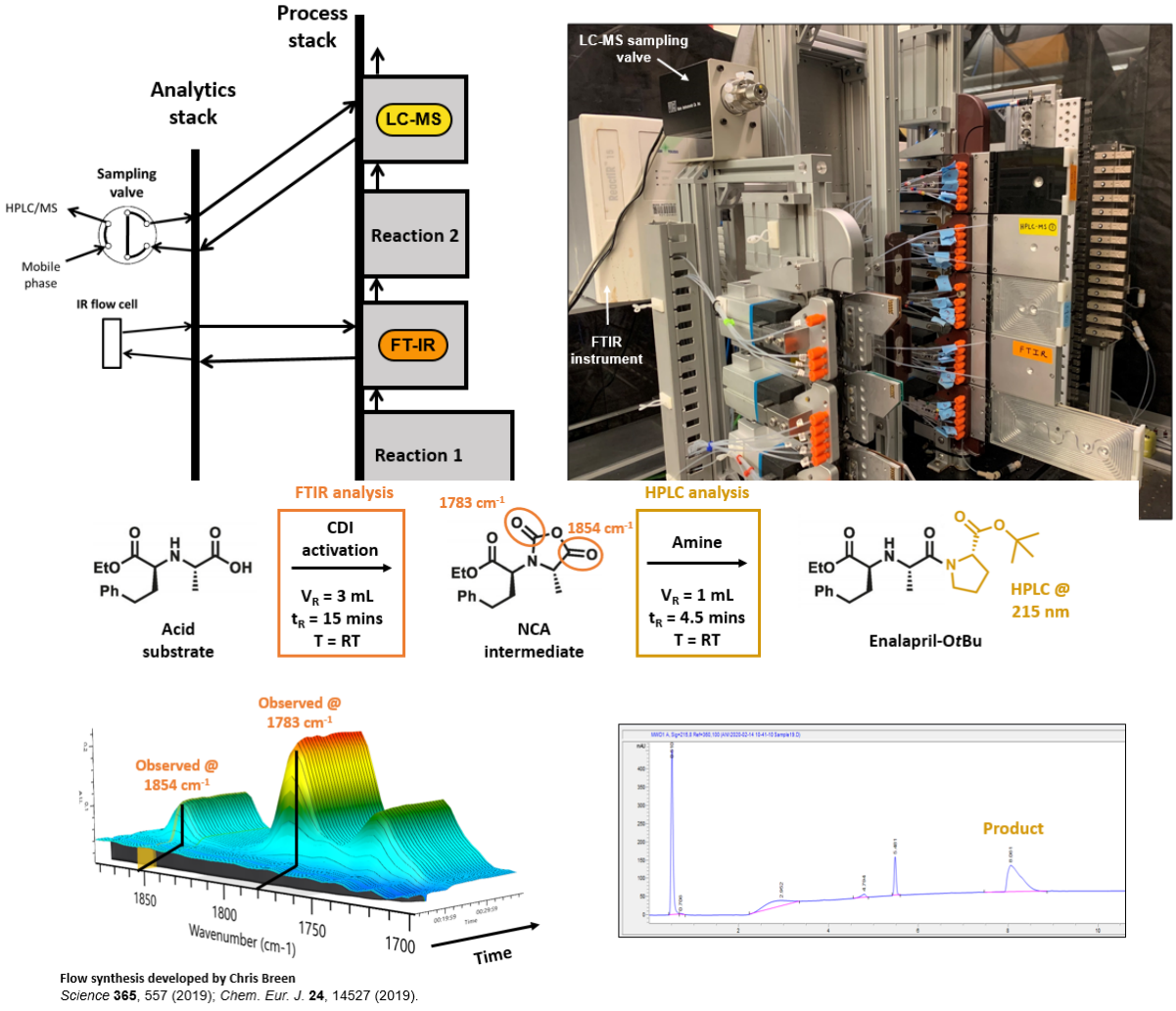


Figure 9. Analytics modules (FTIR and HPLC-MS) for inline reaction monitoring, demonstrated with a two-step synthesis of the ACE inhibitor enalapril.

As illustrated in Figure 9, both modules can be placed inline simultaneously in a multistep sequence for analyzing each reaction independently. This capability was demonstrated for the 2-step synthesis of the ACE inhibitor enalapril. In the first step, the activation of the acid substrate with carbonyl diimidazole

(CDI) produces an N-carboxyanhydride (NCA) intermediate with IR-active functional groups (circled in orange). In the second step, the addition of amine (in this case, proline) results in coupling with the activated substrate and the product can be observed by HPLC at 215 nm. The FTIR spectrum and LC chromatogram shown in Figure 9 reproduced the results reported in the original publication.

3.5 Algorithm for Automated Reaction Optimization

The ability to monitor reaction progress using inline analytics makes it possible to optimize reaction conditions via feedback of the data to an optimization algorithm that guides experimentation (Figure 10). Using synthesis pathways and conditions (e.g., reagents, solvent) generated by ASKCOS for a target compound of interest, automated optimization experiments can help relieve the burden on chemists to perform manual, time-consuming experiments to optimize reaction conditions such as temperature, residence time, etc.

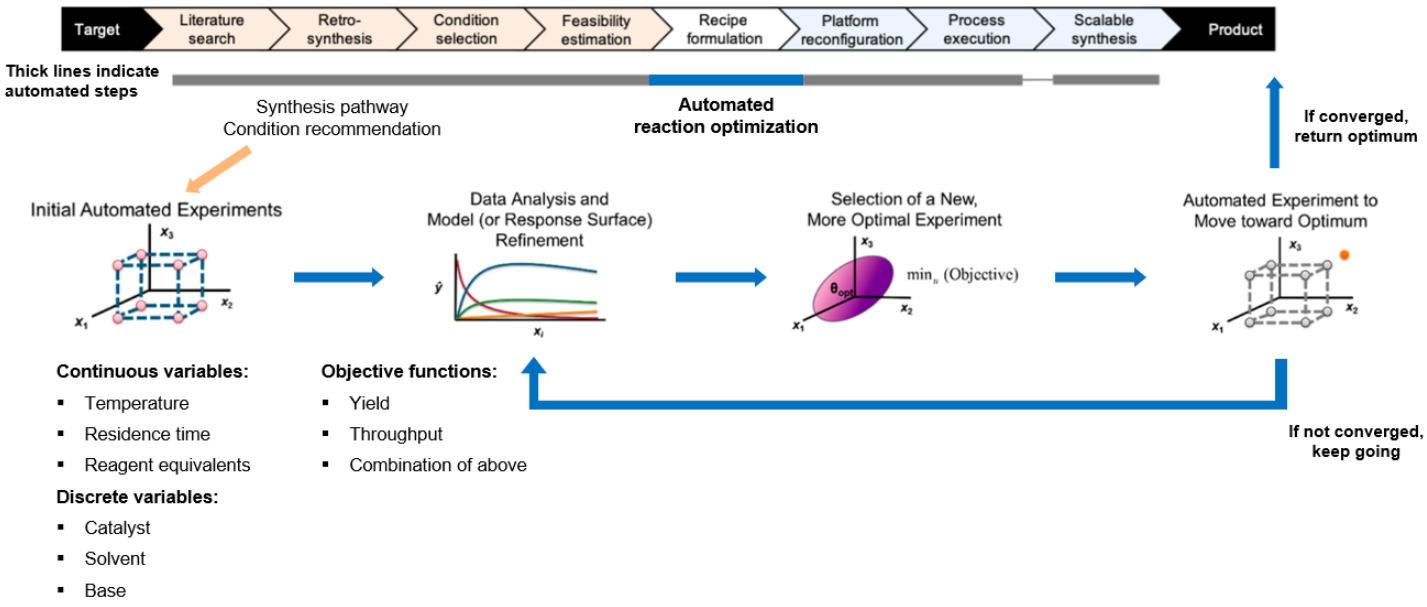


Figure 10. Workflow for developing an optimized synthesis for a target compound. Using synthesis pathways and condition recommendations generated by ASKCOS, the robotic platform runs automated experiments guided by a reaction optimization algorithm. Graphics from: Coley et al., *Science* **365**, 557, 2019, and B. Reizman, K. F. Jensen, *Acc. Chem. Res.* **49**, 1786, 2016.

An algorithm developed in the Jensen Group (*React. Chem. Eng.* **3**, 301–311, 2018) was chosen due to its ability to handle discrete variables (e.g., catalyst, solvent) in addition to continuous variables (e.g., temperature, residence time), as well as different objective functions (e.g., yield, throughput). Experiments are generated using optimal Design of Experiments (DoE). In the first round, a D-optimal design is generated to explore the design space using variable settings that try to maximize the amount of information obtained. Subsequent rounds of experimentation employ a G-optimal strategy where the goal is to minimize the model's uncertainty in the parameter estimates. A branch-and-bound strategy is used to remove poor-performing discrete candidates from subsequent experiments. While the algorithm was originally implemented for a droplet-based system for single-step reactions, we are working on generalizing it to multistep reactions. A graphical user interface has also been developed to allow the user to execute the optimization and monitor progress. Results are emailed to users as they come in, allowing researchers to follow progress remotely. In collaboration with Task 3, chemistry targets to demonstrate the new analytical and optimization capabilities of the platform were identified. Results from optimization campaigns for ribavirin and camostat were reported above under Task 3.

4 Task 4 Systems: Multistep Reaction Optimization in Continuous Flow

Prior work on algorithm-driven feedback optimization in continuous flow has focused almost entirely on single-step reactions. While there are several reports of multistep flow synthesis (J. Britton, C.L. Raston, *Chemical Society Reviews* **46**, 1250, 2017, C.A. Shukla, A.A. Kulkarni, *Beilstein Journal of Organic Chemistry* **13**, 960, 2017) where unit operations are telescoped together, reaction conditions were optimized manually. Apart from recent work by the Bourne group on feedback optimization of 2-unit reaction-separation (Clayton, A. D. *et al. Chem. Eng. J.* **384**, 123340 2020) and reaction-extraction sequences (Clayton, A. D. *et al. J. Flow Chem.* **10**, 199, 2020), an approach for simultaneous optimization of more complex multi-reaction processes involving both continuous and discrete reaction variables has yet to be demonstrated.

The Make-It robotic flow chemistry platform is well suited to carry out multistep processes since it is equipped with multiple reaction bays, process modules for both reactions and separations, and three analytical modules for FT-IR and LC/MS analysis at any point in a multistep sequence. A key capability for multistep optimization is the combination of tubular reactors that are available in different volumes (0.5, 1, 3 mL) and the improved Cartesian robot that can quickly and reliably reconfigure reactor volumes. In multistep flow processes here downstream residence time is fully specified by upstream flow rates and reagent stoichiometry, the ability to increase or decrease the volume of downstream reactors enables independent variation of downstream residence time without modifying upstream conditions.

To handle the increase in the number of variables for multistep processes, an open-source Bayesian optimization algorithm called Dragonfly (Kandasamy, K. *et al. J. Mach. Learn. Res.* **21**, 1 2020) was implemented. Bayesian optimization scales well with the number of variables, involves a mathematical model that is flexible to fit the curved surfaces often encountered in reaction profiles, and is capable of optimizing multiple objectives (e.g., yield, productivity, cost) simultaneously.

The cancer drug sonidegib was chosen as the chemistry example. Figure 11 shows a synthetic route and reaction conditions (reagents, solvent) that were proposed for this target by the publicly available computer-aided synthesis planning (CASP) tool ASKCOS. This multistep process involving four unit operations (S_NAr, nitro reduction, gas-liquid separation, amide coupling) enables utilization of the full range of process modules (homogeneous tubular reactors, heterogeneous packed bed reactor, membrane separator) available on the Make-It system. From an optimization point of view, continuous and discrete (leaving group X for S_NAr, promotor for amide coupling) variables are present, and the goal was to demonstrate that online analytics and feedback optimization can be used to fully specify process conditions (solubility testing is still performed offline to specify concentration, however) for CASP-generated synthetic routes.

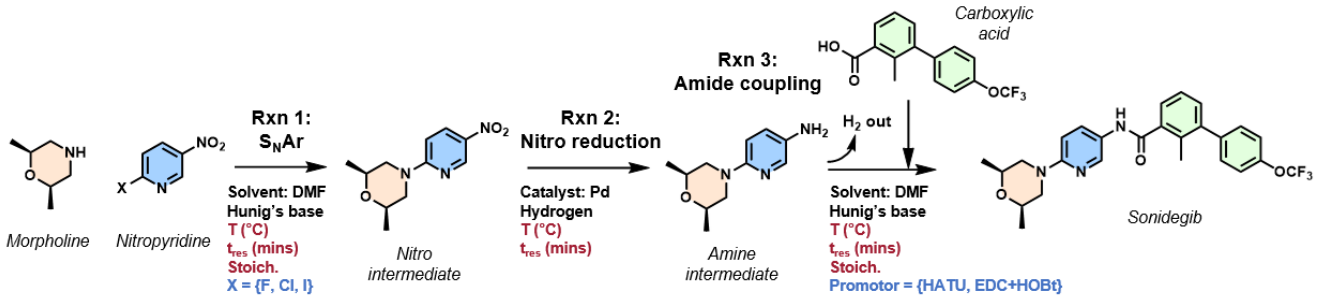
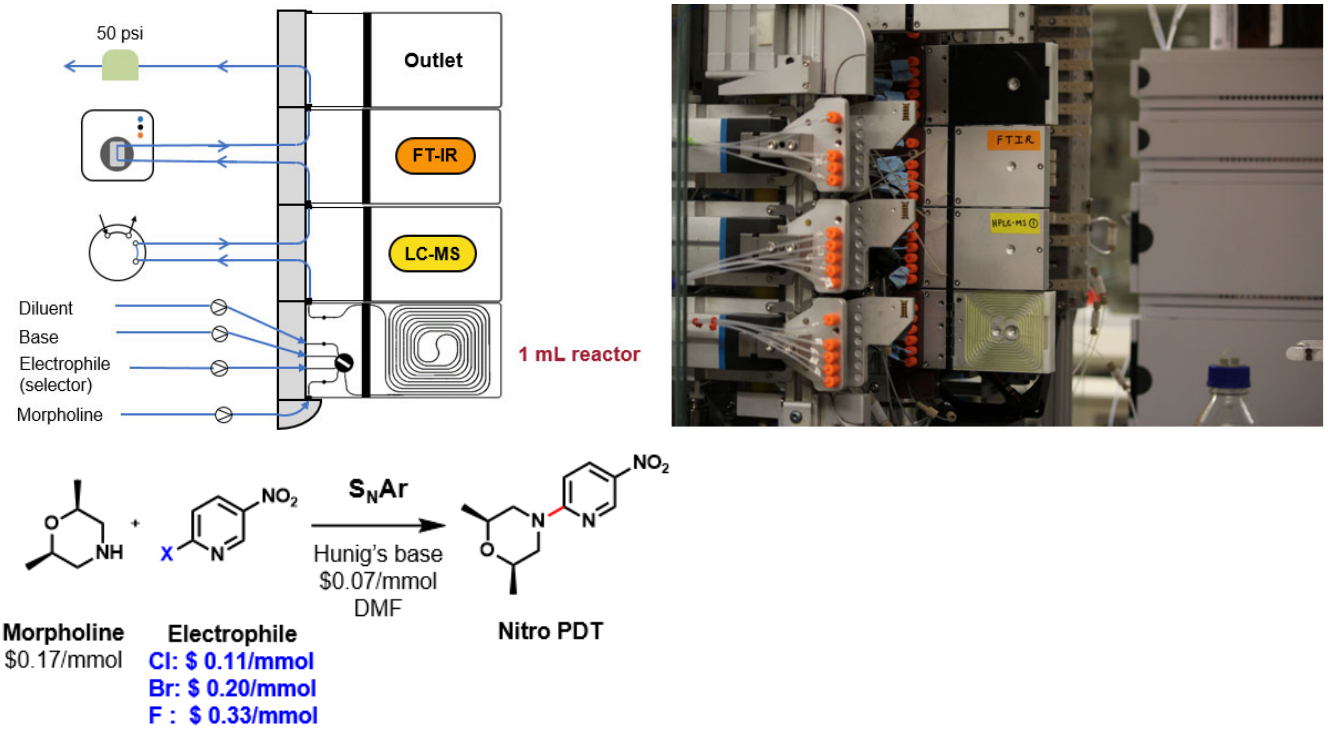


Figure 11. Synthetic route and reaction conditions proposed by CASP tool ASKCOS for sonidegib. **Continuous** and **discrete** variables to be optimized.

We initially attempted to fully telescope all three reaction steps. However, we found that the Pd catalyst for the second nitro reduction step deactivated within an hour in the presence of the base salt byproduct from the first SNAr step. The deactivation was independent of the catalyst support as both the silica and stainless steel supported Pd we tried underwent deactivation. Therefore, it was necessary to split up the three reaction process into a single-step SNAr optimization and a multistep nitro reduction-separation-amide coupling optimization.

Figure 12 shows the optimization setup for the single-step SNAr reaction. The reaction was carried out in a 1 mL tubular reactor with LC/MS for yield quantification followed by FT-IR for steady-state monitoring. An interesting feature of this example is that ASKCOS suggested three different leaving groups (Cl, Br, F) with fluorine being the best leaving group based on the known SNAr mechanism but chlorine being three times cheaper. Therefore, in addition to yield and productivity, we decided to include cost (which is a function of the leaving group and reagent equivalents) per mmol of product made as a third optimization objective since economics becomes relevant during process development.



Variables (5)	Values
Temperature	40–120 °C
Residence Time	1–10 mins
Morpholine Equiv.	1–1.2 eq.
Base Equiv.	1–1.2 eq.
Leaving Group X	Cl, Br, F (per ASKCOS)

Optimization Objectives (3)	Definition
Yield (%)	= PDT Conc. / Max Conc. x 100
Productivity (g/h)	$\propto \frac{\text{Yield PDT}}{\text{Residence Time}}$
Cost ($ reagents/mmol PDT made)	$= \frac{\text{Cost Reagents}}{\text{mmol PDT}}$ $= \frac{(\text{Cost X} * 1\text{ eq}) + (\text{Cost morpho} * \text{Morpho eq}) + (\text{Cost base} * \text{Base eq})}{\text{Yield PDT}}$

Figure 12. Optimization set setup, variables, and objectives for the single-step SNAr reaction.

The objective values for the 30 experiments (9 random initial + 21 refinement experiments) run during the SNAr optimization campaign are shown in Figure 13. The data points color-coded by leaving group are separated along the vertical cost axis, with chloro being lowest (cheapest) and fluoro being highest

(most expensive). While fluoro gave the highest yields (furthest to left), both the bromo and chloro gave yields above 90% at high temperatures and reagent equivalents. Conditions near the shortest residence time of 1 minute led to high productivity with a slight decrease in yield. Considering all three objectives, the best condition employed the chloro (94% yield, 5.7 g/h, $0.41/mmol) which was 3x cheaper but only slightly lower yielding than the fluoro.

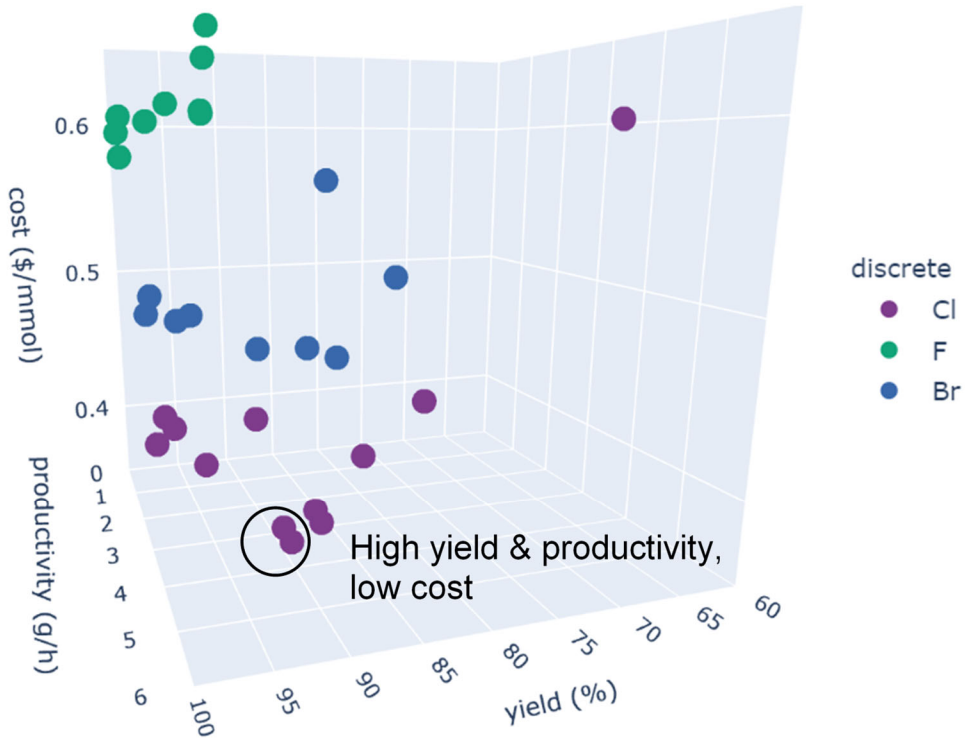


Figure 13. Scatter plot of objective values (yield, productivity, and cost) for 30 experiments performed in SNAr optimization campaign. Points are color-coded by leaving group (**Cl**, **F**, **Br**).

Next, the downstream multistep process was optimized (Figure 14). The process was set up in a convergent configuration where the liquid amine stream after the nitro reduction and hydrogen membrane separation meets the activated carboxylic acid from the parallel activation reactor to form the amide bond in the sonidegib product. The platform's flexibility to accommodate a convergent route was key to controlling the order of addition and minimizing a side-reaction between the amine and coupling promotor. Yield and productivity were optimized with respect to 4 continuous and 1 discrete variable. The residence time of the last amide coupling step was varied by automated robotic reconfiguration between a 1 mL and 3 mL reactor.

The objective values for the 15 experiments (8 random initial + 7 refinement experiments) are shown in Figure 15. HATU was a better coupling promotor than EDC (with 1:1 HOBt as an additive), giving close to 93% yield by LC and 7.5 g/h productivity at the optimum condition with the 3 mL reactor. Crucially, the FT-IR module placed after the nitro reduction showed stable Pd catalyst activity without deactivation over two days of operation. Meanwhile, the FT-IR module after the activation reactor revealed that the formation of the activated ester with HATU as the coupling promotor was complete within 1 minute, whereas activation with EDC was slow in forming the O-acylisourea reactive intermediate.

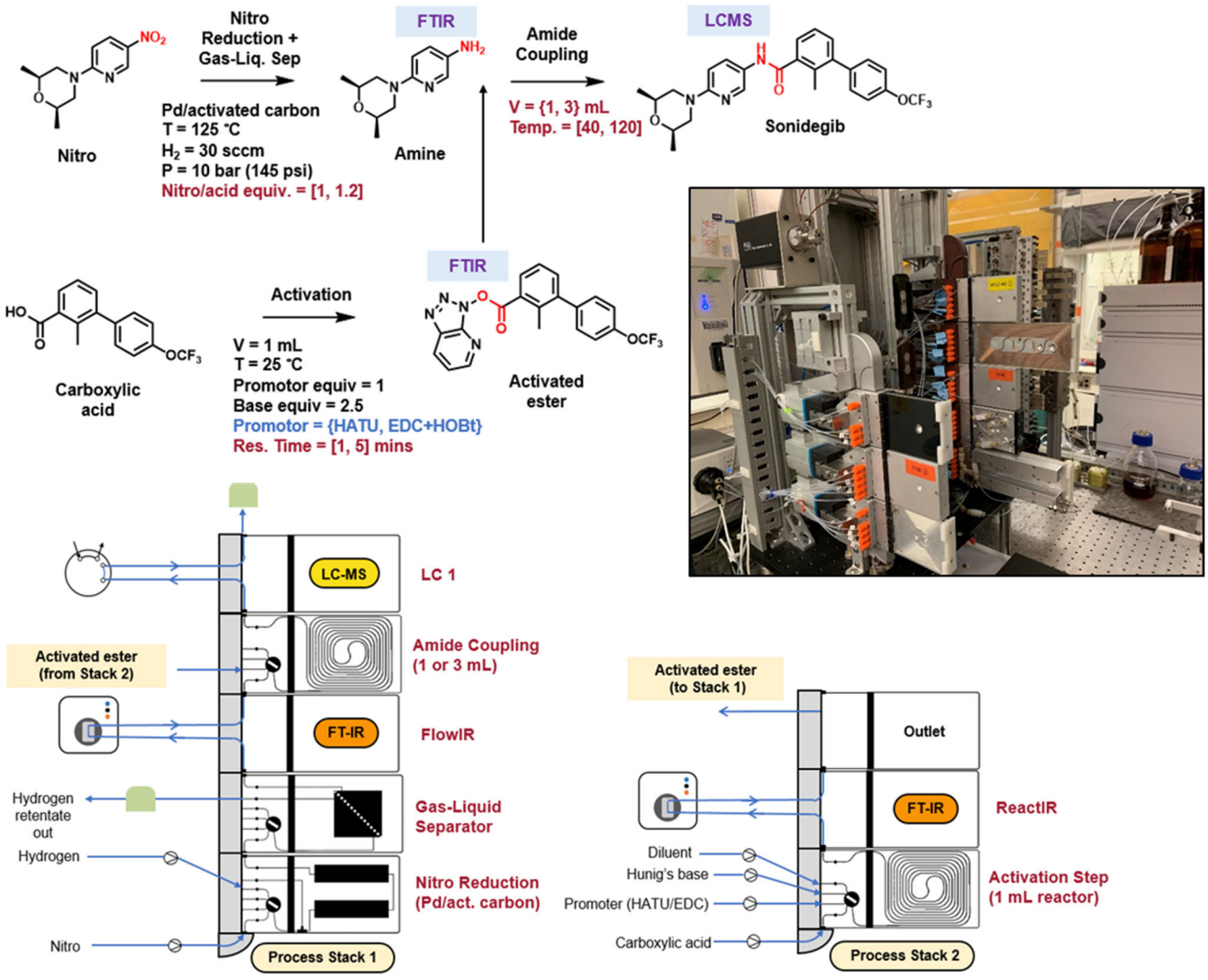


Figure 14. Optimization setup for convergent multistep nitro reduction-separation-amide coupling sequence for sonidegib synthesis

While the conventional approach for collecting data in flow reactors is to wait until steady-state is reached, an alternative approach involves dynamically varying process conditions (flow rate, temperature) in a controlled manner and gathering data during the transient. When applicable, this approach is fast, data-rich and material-efficient. This approach was implemented on the Make-It system for dynamic variation of multiple variables simultaneously (residence time, reagent equivalents, and temperature). A manuscript describing the design rules for dynamic trajectories and experimental results has been published (F. Florit, *et al. React. Chem. Eng.* **6**, 2306-2314, 2021).

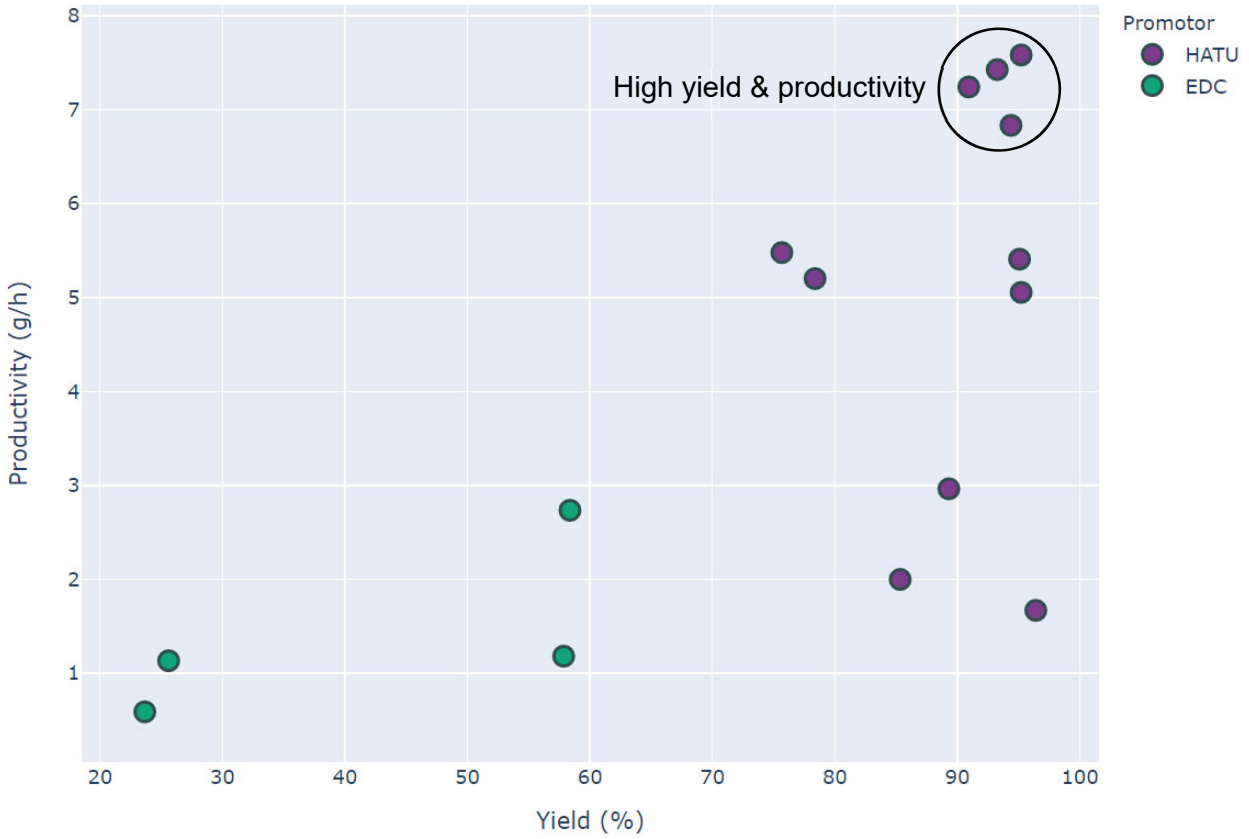


Figure 15. Scatter plot of objective values (yield, productivity) for 15 experiments performed in multistep optimization campaign for sonidegib synthesis. Points are color-coded by amide coupling promotor (**HATU**, **EDC+HOBt**).

5 Task 4 Systems: Army Research Laboratory (ARL) System

A modular system for benchtop continuous flow synthesis that allows user flexibility and reliability in the pursuit of challenging chemistry was developed at MIT and successfully transitioned to the U.S. Army Research Laboratory (ARL) for the purpose synthesizing the energetic materials. Figure 16 shows the completed system.

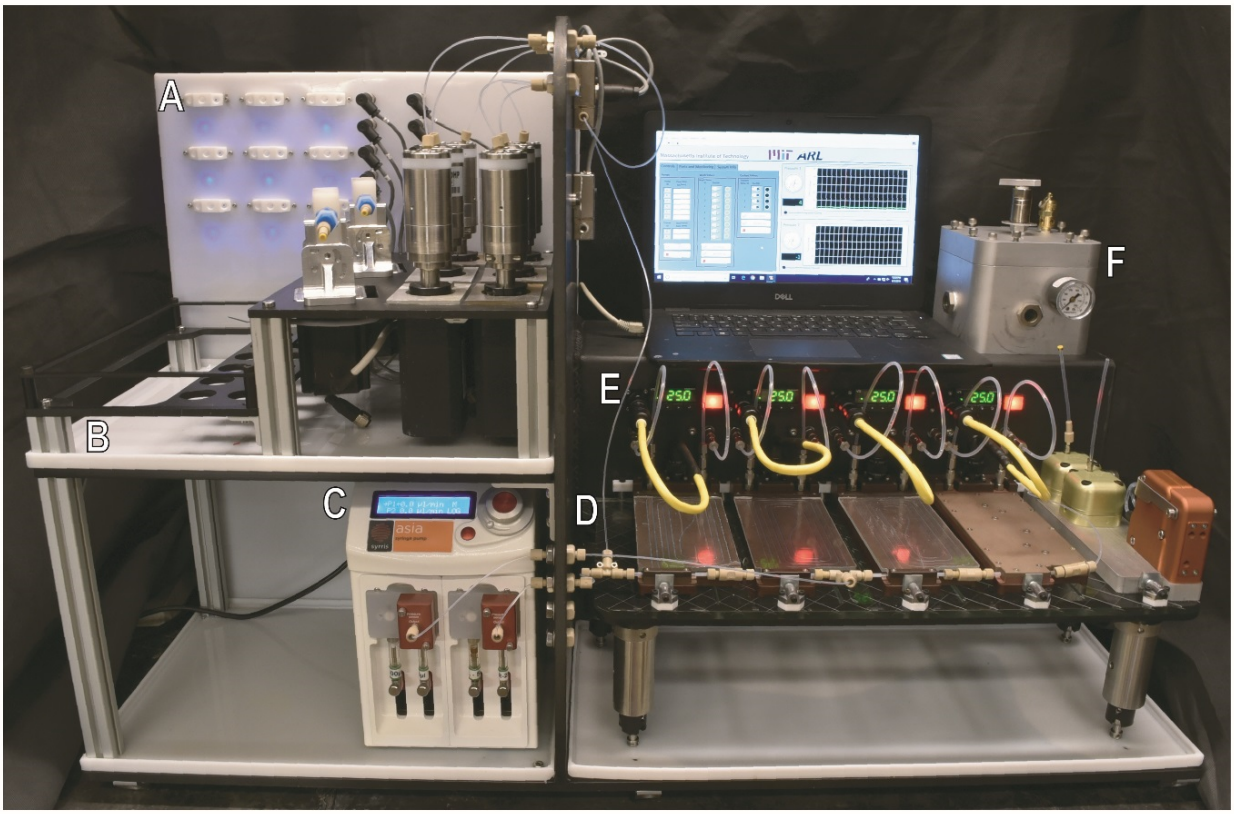


Figure 16. ARL benchtop system (1.02 x 0.46 x 0.66 m) overview. A) Electronics enclosure plus 3-way valves. B) Reagent storage. C) Pumps, including oscillating pumps, continuous positive displacement pumps, and a syringe pump. D) Reaction platform and reactors. E) Thermal management controls. F) Pressurized collection vessel.

The system is equipped with modular reactors that can be connected in series based on reaction needs. Standard reactor modules with different volumes can safely operate in the range of -40°C to 180°C and up to 15 bar. A continuous oscillatory baffled reactor (COBR) was designed to assist with the handling of suspended solids formed during a reaction by providing mixing through the formation of eddies under oscillating flow. The system also contains a membrane separator (Zaiput) for liquid-liquid and gas-liquid separation, and a pressurized collection vessel (PCV) enabling sample collection at elevated pressure and under an inert atmosphere. M6 Milligat positive displacement pumps (VICI) served as the primary pumping method, while a high-end syringe pump (Syrris Asia) with only PTFE and glass wetted parts was used for pumping concentrated acids like fuming nitric acid. A LabVIEW user interface allows the user to control the system hardware and monitor pressure during a reaction. The software has an over-pressurization safety feature that stops pumps in an emergency should the pressure ever exceed the adjustable soft cap or rigid hard cap.

A series of diverse chemical reactions were performed as validation and stress testing for various components of the design, as shown in Figure 17. Reactions were selected in an effort to display the versatility of the system in handling an array of chemical processes that a typical research laboratory would conduct with traditional batch equipment. The completed system was transferred to Dr. Jesse Sabatini at ARL.

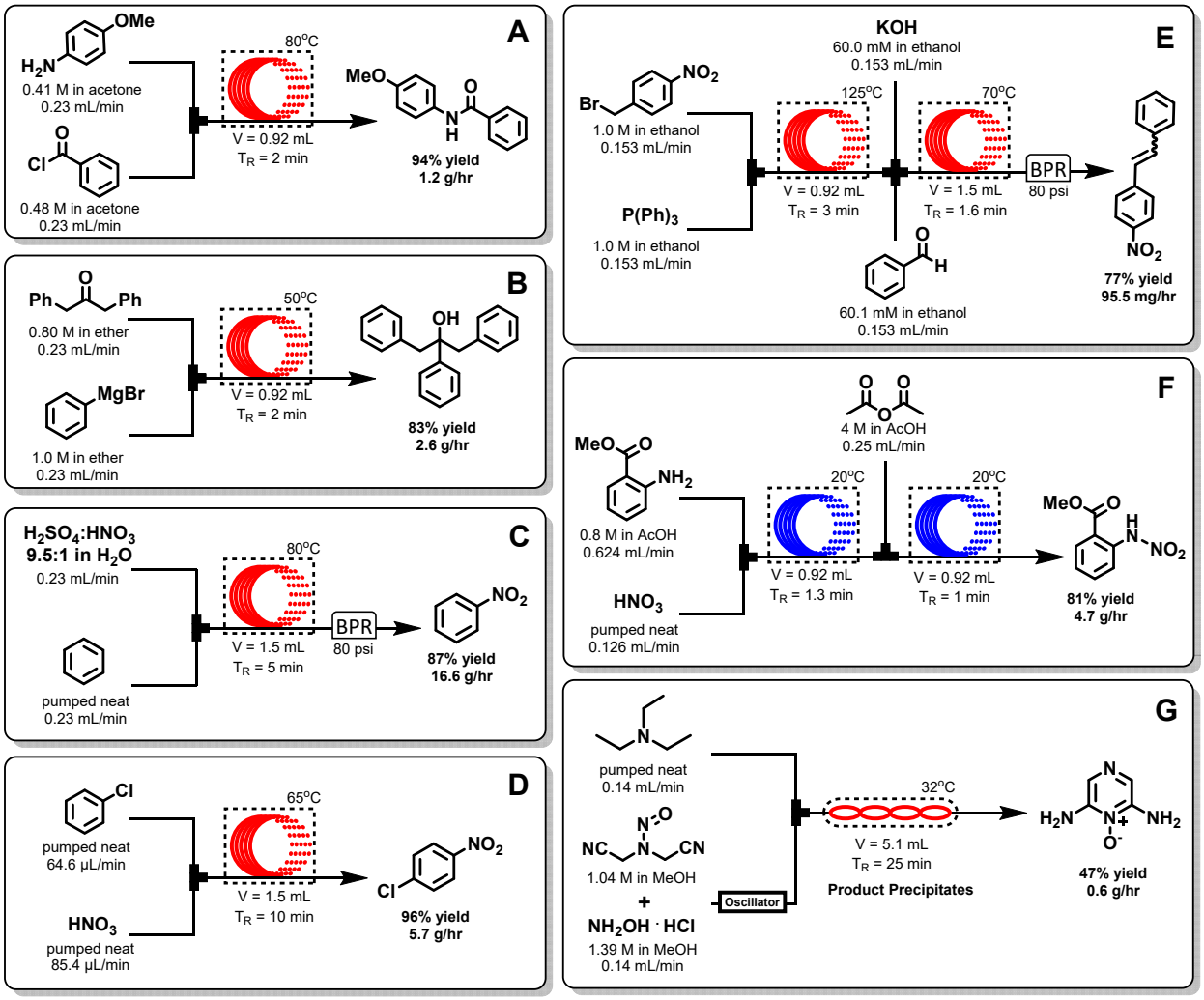


Figure 17. Chemical reactions run on ARL system. A) Amide coupling. B) Grignard reaction under anhydrous conditions. C) Aromatic nitration with acid mixture. D) Aromatic nitration with fuming nitric acid. E) Coupled Wittig reaction. F) Nitroamine formation with fuming nitric acid. G) Synthesis of DAPO (LLM-105 precursor) with COBR reactor.

6 Task 4 Systems: Process Control

6.1 Offset-free stochastic model predictive control based on polynomial chaos theory and symbolic arithmetic

Stochastic model predictive control (SMPC) has the ability to explicitly take model uncertainties and input and output constraints into account in real-time optimal control calculations. We have developed a suite of stochastic model predictive control (SMPC) algorithms that achieve offset-free control, which

we have proved theoretically and demonstrated in simulations.[2] Additionally, we have proved that the past SMPC formulation and other seemingly reasonable SMPC formulations are not offset free. The theoretical proofs are based on deriving the z-transforms of the controllers and assessing analytically whether their denominators contain an integrator (see paper[2] for details). All of the offset-free stochastic model predictive control algorithms have low on-line computational cost and are available for use. We do so by using polynomial chaos theory (PCT), which quantifies the effects of probabilistic uncertainties on the states and outputs of linear and nonlinear dynamical systems. In order to automate the construction of exact PCT representations for dynamical systems described by ordinary differential equations (ODEs), we propose a dual path approach using symbolic methods. The methods automatically construct exact PCT representations for both nonlinear dynamical systems and their linearizations with respect to the states. The implementation is done using the Matlab Symbolic Toolbox and demonstrated in a case study on the SMPC of an automated continuous-flow system for multi-step chemical synthesis.[3]

6.2 Optimization and control of plant startup

Optimization of plant startup is of general interest as a way to reduce waste and achieve on-spec product rapidly. Especially, in on-demand manufacturing where short campaigns often arise due to a relatively low volume of products, optimal startup can be even more significant compared to other types of continuous manufacturing. We demonstrate the successful computational implementation of a startup methodology that consists of an optimization step, where an optimal startup trajectory is obtained offline using dynamic optimization (DO), followed by a closed-loop control step using quadratic dynamic matrix control (QDMC) to drive the plant as close to that trajectory as possible, which can be real-time implementable.[4]

A plant-wide first-principles model can be automatically constructed from the process flowsheet knowing the unit operations involved. Parameter identification can take place in a smaller scale system offline (e.g., a droplet-based system[5]). Such a model can be used to optimize a metric of the plant performance during startup. Since the system is described by partial differential algebraic equations, which need to be discretized, very large model sizes are generated. An optimization of such a system is a nonlinear program (NLP) of very large size which is computationally expensive. However, the NLP for startup can be solved offline given enough time before the beginning of the plant operation. Additionally, the NLP for a system of low discretization can be solved to generate an approximation of the solution in reasonable time, if required. For the objective function we propose a multi-objective formulation that accounts for different performance metrics of the operation. We implement this approach in the case of upstream atropine synthesis[6] (Figure 18) and we consider as metrics the total amounts of product and waste during startup. The obtained trajectory is shown in Figure 19 (black dashed line) and is controlled using QDMC. Additionally, the closed-loop control performance under parametric uncertainty is examined (Figure 20). Without closed-loop control the operation would be significantly suboptimal under parametric uncertainty. The successful implementation of linear control such as QDMC

[2] M. von Andrian and R. D. Braatz, "Offset-free input-output formulations of stochastic model predictive control based on polynomial chaos theory," Proceedings of the American Control Conference, pp. 360-365, July 2019.

[3] M. von Andrian and R. D. Braatz, "Stochastic dynamic optimization and model predictive control based on polynomial chaos theory and symbolic arithmetic," Proceedings of the American Control Conference, pp. 3399-3404, July 2020

[4] A. Nikolakopoulou, M. von Andrian, and R. D. Braatz, "Fast model predictive control of startup of a compact modular reconfigurable system for continuous-flow pharmaceutical manufacturing," Proceedings of the American Control Conference, 2778-2783, July 2020.

[5] Y.-J. Hwang, C. W. Coley, M. Abolhasani, A. L. Marzinzik, G. Koch, C. Spanka, H. Lehmann, and K. F. Jensen, "A segmented flow platform for on-demand medicinal chemistry and compound synthesis in oscillating droplets," Chemical Communications, vol. 53, no. 49, pp. 6649-6652, 2017.

[6] A.-C. Bédard, A. R. Longstreet, J. Britton, Y. Wang, H. Moriguchi, R. W. Hicklin, W.H. Green, and T. F. Jamison, "Minimizing E-factor in the continuous-flow synthesis of diazepam and atropine," Bioorganic & Medicinal Chemistry, vol. 25, no. 23, pp. 6233-6241, 2017.

during a nonlinear operating regime of the plant is enabled by using appropriately selected linear models as we proposed in reference[7].

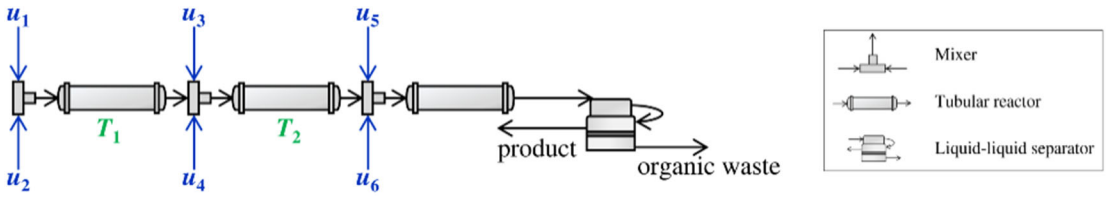


Figure 18. Process flowsheet for atropine synthesis.

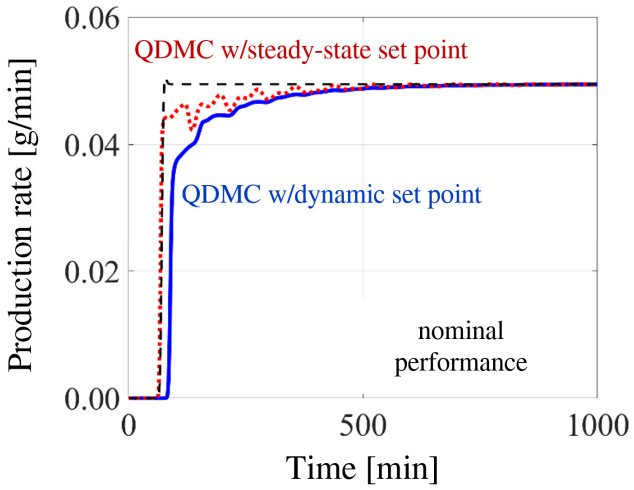


Figure 19. Control of the upstream atropine synthesis nominal plant during startup. The optimal startup trajectory obtained using DO is shown in black dashed line. QDMC is used to control the system during startup either using a dynamic set point (blue solid line) or the steady-state value for the set point (red dotted line). To obtain either set point a DO needs to be solved.

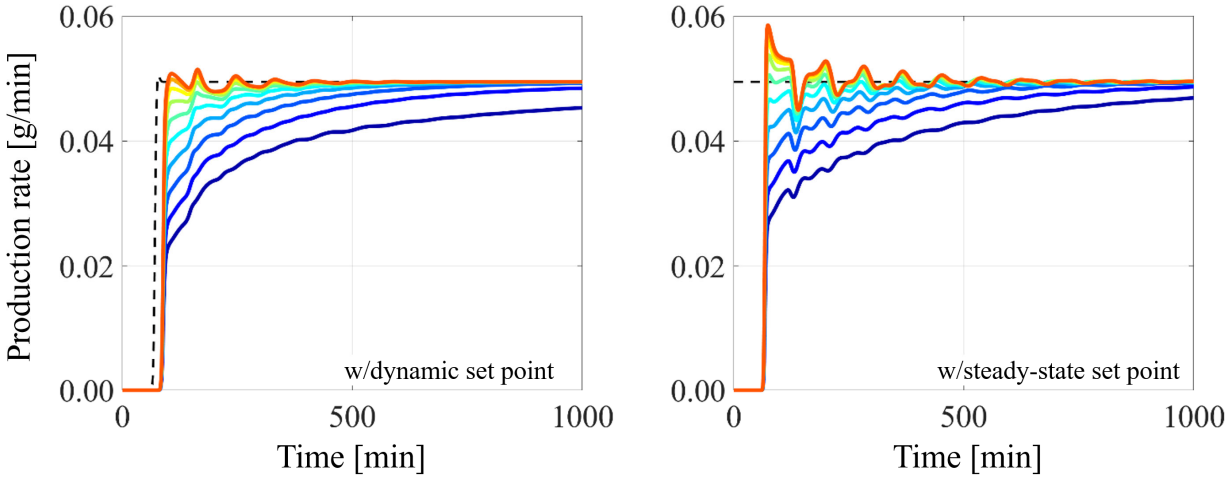


Figure 20. QDMC control for startup in the presence of parametric uncertainty in the kinetic constant of the atropine production reaction, for different values of the uncertain parameter.

[7] A. Nikolakopoulou, M. S. Hong, and R. D. Braatz, "Output feedback control of dynamic artificial neural networks using linear matrix inequalities," Proceedings of the Conference on Decision and Control, December 2020. Accepted

6.3 Nonlinear feedback control and estimations with performance guarantees for Dynamic Artificial Neural Network models

Dynamic Artificial Neural Networks (DANNs) can be used instead of first-principles models to model processes, due to their ability to universally approximate nonlinear dynamics. Controlling nonlinear models such as DANNs is often approached from a nonlinear predictive control (NMPC) perspective. NMPC can be computationally expensive and also comes with limited guarantees, i.e. regarding stability and performance. We propose a way to construct nonlinear feedback controllers for DANNs with guaranteed closed-loop performance and stability. We first transform the mathematical representation of the DANN into a more convenient formulation known as diagonal norm-bounded linear differential inclusion (DNLDI). We then use Lyapunov theory and linear matrix inequalities (LMIs) to come up with sufficient criteria that guarantee closed-loop stability and performance. We also provide criteria for DANNs stabilizing and optimal state and output estimators. We successfully demonstrate the described methodology in a pH control example,[8] also shown in Figure 21.

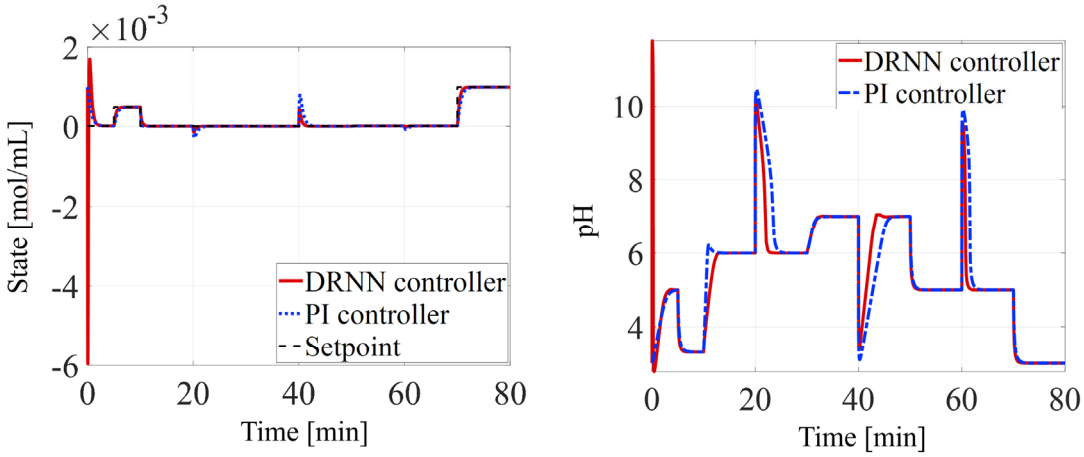


Figure 21. pH control for simultaneous set point tracking and disturbance rejection. The DRNN controller with a proven closed-loop stability guarantee performs very well, with faster tracking and improved disturbance rejection than a standard controller.

6.4 Data-driven nonlinear surrogate modeling

Model predictive control (MPC) is the most widely used advanced control methodology due to its advantages like constraint handling. The type of model used to predict the future behavior of the plant and generate control input sequences determines if the MPC will be linear or nonlinear. Linear MPC such as QDMC is fast but also has some shortcomings when used to control nonlinear systems. Nonlinear MPC (NMPC) on the other hand can be more accurate for nonlinear processes but is very computationally expensive, prohibiting its real-time implementation. We are exploring ways to reduce NMPC associated computational costs. In that direction, we are considering NMPC formulations based on nonlinear input-output models. We are looking into nonlinear autoregressive and moving average with exogenous input (NARMAX) models, with a focus on polynomial models, such as Volterra series. These models have a high number of terms that need to be determined so we are exploiting machine learning methodologies to simultaneously perform regression while reducing the number of terms. Such models, also known as "surrogate models," can be constructed from simulated data obtained from first-principles models. We have successfully constructed surrogate models for unit operations that include a continuous stirred tank reactor (CSTR) and a plug flow reactor (PFR) (Figure 22). We are optimizing sampling to minimize the

[8] A. Nikolakopoulou, M. S. Hong, and R. D. Braatz, "Output feedback control of dynamic artificial neural networks using linear matrix inequalities," Proceedings of the Conference on Decision and Control, December 2020. Accepted

data required for accurate model construction, extending the modeling to multiple-input multiple-output (MIMO) systems, and integrating the models with NMPC.

6.5 Additional results

We completed an extension of fast stochastic model predictive control to unstable dynamical systems,[9] which coupled an automatically designed pre-stabilizing feedback controller with our fast stochastic model predictive control algorithm that applies to stable systems. We completed a thorough investigation of an alternative approach to fast stochastic model predictive control for the control of modular on-demand chemical systems.[10] Although the alternative approach was effective for the atropine manufacturing system, the design procedure was much less automated than our fast stochastic model predictive control algorithms.

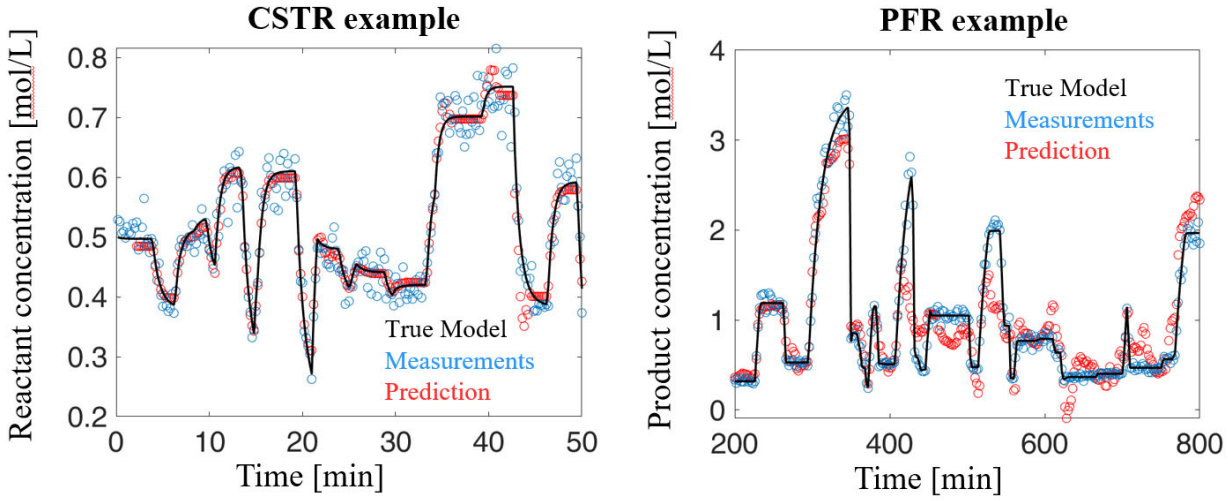


Figure 22. Surrogate model (red) obtained from data generated using a first-principles model (black) with noise (blue) using Volterra series with elastic net regression.

7 Task 4 Systems: Purification and Isolation

The state-of-the-art crystallization process development is still based on the use of extensive experimental screenings and empirical models for process simulation and optimization. Recent advances in crystallization process development focus on acquiring more information on the process by adding new methods and instruments, rather than simplifying the existing approaches to minimize the usage of time and raw materials. This approach is desirable when a tight control over particle size distribution or polymorphism is required, and when the systems are especially challenging. However, a general application for most systems can become wasteful and unpractical. The following work is based on two concepts:

1. Moving from empirical models to mechanistic models can simplify process development and limit the number of experiments required for regression, as the general form of the expression is already known and we know what parameters we are aiming for.

[9] M von Andrian, and R. D. Braatz, "Fast stochastic model predictive control of unstable dynamical systems," Proceedings of the International Federation of Automatic Control World Congress, July 2020.

[10] D. P. Piñeiro, A. Nikolakopoulou, J. Jäschke, and R. D. Braatz, "Self-optimizing control of a continuous-flow pharmaceutical manufacturing plant," Proceedings of the International Federation of Automatic Control World Congress, July 2020.

2. For systems where a tight control of particle size is not necessary (pharmaceutical intermediates, API for liquid formulations, bulk chemicals and some fine chemicals), an accurate determination of crystallization kinetics is not necessary. We can bypass or simplify kinetic screening by using two crystallization stages and a reasonable residence time (over 30 min/stage). Because over 90% of time and material is usually spent on characterizing kinetics, this approach can lead to significant reductions in the time and material requirements for crystallization development.

7.1 Solubility Models for Antisolvent Crystallization

The viability of available mechanistic solubility models was addressed as a first step to minimize the number of solubility data points required for calibration. The most viable expression was deemed to be the two-suffix expression derived by Williams and Amidon:[11]

$$ln\, x_m = z_1\, ln\, x_1 + z_2\, ln\, x_2 + A z_1 z_2 \frac{q_{sol}}{q_1}$$

This expression represents the mole fraction solubility of the solute in a binary solvent mixture (x_m) as a function of the volume fractions of the solvents in solution (z_1, z_2), the solute solubility in the pure solvents (x_1, x_2), the molar volumes of the solute (q_{sol}) and the solvent (q_1), and an interaction term A. In the original papers, the authors recommend estimating the interaction term from vapor-liquid equilibrium data. However, we could not successfully predict solubilities at the level of accuracy that we would need to predict crystallization yields, presumably due to additional interactions in crude systems that are highly non-ideal. Instead, we evaluated the approach of experimentally determining A from regression using the pure solvent solubilities and one or more solubility points at intermediate solvent compositons. Results for the Diazepam system in acetone-water, shown in Figure 23, indicate that a single data point may be sufficient to determine the interaction term. We later used this approach to measure the solubility of several pharmaceuticals in organic solvent mixtures, including Ketoconazole, Ibuprofen and Azithromycin. While this expression works well for most systems, reliance on a single data point is risky and prone to poor predictions due to the propagation of experimental errors. As a rule-of-thumb, we recommend measuring the solubility in pure solvents and regress the parameter A based on 2-3 experimental data points at intermediate solvent fractions between 0.4 and 0.8. This approach ensures that the expression is valid for the considered system and reinforces the accuracy of the solubility models.

7.2 Determination of an Operating Space for Continuous Crystallization

We developed a general method for the selection of an operating range for multistage continuous crystallization, where users are presented with an intuitive plot of attainable yields and suspension densities for the selection of solvent compositions on each crystallization stage. Addition of an antisolvent can increase yield through attainment of a lower solubility. However, because the system is being diluted too, crystallization may not occur at certain conditions, or the suspension concentrations (and thus the productivity and the stability of the crystallizer against supersaturation fluctuations or fouling) can be affected. These methods consider both effects for the selection of the operating space. First, the user selects a solvent composition and a solute concentration for the feed stream (aided by solubility data at solvent volume fraction = 1). Then, the maximum attainable yields and maximum suspension densities are calculated from the solubility data and the crystallizer mass balance (Figure 24):

[11] Williams, N. A.; Amidon, G. L. Excess Free Energy Approach to the Estimation of Solubility in Mixed Solvent Systems I: Theory. *J. Pharm. Sci.* **1984**, *73* (1), 9–13.

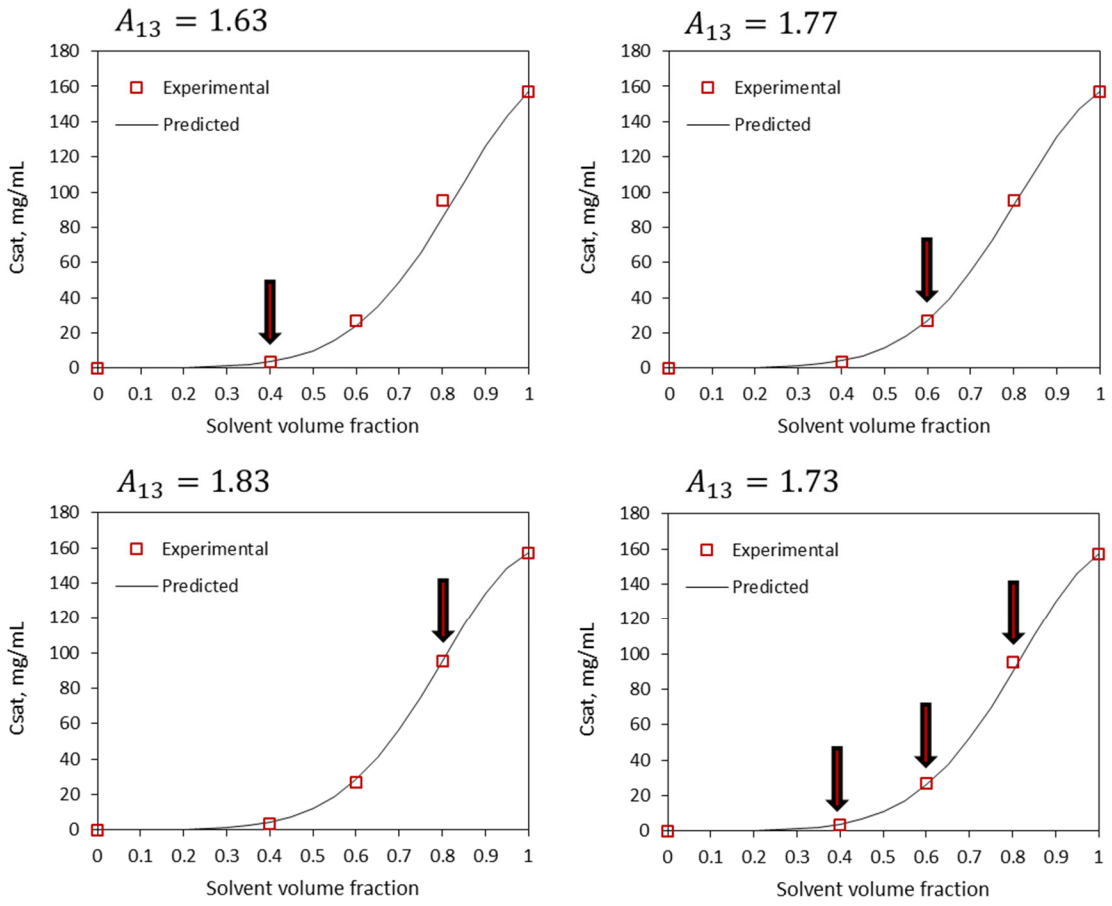


Figure 23. Estimations of mixed-solvent solubility of Diazepam in acetone-water mixtures using the Williams and Amidon expression and one data point at an intermediate solvent composition (top left, top right, bottom left). Results from the regression using all the available data (bottom right).

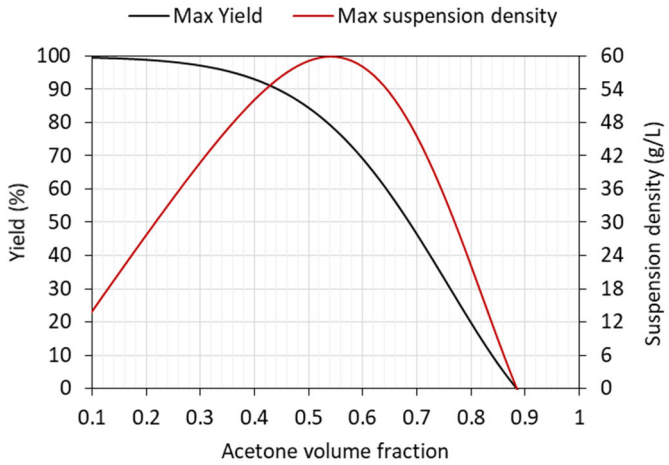


Figure 24. Maximum attainable yields and suspension densities for crystallization of Diazepam from acetone-water. The considered feed contains 140 g/L Diazepam in pure acetone.

For antisolvent crystallization, we consider crystallization at room temperature and select the solvent fraction on each stage. The following rules of thumb are implemented for the selection of operating conditions:

- Crystallization is conducted in two MSMPR stages.
- Start by selecting the second stage (final product) conditions. You can focus on maximizing yield, productivity (max. suspension density), or a reasonable combination of both.
- First stage should have a maximum yield between 0.67 and 0.8 of the second stage yield, meaning that between 2/3 and 4/5 of the API will be crystallized in the first stage. Generally, higher suspension densities in the first stage increase the overall mass deposition rate and productivity.
- Do not use suspension densities below 20 g/L on any stage (very dilute suspensions can lead to fouling and system instability).
- Do not operate at solvent compositions that have shown aggregation or fouling during the solubility measurements.
- Both stages should have a similar residence time, which should be at least 30 min.

Note that the values in Figure 24 are for a hypothetical crystallizer with an infinite residence time (supersaturation tends to 0). In real systems, slow kinetics can lead to the system quickly deviating from equilibrium. However, for antisolvent systems that work with a low solubility, a small variation in mother liquor concentration is enough to significantly affect supersaturation and kinetics. For example, if we were to crystallize Diazepam at room temperature with a solvent composition of 25% acetone and 75% water, our equilibrium solubility would be 0.6 mg/mL. Starting with 140 g/L in the feed, and accounting for antisolvent dilution, we would have a total API concentration of 35 g/L and a maximum suspension density of 34.4 g/L (shown in Figure 24). If our crystallizer were to operate near equilibrium, e.g. at a supersaturation of 10%, the mother liquor concentration would be equal to 1.1 x 0.6 mg/mL, or 0.66 mg/mL. With a supersaturation of 60% (rarely seen in steady-state MSMPR without fouling), the mother liquor concentration would be 0.96 mg/mL. The second condition would have a driving force six times higher with the corresponding increase in mass deposition rate, but the operating yield would only drop from 98.1% to 97.2%.

The chosen set of operating conditions for MSMPR demonstration are provided in Figure 25. For this demonstration, we operated the cascade with a 60 min residence time in the first stage and a 40 min residence time in the second stage.

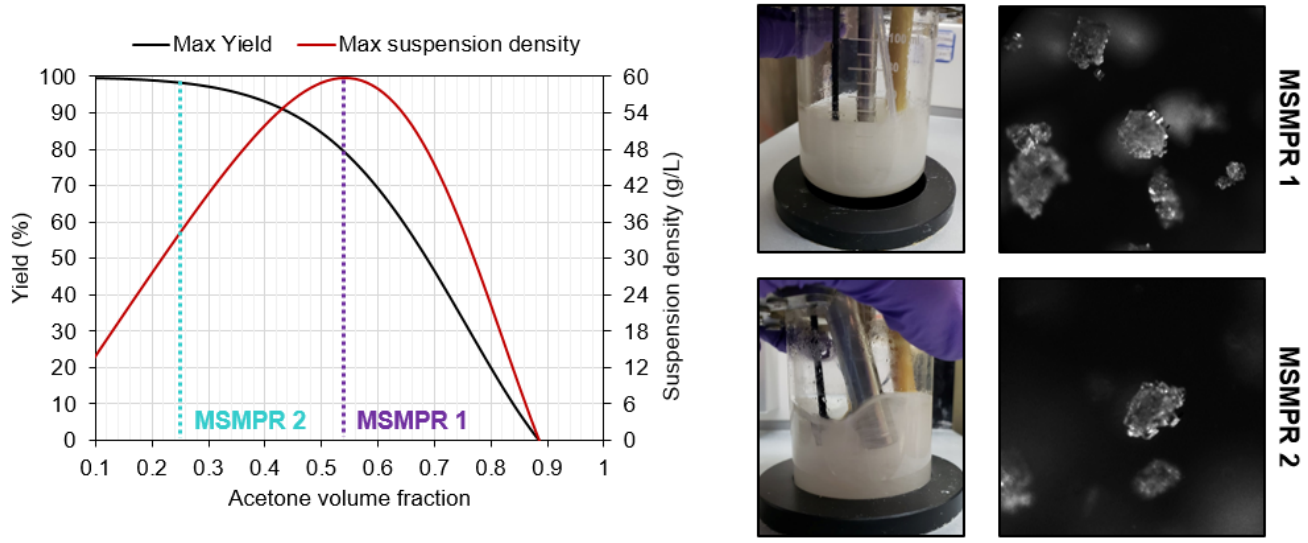


Figure 25. Operating conditions for 2 stage MSMPR crystallization of diazepam and experimental demonstration. As predicted in the diagram, stage 1 has a significantly larger suspension density.

7.3 Predictions of Crystallization Yield

Continuous crystallization processes of the Mixed Suspension Mixed Product Removal (MSMPR) type are self-regulated systems that operate at near-equilibrium conditions. Small variations in kinetics are quickly met with an opposite variation in supersaturation, keeping the system within a narrow operating range. An accurate determination of the roles of nucleation and crystal growth is critical for the prediction of particle size at the steady state, as it has been demonstrated from theory and in past publications.[12,13] However, the same may not be true for crystallization yield due to the low solubilities and the effects of small variations in mother liquor concentration on yield and supersaturation. For the initial demonstration, we bypassed the need for calibrating crystallization kinetics and assumed values for the nucleation and crystal growth constants. This brings the total development time to around one week and the material consumption down to a few grams of API, as only solubility data is needed. We took the model and kinetics (based on commercial Ciprofloxacin) from Capellades *et al* as an example of typical kinetics for organic materials,[13] and incorporated the Diazepam solubility parameters from the regression in Figure 23. A comparison of experimental and predicted results is provided in Figure 26. As expected, using the wrong kinetics led to estimation errors of an order of magnitude in the steady-state supersaturation, and appr. 200% error in estimations of mean particle size. However, because of the low solubility of the system, predicted yields are reasonable even for the first stage. For the second stage, the lower solubilities combined with the typical stability of an MSMPR that receives a seeded feed further improve the prediction of crystallization yield.

Stage 1, steady state:

Parameter	Experimental	Predicted
Solubility (mol/mol)	1.74 E-3	1.68 E-3
Mother liquor concentration (mg/mL)	16.2	18.4
Supersaturation	0.014	0.181
Yield (%)	81.6	79.1
$L_{4,3}$ mean size (µm)	311	130

Stage 2, steady state:

Parameter	Experimental	Predicted
Solubility (mol/mol)	6.61 E-5	6.35 E-5
Mother liquor concentration (mg/mL)	0.86	0.83
Supersaturation	0.03	0.04
Yield (%)	98.1	98.1
$L_{4,3}$ mean size (µm)	320	130

Figure 26. Experimental and predicted results of 2 stage MSMPR crystallization of Diazepam, using regressed solubility and assumed kinetic parameters.

The limitations of this approach were also assessed through a sensitivity analysis based on the assumed kinetics. For this study, we assumed that we had operated with a single stage MSMPR and a solvent composition of 25% acetone, 75% water. With a known feed concentration (140 g/L), the dependence of supersaturation, yield, and mean size on kinetics and residence time has an analytical solution.[13] The assumed kinetic parameters were increased and decreased by an order of magnitude to validate that this prediction is not a consequence of coincidentally similar parameters. Results, as shown in Figure 27, demonstrate our original hypothesis: accurate determinations of crystal size require good kinetic data,

[12] Schall, J. M.; Mandur, J. S.; Braatz, R. D.; Myerson, A. S. Nucleation and Growth Kinetics for Combined Cooling and Antisolvent Crystallization in a Mixed-Suspension, Mixed-Product Removal System: Estimating Solvent Dependency. *Cryst. Growth Des.* **2018**, *18* (3), 1560–1570

[13] Capellades, G.; Wiemeyer, H.; Myerson, A. S. Mixed-Suspension, Mixed-Product Removal Studies of Ciprofloxacin from Pure and Crude Active Pharmaceutical Ingredients: The Role of Impurities on Solubility and Kinetics. *Cryst. Growth Des.* **2019**, *19* (7), 4008–4018.

but for systems presenting a low solubility and reasonable residence times, yields can be predicted based on a simple mass balance and approximated kinetics.

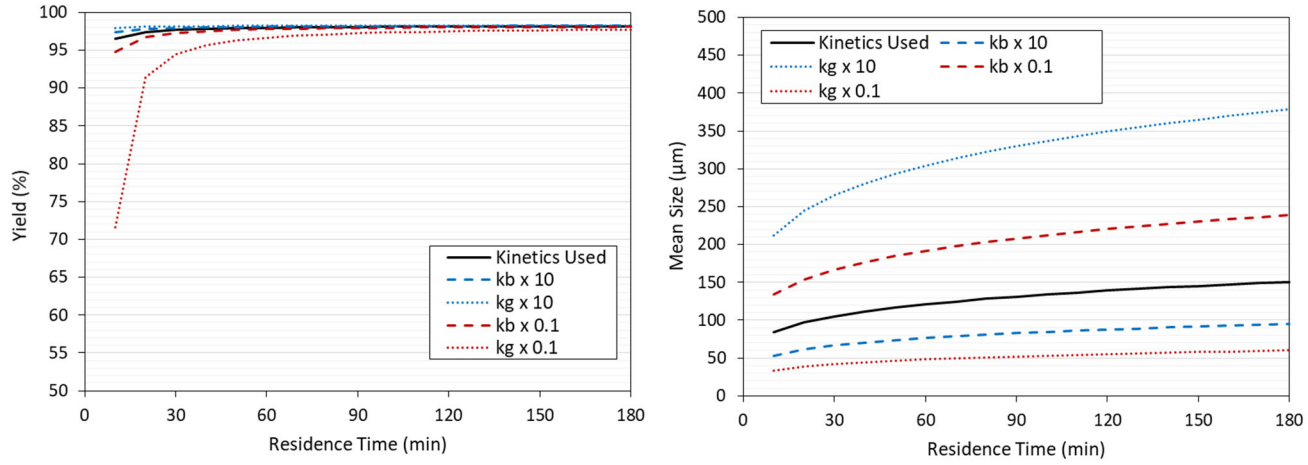


Figure 27. Sensitivity of steady-state crystallization yields and volume-based mean crystal size on the kinetic factors for crystal growth (k_g) and nucleation (k_b)

7.4 Crystallization as a Purification Process

In the context of the Make-It project, it is necessary that crystallization process development accounts for purification on top of attainable yields. At the same time, the design of a nearly-optimal continuous crystallization process will demand some kinetic data, both for the approximation of particle size distributions and for the optimization of residence times for maximum productivity. Consequently, we sought for approaches that allow for the incorporation of purification screening and kinetic screening within the solubility tests, or by reusing material from the solubility work. A workflow was developed for the general development of continuous crystallization processes for purification (Figure 28). State-of-the-art workflows base the selection of operating conditions on solubility alone, and require an extensive kinetic screening of bulk crystallization, often using several kg of API in the process. Our development philosophy is based on maximizing the information collected during the early solubility screenings (mL and µL scale) so that, by the time we select a crystallization solvent and temperature, we have information on impurity incorporation, crystal habit, aggregation/fouling tendency, and solvent-dependent kinetics. By selecting a nearly optimal solvent composition prior to our bulk crystallization studies, we can limit these to a single experiment where the kinetics (dependent on solvent and temperature) are regressed for the prediction of particle size as a function of residence time. Contrary to traditional development methods, we don't have to screen the effect of solvents and temperature on kinetics because we can direct our bulk crystallization efforts to these nearly optimal conditions.

The first step in the workflow is to conduct solubility and distribution coefficient screening in pure solvents (Figure 29). This is done by preparing duplicate samples at 40 °C containing a slight excess of solid crude in a saturated solution. Solid crude is added to the pure solvents until this saturation condition is met. Then, the samples are aged 24 hours to ensure equilibrium and filtered. The filtered solution is analyzed via HPLC and divided in 3 vials. The first will be used for kinetic screening, and the other two for solubility and impurity incorporation screening at 25 °C and 10 °C. The latter samples are cooled from a clear solution (now with known concentrations of solute and impurities) and aged 48 hours. Filtration and HPLC analysis reveals both the solubility of the solute at 25 °C and 10 °C, as well as the impurity incorporation via a mass balance. At the same time, the excess solids is collected for XRPD analysis to elucidate the crystal form that was obtained and relate it to the solubility data.

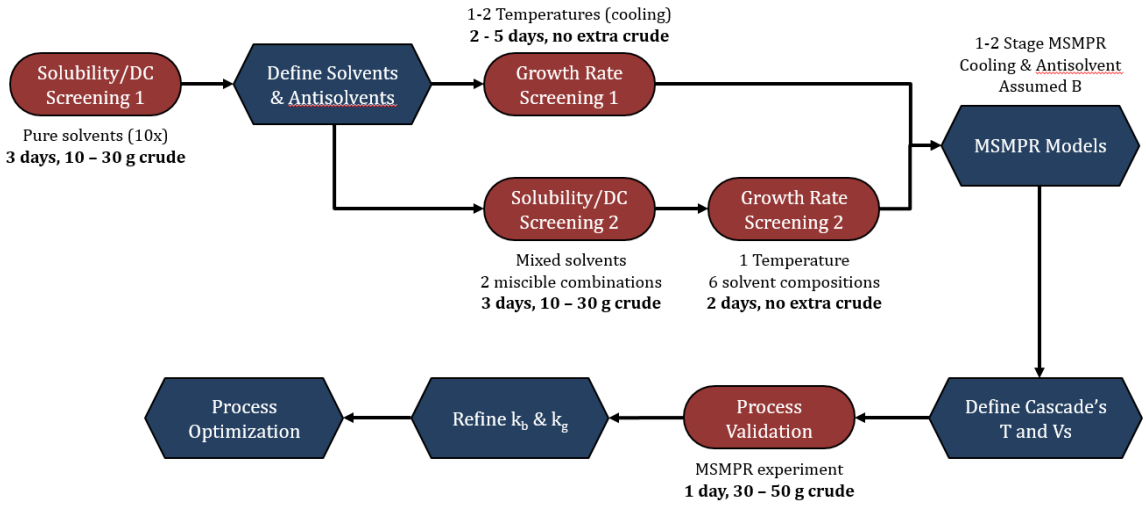


Figure 28. Workflow for the rapid development of continuous crystallization processes for purification

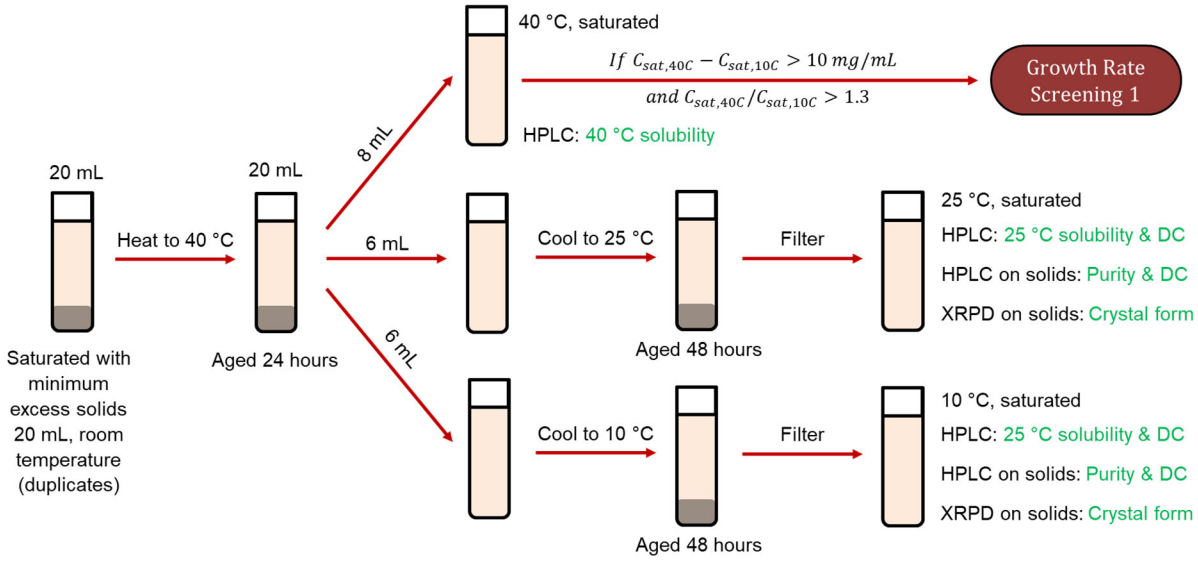


Figure 29. Solubility and distribution coefficient screening for pure solvents

Results from this solubility screening give duplicate measurements of solubility at 3 relevant crystallization temperatures, for each investigated solvent. They also reveal which solvents work best for purification, which are optimal for cooling crystallization (via solubility drop) and which are optimal as solvents and antisolvents for mixed-solvent crystallization (via absolute solubilities and impurity incorporations). If antisolvent crystallization is deemed necessary, a second screening is conducted. Here, the impurity incorporation rate is not just a function of the solvent composition, but also of the amount of antisolvent that has been added, as the addition itself dilutes the crude and limits the available impurity concentrations for incorporation into the crystal. Consequently, the approach for mixed solvent screening is slightly different (Figure 30).

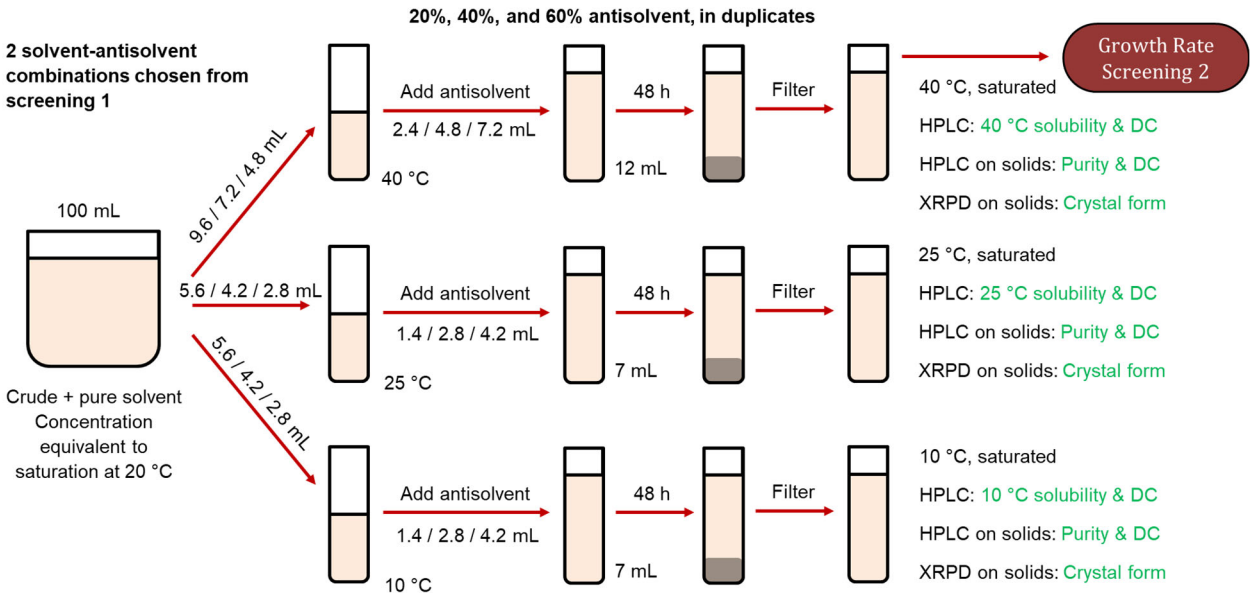


Figure 30. Solubility and distribution coefficient screening for mixed solvents

First, two combinations of solvent and antisolvent are selected. Then, feed solutions are prepared in the pure solvent containing the equivalent crude amount to the solute solubility at 20 °C. This ensures that the feed stays a clear solution without the need for heating the tank and pipes, while maximizing the amount of crude we can dissolve. The prepared feed is divided in nine duplicate samples, to conduct crystallization using 20%, 40% and 60% antisolvent at 40 °C, 25 °C and 10 °C. After aging the samples for 48 hours, they are filtered and analyzed using HPLC and XRPD. The samples at 40 °C are kept for kinetic analysis. This screening yields 9 solubility points for each solvent-antisolvent pair, which can be combined with the data from pure solvents to yield 3 mixed solvent solubility curves (Figure 31) at 40 °C, 25 °C and 10 °C. At the same time, we can obtain distribution coefficient data for each impurity, at the 9 combinations of temperature and solvent composition. These results can be used to generate attainable product plots, which in combination with observations of fouling and impurity incorporation during the prior screenings, can assist in selecting the operating conditions in the MSMPR cascade. First, a set of possible combinations of solvent composition and temperature are selected. Then, the optimal conditions for stage 1 are chosen based on observations of fouling and impurity incorporation, and if necessary, on kinetic information. Crystal growth kinetics can be obtained from microliter-scale kinetic screening using the 40 °C samples from the two prior screenings. These samples are divided into 600 µL samples within 1 mL vials. Then, they are placed in a cooling jacket at 10 °C. A programmable microscope screens through the samples (80 per screening, usually 5 solvent compositions) until the first crystal forms (Figure 32). Then, by tracking crystal growth within the first hour (known supersaturation), we can estimate the crystal growth rate at that solvent composition. Then, by means of the supersaturation, we can estimate the growth rate constant as:

$$k_g = \frac{G}{\sigma}$$

The final value of kg is the average of at least five independent samples. While single crystal growth does not have to be representative of the bulk scale, we previously demonstrated that when accurate crystal size estimations are not critical, only an approximate value of the crystal growth is required to estimate crystallization yields. At the same time, while the effects of scale-up can change the actual growth rates in the system, the dependence on solvent composition (i.e. which solvents generate larger crystals) should remain similar. These methods are useful as a rapid and data-rich way to make a kinetic-

driven solvent selection. A reproducibility study of these methods was conducted for acetaminophen in ethanol (Figure 33).

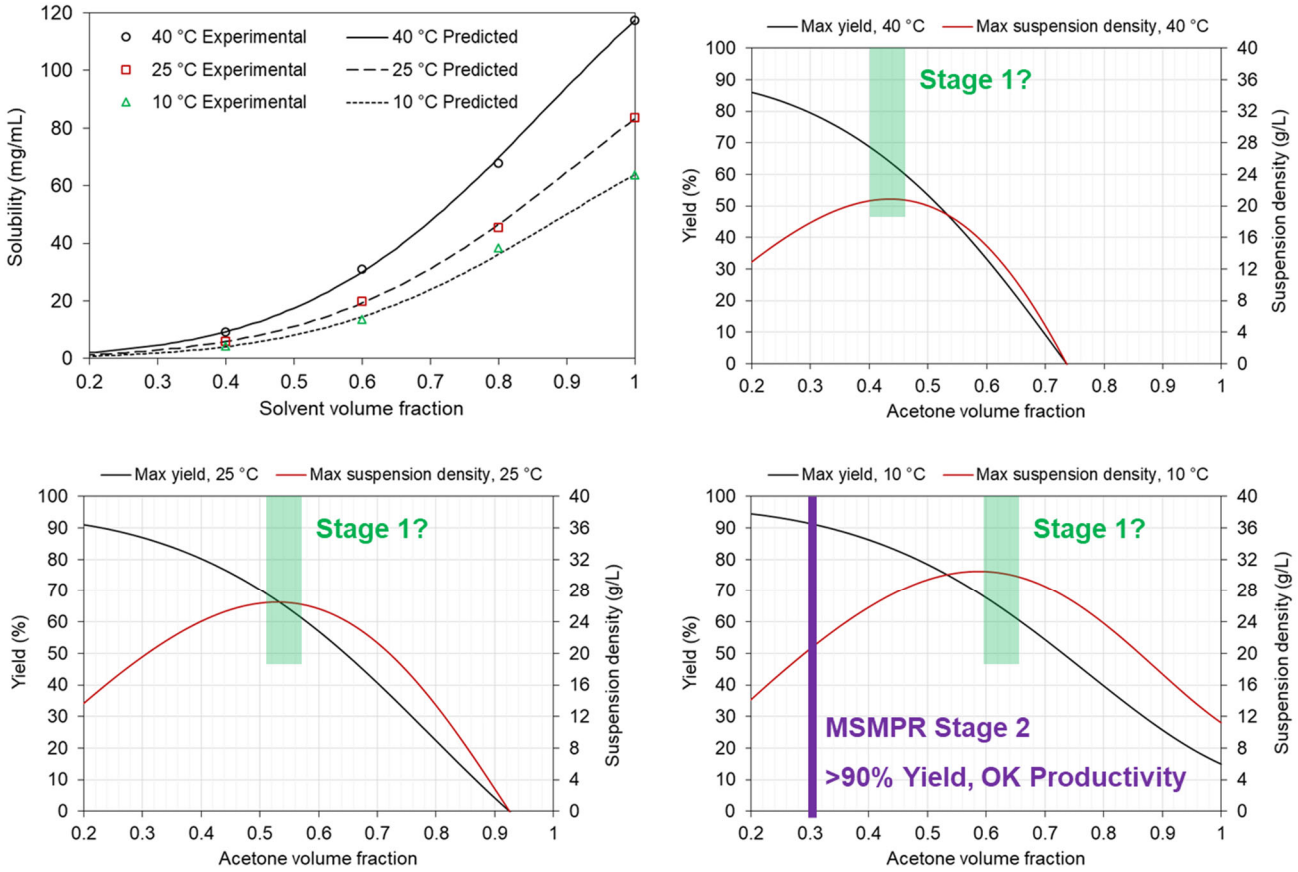


Figure 31. Example for mixed-solvent solubility of acetaminophen crude in acetone-heptane mixtures

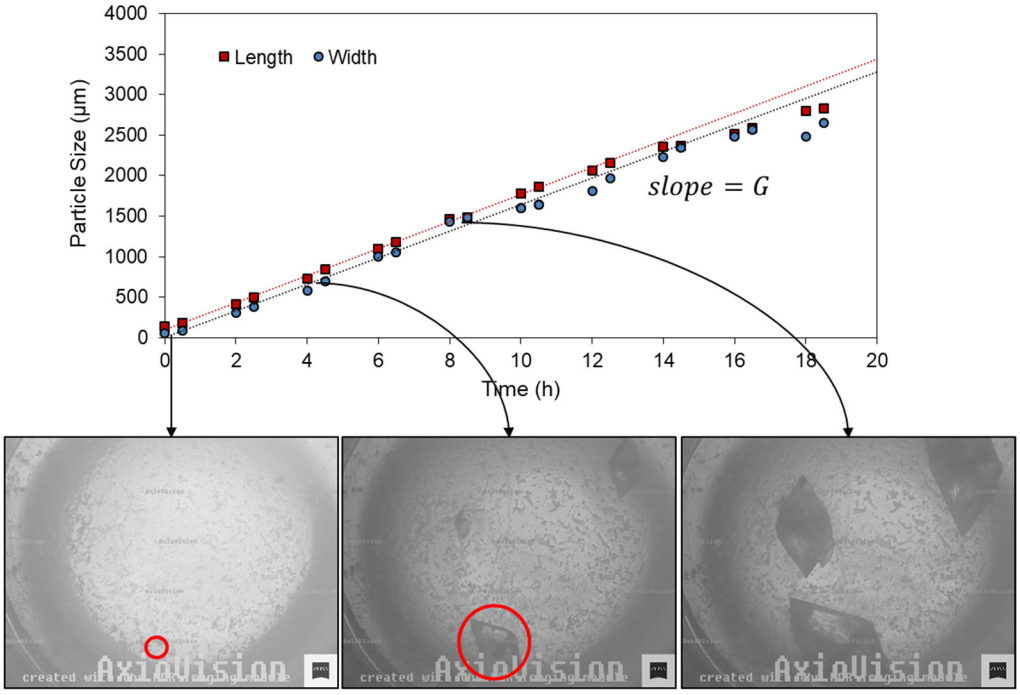


Figure 32. Estimations of crystal growth rate based on microliter-scale screening

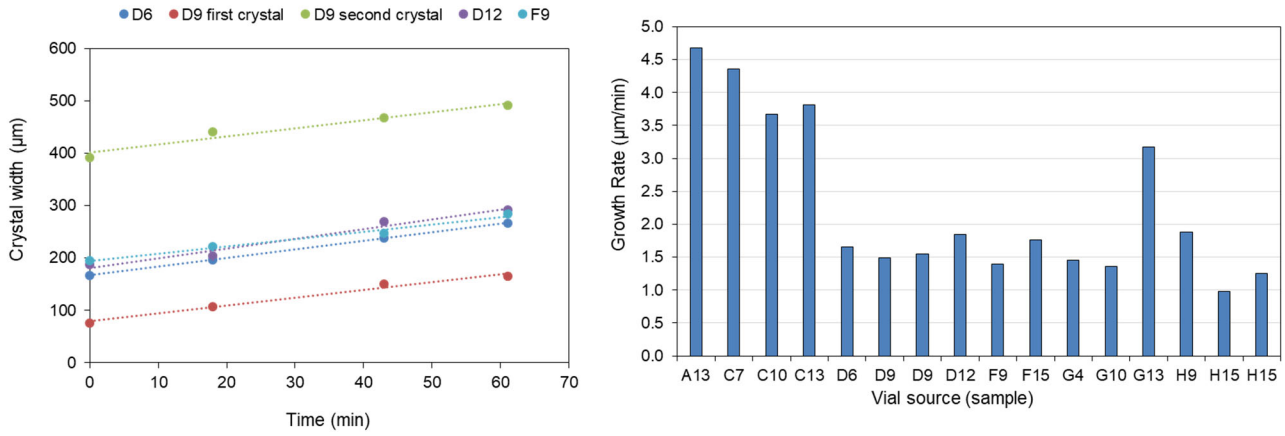


Figure **33**. Reproducibility of crystal growth rate measurements for samples within the same and different vials. The two distinct growth rates in the figure on the right occur due to different 2D projections of the crystal being presented to the microscope