

Computational and Automated Approaches to Reaction Discovery and Chemical-Space Exploration

Yasemin Bilgi-Gadina M.Sc.

Supervisor: Prof. Bartosz A. Grzybowski

Abstract in English

My doctoral research addressed four connected challenges in modern synthetic chemistry: discovering new multicomponent and one-pot reactions using Allchemy software's MECH module; developing *AbSynth*, an open-access total-synthesis database, as a reliable, machine-readable collection of published synthetic knowledge for future computational research; robotically exploring multidimensional reaction spaces; and quantifying via automation large numbers of complex crude reaction mixtures.

In the first stage, as part of a broader team effort, I contributed to the experimental evaluation of *Allchemy-MECH*, which generates mechanistically plausible reaction pathways from selected starting materials. I first validated a computationally proposed one-pot sequence involving two Williamson ether syntheses connected by a [2,3]-Wittig rearrangement to afford branched diallylic ethers. One product was further converted by ring-closing metathesis into a synthetically useful cyclic enol ether. I then used *Allchemy-MECH* to identify conditionally divergent multicomponent pathways from allyl thiol, an aromatic aldehyde, and an allenyl ketone. These afforded hydroxylated monothio- β -diketones under DBU-mediated conditions and 2,3-dihydrothiophenes after addition of acetyl chloride; the MCR product was further converted into a substituted pyrazole, a common structural motif in many drugs. I developed the corresponding experimental procedures. These results – published in *Nature Communications*¹ – showed that mechanistic network analysis can propose experimentally applicable MCR and one-pot sequences while identifying reagent-compatibility constraints, the required order of addition, and condition-dependent divergence into different product classes.

In turn, the *AbSynth* project addressed the need for reliable, machine-readable data on complete total syntheses by assembling more than 3,000 published pathways as an open-access resource. I served as a co-corresponding author on this work (now published in *J. Am. Chem. Soc.*²) and my contributions spanned data curation, its analysis, structural and stereochemical quality control, correction of pathway errors, and coordination of the curation team. This work was essential because errors in structures, stereochemistry, step order, or pathway connections would distort subsequent analyses of synthetic strategy and route design. The resulting collection provides a reliable foundation for large-scale theoretical studies of synthetic-route organization, complexity progression, retrosynthetic logic, synthesis planning, cheminformatics, and chemical artificial intelligence.

The next stage of my doctoral research focused on robotic chemistry and multidimensional reaction-space exploration in the laboratory of the Grzybowski group in South Korea. In the work published in *Nature*³, I studied the E1 and S_N1 reactions of fluorenols as model systems to establish how reactions could be executed robotically, quantified their yields by algorithmic UV–Vis spectral unmixing, and mapped the diverse outcomes across multidimensional condition “hyperspaces”. These studies provided the methodological foundation for applying reaction-hyperspace mapping to more complex chemical systems. The large number of concentration-dependent yield measurements also enabled kinetic and equilibrium models to be fitted to the E1 and S_N1 hyperspaces, providing estimates of kinetic and thermodynamic parameters despite each hyperspace being measured at a single reaction time.

I subsequently designed and carried out a hyperspace study of textbook olefin bromination reactions, now published in *Angewandte Chemie Int. Ed.*⁴. I contributed to defining the reaction hyperspaces, orchestrated on the Robowski platform and then analyzed approximately 1000 reactions, and characterized more than twelve compounds formed across these systems. Automated 80 MHz benchtop NMR acquisition and characteristic-signal fitting enabled hyperspace-wide quantification of the crude reaction mixtures. Unexpected complexity emerged from these textbook brominations, including previously unrecognized products, condition-dependent pathway switchovers, and broader reaction networks than their conventional descriptions suggest.

As an additional extension of this work – not included in the publication cycle constituting the dissertation – I designed and conducted hyperspace mapping of the Chan–Lam coupling, an important C–N bond-forming reaction widely used in synthetic and medicinal chemistry. Because the crude spectra showed extensive overlap and condition-dependent resonance changes, quantitative analysis used NMRanger, a reference-guided deconvolution method that fits selected spectral regions to determine the amounts of individual reaction components. I established the reaction-specific workflow by optimizing the fitting parameters, operating the method independently, and validating the derived yields through ground-truth comparisons, including isolations. This enabled quantitative mapping of coupling products, regioisomers, by-products, and unconsumed reagents across approximately 1,600 conditions. The NMRanger framework was developed by Louis P. J.-L. Gadina, while my contribution focused on its reaction-specific implementation, validation, and application to a complex catalytic system. The corresponding manuscript is currently under revision at *J. Am. Chem. Soc.* and is provided separately as additional material.

The studies constituting this dissertation trace a progression from computationally assisted reaction validation and chemical-data curation to independent expertise in robotic experimentation, reaction-hyperspace design, compound characterization, and automated quantitative analysis. The additional Chan–Lam/NMRanger study further extended these approaches to a complex catalytic system. Together, this work contributed both new chemical knowledge and practical methods for investigating complex reaction systems at a scale inaccessible through conventional experimentation alone.