A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis

Daniel E. Fitzpatrick,*† Claudio Battilocchio, † and Steven V. Ley†

†Innovative Technology Centre, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Supporting Information

ABSTRACT: We have developed a modular software system that enables researchers to monitor and control chemical reactions via the Internet, using any device from any location in the world. It facilitates the automation of synthetic procedures and is able to autonomously self-optimize reaction parameters to find the best conditions meeting customizable, multicomponent optimization functions. In this report, we demonstrate its utility as applied to reaction automation to maximize the output from a fixed volume of catalyst. We also showcase its ability to optimize a three-dimensional heterogeneous catalytic reaction and a five-dimensional Appel reaction against various target functions.

INTRODUCTION

Chemists in research and process laboratories frequently spend significant time performing routine, day-to-day activities such as reaction monitoring and optimization. This is especially the case for research carried out at the discovery level. Recently we have reviewed how contemporary developments in the area of machine-assisted synthesis are able to release these skilled workers from monotonous tasks, enabling them to focus on more productive pursuits. Indeed, in the process environment machine use is more widespread for activities such as scale-up, greatly reducing the time taken for optimization and improving efficiencies.

The adoption of machines at a discovery level in organic synthesis laboratories however has been a much slower process. While other physical science disciplines have enjoyed the increasingly beneficial influence of machines on their work, chemistry is still very much focused on batch processes using glassware which has not changed significantly for many decades. Recent advances in continuous flow chemistry are beginning to break through this historical impasse, creating new opportunities for high temperature, high pressure, multistep, and downstream processing techniques when applied to synthetic challenges. These changes in synthesis practices impose a different approach to solving problems in the area, giving rise to a more holistic understanding of the reactions involved.

Consumer demand has driven a huge surge in the development of new “soft” technologies, primarily focused on producing increasingly smaller computing devices which have the ability to communicate via the Internet. Cloud computing, where Internet services are hosted remotely (such as email services and file storage), are experiencing rapid growth. Some developments in these areas have made their way into the chemistry environment, with examples including Internet-based computational chemistry tools and laboratory management software. Evolving from these trends is a new concept named the Internet of Things (IoT) which envisages a world where every device is connected to the Internet, supplying real-time data to both other devices and central control systems. The application of such a concept in the chemistry environment would greatly alter the status quo of research laboratories: reactors could communicate with detectors which in turn could send data to downstream processing units. This communication would be facilitated and recorded in its entirety by central control systems, allowing chemists to more actively monitor their reactions in real time and peruse data when appropriate.

Many more benefits would arise from such a system, including improved laboratory safety (e.g., automated shutdown sequences if parameters fall outside allowable ranges), greater insight into chemical processes, and the ability to change reaction conditions in real time based on feedback from multiple data sources.

Our group has focused on integrating control systems into synthetic procedures for almost a decade, with our first major venture in this area leading to the fully automated total synthesis of the natural product grossamide in 2006. More recently we have reported a system which enabled a predefined sequence of experiments to be carried out, including those for a Design of Experiment (DoE) study, and, in a separate investigation, allowed a single researcher to manage a telescoped, continuous three-step chemical transformation procedure with three intermediate downstream processing steps. Other research groups have developed computer control solutions on a case-by-case basis, including for the control of reagent addition based on detector feedback, reaction kinetic analysis, and the telescoped continuous preparation of an active pharmaceutical ingredient.

While the bespoke nature of such control systems renders them excellent at performing predefined reactions, their rigidity of operation may limit their suitability in the rapidly changing discovery environment. It takes considerable time to redevelop any new control sequences, leading to a reduction in the number of experiments that can be conducted in one working day. Furthermore, researchers require considerable experience...
of computer programming in order to connect such systems to their wide range of laboratory equipment.

■ DEVELOPMENT OF AN INTERNET-BASED CONTROL SYSTEM

To overcome these limitations, and in a first step toward integrating the IoT into chemistry research, we created a new system that facilitates laboratory control by any researcher regardless of computer knowledge, integrates with detectors commonly found in the laboratory (infrared detectors, mass spectrometers, etc.), records all equipment parameters, enables set point adjustment of equipment (heaters, pumps, etc.), and includes the ability for more advanced control strategies (such as synthesis automation). This system was developed to be Internet-based, thus allowing it to be accessed through any browser on any Internet-connected device, from anywhere in the world, and has been designed to complement the work of researchers in the laboratory where the system is set up.

This software, which we have named LeyLab (this can be configured and/or renamed to suit any laboratory environment), is comprised of four main components: a graphical interface, accessed by the user through an Internet browser; a database in which all experiment, equipment, and user information is stored; an equipment communication module, consisting of various code definitions listing whichever protocols and commands are required to gather equipment data for each manufacturer; and an equipment command module, consisting of code definitions listing commands to be sent to individual equipment to achieve a certain outcome. The creation of modules to integrate new pieces of laboratory equipment takes less than 30 min.

LeyLab operates purely through Internet protocols, with both user–server and server–equipment communication occurring using TCP/IP (Figure 1). As a result of this, the LeyLab server can be located anywhere in the world, as both user-server and server-equipment communication occurs through the Internet.

Having launched an experiment, LeyLab’s server gathers real-time data (on a per-second basis) from all connected equipment and stores it in a database indexed against the relevant experiment. When a user views the experiment page in their browser, all data saved in the database are displayed for each piece of equipment, with an update loop adding new data points as they are collected.

Having real-time control over laboratory equipment through an Internet-accessible device raises clear security considerations. Potential breaches of server security could enable unauthenticated users to modify equipment parameters, view unpublished data, or gain access to other servers connected to the same internal organizational network. A discussion of these issues and other considerations related to remote reactor control have been described in a recent commentary.17

To mitigate security risk in LeyLab, each user is supplied with access credentials which must be entered before access to the system is permitted. The LeyLab server is also placed behind University firewalls, allowing access only from within the department’s network (external access such as from researchers’ homes is possible through a VPN connection).

■ SIMPLE AUTOMATION

To demonstrate the efficacy of LeyLab for simple automation procedures, we explored a new method of conducting catalytic reactions in which a fixed volume of reaction solution is passed multiple times through a small volume of catalyst contained within a packed column. In this case we utilized a commercially

Figure 1. LeyLab communicates with both users and equipment through the Internet (TCP/IP), rendering it a fully cloud-based system.

Figure 2. A four-step process allows users to create a new experiment. The experiment is named in step one; equipment connection settings are collected in step two (IP address and port); users can add any applicable automation scripts in step three; and finally the experiment start time is set in step four (experiments can be started immediately or at a specified later time).
available flow reaction system to provide the required pump, temperature, and switching valve support, alongside an Advion miniature mass spectrometer (MS) for simple reaction monitoring purposes (Figure 3). At the beginning of the procedure, flask A was filled with a fixed volume of solution which was pumped through the packed column to flask B. When the liquid level in flask A fell below a certain level, the position of valve 1 was switched and solution was then pumped from flask B to flask A. This process was repeated until a set number of passes had been completed.

Initially each flask was placed on its own mass balance to allow LeyLab to monitor liquid levels. However, we noticed that changes in ambient atmospheric pressure during the course of a few days led to great changes in balance readings, even when there was nothing placed on the balance tray (refer to Supporting Information). Accordingly, we adopted a camera-based solution\(^1\) in which a webcam connected to a Raspberry Pi computer\(^2\) monitored the location of a colored plastic float positioned on the liquid surface (Figure 4). LeyLab queried the Raspberry Pi device to retrieve positional information on a per-second basis, just as for other pieces of equipment.

For this procedure we chose to carry out the heterogeneous hydration of 3-cyanopyridine to its corresponding amide over manganese dioxide (Figure 5a), as we have explored this particular chemistry in a previous study.\(^3\) Before attempting multiple passes in this system, we wished to test whether the MS was able to monitor the composition of the product mixture effectively when operating under direct injection. Accordingly we configured LeyLab to pump a small plug of reagent solution through the system, while recording the output from the MS which was itself configured to monitor the intensities of two peaks corresponding to the starting material and product. The data collected by LeyLab showed very clearly the relative composition of the reaction mixture (Figure 5b).

Satisfied with these results, we then turned our attention to carrying out multiple passes through the system. LeyLab followed simple procedural logic to achieve this (Figure 6) while monitoring conversion using MS data. In this case, the system was allowed to run continuously with no human intervention for 14 h. As can be seen, distinctive step changes were observed corresponding to a full single pass from one flask to the other (Figure 7).

\section*{AUTONOMOUS SELF-OPTIMIZATION}

Owing to its ability to change experimental set points and determine the effects of these changes on reaction outcomes through the use of in-line or online detectors, LeyLab is ideally placed to autonomously search for optimal reaction conditions without researcher input. Most literature of a practical nature in this area has focused on the use of the simplex algorithm (or modified versions of it) first proposed by Nelder and Mead in 1965.\(^4\) These studies were conducted using bespoke software written with tools such as Matlab or LabView to optimize a predefined reaction.\(^5\) As the method of operation for this algorithm has been described in these publications, discussion here has been kept to a brief overview of its main points.

A modified version of the simplex algorithm known as the Complex Method\(^6\) has been implemented into a control...
module in LeyLab. As an n-dimensional optimization tool, the complex method is able to manipulate any number of reaction parameters to maximize or minimize an evaluation function (e.g., conversion or yield). When a self-optimization experiment is first created, each parameter is defined with its upper and lower boundaries (e.g., temperature between 30 and 140 °C) setting the chemical space within which LeyLab can optimize.

The reaction control steps followed by LeyLab for an optimization process are shown in Figure 8. Before any equipment is accessed n + 1 initial iterations are created (where n is the number of dimensions), each defining a collection of parameter set points corresponding to one experimental trial. These initial points are not randomly selected; instead they are distributed evenly throughout a large portion of the available chemical space. The system conducts experiments for each of these iterations, calculates the iteration performance using the evaluation function after the system reaches steady state (based on the stability of detector output), and then ranks them from best to worst performing before generating a new iteration. This process repeats until the system cannot optimize conditions any further.

One key benefit of simplex-based algorithms when applied in a chemistry environment is that quantitative values of iteration performance (e.g., exact measurements of yield or conversion) do not need to be obtained. Instead, relative performance is used for iteration ranking. As long as there is an improvement from one set of conditions to the next, no matter the absolute value of this improvement, the system knows that it is moving in the right direction and thus will optimize in the anticipated way. Accordingly, measurements from detectors do not need to follow a linear relationship with yield, conversion, or concentration as long as the response increases.

The most basic process followed by LeyLab to generate the set points for a new iteration is shown in Figure 9a. Take, for example, the three-dimensional optimization process shown in Figure 9b. Having ranked the initial four iterations from best to worst, LeyLab takes the worst performing experiment (\(P_w\)) and reflects it through the centroid (shown in red) of the plane intercepting the remaining three iterations (\(P_1\), \(P_2\), \(P_3\)) to find the next iteration’s set points (\(P_{ref}\)). Having carried out that experiment and reranking the iterations, if \(P_{ref}\) represents the best performance the next iteration is generated (\(P_y\)) based on

![Figure 6](image1.png)

**Figure 6.** LeyLab followed a simple procedure when performing the multipass experiment. Counter represents how many passes have occurred and was initialized to 0 at the beginning of the experiment.

![Figure 7](image2.png)

**Figure 7.** Mass spectrometer response collected by LeyLab during the 14 h multipass experiment (the blue line corresponds to the starting material, while the orange line represents the product).

![Figure 8](image3.png)

**Figure 8.** Method followed by LeyLab when performing an optimization experiment. The system can be configured to flush fully any connected apparatus between iterations.
extending the previous reflection. However, if $P_{\text{ref}}$ is the worst performing point, the system creates the next iteration ($P_x$) by retracting the previous reflection. If $P_{\text{ref}}$ is between the best and worst performing iterations, then the system discards $P_x$ and performs another reflection based on $P_{\text{ref}}, P_1, P_2,$ and $P_3$. The system may also follow different routes to generate new iterations based on where each previous iteration falls relative to others (refer to the Supporting Information for more information).

LeyLab stops the optimization process either when a set number of experiments have been carried out (as set at the beginning of or during the experiment) or when the evaluated performance of each iteration converges to one value (as determined by calculating the variance of the current simplex and comparing it to the simplex mean). If these conditions are reached, then LeyLab follows a user-defined experiment shutdown sequence.

**DEMONSTRATION OF THREE DIMENSIONAL OPTIMIZATION**

As a first foray into reaction self-optimization, we decided to carry out a three-dimensional optimization based on the hydration reaction previously used to demonstrate simple sequence automation. In this case, LeyLab was configured to optimize for temperature, residence time, and inlet reaction concentration. The evaluation function was simple, consisting of only the ratio between the product and starting material MS readings (eq 1). While ion suppression effects may affect detector data in this case, a linear relationship linking MS response with conversion and the evaluation function is not required owing to the qualitative nature of the Complex Method.

**Equation 1: Evaluation function for the three dimensional optimization experiment**

$$f(x) = \frac{P}{s}$$

Where $p =$ amide MS response $s =$ nitrile MS response

**Figure 10** shows the equipment layout used for this procedure. A reservoir containing 1.0 M solution of 3-cyanopyridine in H$_2$O was connected to one pump on a Vapourtec R2/R4 unit, with a second reservoir containing just H$_2$O connected to the second pump. The outlet streams from both pumps were mixed at a T junction, before flowing through a column packed with 2 g of MnO$_2$. A sampling valve was placed at the back-end of the reactor to enable online MS analysis of the reaction mixture. Reaction residence time was controlled by the overall flow rate (the sum of the two individual pump flow rates), while inlet concentration was adjusted by modifying the ratio of the two pump flow rates. LeyLab was configured to perform a full flush of the system with solvent between each iteration so as to ensure no intertrial interference. The full operation script used for this procedure can be found in the Supporting Information.

LeyLab found optimal conditions within 12 experiments carried out over 17 h (Figure 11). As can be expected from the

**Figure 10.** Equipment configuration for the three-dimensional optimization experiment. Residence time and concentration were controlled by pump flow rates, while reactor temperature was controlled directly by a Vapourtec R2/R4 unit.

**Figure 11.** Results from the three-dimensional experiment. The initial iterations generated by LeyLab are shown in red, while the optimized conditions are shown in green.
simplicity of the reaction, over time the conditions chosen by the system trended toward higher temperatures, longer residence times, and lower concentrations. It is worth noting that at 2 a.m. during experimentation (marked by the vertical line in Figure 12), the system had shut down, as it detected sharp pressure fluctuations in pump A caused by air bubbles in the input liquid stream reaching the pump head. The following day the pump was primed, thus removing any air in the pump and stabilizing the output flow rate, and the system was allowed to continue to optimize conditions. While this is not a common occurrence, we report it here to emphasize that the safety shutdown sequence operated effectively.

DEMONSTRATION OF FIVE DIMENSIONAL OPTIMIZATION

In the experiment described above, there was very little complexity involved with the reaction itself. Any chemist would be able to state that for a heterogeneous catalytic reaction of that nature, in which no side- or byproducts are formed, reaction conversion can generally be increased by increasing temperature and residence time while decreasing inlet concentration. While LeyLab discovered that for itself without any input from a human, the use of such technology may be considered excessive in this particular case.

Accordingly, we shifted our attention to a more complex example involving the five dimensional optimization of an Appel reaction (Figure 13). For this experiment an in-line IR detector was used to monitor the concentrations of starting materials remaining in the product mixture following the reactor. However, as the IR stretch associated with the substituted product fell below the range visible with the detector, we were not able to directly monitor product formation. Instead, we were able to use the IR reading corresponding to triphenylphosphine oxide as a substitute, as this compound formed only when the Appel reaction was successful.

For the previous optimization experiment, the evaluation function was based on a single target containing just two variables (MS reading of starting material and product intensities). Yet optimizing such a simple, two-term evaluation function does not demonstrate the full benefits that can be realized through the use of automated control. For this experiment we chose an evaluation function that better incorporates two of the ancillary factors that play an important role in any optimization process alongside conversion, namely consumption of starting materials and throughput.

These factors are represented by individual terms in the evaluation function (eq 2). The first term, which we have denoted as the throughput factor, takes into account the residence time in the reactor as well as the overall concentration of material being processed. As residence time decreases and overall concentration increases, this term rewards the system and so causes it to try to optimize in that direction. However, as we wished the system to put more weighting on conversion, this throughput term incorporated a 0.25 multiplier so that its overall importance was reduced.

The second term is a combination of both the conversion and consumption factors. In this particular case, a large weighting was placed on the amount of product formed relative to starting material consumed while a reducing factor was placed in front of the terms representing starting material consumption.
Equation 2: First evaluation function which allowed LeyLab to optimize the five experiment parameters against six conditions; these were split into throughput, conversion, and consumption terms

Equation 3: Second evaluation function with separated conversion and consumption terms

This combination of the conversion and consumption factors in one equation term led to optimization problems, however, as the consumption term greatly skewed the output of the conversion term. As can be seen in Figure 14, while LeyLab did progress toward conditions that produced much greater evaluation responses, these iterations were interspersed with other iterations that reflected no effective improvement on the initial n + 1 iterations. Accordingly we decided to stop the optimization process after 19 iterations.

To mitigate this unsatisfactory pattern, we modified the evaluation function to separate the conversion and consumption terms (eq 3). In this new equation, we adjusted the weighting factors for all terms to better reflect the overall aim of the optimization process. The function now led the system to shift conditions toward those which produced higher conversion, while increasing throughput and reducing unnecessary material consumption.

The results for this equation were excellent (Figure 15). A clear increasing trend over time can be seen, with the vast majority of iterations chosen by the system producing a better result than the initial six iterations. In this case the system was configured to stop after 30 reactions.

Optimal conditions were found to correspond to the last iteration in the process, which resulted in a 92% yield. The set points in this case were a reactor temperature of 111 °C, residence time of 4.3 min, and overall reaction concentration of 0.3 M with 0.87 and 1.72 equiv. (to the alcohol) of CBr₄ and PPh₃ respectively. Operating under these conditions at steady state would enable the generation of 1.9 g h⁻¹ of the bromo-substituted product. It is worth noting that the system was configured to optimize overall concentration between 0.05 and 0.30 M, thus limiting the upper value of this production term.

The ability to view stored reaction data presents a secondary benefit of LeyLab’s simplex module: researchers can gain a more in-depth understanding of how certain experimental parameters affect the outcome of a reaction by looking at raw data plots on their Internet browser (Figure 16). By comparing iterations with similar set points, for example, it was seen that for this experiment the residence time did not play as important a role as temperature. While this information may not be vital at a discovery level, such insights coupled with access to raw experiment data can be extremely valuable when scaling up processes to production.

**CONCLUSIONS**

We have developed a new Internet-based software system (named LeyLab) that allows chemists to monitor and control chemical reactions from any Internet-connected device, anywhere in the world. It facilitates the automation of synthetic procedures and is able to self-optimize reaction parameters to find the best conditions meeting customizable, multicomponent optimization functions. The system does not require significant computer knowledge to use, allowing any chemist to enjoy its benefits regardless of background.

We have showcased the utility of LeyLab for maximizing small volumes of catalyst through a recycling process in which a fixed volume of reaction solution is passed multiple times through a packed column. Through the incorporation of machine vision techniques, the system was able to monitor liquid levels in reservoirs while an online mass spectrometer provided analytical feedback regarding conversion from each pass.

It is clear from our optimization experiments that the use of machines can simplify the investigation of problems containing significant complexity, releasing bench chemists from what would be significant drains on their time. When left to optimize an Appel reaction, LeyLab was able to perform 30 experiments in 10 h to find optimal conditions for five experimental parameters (temperature, residence time, overall concentration, and equivalents of two reagents relative to a third).
Owing to its modular nature, LeyLab is not limited to self-optimization using solely the Complex Method. We are currently exploring the integration of alternative control strategies such as those involving Gaussian Processes. We envisage that LeyLab will present a selection of control and optimization tools to chemists which will allow them to choose the best method relevant to their projects. This modularity also applies to equipment support, enabling devices from a variety of manufacturers to be integrated into LeyLab. In this way, our software system is not limited to flow chemistry applications and one can imagine many other areas of research where such a system could find use.

**ASSOCIATED CONTENT**

* Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.oprd.5b00313.

Characterization of compounds, additional descriptions of Complex Method optimization, and description of flow setup. (PDF)

**AUTHOR INFORMATION**

*E-mail: def27@cam.ac.uk.

**Notes**
The authors declare no competing financial interest. Experimental data relating to this article can be accessed free of charge at https://www.repository.cam.ac.uk/handle/1810/251228.

**ACKNOWLEDGMENTS**

We are grateful to the Woolf Fisher Trust (D.E.F.), Pfizer Worldwide Research and Development (C.B.), and EPSRC (S.V.L., Grant Codes EP/K009494/1, EP/M004120/1, and EP/K039520/1) for financial assistance.

**REFERENCES**


Figure 16. Raw data collected during the course of an experiment by LeyLab can be viewed at any time after the experiment has completed, giving researchers greater insight into their reactions.
(10) Refer to software developed by Chemical Inventory Ltd., http://www.cheminventory.net/, last accessed 29 October 2015.