Mestrelab Research
From instruments and analytical data
to knowledge and decisions

Application note

Dr. Gary Sharman

**Senior scientific director
at Mestrelab Research S.L.**

Reactions, reactions, reactions – data automation to the rescue!

Computational chemistry and AI have allowed us to make great strides in predicting reaction conditions for organic synthesis, providing valuable tools for optimising reactions and reducing the number of trials required. However, there are situations where running reactions under various conditions, altering things like the solvent, base, catalyst, or temperature, remains indispensable. Once such an array of reactions has been completed, it becomes necessary to acquire analytical data to experimentally measure the amounts of products, starting materials, and other components in the reaction. While LCMS offers a rapid method for gathering such data, it also presents a significant challenge in terms of processing and extracting useful information from the results so obtained. This is where Mnova's automation engine (Mnova Gears, or simply Mgears) and the Chrom Reaction Optimization plugin come into play, offering a potential solution to these challenges. In this note, we explain how these tools can assist us in addressing these issues and discuss specific scenarios where they can be applied effectively.

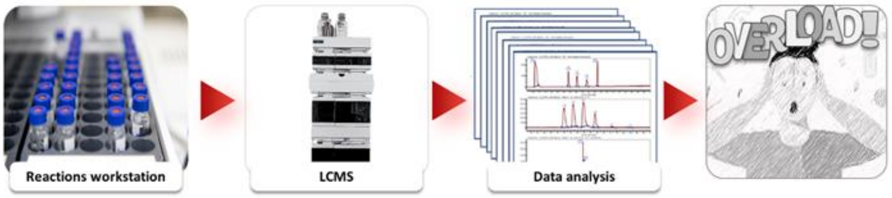


Figure 1 - Modern automation of reactions in arrays and rapid LCMS analysis creates a new headache – analysing the resulting data to quantify components of interest.

The basic problem – finding the same components across a set of samples

The basic problem that needs to be solved in reaction optimisation or monitoring is to find the same set of components (typically starting materials, products, known side products) across a collection of LCMS files. This problem consists of two parts: the first, which is common to any automation process, is to find or somehow specify the relevant data to be operated on. The Mgears engine provides a solution to this part with several ways to select data. *Do you need to search for data on a file share? Or listen for new data as it is generated? Or perhaps provide a list of files in a CSV? Or maybe you have an SMDS or need retrieval from a custom database?* All these modes are supported in Mgears – not just for reaction optimisation, but for any similar automation. In most cases, the desired actions can be achieved through simple configuration options, such as clicking a checkbox to 'find LCMS data'. In cases where the requirements are more exotic, the script engine can always come to the rescue to customise the behaviour according to the specific needs.

Providing multiple ways to define components

Once the data has been found, we must specify the components that require searching, and certain other options for analysis. Components can be defined in various ways: as a structure (mol file or SMILES string), a molecular formula, or simply as an m/z value. If you only have UV-based data, it is also feasible to define components solely according to their retention time. It is also possible to mix and match; in other words, you can potentially define certain components by mass, while others (such as those that do not ionise) can be defined exclusively by their retention time. This approach offers excellent flexibility, enabling the identification of components using the most suitable method for each individual case.

One problem that can arise is that of isomers. Perhaps you have two products with the same molecular weight. You can address this problem using an AND logic – defining a mass and a retention time range to ensure isomers are identified in a consistent manner.

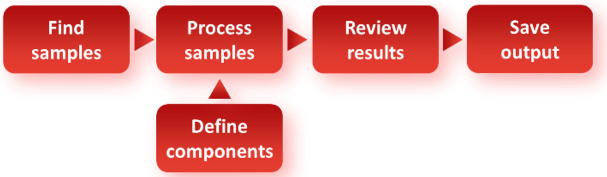


Figure 2 - To analyse reaction optimisation data, we need to find samples, define components, and then quantify them in every sample. We also probably want to review and save the results. Mgears automation makes this process simple.

Providing a simple view of the results – including metadata

Once the relevant data is found and the levels of the defined components have been calculated, we need a way to rapidly navigate and review the results, allowing us to home in on the best cases and see which conditions they correspond to. The Mgears viewer comes to our rescue here, providing a tabular or well plate view of the results that can be quickly navigated through. In addition, specific views for Chrom Reaction Optimization provide a visually appealing method for browsing the data. A pie chart of the components is standard which, together with a consistent colouring scheme between chromatogram and chart, helps to readily spot the best or worst conditions.

Reviewing the results in this way is powerful, but typically what we really want to know are the conditions that go with that good result. If you then have to search or cross reference the sample name against a lookup table of conditions, this can become very tedious and error prone. Thankfully, Chrom Reaction Optimization provides a number of methods to directly display relevant metadata, either from a lookup mapping file such as a CSV, or from parameter values stored in the raw data. This means you can identify the best results and instantly determine the conditions they correspond to.

Providing the ability to fix issues – review by exception

Automation really is the friend of scientists doing high throughput analyses of the type described here. But a common problem with automatic analyses is what to do with edge cases or failures. With the best will in the world, no method of automation will get an analysis correct 100% of the time. If we don't have a quick way to identify problem samples, and perhaps even more importantly a simple way to fix them, we can end up in a situation where the net benefit of the automation has been significantly reduced. In my own experience, I have, in the past, not used automation because fixing the outliers was more problematic than analysing them all manually in the first place.

Fortunately, the results viewer is designed with this in mind. Problematic samples can be identified and loaded into the main Mnova window with a single click. Integrals can be tweaked, missed components added, and misassigned components corrected. On completing this, all graphical summary outputs (such as pie charts) are updated and, very importantly, output pertinent to the batch, such as CSV files, is recreated to reflect the modifications. This process also includes rerunning any custom output scripts that write to other systems or generate summary reports.

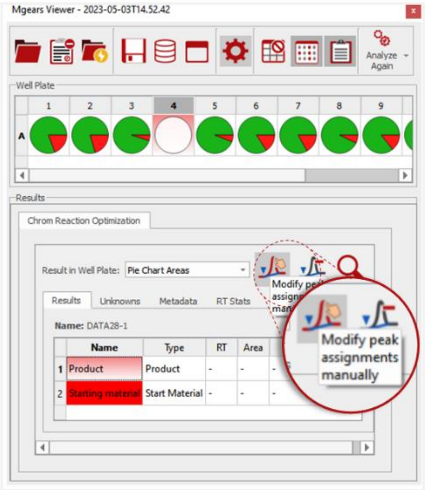
Such a review by exception strategy is essential to the efficient operation of an automated process. The simple yet powerful implementation here means the review is rapid and does not simply form a new bottleneck in the process.

Warnings and controls

In the previous section, we emphasised the importance of quickly identifying and addressing problematic analyses. As a prerequisite to this, we said 'problematic samples can be identified and loaded into the viewer' for fixing. However, we skirted around the question of *how* we identify such problematic samples.

Some cases may be obvious. In the example in Figure 3, it was clear something was amiss as no components were identified in a particular well. But how would we know if something was somehow misassigned without manually checking every sample? This is where the controls feature in Chrom Reaction Optimization becomes valuable.

Controls check statistics at a sample or at a batch level, enabling us to make decisions about whether a sample should be reviewed or not. A good example of this is retention time outliers. In any identification of components by MS, there is the possibility of a false positive, where a target isotope cluster is matched by chance. If this happens to align with a peak in the UV trace (typically used to quantify), we have a false assignment. But such events are likely to be rare in a batch. So, if, for example, a starting material is identified at 1.5 ± 0.02 minutes in 95 samples out of a batch, but there is one sample where it has been identified at 2.7 minutes, it is highly likely that this is a misassignment. Thus, by setting up controls like this and checking the results displayed in the viewer across a well plate, we can quickly identify which samples might need a manual review.



Mgears Viewer - 2023-05-03T14.52.42

Result in Well Plate: Pie Chart Areas

Name: DATA28-1

	Name	Type	RT	Area	
1	Product	Product	-	-	-
2	Starting material	Start Material	-	-	-

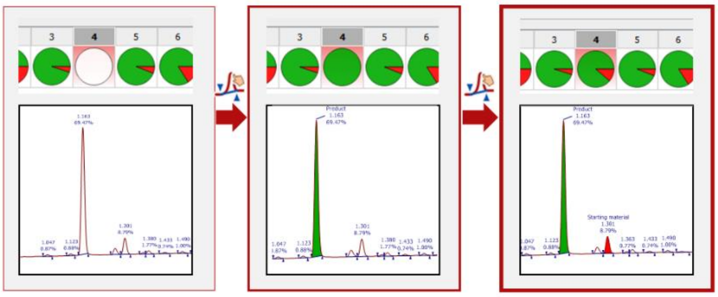


Figure 3 - In this example, one well does not seem to have processed properly. This was because the component definition was based on retention time, and in this chromatogram things moved too much for the window to identify the components. However, it is extremely fast to identify and fix this sample, assigning the two components manually.

Rerunning a batch

One advantage of the post-acquisition analysis approach adopted by Mgears is that it is easy to rerun an analysis whilst making certain modifications. Perhaps on review, you identify an additional component that you had not initially considered when you set up the reaction and the LCMS run. In this paradigm, this is no issue, as the analysis can be easily rerun after modifying definitions or adding, deleting, or changing components.

Handling similar reactions

You might be thinking '*this sounds great, but I have several reactions with different reagents on the same plate*'. Or, in the extreme, perhaps you are doing parallel synthesis with different combinations of starting materials, that is, with related reactions all conducted under the same conditions. If we need to define the components in the batch analysis configuration, how can we accommodate different components in each well?

Although not directly supported in the Chrom Reaction Optimization plugin, this is simple to achieve through a customisation to the script engine. Perhaps the actual components are defined in the metadata of the sample, or perhaps they are in a CSV file, or in a database. In this scenario, we can use the GUI of the Chrom Reaction Optimization plugin to define the generic components, their roles, and attributes (e.g., colour on the plot). Then, with the help of a script (which can be added as a hook in the Mgears engine), these generic components can be changed on-the-fly to the specific values of structures or molecular formulae required for each case. As Mnova is able to understand and render structures, the actual components used can be displayed in the main interface as you review your results.

Other processes where Chrom Reaction Optimization can be useful

In the above discussion, we have focused on optimising reactions, and touched briefly on parallel reactions. It is worth stepping back for a second, though, to consider other analyses where the approach can be useful.

At a slightly higher level, the purpose of the Chrom Reaction Optimization plugin is to find a set of components in a set of different samples. A reaction where conditions such as solvent, catalyst, and temperature are varied is, of course, just one example of where such an analysis could be useful. Following reaction kinetics is another such case. Here, instead of varying reaction conditions, the variable that is changing is time, yet the same workflow can be used to extract the relevant information about components.

Another application might be to look at the best way to clean up a crude reaction mixture. Perhaps you are using crystallisation to isolate a product or remove impurities. Or perhaps you are washing a slurry with different solvent mixtures to achieve the same result. Either way, the same Chrom Reaction Optimization workflow can be used, with the goal of maximising or minimising the product, depending on which way round the separation is working.

Integration into the bigger picture

Automating analyses and rapidly determining optimum reaction conditions for a reaction is a significant time-saver for scientists. However, no process operates in isolation, and almost certainly the results will need to fit into the larger context of a research organisation. Perhaps the results need to be stored in a database for machine learning training. Perhaps they need to be reported in a LIMS. Almost certainly, they must be included in an ELN. The gains in efficiency can feel squandered if the final step involves time-consuming manual copying of data between systems.

Fortunately, Chrom Reaction Optimization and the Mgears engine offer several solutions to this problem. In addition to the individual documents generated for each analysed well, a customisable CSV file is produced. This CSV file can be used to populate another system or can otherwise be attached to an ELN. Mgears also provides built-in support for posting results to several ELN providers or to an Mnova spectral database. The CSV file is also very useful for analysis of the results in data exploration tools such as Spotfire or PowerBI.

However, your integration requirements may be more complex than this. Maybe you need to automatically write the results to an SQL database or send them to an SMDS or LIMS system. The Mnova script engine can provide a straightforward solution to achieve this. The simple JavaScript interface, together with access to objects that can communicate with SQL databases or RESTful web services, enables a high level of integration. This means that with a little customisation you can achieve seamless integration, and so the dreaded '*now I need to write it up in my ELN*' problem is greatly reduced.

Conclusions

Reaction optimisation is a critical step in developing efficient syntheses for final products or intermediates. While many approaches can be used to reduce the number of trial reactions, there is still a need to efficiently handle and analyse the generated LCMS data to quantify components. Mgears and Chrom Reaction Optimization provide a rapid, flexible, and configurable solution to this challenge, potentially saving laboratory scientists significant time. By providing efficient ways to locate data, define components, review results, address issues, and integrate with other systems, these tools streamline the analysis process and significantly enhance the productivity of research organisations.

Mestrelab Research

Find out more at
mestrelab.com/software/mnova-gears-chrom-reaction-optimization/

Contact us for general information or sales inquiries.