A Universal Work Flow Automation Toolbox for Mass Spectrometry

Don Kuehl, Stacey Simonoff, Yongdong Wang
Cerno Bioscience, Norwalk, CT, USA

Presented at Pittcon2018
March, 2018
Orlando, FL

The Objective: Provide automated MS solutions to a diverse market

Mass spectrometry is a versatile, sensitive tool that covers a vast array of applications including environmental, forensics, clinical, food safety, proteomics, drug metabolism and quality control to name a few. This diversity presents tremendous challenges to automating repetitive analytical procedures and the varying needs of each laboratory. While instrument vendors provide a vast number of different software packages to address the most common applications, they are typically limited in their scope and unfortunately, are proprietary to each vendor platform. This also makes it challenging to integrate third party applications which further restricts users from implementing efficient workflows.

Cerno has developed, and made available a number of useful data processing technologies over the past decade. These technologies include a MS calibration technology[1,2,3] (TrueCal™) that provides dramatic improvements on unit resolution LC and GC quadrupoles to ±5 mDa and more importantly the isotope lineshape profiles (Spectral Accuracy) to greater than 99.0% on both high and low resolution instruments. This, and other novel technologies including spectral mixture and chromatogram deconvolution can be used to save time, improve the efficiency and reliability of MS analysis. However, the challenge is to provide these technologies as a integrated software solution that is easily configurable to the vast array of MS applications and instruments.

The Solution: MassWorks Rx: a modular, extensible layered software with a multi-level automation interface

Software development is always a balancing act between speed, flexibility, maintainability, and ease-of-use. For pure speed, C++ working with optimized math libraries are critical to computationally intensive routines such as those necessary for Cerno's technology. The downside is a lack of flexibility to adapt to special needs. A common solution to this is to layer or embed an interpreted language into the product. Agilent Chemstation macros and Thermo's (Galactic) Array Basic are good examples of this approach. However such embedded languages can be problematic to support as the software evolves. In addition, the interpreted nature can make them slow, and integration with third part applications somewhat cumbersome and/or limited. A third level of customization is the "Method". This is typically a high level text or GUI oriented scripting language commonly used in many chromatography data systems.

In our approach, we have tried to combine the best of each solution in a layered application. The bottom layer is a collection of C++ modules that each provide a unique functionality and are optimized for speed and robustness. Adding new capabilities or integrating third party software is simple a matter of creating additional modules that conform to the architecture. The "glue" that creates the application layer is written in Python. Python likely is the most popular and powerful scientific programming language in the world today. It has many advantages including 1) it can easily call into C++ libraries, 2) it's interpreted for rapid prototyping, yet 3) it can be compiled to C++ for speed and security 4) it is a modern, object oriented language 5) it has huge open source support in the scientific community with freely available libraries suitable for many scientific applications.

Finally, for ease of customization by end users, there are powerful method editing capabilities which can easily adapt the application to most any processing or reporting capability. This application with its novel architecture is called MassWorks Rx, with the Rx providing the prescription to easy automation. This architecture is illustrated in Figure 1.

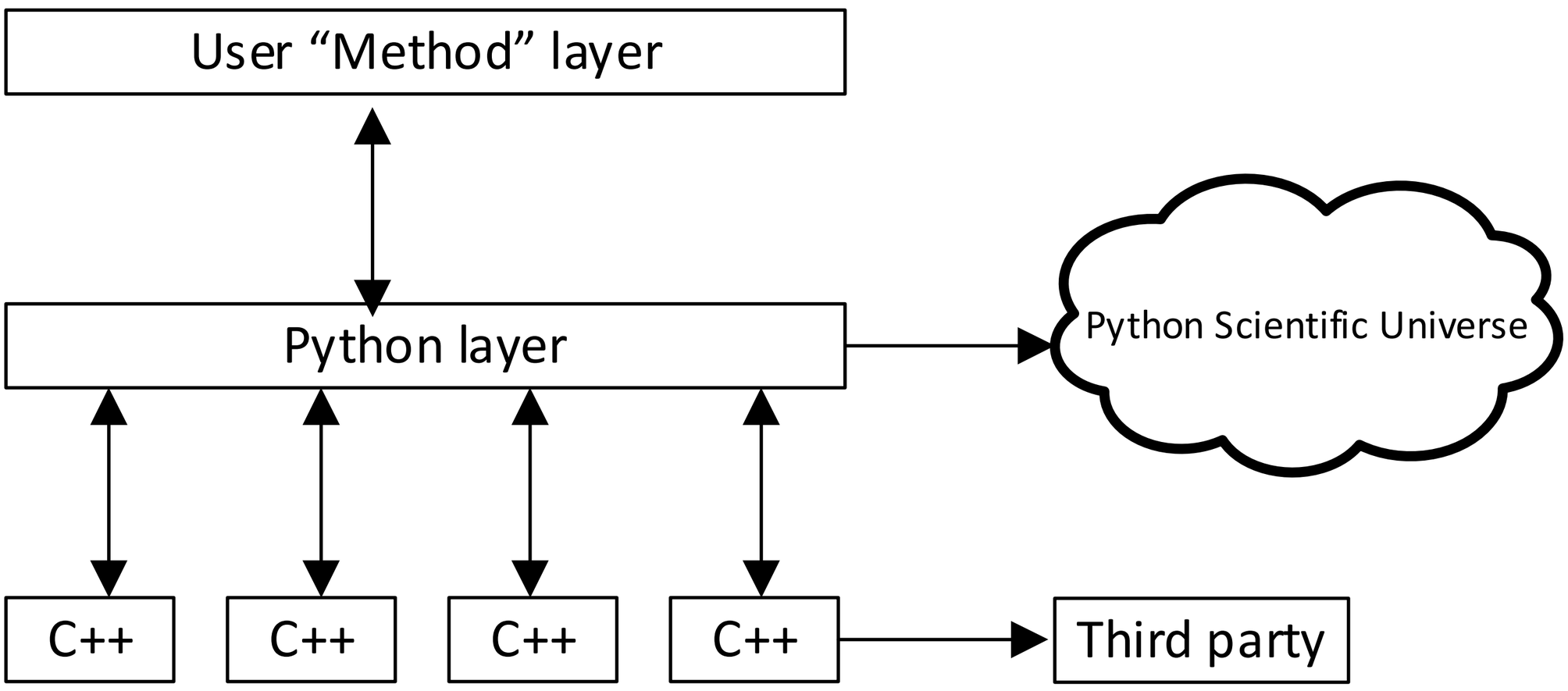
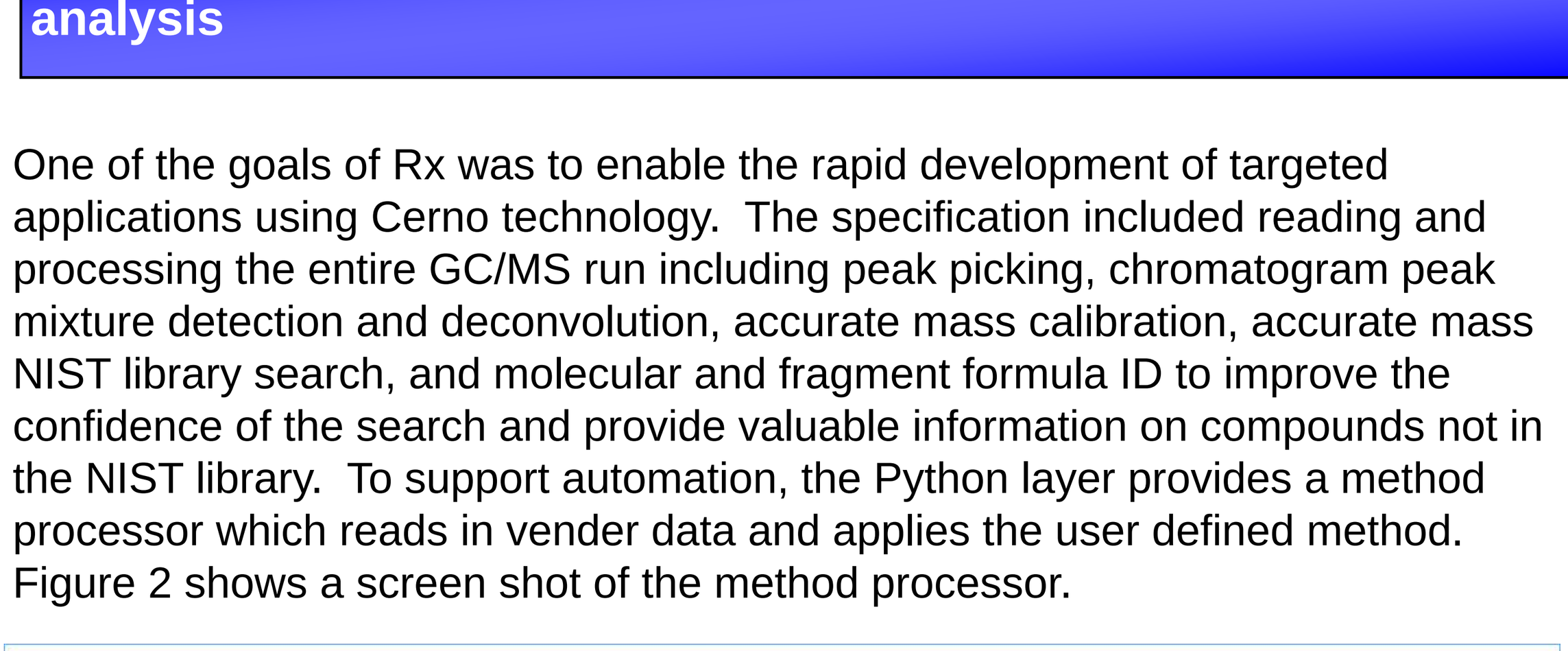


Figure 1. Basic block diagram for MassWorks Rx. The C++ layer provides speed and robustness, the Python layer is highly flexible with huge scientific library support and allows rapid development of application “personalities” while the Method layer give end user flexibility.

Example Application: GC/ID for automated GC/MS qualitative analysis

One of the goals of Rx was to enable the rapid development of targeted applications using Cerno technology. The specification included reading and processing the entire GC/MS run including peak picking, chromatogram peak mixture detection and deconvolution, accurate mass calibration, accurate mass NIST library search, and molecular and fragment formula ID to improve the confidence of the search and provide valuable information on compounds not in the NIST library. To support automation, the Python layer provides a method processor which reads in vender data and applies the user defined method. Figure 2 shows a screen shot of the method processor.



Run Name	Data File	Method	Status	Date and Time	Notes
Calib#33	VOC_JPCerno_ACCMass_5050_081718.D	Basic	Processed	2018/02/05 08:44	RI Calibration
R180205#1002	VOC_JPCerno_ACCMass_5050_081719.D	Basic	PENDING..	2018/02/05 09:58	Start Series
R180212#1001	VOC_JPCerno_ACCMass_5050_081720.D	Basic	PENDING..	2018/02/12 13:16	Rerun 1
R180215#1001	VOC_JPCerno_ACCMass_5050_081721.D	Basic	Processed	2018/02/15 15:42	Series 2
R180216#1002	VOC_JPCerno_ACCMass_5050_081722.D	Basic	Processed	2018/02/16 10:23	Series 3
R180216#1001	VOC_JPCerno_ACCMass_5050_081723.D	Basic	Processed	2018/02/16 10:39	Series 4
R180216#1003	VOC_JPCerno_ACCMass_5050_081724.D	Basic	Processed	2018/02/16 10:41	Series 5
R180216#1004	VOC_JPCerno_ACCMass_5050_081725.D	Basic	Holding	2018/02/16 10:53	Series 6
R180216#1005	VOC_JPCerno_ACCMass_5050_081726.D	Basic	Processed	2018/02/16 11:24	Series 7
R180219#1001	VOC_JPCerno_ACCMass_5050_081727.D	Basic	Processed	2018/02/19 08:41	Series 8
R171229#1004	VOC_JPCerno_ACCMass_5050_081728.D	Basic	Processed	2018/02/19 10:11	Series 9

Figure 2. The Rx method processors queues up the data to be processed according to the method assigned. Data is archived in a secure, accessible database.

Once the data is available, it is automatically processed as follows:

1. Data is read from the native MS vendor format
2. The TIC is used to locate peaks
3. For each peak the number of components is determined
4. The “pure” spectra are extracted from any mixture peaks
5. All pure spectra are searched against the NIST libraries
6. The MS are calibrated to accurate mass (TrueCal)
7. Formulas are determined for the molecular ion and fragments ions to verify the search results
8. Data are archived in a secure but accessible database

Once finished, the job is available for review and a PDF report can be distributed or archived.

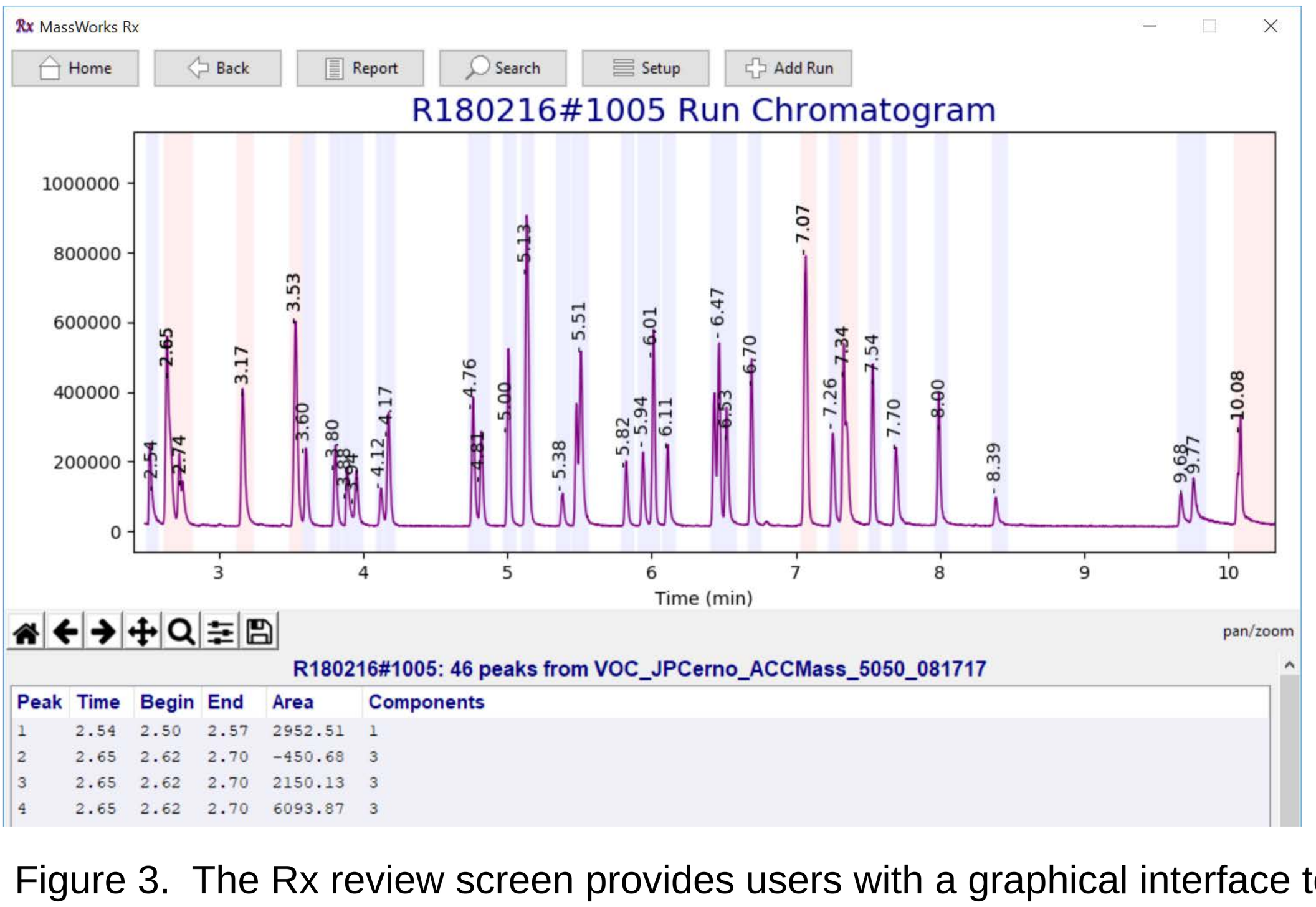


Figure 3. The Rx review screen provides users with a graphical interface to examine all the results processed from the run.

Summary

Rx was designed from the ground up as a vendor neutral, flexible automation processor applicable to the wide diversity of applications in MS. Many other applications can be addressed using Rx by simply changing the Python “personality” and modifying the methods to easily adapt to the desired workflow. For example, deamination studies can be easily automated by using another Cerno technology available in Rx which provides powerful mixture detection and quantitation for the native and deaminated form of the protein. This is an application which might normally take a 90 minute HPLC run but with Rx can be done automatically in half the time.

References

1. Y. Wang, M. Gu., Anal. Chem., **2010**, *82*, 7055.
2. R.J. Abel, Canadian Journal of Forensic Science, **2014**, 47, 74.
3. S. Moldoveanu & K. Kilby, ASMS Annual Meeting, 2014.