Presented at HPLC 2009, Dresden, Germany, June 28–July 2, 2009

Building a Global Chromatography Knowledgebase for Method Development Prediction

Mike McBrien
Advanced Chemistry Development, Inc., Toronto, ON, Canada

Gang Xue, Jeff Harwood, and Yanqiao Xiang
Analytical R&D, Pfizer Global R&D, Groton, CT, USA

Andrey Vazhentsev
Advanced Chemistry Development, Ltd., Moscow, Russia

INTRODUCTION

HPLC is perhaps the greatest workhorse of analytical chemistry. Chemical and pharmaceutical companies collect an astounding amount of data during investigations into the contents of various samples. Over the course of these investigations, an incredible amount of knowledge of chromatography is also being collected. The behavior of each known molecule under a given set of conditions is an additional piece of knowledge that is routinely discarded immediately after collection. Collecting and leveraging this knowledge effectively presents an interesting challenge.

STRUCTURE-BASED PREDICTION OF CHROMATOGRAPHY

There have been a considerable number of publications on the prediction of chromatographic behaviour based on chemical structures, each with their own strengths and weaknesses. Due to the complexities of liquid chromatography, there is not (and perhaps may never be) a global model spanning all potential methods and all potential structures; each model is specific to a given method. For this reason, it is necessary to collect information regarding each chromatographic method of interest, accumulating knowledge for that particular method, before prediction can be done. Once the knowledge has been collected for a sufficient number of compounds, a system can be devised to predict the behavior of new compounds.

ACD/ChromGenius is a system for structure-based prediction of retention times for generic chromatographic methods. The system is designed to actively generate localized prediction models for a given structure by finding and grouping similar molecules, and then relating physicochemical parameters to retention time. This "Federation of Local Models" approach can be extremely accurate, but requires a reasonable sized knowledgebase in order to function effectively. There has to be sufficient knowledge of a given structural area before an accurate prediction can be expected. Clearly it is desirable to have an efficient way to train such a system.

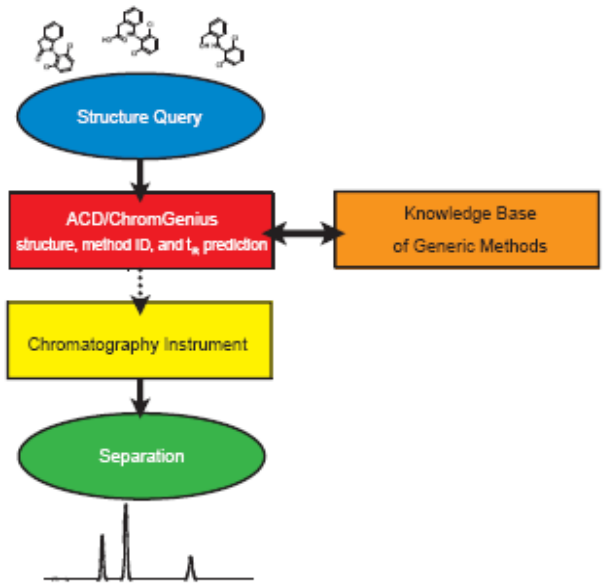

Figure 1: ACD/ChromGenius: a "Federation of Local Models" approach to structure-based chromatographic prediction.

COMET: COMBINING PARAMETER SCREENING AND COMPONENT DETECTION

While optimization of continuous variables such as temperature and gradient are commonplace, discontinuous variables such as organic solvent and column choice are typically screened to identify combinations with desirable selectivities. The screening techniques used are normally the same regardless of the sample involved, thus these are generic methods that are applied to virtually every method development problem. Therefore, method development can be a rich source of non-routine injections under generic method conditions, provided that the information can be extracted efficiently.

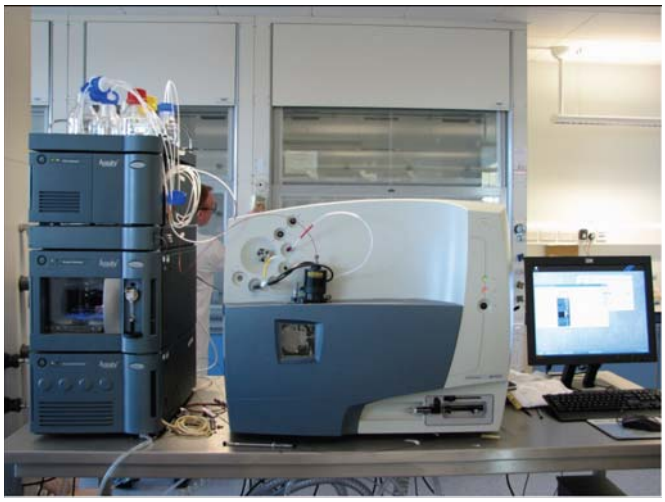

Figure 2: The COMET screening system.

The COMET system is one approach to parameter screening1. Column, pH, and organic modifier are all varied in order to maximize the selectivities investigated, with chemometric elution data extraction. COMET does not preclude use of all 18 of the potential combinations of the systems shown in table 1 (a "symmetrical" screen), but it is possible to do an asymmetrical screen with as few as 3 experiments. Since a huge number of unknown samples are investigated with these methods, there is a considerable amount of organizational knowledge collected.

	BEHRP18	BEH C18	BEH Phenyl
BEHRP18 0.1% HCOOH/ACN		0.558	**0.115**
BEH C18 10 mM NH4Ac/MeOH			0.384
BEH Phenyl 0.1% NH4OH/MeOH			

Table 1: The orthogonal methods used for the COMET screening system.

DATA EXTRACTION

To create a chromatographic prediction model, it is necessary to link chromatographic elution data with the chemical structures that are involved. Manual data interpretation can be a major barrier to creating a knowledgebase. However, a large part of the data reduction required is already done as part of the method development workflow. ACD/AutoChrom MS has the capacity to perform automated peak tracking based on LC/MS data. When performing peak tracking using LC/MS, part of the function of the MS-MAP tracking algorithm is extraction of the weight of the ionized molecule (WIM). The peaks for a given WIM are labeled for all injections that showed ionization. Since these are linked to an overall component table, one simple operation can link a chemical structure to the elution data for each set of conditions.

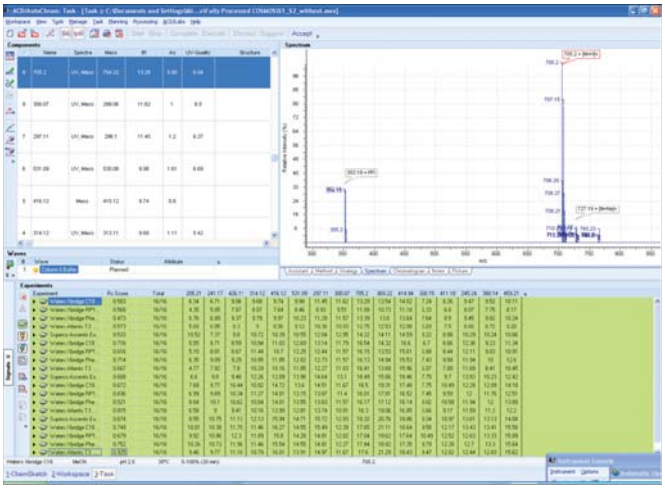

Figure 3: The ACD/AutoChrom project. As part of the standard peak tracking operation, weights of ionized molecules are extracted. These can be used to quickly identify the peaks for known components.

KNOWLEDGEBASE CREATION

Once structures have been attached to the component table in the project management system, transfer of the elution information requires practically no effort. As part of the project sign-off procedure, the user can automatically move all structures, elution information, and methods to the ChromGenius knowledgebase. The system contains complete integrity checks, with comparison between present and past methods for consistency, and hyperlinks back to the original project. Once in the knowledgebase, the methods are grouped, and the elution model is automatically built.

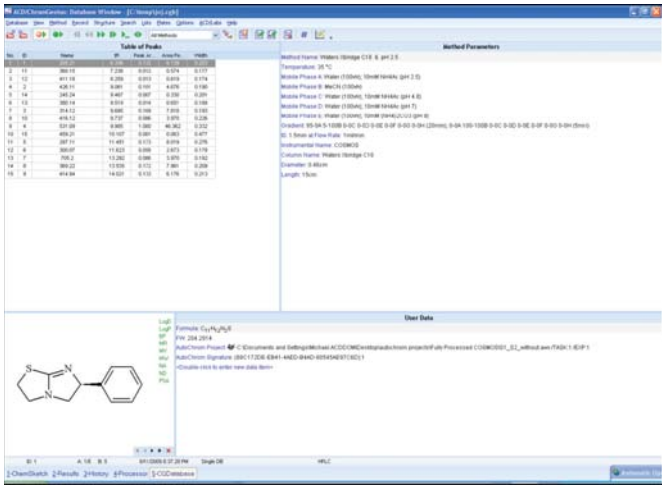

Figure 4: The ACD/ChromGenius record. All structures for the injection are updated to the record.

The training of the system requires almost no additional work beyond the original COMET process. The raw data is moved into the AutoChrom project, and impurities are detected by the MS-MAP system. The user verifies the data extraction, attaches any known structures, and updates to the knowledgebase.

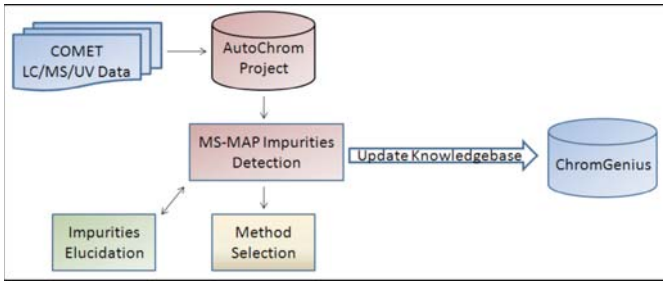

Figure 5: Information flow.

FUTURE WORK

It is interesting to note that in early drug development, few impurities are known. Indeed, one of the objectives of the parameter screening process is the detection of unknown impurities. Elucidation may take place weeks or months after the chromatographic measurements have been completed, but these comprise a great deal of the knowledge collected. Today, it is necessary to return to the project when elucidation is complete in order to transfer data for impurities that were not known at time of screening. In the future, it is anticipated that the elucidation process will be directly linked to the chromatographic project, meaning that as structures are elucidated, the legacy chromatographic knowledge will be automatically transferred to the knowledgebase.

CONCLUSION

Connection of COMET to ACD/AutoChrom and ACD/ChromGenius demonstrated the capacity to create a structure-based chromatographic prediction system using knowledge gained from standard method development practices. Manual work beyond that which must be performed for method development is minimal, with automated data transfer and integrity checks.

REFERENCES

1. Gang Xue , Anne D. Bendick, Raymond Chen and Sonja S. Sekulic, *Journal of Chromatography A*, **1050** (2): 159–171, 1 October 2004.