

Processing Results of Polymer Liquid Chromatography Using Multi-Agent System

STANISLAVA LABÁTOVÁ
DUŠAN BEREK *
Institute of Informatics
*Polymer Institute
Slovak Academy of Sciences
84507 Bratislava, Dúbravská cesta 9
SLOVAKIA
www.sav.sk

Abstract: In this contribution we apply the multi-agent system approach to treat selected SEC (Size-excluded chromatography) and enthalpy assisted SEC calibration curves. Individual agents represent approximation of experimentally measured calibration curves by different types of empirical functions. Decision agent detects the best model of simulation. The genetic programming is suggested for the approximation of calibration curves by arbitrary function.

Keywords: multi-agent systems, genetic programming, modeling, liquid chromatography of synthetic polymers

1 Introduction

Software products that represent automation of traditional human activities are codified as agent software. 'Agent' software may realize various human functions. In real life, individual people collaborate with each other to reach some goals. Multi-agent system represents a computer simulation of certain part of social community that wants reaching identical goals. Thus, each agent represents a profession – handicraft. It is evident that when agents represent activities of people, there exists similarity and connectivity between agent systems and biological systems. It may be of interest to utilize ideas from biological systems – genetics – to create agent's software. The idea to utilize inspiration from biology in information technology is not new: in early seventies of the last century Holland utilized genetics to implement the well-known genetic algorithm. Genetic algorithm mimics the process of evolution in nature. Evolution produces many optimal forms despite its randomness, and absence of directing intelligence.

In this contribution we apply the agent system approach for processing results of polymer liquid chromatography.

2 Multi – Agent system

Multi-Agent System (MAS) is a computer simulation of certain part of social community where each agent has some own goals and where a common aim exists, which the system wants to realize. MAS is based on the generation of autonomous agents. Agents are chosen so that the goals of MAS can be reached in optimal way. General characteristic of MAS are as follows [4]:

- The system is decentralized, i.e. agent, or the group of agents is available to evoke some actions of other agents.
- The system is distributed, i.e. behavior of the system is conditioned by behavior of single agents.
- Next states of agents depend only on previous action.
- The system is open, i.e. new agent activation evokes change of mode to obtain the final state.

3 Polymer liquid chromatography

End-use properties of synthetic polymers depend on the molecular characteristic of polymer molecules (macromolecules), on the arrangement

of macromolecules in the solid system and on amount and nature of additives. It is evident that for optimization of polymer properties it is necessary to precisely determine molecular characteristics of polymers. These are molar mass, chemical structure (composition, functionality) and physical architecture. All man made polymers exhibit certain distribution of their molecular characteristics. As result, molecular characteristics of synthetic polymers are described not only by the average (mean) values but also by distributions. It is useful to know both of them because they both affect polymer properties. For example, the higher molar mass and the narrower molar mass distribution, the higher is the mechanical strength of a polymer. However, very high molar mass materials are difficult to process and therefore often also smaller macromolecules must be present. The optimum is to prepare systems with a bi-modal distribution, where large species are responsible for mechanical properties and smaller species allow easy processing.

At present, the most common method for molecular characterization of synthetic polymers is size exclusion chromatography (SEC). SEC belongs to the family of high performance liquid chromatography (HPLC) procedures and enables to determine molar masses of polymer species both averages (M) and distributions (MMD). SEC separates macromolecules according to their size in a solution. Dissolved polymer sample is injected into an SEC column packed with microspherical porous particles and eluted with an appropriate eluent, (mobile phase). The result of SEC separation is called chromatogram. The chromatogram is a dependence of polymer concentration on retention volume (V_R), that is on the volume of mobile phase, which is needed for elution of macromolecules of certain size. V_R is a function of the size of macromolecules. Unfortunately, both structure of pores and behavior of macromolecules in a solution are complicated and therefore it is so far impossible to formulate a quantitative theory of the SEC process and to calculate M and MMD directly from chromatograms. Therefore, the SEC separation column must be calibrated with a series of polymer standards with known molar masses. The $\log M$ vs. V_R dependences are

constructed and used for SEC data processing. They are called SEC calibration dependences. The calibration dependences are not linear. They are often described by polynomials. Another drawback of SEC is its low selectivity of separation. Moreover, raw SEC data cannot be used for molecular characterization of complex polymer systems, which exhibit more than one distribution in their molecular characteristics that is not only in their molar mass but also in their chemical structure and/or physical architecture. Typically, these are copolymers, polymer blends, functionalized polymers, stereoregular polymers, etc. This is because in the case of complex polymers the V_{RS} of macromolecules depend on at least two characteristics. Quantitative interpretation of such chromatograms is not possible. In order to solve above problems novel approaches to liquid chromatography separation of synthetic polymers were elaborated, which combine two or even several separation mechanisms. These approaches are called *coupled method of polymer high performance liquid chromatography (polymer HPLC)* [3]. The coupled methods of polymer HPLC combine entropic (exclusion, SEC) separation mechanism with the appropriate enthalpic retention mechanism(s), such as adsorption, enthalpic partition, phase separation or ion interactions. The aim is to either enhance or suppress separation of macromolecules according to one molecular characteristic so that the effect of other characteristic(s) can be neglected. In this case, the resulting separation can be interpreted in terms of one single molecular characteristic. The suppressed molecular characteristic of analyzed polymer is usually its molar mass. The resulting dependence of $\log M$ vs. V_R is vertical.

The enhancement of liquid chromatography separation selectivity of synthetic polymers is important also for simple polymers (homopolymers), which exhibit single distribution in molar mass. The precision of analyses increases and the heterogeneities in molar mass distribution, which cannot be disclosed by conventional SEC may be visualized. The corresponding coupled polymer HPLC methods are called *enthalpy assisted SEC* because the basic courses of $\log M$ vs. V_R dependences are the same as in SEC (SEC-like

situation, retention volumes decrease with increasing polymer molar mass) but their slopes are reduced – larger difference in V_R is caused by smaller change in M .

Evidently, it is much more difficult to describe mathematically the complex courses of $\log M$ vs. V_R curves for coupled polymer HPLC than for simple SEC and to further process the parameters of resulting dependences. In this contribution, we will apply the multi-agent system approach to treat selected SEC and enthalpy assisted SEC calibration curves.

4 Evaluation of SEC calibration curves

An important step in the treatment of data obtained by SEC and enthalpy assisted SEC to detect MMD or M is modeling $\log M$ vs. V_R calibration curves (CC). In regression analysis the y variable is used usually as a dependent (measured variable) while x is the independent variable. In the calibration model the x value is an unknown sample. In practical SEC calibration retention volume (V_R) and independent variables $\log M$ are exchanged. The reason for this is not known, may be it is historical. Examples of the calibration curves for a coupled polymer HPLC are shown in Figs. 1–5. Narrow molar mass distribution polystyrenes (PS) were eluted from a series of columns packed with different materials to determine how the enthalpic retention mechanism affects retention of macromolecules. The column packings matrix was porous silica gel, either bare or modified with chemically bonded C-18 alkyl groups. The basic exclusion retention mechanism of polystyrene dominated polymer elution from the non-modified porous silica gel. (Figure 1) Matrices of further four column packings were identical silica gels but they were modified with different amount of alkyl groups (characterized by percentage of carbon). As eluents, pure tetrahydrofuran (THF), dimethylformamid (DMF) and their mixtures were applied. THF is a thermodynamically good solvent for PS and macromolecules elute in the SEC mode irrespectively of presence of C-18 bonded groups. On the other hand, DMF is a poor solvent for PS. This solvent „pushes“ polymer species into the bonded alkane phase and strongly promotes their enthalpic partition, which results

in their retention and increase of their V_{RS} . As result, adding DMF to THF eluent leads to coupling of exclusion and enthalpic partition. Shapes of CC are changed and at certain DMF concentration, the enthalpic retention mechanism may prevail (V_R increase with increasing polymer molar mass).

It was of interest to see whether polystyrene species would trace the alkyl groups and whether the SEC-like separation selectivity can be improved in this way. At a low degree of coverage of silica surface with alkyl groups (4% of carbon) the enthalpic partition plays little role and the CCs exhibit SEC shape with a rather low separation selectivity. At 8% of carbon (Fig.3), the CCs change their courses rather dramatically in dependence on the eluent composition. A slight increase of separation selectivity is observed at 15 wt.% of THF. This tendency is pronounced at 12 % (Fig.4) and even more at 18 % (Fig.5) of carbon. The effect of enthalpic partition of polystyrene species and consequently the SEC-like separation selectivity strongly increases.

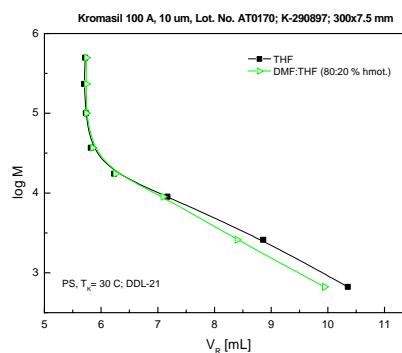


Fig.1. CC for Kromasil 100A

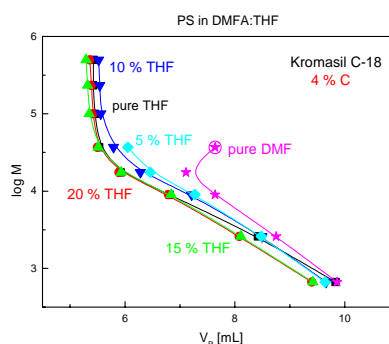


Fig. 2. CC for Kromasil C-18, 4% C

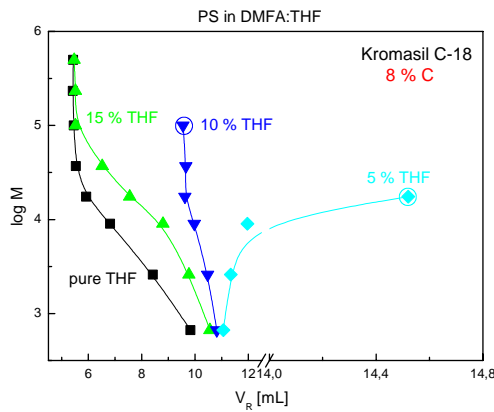


Fig.2. CC for Kromasil C-18, 8% C

For creating a model describing the determined calibration dependences in first level we suppose that empirical formula has the form:

$$y = Ae^{Bx} \quad (1)$$

where the exponential part is represented by straight line. From (1) we obtain equation:

$$\ln y = \ln A + Bx, \quad C = \ln A \quad (2)$$

Experiment SEC generates a set of points:

$$\{x_i, y_i\}, \quad i = 1, 2, \dots, n \quad (3)$$

For determination of coefficients B, C, from (2) this method is used to minimize function Φ given by discrete set of points (least square regression procedure).

$$\Phi(C, B) = \sum_{i=1}^n (C + Bx_i - \ln y_i)^2 \rightarrow \min \quad (4)$$

so we obtain equations:

$$\frac{\partial \Phi(C, B)}{\partial C} = \sum_{i=1}^n (C + Bx_i - \ln y_i) = 0$$

$$\frac{\partial \Phi(C, B)}{\partial B} = \sum_{i=1}^n (C + Bx_i - \ln y_i)x_i = 0 \quad (5)$$

Solution of this system of linear equations determines coefficients B, C, where for coefficient B

$$B = (n \cdot S_{xy} - S_y \cdot S_x) / (n \cdot S_{2x} - S_x^2) \quad (6)$$

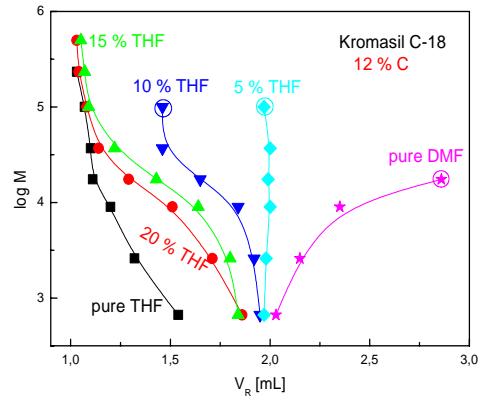


Fig. 3. CC for Kromasil C-18, 12% C

where

$$\begin{aligned} S_y &= \sum_i \ln y_i & S_{2x} &= \sum_i x_i^2 \\ S_x &= \sum_i x_i & S_{xy} &= \sum_i x_i \end{aligned} \quad (7)$$

To calculate coefficient C a similar method is used. This activity, i.e. the calculation of coefficients for approximation of calibration curves in form (2), is represented by agent 1. As the calibration curves have different both shape and complexity, we used different empirical formulas for approximation of calibration curves. The empirical formula

$$y = Ae^{(Bx^2 + Cx + D)} \quad (8)$$

where exponent is polynomial of 2nd order and is base of activity of agent 2 and the empirical function

$$y = Ae^{(Bx^3 + Cx^2 + Dx + E)} \quad (9)$$

represents agent 3.

As an example, coefficient A in equation (8) is

$$A = (S_{2xy} \cdot S_{2x} \cdot n + S_{3x} \cdot S_x \cdot S_y + S_{2x} \cdot S_{xy} \cdot S_x - S_{2x} \cdot S_{2x} \cdot S_y - S_{3x} \cdot S_{xy} \cdot n - S_{2xy} \cdot S_x \cdot S_x) / M \quad (10)$$

$$B = (S_4 \cdot S_{xy} \cdot n + S_{2xy} \cdot S_{2x} \cdot S_x + S_{2x} \cdot S_{3x} \cdot S_y - S_{2x} \cdot S_{xy} \cdot S_{2x} - S_{2xy} \cdot S_{3x} \cdot n - S_4 \cdot S_x \cdot S_x) / M \quad (11)$$

$$M = S_4 \cdot S_{2x} \cdot n + 2 \cdot S_{3x} \cdot S_x \cdot S_{2x} - S_{2x} \cdot S_{2x} \cdot S_{2x} - S_{3x} \cdot S_{3x} \cdot n - S_4 \cdot S_x \cdot S_x \quad (12)$$

where individual terms in equations (10) – (12) are expressed similarly as in (7).

To decide which model satisfies the experimental conditions we use the decision agent based on the root mean squared error (RMSE) method.

RMSE is a measure of the average difference between measured y value and the value predicted by model y' .

We obtain different models for individual calibration curves, and the model is chosen, which fits better to the data.

The error is defined as:

$$E = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n}} \quad (13)$$

The decision agent detects which agent from $\{A_i\}$, $i=1, \dots, n$ is chosen and fulfils the given conditions. A problem occurs when no agent satisfies the given conditions and we have to find another model with better fit. We suggest to use genetic programming to solve this problem.

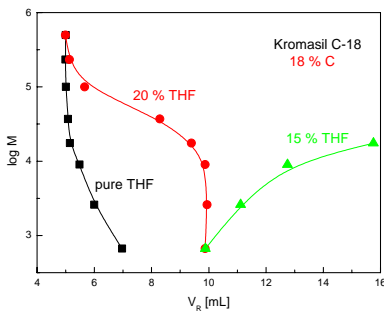


Fig.5. CC for Kromasil C-18, 18% C

For classification characteristics of polymers we need to specify functional coherence among individual data obtaining from SEC. Success in this area depends on model exactness that expresses individual experiments. Functional dependencies of model determine the characteristics of polymers. From this point of view, in the sequel we describe the approach for obtaining different models for different experiments.

5 Genetic Programming

Genetic programming was founded by John Koza [2]. While genetic algorithms are most suitable for the problems of parametric optimizations – to find out optimal parameters for strictly defined structures, in genetic programming the object of optimization is represented by internal structure of objects, the connectivity among elements. In this way the dimension of the problem becomes complex and extensive.

5.1 Symbolic regression

The goal of symbolic regression, or system identification, is to find out the function describing the relation in finite set of data (measured points), i.e. the identifications of functions by deviation-controlled evolution.

In our case we have a measured set of points and we want to find out the regression function that approximates the dependence among those data points. In the previous approach we have showed the method of exponential regression, which detected the coefficients of approximated polynomial, however, in many cases this approximation does not find satisfactory solution. As the SEC creates calibration curves of different shapes, we suggest a solution where approximation function may be arbitrary function, which may consist of combinations of different functions (+, -, *, /, sin, cos, ln, exp, and, or,...), as well as various parameters and constants. We limit not only the number of single functions but also their type and mutual connectivity. In this case individual members of the population may be structures with variable types and lengths. Objects function is the measure for obtaining behavior or succeeding in competition with others.

5.2 Representation of individuals

Transformation of potential solutions (phenotype) to the strings (genome) is a difficult problem. The most frequent method of representation of individuals in genetic programming is represented by tree structures or linear structures. In tree structure we distinguish nodes connected by edges. Two kinds of nodes are used, namely terminals and functions. Terminals are variable or constant, they bring values to the tree. Functions are located in the tree and transform variables and

constants. Then the tree structure has to be transformed into linear structure, for example, by Read linear code where a number of edges is assigned to the node, which entered the node (valence), and then by bottom-up concatenation the next valence is added to the string. The Read code contains complete information about the tree structure. To realize genetics operations – crossing and mutation – special operators are created which are suitable for the given task. The method of presence of the best individual or that of tournament is accepted for selection.

6 Conclusion

We suppose that genetic programming is suitable for solving the problem of approximating different kinds of curves obtained by polymer HPLC. Many problems of optimization tasks were solved by genetic algorithms when classical methods have failed. Inspiration from biological system for solving many kinds of synthetic problems is very useful.

Acknowledgement

This contribution has been prepared within the projects VEGA 2/4148/04: Modeling and Control of Distributed Processes on the Base Multi-Agent Systems, and APVT –51-013204: Analysis of minor components in complex polymer systems.

References:

- [1] 75. J.H.Holland: Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975
- [2] 92. J.R.Koza: Genetic Programming, MIT Press, Cambridge, MA 1992
- [3] 00. D. Berek: Progr. Polym.Sci. 25(2000) 873-908
- [4] 94. S. Labatova: Multi-Agent Technology Based on Genetic Processing, WSEAS Transactions on Information Science and Application, Vol. 1, Nov. 2004, 1170-1174