Author Query Form

Dear Professor,

All queries pertaining to chapter_____3_____ are listed below. These require your responses. Please provide your responses in the last column by using the PDF reader so that we can make the necessary amendments as per your advice.

Queries pertaining to chapter/chapters:

Page No	Queries	Author Response
59	AQ01: Should the type of chromatography be specified here?	
60, 65, 72, 73, 83, 84	AQ02: Figures **[3.3, 3.8, 3.10, 3.11, 3.13 and 3.14]** are of low resolution. Please provide better resolution figures.	
69	AQ03: Please check the sentence for clarity.	
91	AQ04: Please rephrase the sentence for clarity.	

Chapter 3

ChromSword®: Software for Method Development in Liquid Chromatography

Sergey V. Galushko*‡, Irine P. Shyskina*, Evalds Urtans† and Oksana Rotkaja†

*ChromSword, Dr. Galushko Software Entwicklung GmbH Im Wiesengrund 49B, 64367, Muehltal, Germany

†ChromSword Baltic, Antonijas 22-1, Riga, LV-5041, Latvia

‡galushko@chromsword.de

3.1 Introduction

Method development in chromatography can be considered as a process studying the empirical relationships between the quality of a chromatogram and the chromatographic conditions. A chromatographer changes conditions to find an acceptable method to achieve separation in a reasonable time. The time required to find optimal conditions or to make any conclusion can be substantially reduced by using computer programs for method development. HPLC method development programs can be utilized interactively (off-line) and for automatic optimization (online). ChromSword® for off-line computer-assisted method development was launched in 1994 as an extension of ChromDream® software [1]. During 1998–2000, the first version for unattended method development was started [2]. The latest version of ChromSword® combines different technologies of method development in one software platform:

- Computer-assisted
- Automated optimization

- Automated robustness studies
- Scouting to screen different column, solvents, buffers and methods

It is possible for a chromatographer to use only the computer-assisted (off-line) or automated method development approach or to use both interactive and unattended optimization.

ChromSword® off-line can be used for optimizing separations in reversed-phase (RPLC), normal-phase (NPLC) and ion-exchange (IEX) liquid chromatography (LC). In the off-line mode, chromatogram simulations and optimizations as a function of one or more variables are possible. The off-line mode includes two possibilities for optimization in RPLC.

The approach which takes into account the characteristics of compounds and column/solvent properties is the solvatic or solvophobic model of RPLC.

The traditional method for optimizing separation using only retention data of analytes is the linear solvent strength (LSS) model and other polynomial models.

In the automated mode, the software operates as a chromatography data system controlling HPLC instruments and executes a sequence of runs. The user can predefine such a sequence of runs — this is a scouting approach to screen different stationary phases (SPs) or mobile phases (MPs) or statistical design of experiments (DoE) according to some statistical rules to study the effect of method variables on the separation. This method is defined as robotic process automation. Another approach is intelligent automation. Intelligent automation automates non-routine tasks like optimizations involving complex data processing and reasoning. ChromSword® supports both types of automation to assist chromatographers for routine and intelligent method development workflow.

To support various method development workflows ChromSwordAuto® package contains modules dedicated to different scenarios and tasks:

ChromSword®	for computer-assisted method development
ChromDraw®	chemical editor for drawing and processing structural formulae

ColumnViewer®	reversed-phase column properties data base
ChromSword® Scout	for automated method screening
ChromSword® Developer	for automated method optimization
AutoRobust®	for automated robustness study and method transfer
ReportViewer®	for data browsing, chromatogram and spectra processing, project management and report generation

3.2 Automated Method Development

Most automated HPLC method development approaches can be divided into three classes:

- Mechanistic or model-based optimization.
- Statistic or direct process optimization.
- Screening or running a large number of column/solvent/method combinations to identify those with a reasonable separation.

In the model-based optimization, mathematical models are utilized to reduce the number of experiments. The development of mechanistic models requires good chromatography understanding, reliable tests for parameter estimations and peak tracking. Limiting factors are computational time and reliability of the models that are applied for simulation and optimum search. The determination of mechanistic model parameters can be complicated for computer-assisted (off-line) method development and requires time and operator qualification for optimization of multi-component mixtures. Automatic optimization with mechanistic DoE incorporates engineering knowledge in the form of constrains, expert-rules and known fundamental relationships of LC; therefore, this technology can find optimal conditions faster than the off-line approach. One of the main advantages of the automatic optimization is that a chromatographer can avoid complex tasks of the off-line computer-assisted optimization — peak tracking, data input, method and sequence specifications and other routine and non-routine operations. It should be noted that in the

recent final guidance for industries with regard to the analytical method development, the U.S. Food and Drug Administration (FDA) recommends submission of data to indicate a mechanistic understanding of the basic methodology [3].

An alternative to the mechanistic model-based approach is to directly identify process optima based on the results of experiments that are planned by statistical software such as repeated DoE. In contrast to model-based strategies, no mathematical process model is required, which is a significant advantage for many operators, and it is also better to use when the theory of LC and separation process interactions are not yet fully understood. Unfortunately for complex mixtures, when retention models cross each other in different regions of method variables, the direct approach can find the optimum only accidently. Usually, this type of DoE is used in a case where no, or little, prior process knowledge is available. However, for separation processes where a high degree of knowledge is available, statistical DoE is often not the most efficient strategy. Nevertheless, experimental results from the direct approach can be successfully used to identify a local optimal separation region for simple mixtures and to estimate the sensitivity of method quality to specific parameter changes within the design space (DS). Special software that include both features to create DoE and control of LC instruments to execute the DoE have substantial advantages against statistical software which have only options to plan DoE.

An alternative to the mechanistic and statistic approaches is to run the high-throughput screening to test combinations of method variables and factors — columns, solvents, buffers, gradients, etc. In contrast to the model-based and the statistical strategies, neither mathematical process model nor statistical DoE is required for the scouting approach. A chromatographer needs to only create a large sequence and then run it for new samples, thus relying on these few combinations of method variables and factors that will provide practically reasonable separations. The scouting approach is used frequently for chiral separations and samples when specific optimization is not necessary. Specialized software for automated method scouting are practically useful to create and edit long sequences rapidly and run them automatically.

For analytical method development, all three approaches proved to be practically useful, and any combination of them increase the probability of finding more suitable methods. To support various automated method development workflows, ChromSwordAuto® can operate in three modes: scouting, model-oriented optimization and statistic (direct optimization). Each mode can be applied separately or in various combinations depending on the preferred strategy of method development at a particular laboratory and project stage. Each mode is operated with a dedicated module.

3.2.1 *Instrument control and software configurations*

ChromSwordAuto® can operate as a chromatography method development data system (CDS) or as a third-party software. Functioning as the CDS ChromSwordAuto® controls Agilent, Waters and Hitachi HPLC and UHPLC systems. To control these instruments, no other CDS is necessary, and a stand-alone or a client–server configuration of ChromSwordAuto® can be chosen during installation. For the client–server configuration, data are collected on the local network or the internet file server (Fig. 3.1). The client–server configuration satisfies the requirements for data integrity with regard to applicable regulations like FDA 21 CFR Part 11.

Operating as a third-party software, ChromSwordAuto® controls Agilent, Waters and Dionex instruments thorough OpenLab/ChemStation, Empower or Chromeleon CDS. These CDS can work in the stand-alone, network or client–server environments.

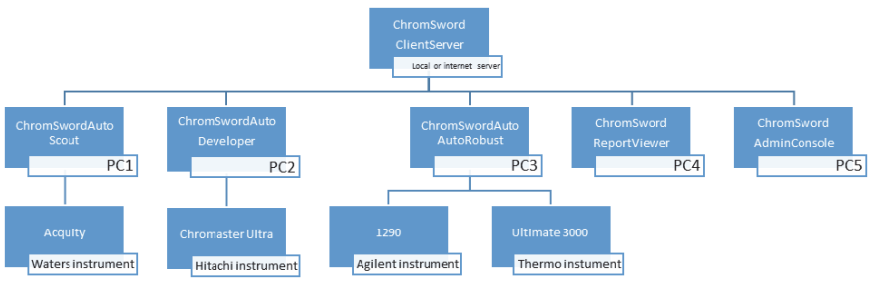

Figure 3.1: ChromSwordAuto® client–server configuration.

Different configurations of HPLC and UHPLC instruments can be used for automated method development. The most simple method development system consists of a binary pump, UV detector and autosampler; however, typically, method development systems contain 4–8 columns and 2–6 solvent channels to test different stationary and MPs.

ChromSwordAuto® incorporates automation of routine operations: column equilibration, column wash-out methods, system purging and column and solvent switching sequences.

3.2.2 *Strategies of automated method development*

Different strategies can be applied for automated method development. Strategies can combine screening, optimization and robustness study steps. One of the successful strategies for development of RPLC methods with ChromSwordAuto® has been used for dug candidates. It includes an automated screening step to identify the best column and solvent followed by an optimization step to fine-tune the separation [4, 5]. A similar strategy was used to apply ChromSwordAuto® for optimization of chiral separations in NPLC [6] and RPLC [7]. In another approach, the rapid optimization mode can be used for several predefined SP and MP combinations which are accepted at a lab as a standard method development column set, and then the fine optimization mode is applied for the most promising combination. Robustness studies can be included optionally for late-stages projects or methods to be transferred to other laboratories. The steps of such a strategy are shown in Fig. 3.2.

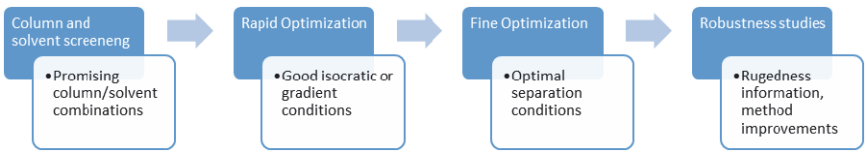

Figure 3.2: The strategy of method development for the latest stages of product developments.

3.2.3 *Automated method screening with ChromSwordAuto® Scout*

Automated screening of SP and MP are used to find practically a acceptable separation and run time when full optimization is not necessary. The screening can also be the first step in a multi-step method development strategy to identify promising combinations of columns and MPs.

ChromSwordAuto® Scout screening module generates sequences automatically and runs them to scout different gradients, columns, solvents, buffers, temperatures and other method variables for one or several samples. For multi-column and multi-solvent instruments, ChromSwordAuto® Scout controls several column compartments with 4–8 columns in each compartment and several (4–12 position) solvent switching valves connected to a binary or a quaternary pump. ChromSwordAuto® Scout analyzes 2D and 3D data acquired from two detectors simultaneously.

ChromSwordAuto® Scout application incorporates automation of column equilibration, column wash-out methods, system purging and column and solvent switching sequences for changing solvents, buffers, columns and other chromatographic process variables and factors.

3.2.4 *Automated model-based method optimization with ChromSwordAuto® Developer*

ChromSwordAuto® Developer module can be used for automated method optimization in RPLC, NPLC, IEX, HIC, HILIC, size exclusion chromatography (SEC) and supercritical fluid chromatography (SFC). For SEC, ChromSwordAuto® optimizes isocratic conditions, and for another type of chromatography, both isocratic and gradient separations can be optimized. Retention models that are used for different type of LC are described in Section 3.3.

ChromSword® is used for automated optimization of various mixtures; however, most frequently, it is applied for method development in the pharmaceutical industry. Typical applications are the development of stability-indicating and quality control methods (e.g. impurity profiling,

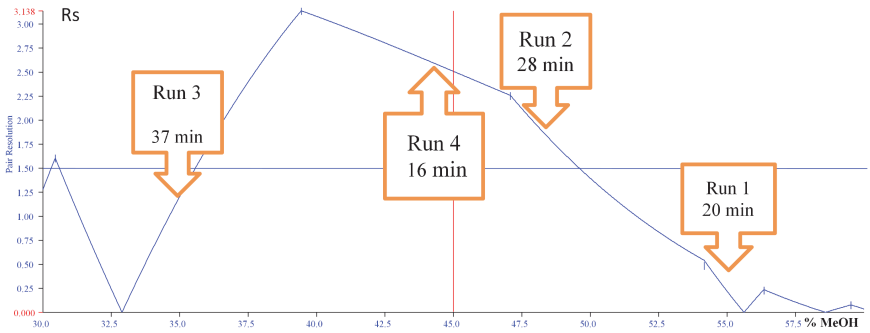

Figure 3.3: Runs shown on the resolution map that the software performs searching for optimal conditions in the unattended mode. Method development for a mixture of nine beta-blockers. Column: Purospher RP 18e, 5 μm, 150×4 mm. Mobile phase: 0.05 M phosphate buffer, $\mathrm{pH} = 3.0$ — methanol. The goals: $R_s \geq 2.0$ and run time ≤ 20 min.

assay, cleaning control, etc.). For automatic optimization, a user should specify the starting conditions: the column, solvent, flow rate, injection volume and the task type — rapid or the fine optimization. A chromatographer can also specify the development of either isocratic or isocratic and gradient methods. For both procedures, the optimization process includes the study of a sample to build retention models followed by application of the optimization procedure to find the optimal conditions. For planning new runs, the software processes the results of the previous runs and takes them into account. In Fig. 3.3, the method by which the software searches for optimal conditions developing the isocratic methods is shown.

For optimizations of gradient methods, both the studying and optimization runs can be linear and multi-step gradients. For optimization of separation, the Monte Carlo, genetic algorithms and the neural network methods are used. For the rapid optimization algorithm, the software performs 3–4 runs (Figs. 3.4–3.6), and for the fine optimization algorithm more runs are executed to study a sample and optimize the separation.

3.2.4.1 *Method development for large molecules*

Large molecules like proteins exhibit substantially different retention behavior than small analytes [8]. For these samples a small shift in chromatographic conditions can lead to high changes in retention and

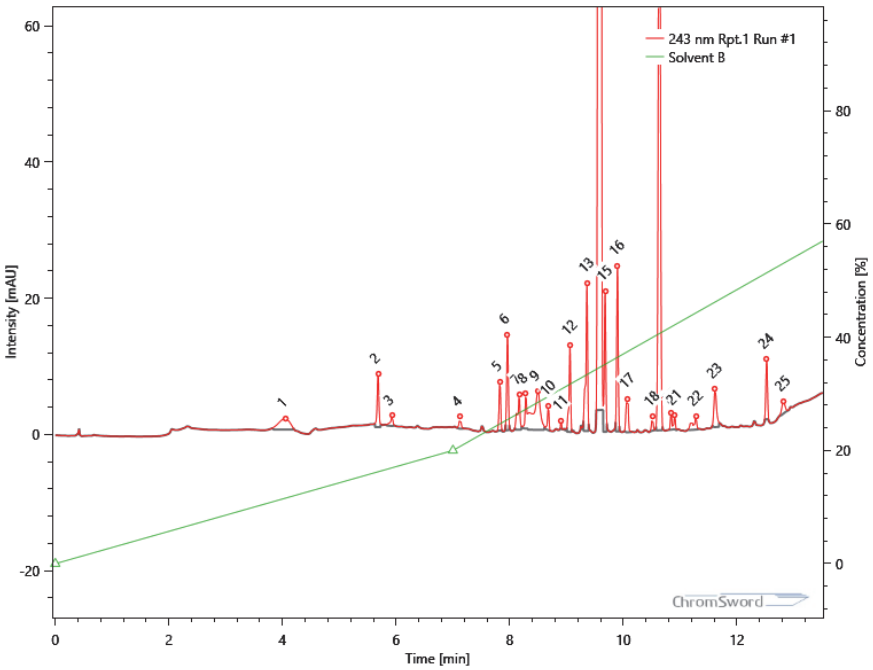

Figure 3.4: The first run of the automatic rapid optimization of the force degradation test mixture. Column: Zorbax Eclipse C18, 1.8 μm, 50×2.1 mm, flow rate 0.6 mL/min.

efficiency. The other point is that these compounds have practically identical UV spectra and cannot be used for peak tracking. Recently computer-assisted (off-line) method optimizations were reported for monoclonal antibodies (mAbs) and their domains in RPLC and IEX using 2D model as the gradient time–temperature model [9, 10]. It should be noted however, that the computer-assisted method optimization can be a time consuming process when many samples, columns and effects of different method variables require evaluation. An effective approach to circumvent and increase productivity is automated method development. In this instance, an analyst defines a strategy and an '*intelligent*' chromatography method development data system plans and performs many routine and optimization experiments autonomously. Various strategies of automated method development for mixtures of large molecules can be realized with ChromSwordAuto®. These can combine automated screening experiments with unattended

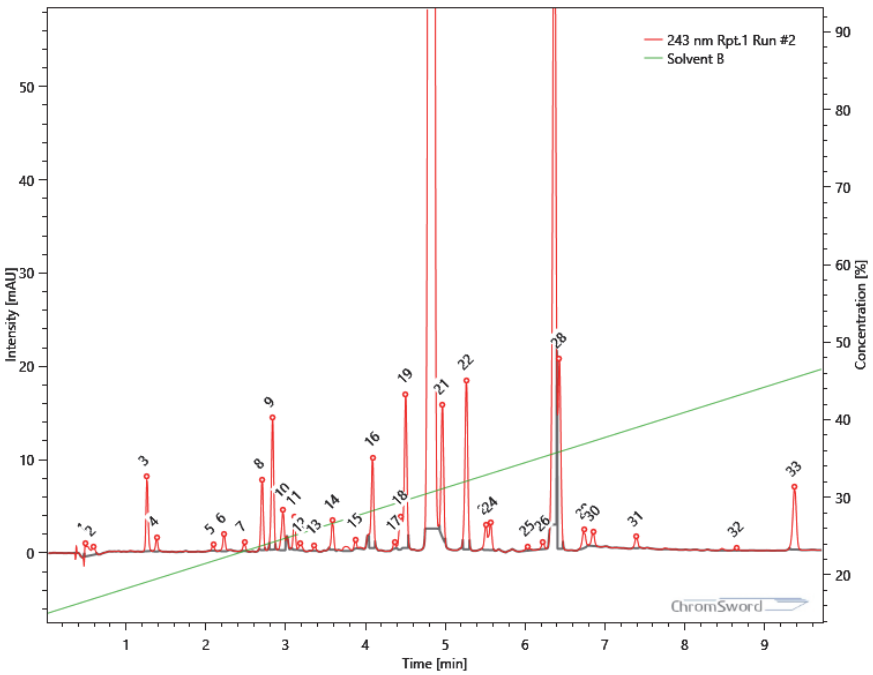

Figure 3.5: The second run of the automatic rapid optimization. Conditions are same as described for Fig. 3.4.

optimization, which is then followed by robustness studies using different DoEs. Results can also be used for off-line simulation and optimization. Such a strategy is used in different laboratories for automated RPLC method development using ChromSwordAuto® for the separation of variants and degradation products of the recombinant mAbs. The aim of method development for such projects is to study the domain-specific oxidation and develop stability-indicating methods that separate degradation products. For complex mixtures the optimization program can run multi-step gradients to separate more components (Fig. 3.7).

An important point to be considered is the column length for optimization of small and large molecules. It is known that the column efficiency for small compounds like peptides, after the digestion of proteins, is improved by increasing the column length. In contrast, the retention behavior of large proteins is different, and their bandwidth can be almost

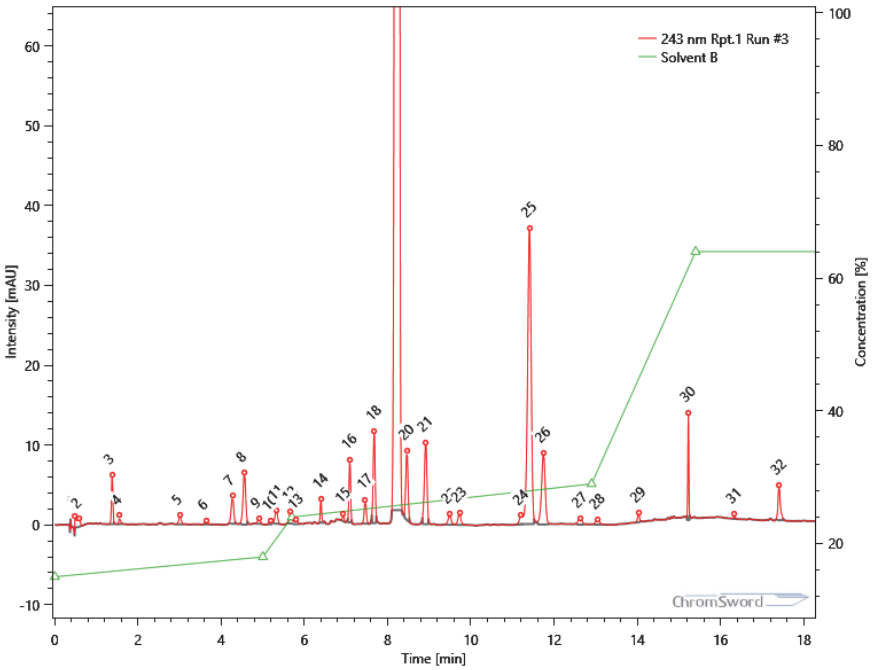

Figure 3.6: The third run of the automatic rapid optimization. Conditions are same as described for Fig. 3.4.

constant for all practical column lengths in the range 50–250 mm [11]. For such samples, longer columns do not provide higher separation efficiency [11], and therefore a short column can be a good alternative. Results in Figs. 3.7 and 3.8 show that the automated procedure can successfully find conditions to separate proteins on small columns. It should be noted that the optimization procedure is not related strictly to the column length. It is related to the target resolution and practical run time; therefore, shorter run times can be obtained on a long column and longer run time on a short column. In Fig. 3.8(a) the initial three study runs and in Fig. 3.8(b) the final gradient run are shown to separate monoclonal antibodies, under RPLC conditions. It should be noted that no optimal linear gradient for this mixture could be found in the temperature range of 70–80°C where reasonable peak width is observed and the column can be operated.

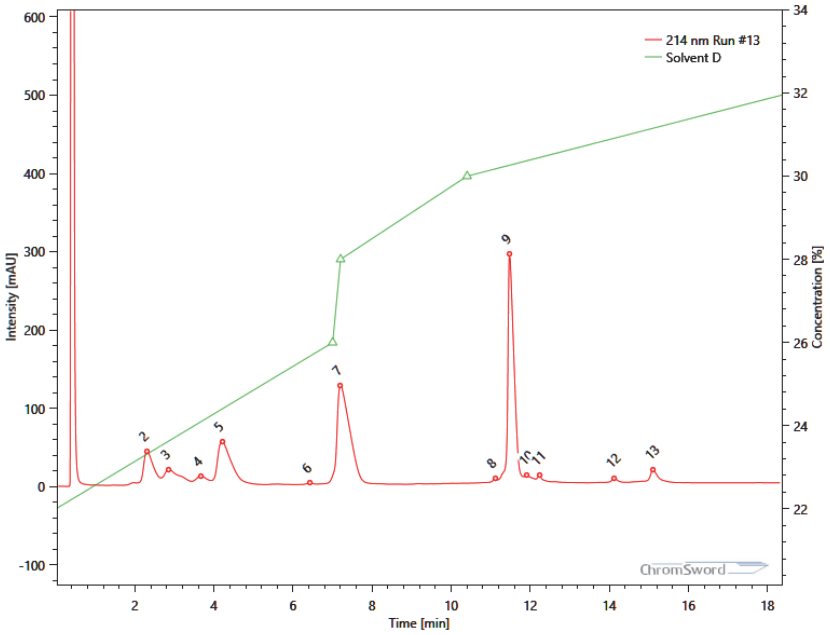

Figure 3.7: Partially digested (using IdeS) and reduced (using dithiotreitol, DTT) mAb sample. Peaks 2–4 — oxidation products of the crystallizable fragment (Fc/2); peak 5 — (Fc/2); peak 7 — the light chain (LC); peak 9 — the N-terminal half of one heavy chain (Fd). Column: 50 mm × 2.1 mm AdvanceBio RP mAb C8. Mobile phase A: Water + 0.1% TFA, B: ACN + 0.1% TFA. Temperature was set to 70°C, flow rate = 0.3 mL/min.

3.2.5 *Automated robustness studies and statistical DoE with ChromSword® AutoRobust*

ChromSword® AutoRobust is a specialized application for automatic evaluation of robustness of HPLC methods. According to the ICH guidelines [12] "Validation of Analytical Procedures: Methodology (Q2B)," the robustness of an analytical procedure is defined as a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage. The robustness should be considered at an appropriate stage in the development of the analytical procedure [12]. AutoRobust is a software tool for automation of robustness experiments to study the influence of variations in method parameters on chromatographic results.

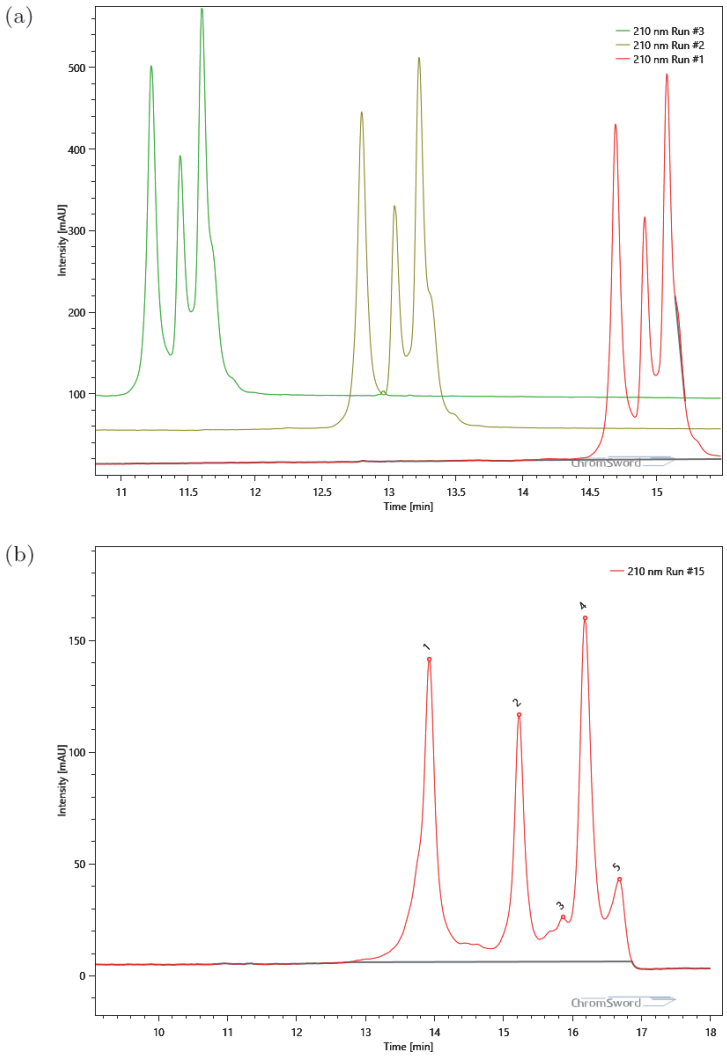

Figure 3.8: Column: 50×2.1 mm Zorbax 300 SB-Diphenyl. Mobile phase A: water + 0.1% TFA, B: ACN + 0.1% TFA. Flow rate: 0.25 mL/min; Temperature: 80°C. Sample: test mixture of mAbs (mAb1, mAb2 (confidential), Erbitux and Avastin). (a) Initial study runs of unattended optimization for separation; gradients: 1. 30–70% B in 25 min; 2. 36–66% in 22 min; 3. 36–66% in 19 min. (b) The final run of the unattended optimization; gradient: 0 min — 50% B in 2.2 min — 51% B; 16.6 min — 54% B; 18 min — 55% B.

Robustness of a method is extremely important for providing method transfer to other laboratories and instruments. Typically, robustness tests are performed at late stages of drug development projects; however, performing robustness tests at later stages involves the risk that when a method is found to not be robust, it should be redeveloped and optimized. Therefore, it is better to perform robustness tests at an earlier stage of method development. Different critical quality attributes (CQAs) of a method can be tested — including area, area%, retention time, resolution and other CQAs. One of the most important CQAs for HPLC methods is the resolution between peaks of target compounds. The resolution characteristic of a method should be within appropriate limits to ensure the drug product quality.

The following steps can be identified for robustness tests projects:

(1) selection of the factors to be tested,
(2) selection of the experimental design,
(3) definition of the different levels of the factors,
(4) creation of the experimental set-up,
(5) execution of the experiments,
(6) calculation of effects,
(7) statistical and graphical analysis of the effects,
(8) drawing conclusions from the statistical analysis and
(9) if necessary, improving the performance of the method.

These different steps are considered in more detail below.

3.2.5.1 *Selection of the factors*

For robustness tests, different operation factors can be considered. The selected factors can be quantitative (continuous) like the temperature or the concentration or qualitative (discrete) like the column batch. These factors should represent those that can be changed when a method is transferred between laboratories, analysts or instruments and that potentially could affect the response of the method. Typically, the following factors can be included in the robustness tests:

- gradient time and slope of linear gradients,
- initial and final concertation of linear gradients,

- time and concentration of each gradient node (step) for multi-step gradients,
- flow rate,
- column compartment temperature,
- pH of the MP,
- wavelength,
- column batch,
- method equilibration time,
- injection volume.

All these parameters and factors are supported by automated DoE with ChromSword® AutoRobust module. A chromatographer can optionally specify all or several factors to be included in the DoE.

The difference in flow rate, concentration and gradient time affect the resolution when different type of pumps (low- or high-pressure mixing systems), different solvent mixers and pumps from different manufacturers are used. The effective temperature inside a column can be different due to the difference in construction of compartments (forced air or still air oven). The small difference in glass electrodes and standard buffers can lead to differences in pH of a MP and selectivity of separation of basic and acidic compounds. If concentration of a sample is too low or too high, then increasing the injection volumes can lead to peak distortion.

3.2.5.2 *Selection of the experimental design*

The one-factor-at-a-time (OFAT), full factorial design (FFD) and the Plackett–Burman partial factorial design (PBD) can be used for robustness tests. The OFAT is the fastest design; however, it cannot estimate interactions of different variables without preliminary studies. The FFD is the most comprehensive design to determine interactions of factors and describe the response surface for finding optimum factor-values; however, it requires substantially more experiments. The PBD can be used as an alternative to FFD, but arrays of data points after the PBD cannot typically be used to solve the system of equations to determine chromatographic retention model parameters. In this case, a less reliable, simplified model is usually used to calculate response; however, deviations between the predicted and experimental value of a critical quality parameter can be too high. Another

problem is a possible confounding of effects due to reducing the number of runs in PBD. In this case, the effects of different factors or interaction factors cannot be evaluated individually and the interpretation of the results becomes difficult and even incorrect.

We consider that the robustness projects should include two designs:

(1) The OFAT design which can rapidly identify which of tested variables has a significant effect on the response.
(2) The FFD of the critical variables which were identified in (1).

Both steps can be executed in a completely automatic manner with a reasonable number of experiments. The PBD can be planned when the number of runs is too high and it is not practically reasonable to run the FFD designs.

3.2.5.3 *Definition of the levels for the factors*

The factor levels of variables to be tested should be set around the nominal values specified in the operating (basic) method. The interval chosen between the extreme values represents the limits between which the factors are expected to vary when a method is transferred. It should be noted that the levels should be defined by the analyst according to the results of a preliminary study of chromatographic retention behavior of compounds and instrument specifications taking into account the precision and the uncertainty with which a factor can be set and reset. To define the factor levels for the temperature, concentration and time of gradient steps, it is recommended to study the effect of these variables in more detail.

3.2.5.4 *Creation of the experimental set-up*

Each variable is studied in the experimental design, which is selected as a function of the number of factors and of levels to investigate. Two-level screening designs are a simple approach that can screen a relatively large number of factors in a relatively small number of experiments. More informative are the two-level designs with center points for effects of concentration and gradient time or the four-level designs with center

points for effects of flow rate and temperature. Such designs are optional in AutoRobust and allow the analyst to establish a linear or nonlinear retention model. Creation of the experimental design manually takes substantial time, even for OFAT. For planning FFD and PBD, normally special statistical software are used and then the design plan should be transferred into a sequence of runs of a chromatography data system. This is also a time-consuming process, and is practically very important that robustness test software can create DoE and transfer it into a sequence of runs automatically. The AutoRobust software module in ChromSword® provides a simple and rapid automated set-up of up to eight variables with 2–7 levels for OFAT, FFD and PBD. An unlimited number of qualitative factors (column, solvent batches, etc.) can also be included in the DoE.

3.2.5.5 *Execution of experiments*

It is important for reproducible robustness experiments to provide constant parameters both for injection and conditioning runs. Column and instrument wash-out, and purging and conditioning runs should be set up according to the instrument and column specifications. Adequate time for column equilibration, not less than 10 column volume have a paramount importance especially for large proteins to obtain reproducible results. For more confidence, it is recommended to include the column equilibration time as a variable in the robustness tests DoE.

The planned DoE is executed automatically with AutoRobust. The method development system performs these runs while interacting with a chromatography data system or directly with the modules. For estimation of time effects and stability of the instrument and the column, a number of additional experiments at nominal levels can be added to the planned DoE. These replicate experiments are performed before, at regular time intervals between, and after the robustness test experiments. These experiments allow checking whether the method performs well at the beginning and at the end of the experiments and to estimate for drift and column stability.

The results of runs are used to calculate effects of variables and determine the response.

3.2.5.6 *Calculation of effects and response determined*

From the performed experiments, a number of responses can be determined. For chromatographic methods, responses describing a quantity such as the content of main substance and by-products and effects of variables on peak area% and areas should be evaluated. The responses determined during the robustness test can be one of the following: the resolution between each pair of neighboring peaks, the retention time, the area and the area% of compound peaks. These parameters allow for evaluating the quality of a method and the effects of variables and factors.

The automated data processing procedure additionally calculates the relative retention, the peak asymmetry, the peak height and number of theoretical plates, which can also be included in the robustness study results.

3.2.5.7 *Numerical and graphical analysis of the effects*

One of the most important CQAs for HPLC methods is the resolution between peaks of target compounds. The resolution characteristic of a method should be within appropriate limits to ensure the drug product quality. As mentioned earlier, two approaches can be used to evaluate the effect of method variables on resolution — descriptive and mechanistic. Traditional statistically based software uses the descriptive approach and models the response surfaces with quadratic polynomials [12]. The main advantage of this approach is the simple and easy data processing procedure. This approach does not use physical models of the separation process and peak tracking from run to run. However, from the theory and practice of computer-assisted HPLC method development, it is well known that the quadratic dependence between resolution and method variables (concentration of organic modifier, gradient profile, temperature, pH) is more an exception rather than a rule for complex mixtures with irregular retention models [8]. Retention models of compounds can cross each other, and dependences $R_s = f$ (temperature, concentration, gradient time, pH) can have one or several maxima and minima. Figure 3.9 shows the resolution plots for limited pairs of a mixture of nine beta-blockers as a function of the concentration of methanol in the mobile phase. It is obvious that

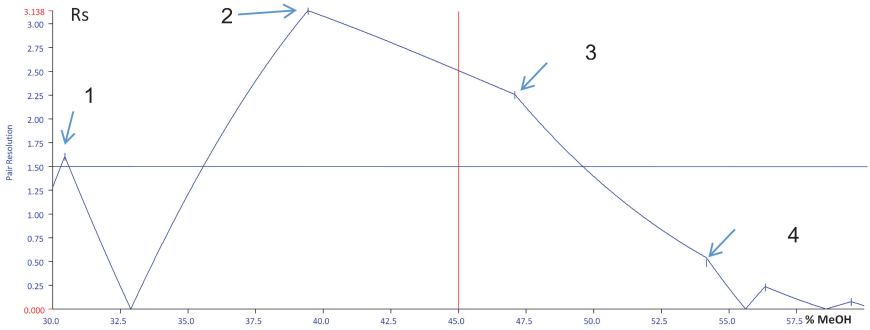

Figure 3.9: Resolution map: Effect of methanol concentration in MP on resolution of a mixture of nine beta-blockers. The arrows show the change of the limited pair in different regions of methanol concentration. 1 metipranolol/alprenolol; 1–2 propranolol/metipranolol.; 2–3 carazolol/celiprolol; 3–4 metoprolol/celiprolol; alprenolol — carvedilol.

modeling of the resolution response without peak tracking in this case will lead to wrong conclusions regarding optimal conditions and robustness of the method. The mechanistic approach uses parameters of the chromatographic process responsible for the response; however, retention behavior of the compounds must be studied to describe the effect of variables on the resolution. These include peak tracking from run to run, evaluation of parameters of retention modes in gradient elution and under different temperatures, and building a system of equations and solving them.

The mechanistic approach that applies relations from the theory of LC is supported in the AutoRobust software. After the design of experiments is created and performed in automated mode, data are processed for statistical and graphical analysis of responses. Method variables can have a substantial effect on resolution, and knowledge of the effect of the combination of these variables is necessary to study the robustness and to build up a DS of the method. The example of the effect of two variables with a fixed nominal value for two other variables is shown in Fig. 3.10.

3.2.5.8 *Improving the performance of the method*

Analysis of the resolution maps for a combination of three different variables enables visualization of areas where resolution can be increased or decreased. For example, the resolution map shows that temperature

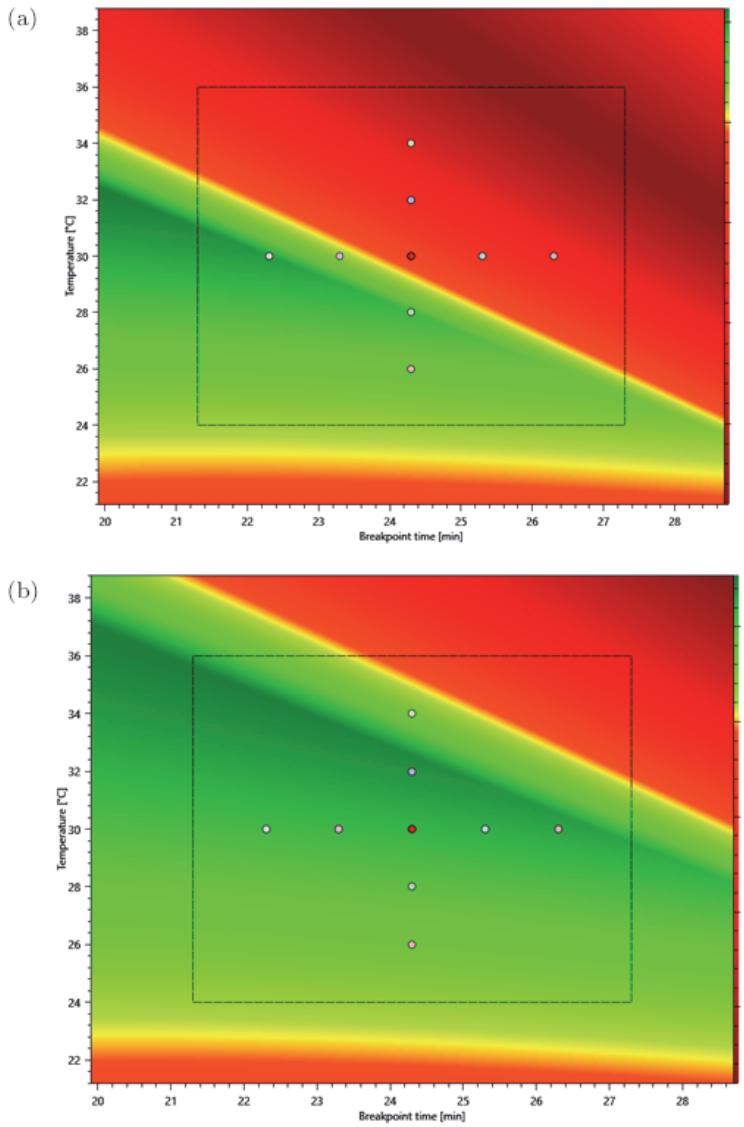

Figure 3.10: Resolution maps: Effect of the temperature and the gradient breakpoint time on resolution of a limited pair at the flow rate of 1.0 mL/min (a) and 0.8 mL/min (b). Mixture: 10 hair dyes. Column: ACE Excel C18-Amide 100×4.6 mm, 3 μm.

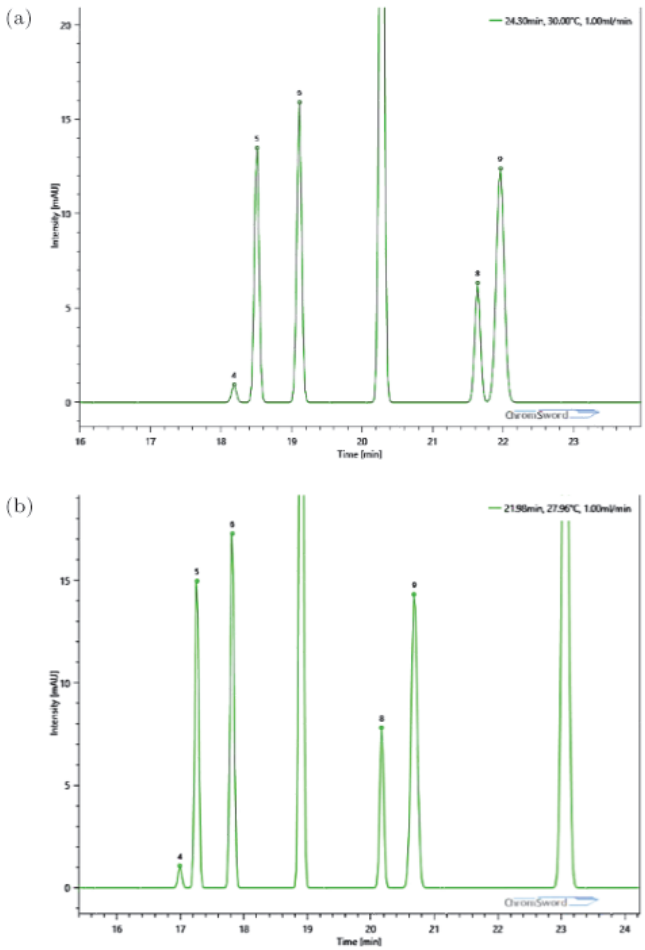

Figure 3.11: Chomatograms at a temperature 30°C, flow rate of 1.0 mL/min and gradient time of 24 min (a) and at 28°C, 0.80 mL/min and 22 min, respectively (b).

at 28°C, the flow rate of 0.80 mL/min and the gradient time of 22 min will provide a more robust method with higher resolution than one that was used after optimization (30°C, 1.0 mL/min and 24 min, respectively) (Figs. 3.10(b) and 3.11). Thus, robustness studies can also be considered as an additional tool to improve the performance of the method.

3.3 Computer-assisted Method Development

ChromSword® in the off-line mode can be used for optimizing separations in RPLC, NPLC and IEX.

If the structural formulae of compounds are known then, ChromSword® can predict the conditions of isocratic or gradient elution for acceptable retention to be obtained. No preliminary experiments need to be performed for the virtual chromatography. If the structural formulae of compounds being separated are known, then it is possible to start optimization of resolution after the first run. In this case, after inputting the experimental retention data for the first run, parameters of solutes will be refined to predict the best conditions for the separation. Entering experimental retention data for the second and the following runs makes possible a more precise prediction.

For solutes with unknown structures, ChromSword® can determine, from chromatographic experiments, their characteristics (molecular volume, the energy of interaction with water, nature (acid, base, neutral, p*Ka* value) and then predict their retention times on different reversed-phase columns and with different MPs.

Prediction is the first step in method development. The subsequent steps are optimization of retention and separation. ChromSword® enables a user to optimize the concentration of a modifier in a MP, pH value, temperature, gradient profile and column coupling. To optimize the separation of a mixture in gradient elution mode, stochastic methods like Monte Carlo and genetic algorithms are used.

For NPLC, it is possible to optimize the concentration of a stronger solvent in a weaker one when the retention data for two or more runs are entered. For IEX, the buffer or salt concentration in a MP can be optimized. Optimization of temperature is possible both for NPLC and IEX.

Optimization of method variables are organized in different modules of the software. The results depend on the information that a user enters into the software (Table 3.1).

ChromSword® can work with massive amounts of data. One sample file can contain up to 100 compounds including structural formulae and the data for up to 20 runs in the each module.

Table 3.1: Input/output of ChromSword® in the off-line mode.

Minimal input	Expected output
Structural formulae are considered	
Structural formulae (up to 100 in a file)	Starting conditions for RPLC: column type, eluent.
Structural formulae and data of one run	Optimal eluent for separation of a mixture in isocratic RPLC on a column being used. Optimal gradient profile. Starting conditions of RPLC for other column types and an eluent.
Structural formulae are not considered	
Data of two runs with different concentrations of an organic solvent in a MP (RPLC)	Optimal eluent for separation of a mixture in isocratic RPLC on a column being used. Starting conditions of RPLC for other column types and an eluent. Evaluation of the analyte parameters (molecular volume, polarity).
Data of two runs with different concentrations of an organic solvent or a buffer in a MP	Optimal eluent for separation of a mixture in isocratic RPLC, NPLC and IEX. Optimal gradient profile.
Data of two runs with different gradient profiles	Optimal gradient profile for separation of a mixture in gradient HPLC. Optimal eluent for separation of a mixture in isocratic HPLC.
Data of two runs with different temperatures of a column	Optimal temperature for separation of a mixture in isocratic HPLC. Enthalpy sorption of analytes.
Data of two runs with different pH of a MP	Optimal pH for separation of a mixture in isocratic RPLC.
Data of three runs with different pH of a MP	Optimal pH for separation of a mixture in isocratic RPLC. Nature of analytes (base, acid, neutral). pK value of analytes.
Two variable optimizations	
Data of three and four runs with different concentrations, pH, temperatures, columns, solvents, gradient profiles	Optimal gradient profile and temperature; concentration and pH; concentration and temperature; pH and temperature; concentration of two different organic solvents; optimal connection of two columns with different selectivity and concentration, gradient profile, pH or temperature.

3.3.1 *Concepts and procedures for developing HPLC methods*

The central idea of the computer-assisted method development is to input information about the mixture to be separated and then to apply a computer simulation to predict results for different chromatographic conditions, thus finding the acceptable conditions for separating the mixture.

One of the options is to use structural formulae as input for a computer program and to predict acceptable chromatographic conditions by analyzing information concerning their structures. It is an easy way for the user, but it is one of the most complicated problems in chromatographic science to predict acceptable conditions from a chemical structure. A much less complicated problem is to predict the results of chromatographic experiments by analyzing the results of several experiments previously performed.

It is understandable that the less information a computer program receives, the less precise the prediction that is obtained. If the input is only the structural formulae of compounds, the level of predictability is much less than that we would have after entering the results of several chromatographic experiments and their conditions. On the other hand, the fewer experimental results the computer program requires to produce acceptable prediction, the less time we have to spend developing the method.

It is hard to obtain an exact prediction of the retention time values from the structural formulae. The task of working with structural formulae is not to enable the precise prediction of retention in the first-guess experiment but to predict the concentration of an organic solvent in a MP (or a gradient profile) for acceptable retention to be obtained. Successful prediction of the concentration or the gradient profile will save time and the amount of solvent used in the experimental work. From a practical point of view, it is not important at this stage to predict the retention factor values precisely. The most important issue is to obtain these values within the acceptable practical limits of 1–20.

A practically reasonable approach is to start method development with only the information about structure, to receive the first prediction of chromatographic conditions (the first-guess method), to inject the sample and then to use experimental retention results for correcting the first-guess

prediction. In this case, a good chance exists to find acceptable conditions within a minimal amount of time. However, in many cases, a chromatographer has no information about compounds in a mixture or the structure parameters are not known. This situation is typical for developing stability indicating methods, reaction monitoring, separation of bio-mixtures and large molecules. In this case, it is necessary to obtain retention times for two or more experiments and then start computer experiments.

3.3.2 *Retention models*

The retention model in ChromSword® is defined as a type of a mathematical equation which describes the relationship between the retention of a compound and its properties as well as the conditions appertaining to the chromatographic experiments.

It is the focal point in method development software to determine retention models that adequately describe the effect of chromatographic conditions on the retention of compounds in a sample. In this case, based on only a few experiments, the software can predict the results of many other experiments under different conditions, thus allowing a chromatographer to simulate experiments with a computer and find the conditions for acceptable or best separation.

ChromSword® supports two approaches for the determination of retention models in RPLC. These are as follows:

(A) A traditional formal approach which applies linear, quadratic, cubic or other polynomial models for describing the relationship between the retention of solutes and the concentration of an organic solvent in a MP:

$$\ln k = a + b(C) \tag{1}$$

$$\ln k = a + b(C) + d(C)^2 \tag{2}$$

$$\ln k = a + b(C) + d(C)^2 + e(C)^3 \tag{3}$$

where k is the retention factor of a compound, C is the concentration of an organic solvent in a MP and a, b, d, and e are parameters of equations that must be determined by the software for each compound from the retention data obtained by using different concentrations of an organic solvent in a MP.

The simplest is the first linear model, which is known as the LSS model. It requires two initial experiments to start the optimization, but sometimes it does not completely predict correctly the effect of concentration of an organic solvent in a MP. This can be observed for basic and acidic compounds that contain highly polar and charged structural fragments. Such fragments are typically observed in natural and pharmaceutical compounds, and retention models for such compounds are nonlinear in many cases. Additional experiments as a rule do not lead to improvement in the accuracy of the linear model when it is applied for nonlinear functions.

The quadratic model describes retention more adequately. Additional experiments improve the accuracy, but three initial experiments are required to start computer optimization. The higher the power of a model, the more complex retention behavior can be described and the more initial experiments must be performed to start optimization of separation.

ChromSword® supports optimizing separation for polynomial models up to power 6. A chromatographer optionally can choose from powers 1 to 6. Typically, the powers 1–3 are most commonly used; however, the most complex retention can be described and separation optimized with the higher polynomial powers.

All polynomial models predict the retention of solutes rather precisely in the interpolation region of those concentrations studied. These models are less reliable in the extrapolation region. For example, if experiments were performed with 40% and 50% of the organic solvent in a MP, one can expect rather a good prediction of retention and separation in the region between of these concentrations and less accuracy in the regions of 30–35% and 50–55%. Extrapolation within wider limits very often leads to substantial deviations between predicted and experimental data.

(B) An approach that takes into account both the features of solutes being separated and the characteristics of the stationary and MPs being used:

In this method, the two-layer continuum solvatic retention model was proposed [14, 15] as an extension of the solvophobic model of RPLC [16]:

- The surface of a modifier sorbent in RPLC has a surface layer that involves hydrocarbon radicals and some of the components of a MP.

- The surface layers are assumed as being quasi-liquid having their own physical characteristics i.e surface tension and dielectric permittivity.
- The surface characteristics vary with varying the MP composition and SP properties.
- Molecules of retained substances penetrate into the surface layer.
- The retention is determined by the difference in molecule solvation energies in the mobile and SPs.

In this model, the retention of a solute is derived as

$$\ln k = a(V)^{2/3} + b(\Delta G) + c \tag{4}$$

where V is the molecular volume of a solute, ΔG is the energy of interaction of a solute with water, and a, b and c are the parameters which are determined by the characteristics of a reversed-phase column in the eluent being used, i.e. surface tension, dielectric permittivity and others. This approach works more precisely and rapidly than that based on formal linear and quadratic polynomial models, but it requires that both the parameters of the solutes (volume and energy of interaction with water) and the characteristics of the reversed-phase column under experimental conditions be known.

The characteristics of different commercially available RPLC columns were experimentally determined initially in a wide range of concentrations of methanol and acetonitrile in water. ChromSword® contains a database of characteristics for more than 150 commercially available reversed-phase columns in these eluents; they load automatically when a column and an eluent are chosen from the software menu.

ChromSword® calculates the parameters of compounds from the structural formulae. If structural formulae of the compounds being studied is not known or a user decides not to draw them, these parameters can be determined by ChromSword® from the two chromatographic experiments with different concentrations of an organic solvent in a MP.

This approach enables ChromSword® to predict regular or irregular retention behavior of solutes separated and enables a chromatographer to move rapidly to achieve maximal separation in minimal time. Each additional experiment leads to an improvement in the predictability.

Thus, this approach enables a chromatographer to start optimizing retention without any preliminary tests if the structural formulae of the compounds are known and also enables one to start optimization of separation on entering the retention data for only one run.

For solutes with unknown or undefined structures, this approach can also be used after entering the retention data and chromatographic conditions for two runs.

The main advantage of the structure and column properties related approach is that it "fills" both a column and compound features. It works precisely in the interpolation region and reliably in the extrapolation region. Figure 3.12 and Table 3.2 show that the solvatic model provides a

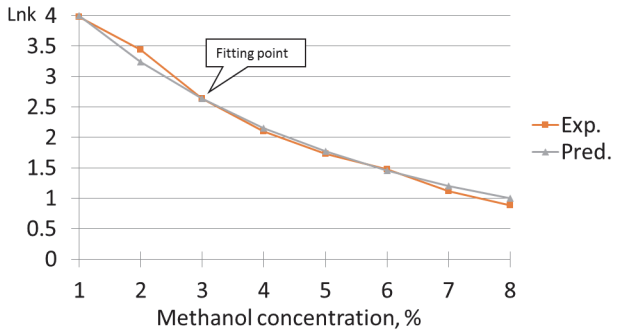

Figure 3.12: Adenosine monophosphate: predicted and experimental retention. Input: structure and data of one run at 3% MeOH. Column: Purospher RP-18e, 5 μm. MP: MeOH $-$ phosphate buffer, $\mathrm{pH} = 2.5$.

Table 3.2: Predicted and experimental retention of the beta-blocker carazolole in the extrapolated region of concentration of MeOH in a MP.

MeOH (%)	k_{exp}	K_{linear}	Dev (%)	$K_{\mathrm{quadratic}}$	Dev (%)	k_{Solvatic}	Dev (%)
60	4.62	4.62		4.62			4.62
50	6.33	6.33		6.33			6.33
45	8.83	7.71	-12.7	8.83		8.26	-0.33
30	33.57	19.70	-41.3	38.74	15.4	31.90	-4.97

Note: Retention values at 60% and 50% were used as input for the linear and solvatic models and at 60, 50 and 45% for the quadratic model. Column: Purospher RP 18e, 5 μm, 150×4 mm. MP: MeOH $-$ 50 mM phosphate buffer, $\mathrm{pH} = 3.5$.

good enough prediction of retention behavior for highly polar compounds that contain both uncharged and charged highly polar fragments.

3.3.3 *Procedure for optimizing pH in RPLC*

When a sample contains basic or acidic compounds with ionizable atoms or groups, pH is a very effective tool for optimizing the separation. ChromSword® supports two mathematical procedures for optimizing pH in RPLC. The first procedure is based on applying polynomials with powers up to 6 and the second procedure determines, using the retention data obtained with different pH values of a MP, the nature of solutes (neutral, acidic, basic), their p*Ka* value and then builds their retention models.

3.3.3.1 *Polynomial models*

The first three members are:

$$\ln k = a + b(pH) \tag{5}$$

$$\ln k = a + b(pH) + d(pH)^2 \tag{6}$$

$$\ln k = a + b(pH) + d(pH)^2 + e(pH)^3 \tag{7}$$

The powers 4–6 optionally can be employed for describing the most complex dependencies between retention and pH value of a mobile phase.

In order to optimize pH, a user must enter experimental retention data for two or more isocratic or gradient runs with different pH value of a MP. By analyzing retention data, ChromSword® determines and then refines the parameters of the retention model for the column being used and predicts the conditions for the best separation.

Tasks of a user are the same as that for optimizing separation in RPLC using a polynomial model and is described in Chapter 2 “procedure” for method development in HPLC using polynomial models.

3.3.3.2 *Fit pKa optimizing procedure*

This procedure determines, using the retention data obtained with different pH values of a MP, the nature of solutes (neutral, acidic, basic), their p*Ka* values and then builds their retention models:

$$k = k(0) + k(i)/(1 + F) \tag{8}$$

where $k(i)$ is the retention factor of an ionic form of a solute, $k(0)$ is the retention factor of a molecular form of a solute, and F is $Ka/[H^+]$ for acids and $[H^+]/Ka$ for bases, where *Ka* is the dissociation constant of a solute.

In order to optimize the pH value using the fit p*Ka* procedure, a user must enter experimental retention and efficiency data for three or more isocratic or gradient runs with different pH values of a mobile phase. By analyzing retention data, ChromSword® determines the nature of the compounds (base, acid, neutral) studied at pH intervals, calculates the p*Ka* values and then refines the parameters of the retention models for the column being used (Table 3.3, Fig. 3.13).

Substantial differences can be seen for retention time of basic and acidic compounds predicted by the p*Ka* and quadratic retention models. The p*Ka*-related model typically predicts retention for acidic and basic compounds better (Table 3.4).

Deviations in predicted retention can lead to a substantial difference in predicted optimal pH value for separation of a mixture with basic and acidic compounds. In Figs. 3.14 and 3.15, the resolution maps as functions of the quadratic and the fit p*Ka* models are shown for optimization of separation of a mixture of sweeteners and preservatives.

The Fit p*Ka* procedure enables a user to not only optimize the separation but also determine the nature of the compounds and evaluate their p*Ka*

Table 3.3: The p*Ka*-related model parameters determined for mixtures of nucleobases and nucleosides.

	Compound	Nature	k_0	k_i	*pKa*
1	Uracil	Neutral	1.12		
2	Cytosine	Base	0.78	0.51	5.63
3	Thymine	Neutral	3.77		
4	Uridine (U)	Neutral	3.10		
5	Cytidine (C)	Base	2.06	1.34	4.45
6	Ara-U	Neutral	4.43		
7	Ara-C	Base	2.68	1.68	4.17
8	6-azauridine	Acid	1.54	1.20	5.62
9	6-azacytidine	Neutral	0.98		
10	5-azacytidine	Base	2.21	1.47	4.04

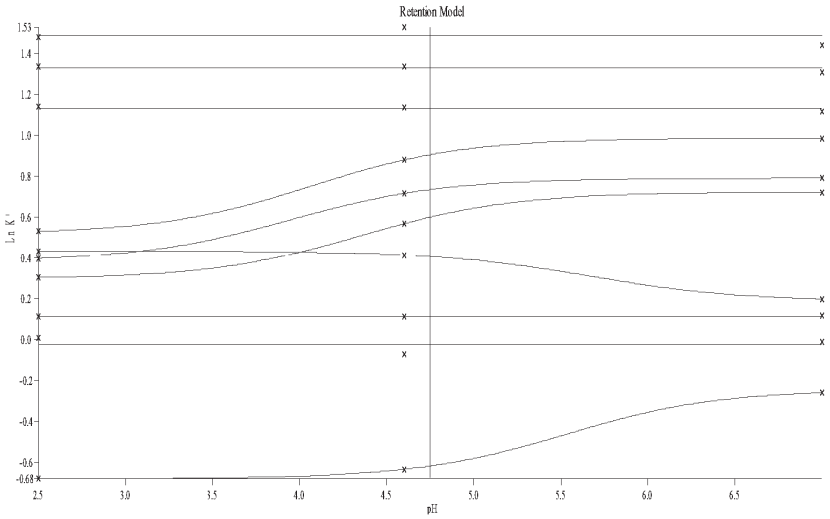

Figure 3.13: Retention models $(\ln \mathbf{k} = f(\mathbf{p}H))$ built with the Fit p*Ka* procedure for the compounds listed in Table 3.3 Column: Purospher RP18e, 5 μm, 125×4 mm. MP: 20 mM phosphate buffer $\mathrm{pH} = 2.5$; 4.6, 7.0. Flow rate 0.8 mL/min, $T = 35°$C.

Table 3.4: Predicted retention time with the quadratic ($\mathrm{RT_q}$) and p*Ka*-related ($\mathrm{RT_{pK}}$) model. $\mathrm{RT_e}$ — experimental values.

	Compound	$\mathrm{RT_q}$	$\mathrm{RT_{pK}}$	$\mathrm{RT_e}$	p*Ka*
1	Sorbic acid	7.54	10.00	10.00	4.67
2	Benzoic acid	4.76	5.41	5.37	4.19
3	Acesulfame	2.63	2.61	2.64	
4	Saccharine	3.41	3.45	3.43	
5	Aspartame	14.18	14.41	14.35	
6	Caffeine	7.07	7.06	7.08	

values under the conditions of a chromatographic experiment. In Tables 3.3 and 3.4, the p*Ka* values calculated from the experimental data are listed. It should be noted that the chromatographic method for the determination of p*Ka* values has advantages over other methods because it can be applied for mixtures and requires only a small amount of compounds.

It is necessary to take into account that the fit p*Ka* procedure assumes solutes to be monoprotic; therefore, for diprotic (and more) solutes as

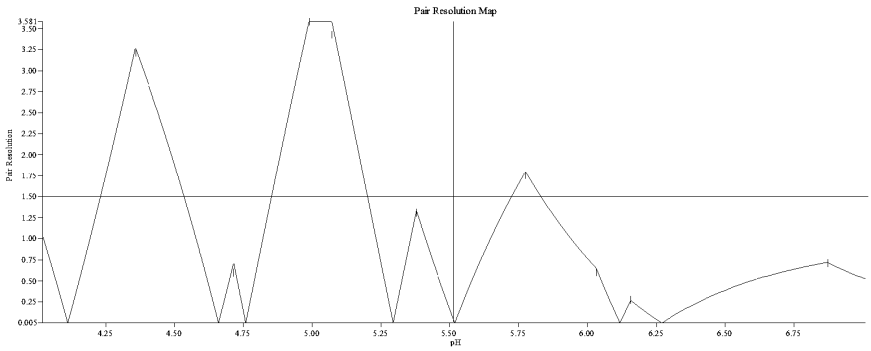

Figure 3.14: Resolution map built with the Fit p*Ka* procedure. Separation of the caffeine, acesulfame, saccharine and benzoic and sorbic acids. Column: Purospher RP18e, 5 μm, 125×4 mm. MP: 10% ACN/90% 20 mM phosphate buffer, $\mathrm{pH} = 7.01$; 4.02, 5.75. Flow rate = 0.8 mL/min, $T = 30^{\circ}$C.

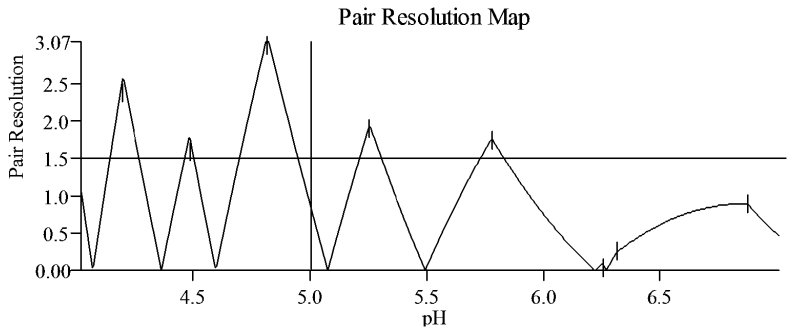

Figure 3.15: Resolution map built with the quadratic model. Conditions and mixture as described for Fig. 3.14.

well as for zwitterions, p*Ka* values can be considered as conditional. Nevertheless, this procedure can give valuable information about unknown compounds.

3.3.4 *Optimization of NPLC methods*

For optimization of the separation in NPLC, ChromSword® now supports only polynomial retention models. Retention in the NPLC can be described rather adequately by bilogarithmic models. ChromSword®

supports polynomials up to a power of 6. The first three are the following:

$$\ln k = a + b(\ln C) \tag{9}$$

$$\ln k = a + b(\ln C) + d(\ln C)^2 \tag{10}$$

$$\ln k = a + b(\ln C) + d(\ln C)^2 + e(\ln C)^3 \tag{11}$$

where C is the concentration of the stronger solvent in the mobile phase.

The powers 4–6 can be employed for describing the most complex dependencies between retention and concentration of a modifier in a MP.

In order to optimize a separation in NPLC, it is necessary to enter experimental retention and efficiency data for two or more runs with different concentrations of a strong solvent in the MP. By analyzing the retention data, ChromSword® determines and then refines the parameters of the retention model for a column being used and predicts the conditions for the best separation.

User tasks are the same as for optimizing separation in RPLC by using polynomial model and described in Chapter 2 "*procedure*" for method development in HPLC using polynomial models.

3.3.5 *Optimization of IEX methods*

The effect of the buffer concentration in the MP on retention in IEX can be described adequately by the same functions as for NPLC. Thus, a user can utilize the same procedure both for normal-phase and for IEXLC.

In order to optimize a separation in IEXLC, the user must enter experimental retention and efficiency data for two or more isocratic or gradient runs with different concentrations of a counter-ion in the MP. By analyzing the retention data, ChromSword® determines and then refines the parameters of the retention model elution for the column being used and predicts the conditions for the best separation.

3.3.6 *Optimization of the temperature*

Optimizing the temperature can be an effective tool if the conformation of solutes changes with temperature. This phenomenon can be observed

rather often in the case of large molecules such as peptides, proteins or for molecules with bulky substituents. In general, the effect of temperature on the logarithmic retention factor can be described by the simple equation $\ln k = a + b(1/T)$ for any mode of chromatography including gas chromatography. But if a solute changes its conformation, the function $\ln k = f(1/T)$ can be much more complex.

To optimize the temperature of a chromatographic separation, ChromSword® uses up to six power polynomials. The first three are the following:

$$\ln k = a + b(1/T) \tag{12}$$

$$\ln k = a + b(1/T) + d(1/T)^2 \tag{13}$$

$$\ln k = a + b(1/T) + d(1/T)^2 + e(1/T)^3 \tag{14}$$

where T is the temperature of the MP.

For optimizing the temperature, the same procedure as for optimizing the concentration of a modifier in RPLC, NPLC and IEX can be used.

In order to optimize a separation, the user must enter experimental retention and efficiency data for two or more runs with different temperatures of the MP. By analyzing the retention data, ChromSword® determines and then refines the parameters of the retention model elution for the column being used and predicts the conditions for the best separation.

If the model with the power one is applied, then ChromSword® also determines the enthalpy of sorption from the retention model:

$$\ln k = \ln k_0 + \Delta H/(RT) \tag{15}$$

where ΔH is the enthalpy of sorption of a solute in kJ/mol and R is the universal gas constant.

Thus, ChromSword® can be applied not only for optimizing a separation but for physico-chemical studies of compounds. For unknown compounds, ΔH values can be useful for elucidation of their structure.

3.3.7 *Optimization of the gradient*

There are different approaches to optimize gradient profiles after the determination of the retention models. The most frequently used approach is the

optimization of linear gradient profiles when two runs with linear gradients and different gradient times are used as input. These runs are used to build retention models. The initial and final concentrations of a modifier are fixed both for input and optimization. In this case, only the gradient time is optimized. This is simple approach that can easily be combined with the optimization of other variable like the temperature of a MP. However, complex mixtures in many cases can be separated only with multi-step gradient profiles. These include natural samples or samples after force degradation tests in pharmaceutical research and development laboratories. Every gradient node can be characterized by two parameters — time and concentration — and the position of every node in the time and the concentration dimensions should be optimized. Such multi-step gradients can be optimized by simulating chromatograms for different multi-step gradient profiles; however, this is not a fast method.

To build retention models, ChromSword® can process two or more runs with linear or (and) multi-step gradients. In this case, every new run can be used to refine retention models. For the optimization of both linear and multi-segment gradient profiles, the Monte Carlo and genetic algorithms are used. A user needs to enter the parameters of optimization, desired run time, separation and target peaks to be separated, and the stochastic procedure will find the best gradient profile automatically, assuming the separation is possible. The more segments on the gradient profile and compounds in a sample, the more time for optimizing is necessary. Typically, ChromSword® spends only a few minutes with conventional PCs finding the best multi-segmented gradient profile.

3.3.8 *Optimizing two variables simultaneously*

Optimization of two variables is an effective tool for improving and developing HPLC methods. ChromSword® provides all necessary interface and mathematical procedures for optimization of two chromatographic variables simultaneously. The following two variables can be optimized with ChromSword®:

Using one column:

- gradient profile and temperature
- concentration of a modifier in a MP and temperature

- pH and temperature
- concentration of an organic solvent and pH
- concentration of two different organic solvents

Using up to four connected columns with different selectivity (*column coupling, column combination*):

- gradient profile and ratio of columns
- concentration of an organic modifier and column ratio for RPLC
- concentration of an organic modifier and column ratio for NPLC
- pH and column ratio for RPLC
- temperature and column ratio for RPLC and NPLC

3.3.9 *Simultaneous optimization of a gradient profile and temperature*

Gradient and temperature optimization procedure allows the user to predict retention and to optimize the separation in gradient elution by entering retention data and experimental conditions for three or more gradient runs with different slopes and temperature. It is practically useful that the gradient profiles can be both linear and multi-step. One of the possible plans of the experiments can be for a user to perform two linear gradients with different slopes and same temperature and the third linear gradient with a different temperature. The slopes should be substantially different.

- Run 1: 20 min linear gradient with concentration of an organic solvent ranging from 5% to 95% at temperature 30°C.
- Run 2: 40 min linear gradient with concentration of an organic solvent ranging from 5% to 95% at temperature 30°C.
- Run 3: 40 min linear gradient with concentration of an organic solvent ranging from 5% to 95% at temperature 40°C.

The difference in temperature should be, in the majority cases, not less than 10°C between gradients.

When the user inputs data of experimental runs (retention, efficiency, area) and conditions (gradient profiles, temperature, column dead time, the dwell time of the HPLC system), ChromSword® builds retention models and the user can compute simulate experiments with different profiles

and temperatures. It is also possible to search for optimal gradient profile and temperature using the automatic procedure. The simplest approach that is used in different method development software is to build resolution maps where the resolution is a function of the gradient time and temperature. In this case, the initial and final gradient time values are fixed and cannot be optimized automatically. The user should change the initial and final MP compositions and observe their impact on the resolution map. This manual procedure takes substantial time, even for simple linear gradients. For example, to study the effect of initial and final concentrations for $+/-5\%$ it is necessary to simulate 100 resolution maps for all combinations of the initial and final concentration. For multi-step gradients, the number of computer experiments to simulate the position of every gradient point and their combinations is enormous. Automated optimization procedures that are implemented in ChromSword® have no such limitations and enable a user to optimize simultaneously the initial and final concentrations, gradient time and temperature for linear gradient profiles or the temperature and position of all nodes in multi-step gradient profiles.

When a user finds a promising gradient profile with ChromSword® and performs the run, it is also possible to input the obtained retention data to refine retention models and then repeat the computer simulation and optimization.

3.3.10 *Optimization of separation using supervised machine learning*

In recent years, machine learning-based models have been able to solve problems that previously could be resolved only by experts [17, 18]. Deep machine learning models on limited datasets were applied for the prediction of retention time of peptides in RPLC [11]. In earlier publications, outdated artificial neural network methods were utilized to predict retention time of simple samples and a few linear gradients [19, 20]. None of these contributions attempted at finding multi-step solvent gradient for separation of compounds. We applied machine learning as one of the optimization methods in ChromSword®. The deep machine learning technology was not utilized widely in chromatography. We consider that some

information on its possibility in gradient optimization can be interesting for both computer scientists and specialists in computer-assisted method development.

For the deep learning model, we used the recurrent neural network (RNN). An efficient algorithm for RNN is the long short-term memory (LSTM) cells [21]. The LSTM cell in an RNN-based model is a recursive function that uses a set of sub-functions. This function receives input data from the training set for every time step. In our case, the time steps are training runs in a sequence of runs. Then, this function tries to forecast the desired result as an optimal solvent gradient to achieve good separation of compounds. Parameters of sub-functions inside the LSTM cell are trained using modern variations of stochastic gradient descent (SGD) algorithms. It should be noted that the LSTM cannot be applied directly to produce usable method conditions because the resulting value will be in a range from -1 to 1 (*tanh* function). To use LSTM layers, we need to normalize input and output data vector to appropriate scale or to use as a last layer linear regression of deep learning model. The linear regression layer would then produce usable values for concentration in a range between 0 and 100. As for the input part of the LSTM, we use the convolutional neural network (ConvNet) [22] to embed features of scouting runs like data points of the chromatogram, spectra, retention time of compounds, solvent concentration gradient, temperature, etc. A very promising development in machine learning research in recent years has been made in the field of deep reinforcement learning [23]. These algorithms use a model that learns regression task when it tries to forecast the cumulative reward of the whole trajectory of actions to perform a predefined task. It means that we can train a model to generate method conditions for a sequence of runs that will gradually lead to the best separation of compounds. For each run, the quality of the result (reward) can be estimated using the sum of pair resolution values for each peak in a run. The model calculates cumulative reward value for each run in a sequence. Using these rewards, the model learns to construct a gradient, extract knowledge from the acquired chromatogram and then construct the next gradient that will have a higher reward value

$$R = \sum_{t=0}^{n} \gamma^t r_t \tag{16}$$

Cumulative reward R is calculated by summing all rewards r_t of runs multiplied by discount constant γ^t that reduces the importance of future rewards at the present state. We cannot use R value directly to train our model, because it takes into account only executed actions. For example, it calculates rewards from runs with method conditions in a specific sequence, but we would like to construct utility function to train our model to include more possible method conditions. To realize this, we can construct the quality method value (Q)-based model using Bellman's equation Q_π that takes advantage of partial Markov decision process property:

$$Q_\pi(s_t, a_t) \leftarrow Q_\pi(s_t, a_t) + \alpha\Big(r_t + \max_a Q_\pi(s_{t+1}, s_{a+1}) - Q_\pi(s_t, a_t)\Big) \tag{17}$$

State s_t contains retention time, width of peaks, pair resolutions and other important method quality characteristics. Action a_t contains proposed a concentration gradient and other method conditions. We try to maximize Q-value that is approximated cumulative reward by changing method conditions.

To train the deep reinforcement learning model, we used physical retention models generated by ChromSword® as a training environment. The retention models were determined from retention behavior of different families of compounds, like small molecules and proteins. Then, a special procedure generated a large dataset of runs and simulated chromatograms for the training. In fact, the pattern of chromatograms as a function of solvent gradients and other conditions like temperature or pH can be used for the training. When beginning the training set, the Q-value model produces random method conditions; however, after training — using distributed computing — it can be applied to new samples. Our results showed that after training with simulated samples, the procedure can process the results of scouting runs of real samples and predict gradient profiles to provide a reasonable separation.

3.3.11 *Column coupling*

ChromSword® provides support in the case of the most complex mixtures when no acceptable conditions were found with several types of columns. In this case, the chromatographer can try to separate a mixture by coupling

columns with different selectivities. To optimize separation on coupled columns, it is possible to use data that were obtained separately for different single columns. Typically, columns with 2, 5, 7.5, 10, 15 and 25 cm lengths are commercially available and can be easily combined by using dead volume connectors or column cartridges. In this case, the generic procedure can be applied. This is done as follows:

- Make several runs with different concentrations of a modifier or gradient profiles in column 1.
- Input data of the runs for the column 1 page.
- Build retention models for compounds being separated.
- Build the pair resolution map, search for promising regions and simulate chromatograms.
- If no acceptable conditions are found, a user has choice for the next step:

  - Try an other type of column (columns 2, 3, 4).
  - Try an other solvent and pH or/and temperature with column 1.

If the chromatographer chooses the first option (change a column), it is possible to repeat the same steps 1–4 to try to optimize the concentration of a modifier in the MP or the gradient profile with that of the column 2. The other conditions must be the same as used for column 1 (solvent type, temperature, pH). If no good separation was found with column 2, the user can perform a computer simulation on:

- Coupling of columns 1 and 2 (a maximum of four columns can be virtually coupled) and optimizing the ratio of column lengths or column segments.
- Effect of the concentration of organic modifiers or the gradient profile on the separation for coupled columns.

The same procedure can be used for optimizations of pH or temperature and column coupling simultaneously.

3.4 Conclusion

ChromSwordAuto® is a software package which includes a chromatography method development data system and ChromSword® module for off-line computer-assisted method development.

ChromSwordAuto® is used for automatic method development of small and large molecules and supports mechanistic and statistic approaches for the optimization of method variables. ChromSwordAuto® also contains a module for high-throughput screening of many SP and MP combinations.

ChromSword® and ChromSwordAuto® are used for method development and optimization in practically all types of LC.

References

[1] S.V. Galushko, A.A. Kamenchuk, G.L. Pit, The calculation of retention and selectivity in RP LC. IV. Software for selection of initial conditions and for simulating chromatographic behaviour, *J. Chromatogr.* 660 (1994) 47–59.

[2] S.V. Galushko, V. Tanchuk, I. Shishkina, O. Pylypchenko, W.D. Beinert, ChromSword® software for automated and computer-assisted development of HPLC methods, In *HPLC Made to Measure: A Practical Handbook for Optimization*, ed. Stavros Kromidas, WILEY-VCH Verlag GmbH & Co. KgaA, 2006, pp. 557–570.

[3] Industry Analytical Procedures and Methods Validation for Drugs and Biologics. https://www.fda.gov/downloads/drugs/guidances/ucm386366.pdf.

[4] E. Hewitt, P. Lukulay, Implementation of a rapid and automated high performance liquid chromatography method development strategy for pharmaceutical drug candidates, *J. Chromatogr. A* 1107 (2006) 79–87.

[5] K.P. Xiao, Y. Xiong, F.Z. Liu, A.M. Rustum, Efficient method development strategy for challenging separation of pharmaceutical molecules using advanced chromatographic technologies, *J. Chromatogr. A* 1163 (2007) 145–156.

[6] S. Larson, G. Gunawardana, M. Preigh, Automated method development in HPLC. Evaluation of the ChromSword® software package. HPLC 2007 Abstract book, P23.06. http://www.chromatographyonline.com/efficient-chiral-hplc-method-development-using-chromsword-software.

[7] F. Vogel, S.V. Galushko, Automated development of reversed-phase HPLC methods for separation of chiral compounds, *Chromatogr. Today* 8 (2015) 54–55.

[8] L.W. Snyder J.W. Dolan, *High Performance Gradient Elution*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2007, p. 228.

[9] S. Fekete, S. Rudaz, J. Fekete, D. Guillarme, Analysis of recombinant monoclonal antibodies by RPLC: Toward a generic method development approach, *J. Pharm. Biomed. Anal.* 70 (2012) 158–168.

[10] S. Fekete, A. Beck, J. Fekete, D. Guillarme, Method development for the separation of monoclonal antibodycharge variants in cation exchange chromatography, Part I: Saltgradient approach, *J. Pharm. Biomed. Anal.* 102 (2015) 33–44.
[11] J. Koyama, J. Nomura, Y. Shiojima, Y. Ohtsu, I. Horii, Effect of column length and elution mechanism on the separation of proteins by reversed-phase high performance liquid chromatography, *J. Chromatogr.* 625 (1992) 217–222.
[12] ICH Topic Q 2 (R1) "Validation of Analytical Procedures"
[13] http://www.smatrix.com/products.html.
[14] S.V. Galushko, The calculation of retention and selectivity in RPLC, *J. Chromatogr.* 552 (1991) 91–102.
[15] S.V. Galushko, The calculation of retention and selectivity in RPLC. II. Methanol–water eluents, *Chromatographia* 36 (1993) 39–41.
[16] Cs. Horvath, W. Melander I. Molnar, Solvophobic interaction in liquid chromatography with nonpolar stationary phases, *J. Chromatogr.* 125 (1976) 129–140.
[17] M. Ren, R. Kiros, R.S. Zemel, Exploring models and data for image question answering, arXiv:1505.02074.
[18] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, arXiv:1411.4389.
[19] N.H. Tran, X. Zhang, L. Xin, B. Shan, M. Li, De novo peptide sequencing by deep learning, *PNAS* 114 (2017) 8247–8252.
[20] T. Bolanča,Š. Cerjan-Stefanović, M. Novč, Application of artificial neural network and multiple linear regression retention models for optimization of separation in ion chromatography by using several criteria functions, *Chromatographia* 61 (2005) 181–187.
[21] H. Wang, W. Liu, Optimization of a high-performance liquid chromatography system by artificial neural networks for separation and determination of antioxidants, *J. Sep. Sci.* 27 (2004) 1189–1194.
[22] Y. Li, Deep reinforcement learning: An Overview, http://arxiv.org/abs/1701.07274.
[23] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, A brief survey of deep reinforcement learning, arXiv:1708.05866.