Adaption of Retention Models to Allow Optimisation of Peptide and Protein Separations

Patrik Petersson,[1] Jette Munch,[1] Melvin R Euerby,[2] Andrey Vazhentsev,[3] Michael McBrien,[3] Sanjivanjit K. Bhal,[3] and Karim Kassam[3]

1. Novo Nordisk A/S, Måløv, Denmark.

2. Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK.

3. Advanced Chemistry Development (ACD/Labs), Toronto, Canada.

While the current percentage of biopharmaceutical drugs approved and used as human medicine is small compared with small molecule drugs, EvaluatePharma® finds that 'the percentage of sales from biotechnology products (bioengineered vaccines & biologics), within the world's top 100, is set to increase from 39% in 2012 to 51% in 2018. In the broader market, sales from biotechnology products are expected to account for 25% of the world pharmaceutical market by 2018, versus the current share of 21% in 2012 [1]. Growing interest in biopharmaceuticals has led to proteins and peptides becoming analytes of increasing importance in the analytical laboratory.

The most commonly used analytical technique for the analysis of protein and peptide purity is reversed phase chromatography (RPC) in combination with UV detection and/or mass spectrometry. As the molecular weight of the protein increases, the selectivity of the RPC separation decreases. Consequently it becomes necessary to introduce complementary separation techniques, e.g., ion exchange chromatography (IEC) for larger proteins.

Retention modelling has been successfully used for the method development and optimisation of analytical scale separations of small molecules for 30 years [2-4] and several commercial software packages are available, for example DryLab, ACD/LC Simulator, ChromSword, and Osiris.

A common method development strategy involves a screening of columns and mobile phases that are known to generate large differences in selectivity. The most promising combination of column and mobile phase is then selected and a limited number of experiments conducted in order to build retention models. Subsequently, these models are applied to find an optimal temperature and gradient shape *in silico* and to assess method robustness.

An important advantage with retention modelling based on theoretical rather than statistical models (i.e., polynomial models based on factorial designs often referred to as DoE), is that a significantly smaller number of experiments are required to fit the models and, in addition, more advanced predictions can be made. For example, it is possible to predict the appearance of an entire chromatogram rather than simply a numerical value which describes the quality of the separation.

When defining a method development strategy for peptides and proteins involving retention modelling of RPC and IEC it was, however, realised that existing commercial software programs were not capable of producing accurate predictions for peptides and proteins.

A literature search revealed that relevant models had been published that account for protein retention as a function of solvent strength [2-4] as well as temperature in various types of chromatography [5-8]. It appears, however, that these had not been implemented into commercial retention modelling software programs at the time this study was conducted in 2011.

As a collaborative effort, the authors set out to adapt and validate a commercially available software program (ACD/LC Simulator [9]) to accurately model retention and peak width of proteins and peptides in analytical scale reversed phase and ion exchange chromatography.

1.1. Solvent Strength Retention Models

As described by Snyder [1] the following isocratic relationships are required in order to account for isocratic retention of peptides and proteins:

$$\ln k = a + b\,x \qquad (1)$$

where k is the isocratic retention factor, a and b are system and analyte specific constants and x the fraction of the strong solvent. Eqn. 1 is valid for reversed phase chromatography (RPC) and hydrophobic interaction chromatography (HIC). Often it is extended with a 2nd order term to account for non-linearity.

$$\ln k = a + b\,x + c\,x^2 \qquad (2)$$

In order to account for ion exchange chromatography (IEC) and hydrophilic interaction chromatography (HILIC) the following equation is needed.

$$\ln k = d + e \ln x \qquad (3)$$

where d and e are system and analyte specific constants and x the fraction of the strong solvent.

Peptides and proteins respond more strongly to changes in solvent strength than small molecules. The response increases with increasing molecular weight [10]. In order to develop selective and robust methods it is therefore commonplace to employ very shallow and long gradients. The development of such gradients without retention modelling is an iterative and time consuming task.

Based on the isocratic models described above it is possible to derive equations that account for retention during linear gradients [2-4]. Segmented gradients that are commonly used do, however, require numerical solutions where a large number of isocratic segments are combined to account for retention.

1.2. Temperature Retention Models

For small molecules it has been observed that simultaneously modelling the gradient shape and temperature is a very effective approach to optimise the separation selectivity. This is true also for peptides and proteins. The relationship which is normally used for small molecules (Eqn. 4) was, however, found to be insufficient for proteins.

$$\ln k = f + g / T \qquad (4)$$

where f and g are analyte and system specific constants and T the column temperature typically, but not necessarily, expressed in Kelvin.

As shown in the plots in Figure 1, when plotting retention factor versus temperature, small molecules such as ibuprofen and toluene exhibit a linear relationship where retention increases with increasing temperature. Proteins, however, do not exhibit the same linear behaviour. To accurately account for the retention of proteins it was necessary to add a second order term (Eqn. 5).

$$\ln k = f + g / T + h / T^2 \qquad (5)$$

A literature study showed that others had previously made the same observation [5-7]. The difference in behaviour can be explained by the fact that the structure of proteins changes when heated. At low temperature the protein is folded and many functional groups are hidden within the protein and cannot interact with the stationary phase. As the temperature is increased, the protein unfolds and more groups are exposed which can interact with the stationary phase and thereby the retention increases with increasing temperature. At high temperature the protein becomes completely unfolded and its retention behaviour now mimics that of a small molecule, i.e., the retention starts to decrease with increasing temperature (Figure 2). One could imagine that a molecule flips from one conformational form to another as the temperature changes and that this would result in multiple linear regions [11]. However, based on our experience and what we can find in the literature proteins seem to display a gradual change in conformation and retention behaviour that can be nicely fitted by a 2nd order polynomial.

1.3. Combined Solvent Strength and Temperature Models

In order to numerically fit a model that accounts for the influence of both gradient shape and temperature, a bilinear combination of the relevant solvent strength model (Eqn. 1, 2 or 3) and the temperature model (Eqn. 5) was employed. Thus, combining Eqn. 1 and 5 resulted in the following model (Eqn. 6):

$$\ln k = a_{10} + a_{01} x + a_{10} / T + a_{11} x / T + a_{20} / T^2 + a_{21} x / T^2 =$$
$$(a_{00} + a_{01} x) + (a_{10} + a_{11} x) / T + (a_{20} + a_{21} x) / T^2 =$$
$$(a_{00} + a_{10} / T + a_{20} / T^2) + (a_{01} + a_{11} / T + a_{21} / T^2) x \qquad (6)$$

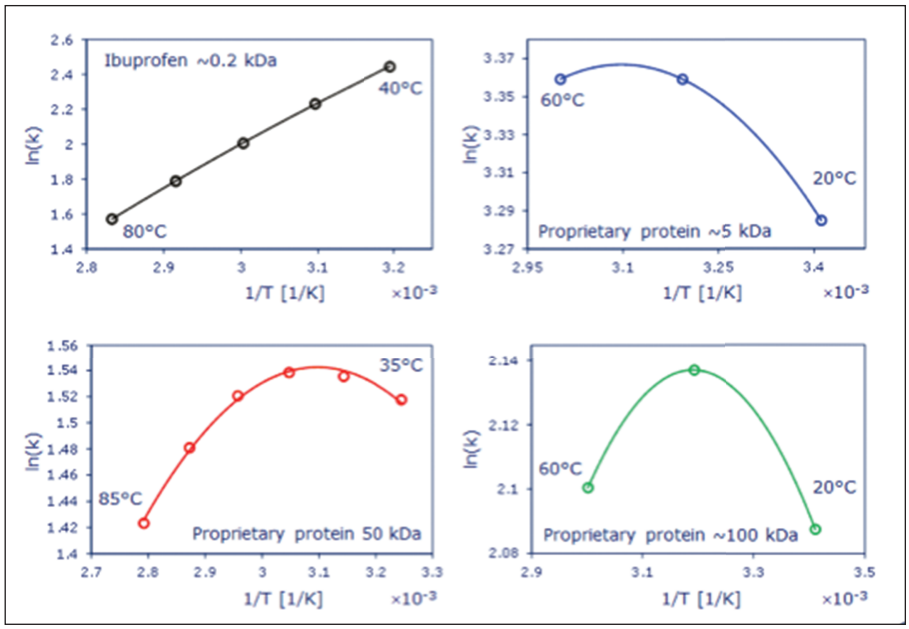

Figure 1. Retention as function of temperature for a small molecule ibuprofen and three proprietary proteins with a molecular weight ranging from approx. 5 to 100 kDa.

It should be noted that for fixed values of solvent strength or temperature the model is consistent with Eqn. 1 or 5, respectively.

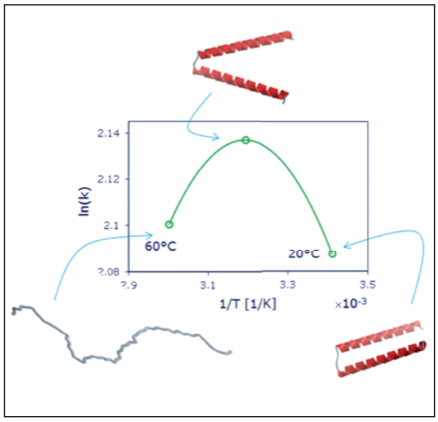

Figure 2. Potential explanation to the temperature dependent retention behaviour of proteins. An increasing temperature unfolds the protein and exposes more groups that can interact with the stationary phase. When completely unfolded, the retention decreases with increasing temperature as for small molecules.

2. Experimental

RPC and IEC retention and peak width data was collected for six proprietary proteins with a molecular weight of approx. 25 kDa. For each separation mode, data was collected for six or nine different combinations of gradient slope and temperature in order to fit the models. In addition, data was also collected for nine to 13 different linear and segmented gradients in order to validate the applicability of the fitted models.

RPC data was collected using an Acquity H-class system, a BEH300 C4 100 x 2.1 mm 1.7 um column, a flow rate of 0.4 mL/min and mobile phases mixed from A and B solvents consisting of 0.1% TFA in water and 0.1% TFA in acetonitrile, respectively.

IEC data was collected using a Protein Pak High Res Q 100 x 4.6 mm 5 µm, a flow rate of 0.5 mL/min and mobile phases mixed from A and B solvents consisting of 5 mM bis-tris propane pH 8.6 and 25% acetonitrile without and with 400 mM NaCl.

Calculations were initially made in Excel and later using an alpha version of a commercial software ACD/LC simulator [9] which contained the modified models. The latter had been modified to allow incorporation of custom gradient models (e.g., Eqn. 3 for IEC) in combination with 2nd order temperature models (Eqn. 5). ACD/LC Simulator is a commercially available software package that aids in the optimisation of chromatographic methods (gradient optimisation, additive

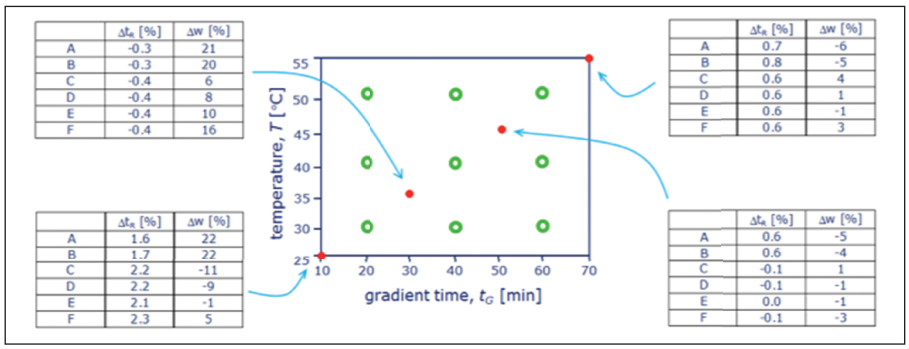

Figure 3. Experimental design used for generation (green circles) and validation (red dots) of combined solvent strength and temperature models for six proprietary proteins A - F. T corresponds to the column temperature, t_G gradient time for a linear gradient, Δt_R prediction error for retention time, and Δw prediction error for peak width.

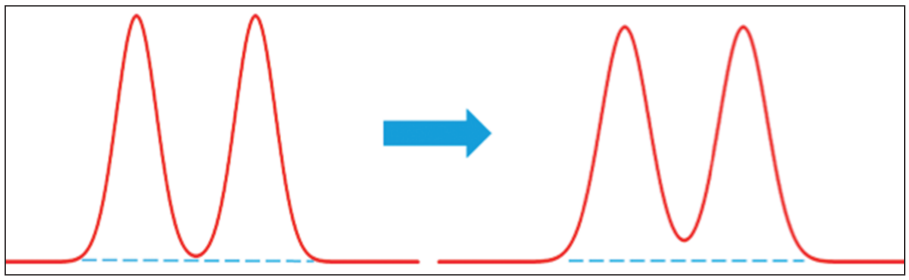

Figure 4. An illustration of the impact of a 20% prediction error in peak width. For symmetric peaks of similar size the impact on resolution is perfectly acceptable for optimisation purposes.

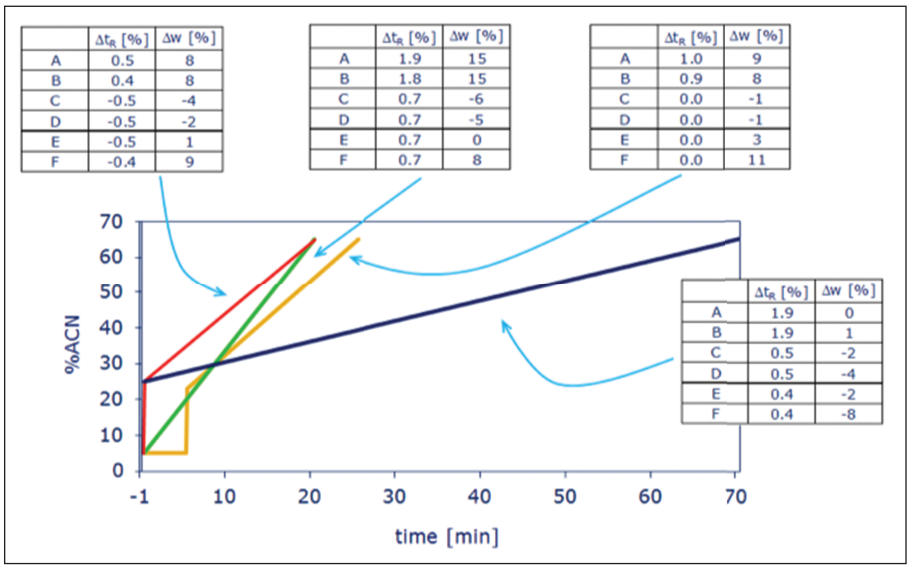

Figure 5. Evaluation of predictions made for multi-step RPC gradients using models built with single step gradients for six proprietary proteins A – F (Figure 3). Δt_R andΔw corresponds to prediction errors for retention time and peak width respectively.

concentration, temperature, pH, and more). It provides a unified environment for processing chromatographic data from different vendor instruments and formats; predicts retention times, carries out automatic peak matching, and predicts chromatograms based on method conditions.

Results

The combination of the 1st order solvent strength models (Eqn. 1 or 3) with the 2nd order temperature model (Eqn. 5) were found to give similar results for RPC and IEC. For the current dataset the 2nd order solvent strength model (Eqn. 2) did not increase the accuracy for predictions. Figure 3 shows the design used for generation and evaluation of models for RPC. A similar design was used for IEC. Green circles represent experimental data used to build the model. Red dots represent conditions for evaluation of predicted vs. experimental retention (t_R) and peak width (w). For interpolations, the deviation between calculated and experimental retention time were less than 1% for both RPC and IEC. This is comparable to what previously has been reported for small molecules [12,13]. For extrapolation to shorter gradient times, the retention error increased up to 2% for RPC and 10% for IEC. As previously reported, it is important to have a certain difference in retention time between the gradients used to build the retention models. A ratio in gradient time between the longest and shortest gradient of three to four has been proposed by Snyder et al [14], e.g., 20 and 60 min gradients.

The deviation between calculated and experimental peak width is less than 22% for both RPC and IEC. This is similar to what previously has been reported in the literature for small molecules [4, 15, 16]. A deviation in peak width of up to 20% may appear excessive but for peaks of similar size, the impact on resolution should be perfectly acceptable as illustrated in Figure 4.

The RPC and IEC models fitted to data from the linear gradients described above were subsequently challenged by the prediction of retention time and peak width for more complex, multi-step gradients. Figure 5 depicts the gradients evaluated for RPC. Similar gradients were evaluated for IEC. For both RPC and IEC the prediction errors for retention time and peak width were similar to what was obtained for linear gradients (i.e., error in retention time and peak width were less than 2% and 15% respectively). It should be stressed, however, that it is important to start the gradient at a solvent strength that results in a strong retention of the analytes. If not, significant errors in peak width can be expected due to poor focusing of the sample.

3. Conclusions

It can be concluded that RPC and IEC gradient chromatography at different temperatures can be modelled with the same accuracy for proteins as for small molecules. Presumably due to the unfolding of proteins at higher temperature, a 2nd order temperature model is needed in order to correctly model the retention behaviour of proteins as a function of temperature.

Since proteins respond much more strongly to small changes in solvent strength than small molecules [6], we believe that the use of retention modelling will facilitate the development of chromatographic methods for proteins not only in order to find an optimal selectivity but also to quickly and conveniently find a gradient that gives a suitable retention.

The potential to define custom gradient models in combination with 2nd order temperature models is now available in

the current commercial version of ACD/LC Simulator (version 2014). It is thereby possible to accurately model and optimise protein separations based on both RPC and IEC. It should, in principle, also be possible to model HILIC and HIC (Eqns. 3 and 1 or 2 respectively [2]) although this has not been evaluated using ACD/LC Simulator.

4. References

[1] Evaluate Pharma, World Preview (2013), Outlook to 2018—Returning to Growth, 2013. June, p. 13.

[2] M.A. Stadalius, H.S. Gold, L.R. Snyder, J. Chromatogr. A, 327 (1985) 27.

[3] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, Hoboken, NJ: Wiley, 2007, pp. 228.

[4] P. Jandera, J. Chromatogr. A, 1126 (2006) 195.

[5] R.I. Boysen, A.J.O. Jong, M.T.W. Hearn, J. Chromatogr. A, 1079 (2005) 173.

[6] A.W. Purcell, G.L. Zhao, M.I. Aguilar, M.T.W. Hearn, J. Chromatogr. A, 852 (1999) 43.

[7] M.T.W. Hearn, G.L. Zhao, Anal. Chem., 71 (1999) 4874.

[8] S. Fekete, S. Rudaz, J. Fekete, D. Guillarme, J. Chromatogr. A, 70 (2012) 158.

[9] ACD/LC Simulator, version 12 (12.02), Advanced Chemistry Development, Inc., Toronto, On, Canada, www.acdlabs.com, 2013.

[10] M.A. Stadalius, H.S. Gold, L.R. Snyder, J. Chromatogr. A, 296 (1984) 31.

[11] G. Vanhoenacker, P. Sandra, J. Chromatogr. A, 1082 (2005) 193.

[12] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, Hoboken, NJ: Wiley, 2007, pp. 399.

[13] M.R. Euerby, F. Scannapieco, H.-J. Rieger, I. Molnar, J. Chromatogr. A, 1121 (2006) 219.

[14] B.F.D. Ghrist, B.S. Cooperman and L.R. Snyder, J. Chromatogr. A, 459 (1988) 1.

[15] N. Lundell, J. Chromatogr. A, 639 (1993) 97.

[16] J.D. Stuart, D.D. Lisi, L.R. Snyder, J. Chromatogr. A, 485 (1989) 657.