Developing a Numerical Metric for the Quality of a Chromatographic Separation

Greg Kehrier
W. Jake Kisiel
Ana Silva
Alexandria Theakston
Yan Wei

April 25, 2007

Abstract This report presents an Excel application useful for comparing a set of High Performance Liquid Chromatography (HPLC) chromatograms and numerically rating them based on separation quality. The assigned quality scores are dependent upon three characteristics: the number of peaks detected in the sample, the resolution of all adjacent peak pairs, and the total experimental run time. The application enables users to customize the metric by prioritizing these three quality attributes as desired. Using this application would drastically reduce the need for visual inspection, making it very useful for interpreting the results of chromatographic optimization experiments. A second Excel application has been developed that enables users to easily simulate chromatograms.

This project was completed for Pfizer Inc. under the direction of Dr. Patrick Lukulay, in partial fulfillment of the requirements of Michigan State University MTH 844, advised by Professor Jay Kurtz.

Table of Contents

Nomenclature 1
Introduction 2
Background 2
Part 1. Simulate Chromatographic Peaks 4
  Analysis 4
  Results 7
  Discussion 9
Part 2. Numerical Metric 10
  Introduction 10
  Analysis 11
  Results 15
  Discussion 18
Future Work 19
Conclusion 20
References 21
Appendix A Terminology 22
Appendix B Test Case Chromatograms 23
Appendix C Derivations 27
Appendix D User Guide – Simulation 31
Appendix E User Guide – Metric 34

Nomenclature (According to their order of appearance)

Symbol	Description
$w_{z\%}$	Peak width at z% of the peak height
w	Band width (peak width at baseline)
t_R	Retention time
h	Peak height
R_k	Resolution between the peaks k and $k+1$
μ	Mean
σ	Standard deviation
$f(x;\mu,\sigma)$	Probability density function of the normal distribution
As_{10}	Peak asymmetry (10%) equation
N	Plate count
$w_{t,x\%}$	Peak width at baseline determined by tangents drawn to x% of peak height
$c_h(t),\ c_A(t)$	Function that simulates a peak with a specific height/area
$Area_n$	Area of the nth peak
$C(t)$	Function that simulates a chromatogram with n peaks
Q_i	Quality score of the ith chromatogram
$p_1,\ p_2,\ p_3$	Punishment functions for number of peaks, resolution, and run time
$p_{1,i},\ p_{2,i},\ p_{3,i}$	Punishment function values of the ith chromatogram
$c_j,\ c_{j,k}$	Prioritization constants (user-defined)
n_i	Number of peaks in the ith chromatogram
$n_{\max},\ n_{\min}$	Maximum/minimum number of peaks among the set of chromatograms
$y_k(R_k)$	Punishment function given to a specific resolution R_k
t_i	Total run time of the ith chromatogram
τ_i	Total run time per peak of the ith chromatogram
$\tau_{\max},\ \tau_{\min}$	Maximum/minimum total run time per peak among all chromatograms
N_{chr}	Total number of chromatograms in the set
$Area^B,\ Area^A$	Area of a peak to the left/right of the retention time
$u_1,\ u_2$	Random variables used for the baseline noise function
d	Noise intensity parameter
x	Random parameter used for the baseline noise function

Introduction

The main objectives of this project were to develop a numerical metric for measuring the quality of an HPLC (high performance liquid chromatography) separation, and to implement that metric into a Microsoft Excel application. When an HPLC procedure for a given mixture is to be developed, the goal is to find a set of experimental conditions (solvent, temperature, type of stationary phase or resin, and so on) that maximizes the separation quality. This is done by performing a series of chromatographic separations on the given mixture using various combinations of parameters, perhaps in a factorial design type of experiment. A chromatogram for each combination of HPLC parameters is obtained. Currently, the separation quality of each run is determined by visual inspection of the chromatograms. The purpose of the numerical metric application is to diminish the need for visual chromatogram inspection. Instead, the application assigns each chromatogram a "separation quality score" relative to the other chromatograms in the set. The score for each chromatogram is based on three quality factors:

- the number of peaks,
- a statistic describing the average resolution of the peaks, and
- the experimental run time.

The user may wish to select the experimental configuration with the highest quality score, or to visually inspect only those chromatograms that received sufficiently high quality scores before selecting an experimental configuration.

A secondary goal of the project was to develop an application that can simulate a chromatogram with specified peak characteristics such as retention time, peak width at half-height, peak height, and peak area. This chromatogram simulator application was created in order to generate a set of sample chromatograms (test cases) that were used to develop the metric application. The simulator is also useful because it produces realistic chromatograms that could be used for technical communication purposes.

Background

Chromatography is used to separate components in a mixture. The HPLC process is shown in Figure 1. A solvent is pumped through a column packed with a resin called the stationary phase. After the column and solvent have equilibrated, the sample mixture is injected, and the various components are carried by the solvent into the column. The material packed in the column is called the stationary phase, and the various sample components interact with this material to a varying degree. The greater the interaction, the slower that particular compound will pass through the column. As the components migrate through the column, they undergo a series of equilibration steps between the stationary phase and the mobile phase so that the separation becomes more pronounced as compounds pass through the column. Ideally, these differences in interaction are sufficient to allow all the components in the sample to be completely separated (resolved)

from each other. As the components elute from the column, they pass through a detector, whose signal is processed and plotted versus time to form the chromatogram.

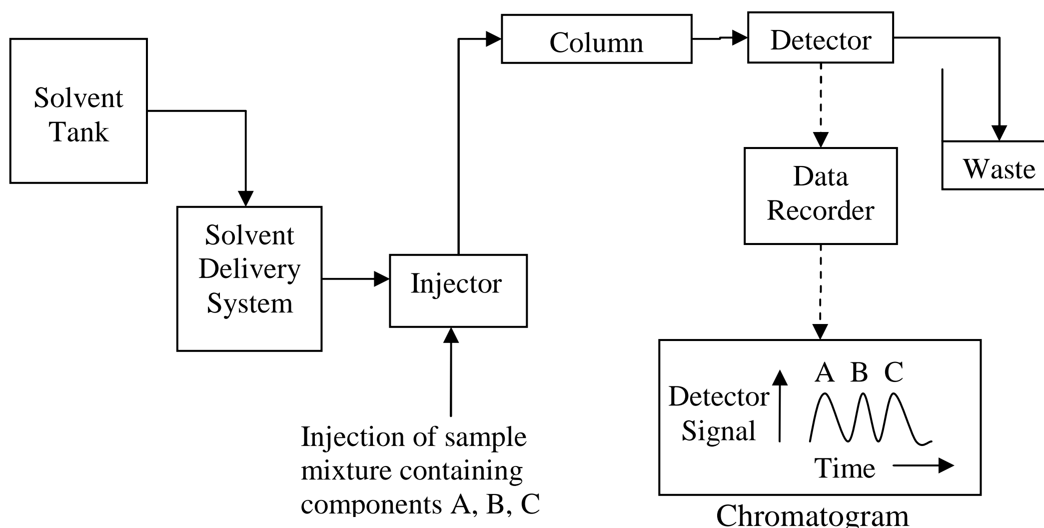

Figure 1. How chromatography works.

Figure 2 shows the most important parameters in a chromatogram. The retention time of a specific component is the time elapsed from sample injection to the time when the component elutes from the column. The other important chromatographic parameters relate to the width of the peaks. There are two peak widths to consider: the peak width at half height and the band width (or width at baseline concentration). Both of these width characteristics are used to quantify the separation quality.

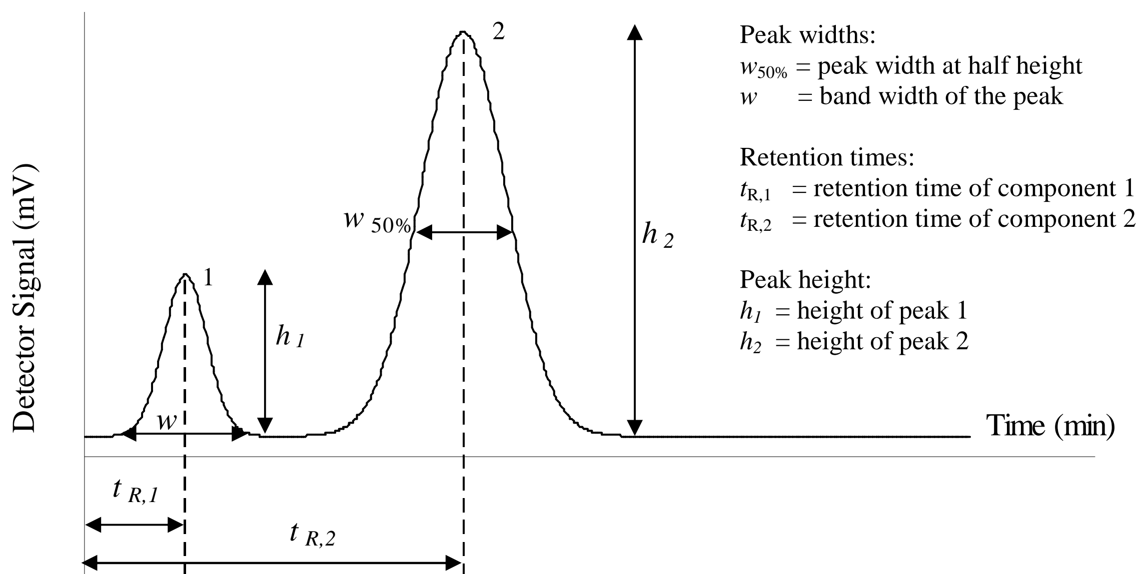

Figure 2. Sample chromatogram representing the most important parameters.

The area under a peak directly corresponds to the amount of the corresponding component in the sample mixture. For example, in Figure 2, the sample mixture contains more of the component corresponding to peak 2 than the component corresponding to peak 1. Therefore it is important to compute the area under each peak. This is difficult if the peaks overlap each other—that is, when the components are incompletely resolved. A measure of how well two components are separated is provided by determining the resolution. The United States Pharmacopeia (USP) resolution R equation at half-height between peaks 1 and 2 is defined as

$$R_1 = \frac{2(t_{R,2} - t_{R,1})}{1.7(w_{50\%}^{(1)} + w_{50\%}^{(2)})}, \tag{1}$$

where $w_{50\%}^{(1)}$, and $w_{50\%}^{(2)}$ are the peak width at 50% of the peak height for peak 1 and peak 2, respectively [3].

The optimal resolution value for a symmetric peak is $R = 1.5$, as demonstrated in Appendix C. This corresponds to a pair of peaks that do not overlap and both reach baseline height without additional time between them. When $R < 1.5$, the peaks partially overlap, and the corresponding components are said to be incompletely resolved. On the other hand, if $R > 1.5$, the peaks are completely resolved, but some amount of time passes between the moment when peak 1 returns to baseline and the moment when peak 2 begins to form. This constitutes a waste of time. Although this is usually considered less of a problem than incomplete resolution, it is still not ideal.

Part 1. Simulate Chromatographic Peaks

Analysis

In this part of the project a software program to simulate chromatographic peaks was developed. In order to generate the peaks, the normal distribution was used, since the shape of symmetric peaks can be approximated by a Gaussian curve. The probability density function of the normal distribution with mean μ and variance σ^2 is a Gaussian function,

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{2}$$

In Appendix C, it is shown that the width at half height of a Gaussian curve is related to the standard deviation by

$$w_{50\%} = 2\sigma\sqrt{2\ln 2}\,. \tag{3}$$

The normal distribution is a simple model for a chromatographic peak, as it does not account for tailing effects. Figure 3 shows the distribution profiles for an ideal and a non-ideal peak. A single normal distribution could accurately model the peak on the left, but not the peak on the right. The profile on the left, in Figure 3, represents an ideal case where the equilibrium is instantaneously achieved. The profile on the right represents a the case in which incomplete equilibration causes peak tailing. The amount of tailing depends on the experimental conditions.

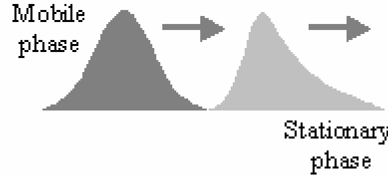

Figure 3. Distribution profiles for an ideal (left) versus a non-ideal (right) chromatographic column.

The USP measurement for peak asymmetry is given by the asymmetry ratio

$$As_{10} = \frac{A}{B}, \tag{4}$$

where A is the time from t_R to width end point at 10% of peak height, and B is the time from width start point at 10% of peak height to t_R [3] as shown in Figure 4. When the asymmetry ratio is one the peak is symmetric. To simulate both symmetric and asymmetric peaks, the simulation application uses two Gaussian distribution curves for each peak, one for the left side and one for the right side of the mean.

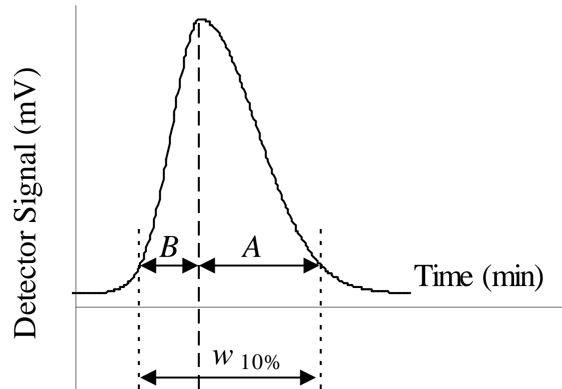

Figure 4. Asymmetry (10%).

The input parameters necessary to generate each peak are the retention time, the width at half height, the peak asymmetry, and either the height of the peak or the area under the peak. The function that generates the nth peak with a specific height is defined as

$$c_h(t) = \begin{cases} f(t;t_{R,n},\sigma_{B,n})\sqrt{2\pi}\cdot h_n\sigma_{B,n} & \text{if} \quad t \le t_{R,n} \\ f(t;t_{R,n},\sigma_{A,n})\sqrt{2\pi}\cdot h_n\sigma_{A,n} & \text{if} \quad t > t_{R,n}, \end{cases} \tag{5}$$

where $t_{R,n}$, $\sigma_{B,n}$, $\sigma_{A,n}$ and h_n are the retention time, the standard deviations of the left and right side of the peak, and height of the nth peak, respectively. The function f is the normal distribution function given by equation (2).

In Appendix C it is shown that the area under the nth peak is given by

$$Area_n = \frac{\sqrt{2\pi}\cdot h_n\sigma_{B,n}}{2} + \frac{\sqrt{2\pi}\cdot h_n\sigma_{A,n}}{2}. \tag{6}$$

Using this equation, the function that simulates a peak with a specific area is

$$c_A(t) = \begin{cases} f(t;t_{R,n},\sigma_{B,n})\dfrac{2\cdot Area_n}{\sigma_{B,n}+\sigma_{A,n}}\sigma_{B,n} & \text{if} \quad t \le t_{R,n} \\ f(t;t_{R,n},\sigma_{A,n})\dfrac{2\cdot Area_n}{\sigma_{B,n}+\sigma_{A,n}}\sigma_{A,n} & \text{if} \quad t > t_{R,n}. \end{cases} \tag{7}$$

The function that simulates a chromatogram with n peaks is

$$C(t) = \sum_{n=1}^{n} c_{h/A} + baseline, \tag{8}$$

where $c_{h/A}$ is the function that generates one peak considering the height or the area. The baseline is either a horizontal line or a line with a specified slope.

The simulation application outputs the resolution for each adjacent peak pair, as well as the area or height of each peak (whichever was not specified). Another output is the plate count N, which is related to column efficiency. The USP plate count is given by

$$N = 16\left(\frac{t_R}{w_{t,61\%}}\right)^2, \tag{9}$$

where $w_{t,61\%}$ is the peak width at the baseline determined by tangent lines drawn at 61% of the peak height.

The simulation application also has the capability to simulate baseline noise. The functions used to do this are described in Appendix C.

Results

When the chromatogram simulator application is opened, the simulation screen shown in Figure 5 appears. There are three tables: Options, Inputs and Outputs. The Options table (enlarged in Figure 6) allows the user to choose peak area or peak height as an input parameter. It also gives the user the ability to include baseline noise and specify the height and inclination of the baseline.

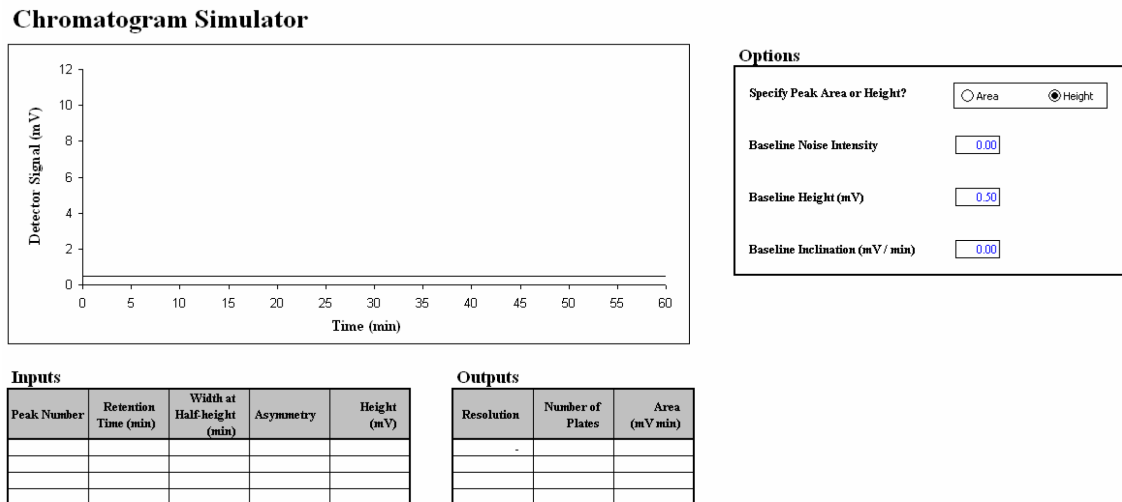

Figure 5. Screenshot of the simulation application.

Options

Specify Peak Area or Height?	Area / Height
Baseline Noise Intensity	0.00
Baseline Height (mV)	0.50
Baseline Inclination (mV / min)	0.00

Figure 6. Screenshot of the Options table.

An example of the input table is shown in Figure 7. In this case the user has selected in the Options table to specify height instead of area as an input parameter, so the inputs in the last column are peak heights. The user then entered peak data for six peaks as shown. Note that the peaks have been entered in order of increasing retention times—this is a requirement of the application.

Inputs

Peak Number	Retention Time (min)	Width at Half-height (min)	Asymmetry	Height (mV)
1	8.0	2.0	1.4	10.0
2	10.0	1.2	1.0	6.0
3	20.0	1.0	1.0	4.0
4	30.0	0.5	1.0	2.0
5	35.0	1.0	2.0	1.0
6	48.0	2.0	1.0	7.0

Figure 7. Screenshot of the Inputs table.

The chromatogram generated by the inputs in Figure 7 is shown in Figure 8. The user can modify the input values as desired, and the graph is automatically updated with each change.

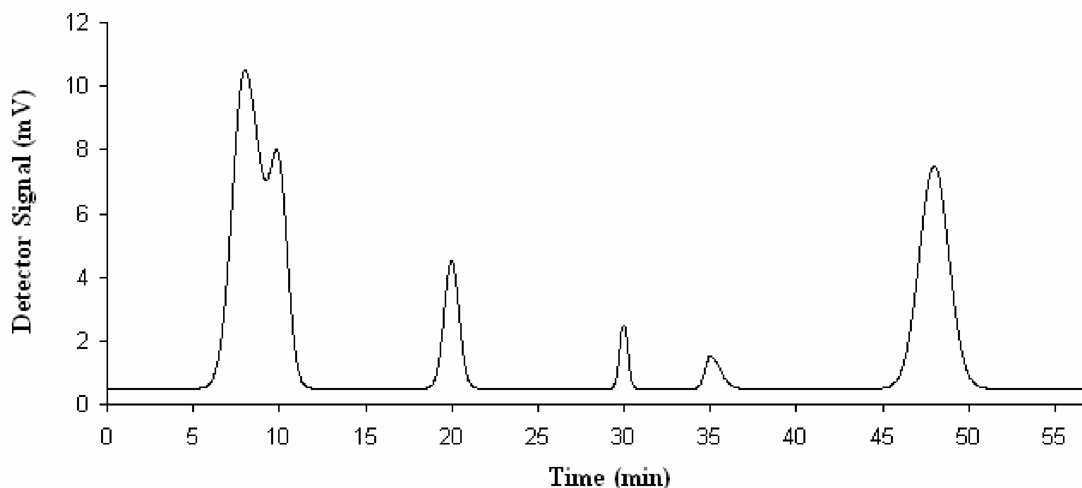

Figure 8. Chromatogram produced by the input values from Figure 7.

The application also outputs the resolution for each peak pair, the number of plates corresponding to each peak, and the peak height or peak area (whichever was not specified). The Outputs table generated by the input values in Figure 7 is given in Figure 9.

Outputs

Resolution	Number of Plates	Area (mV min)
-	88.52	21.29
0.74	384.20	7.66
5.35	2212.99	4.26
7.84	19916.90	1.06
3.92	6777.28	1.06
5.10	3186.70	14.90

Figure 9. Screenshot of the Outputs table corresponding to the inputs from Figure 7.

Discussion

In this part of the project, a software program that simulates realistic chromatograms was developed. The simulation was built using the Gaussian function (the normal distribution probability density function) to simulate peaks. Using the normal distribution function, it is possible to generate symmetric peaks with an area of 1.0 for a given mean and a given standard deviation. To simulate chromatograms with peaks of varying sizes, either the height of the peak or the area under the peak must be used as an additional input value. Also, to simulate tailing effects, two Gaussian curves were patched together at the mean in order to simulate asymmetric peaks. The software thus has four input parameters: the retention time, peak width at half height, peak asymmetry, and peak height or peak area.

There application has some basic limitations. In order to generate realistic peaks, a large number of data points are needed, and therefore the application was designed to simulate chromatograms with at most 20 peaks so that the file size was not inordinately large. Also, for programming simplicity and project time constraints, the application was designed with the requirement that peaks are entered by increasing retention times. If the user enters peaks in another order, the chromatogram will not appear until the problem is corrected. However, this is an unimportant limitation, since it is a trivial task for the user to reorganize the input values so the retention times are entered in increasing order.

The simulation application generates realistic chromatograms using four input parameters. The software not only allows the user to define the baseline properties, but also gives information about the peak resolution, number of plates, and peak area or peak height.

Part 2. Numerical Metric

Introduction

In this part of the project, a quality metric was developed to numerically rate the quality of a chromatographic separation. The metric assigns a quality score between zero and one to each chromatogram (of the same substance) based on the following criteria.

- Number of peaks: more peaks correspond to a better chromatographic separation. The maximum possible value is the total number of components in the sample.

- Resolution: the value of the resolution for each pair of adjacent peaks should be as close as possible to the optimal value of 1.5 (if less than 1.5, the peaks overlap).

- Total run time: when all the peaks are completely resolved less total run time is certainly desirable.

Figure 10 and Figure 11 show different chromatograms of the same substance containing six components. The chromatogram shown in Figure 10 is of higher quality than the one shown in Figure 11, because the first chromatogram has more peaks, a better resolution profile, and a shorter total run time.

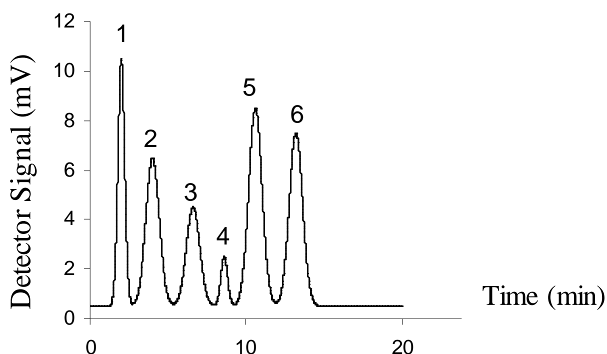

Figure 10. Sample chromatogram representing a high quality chromatographic separation.

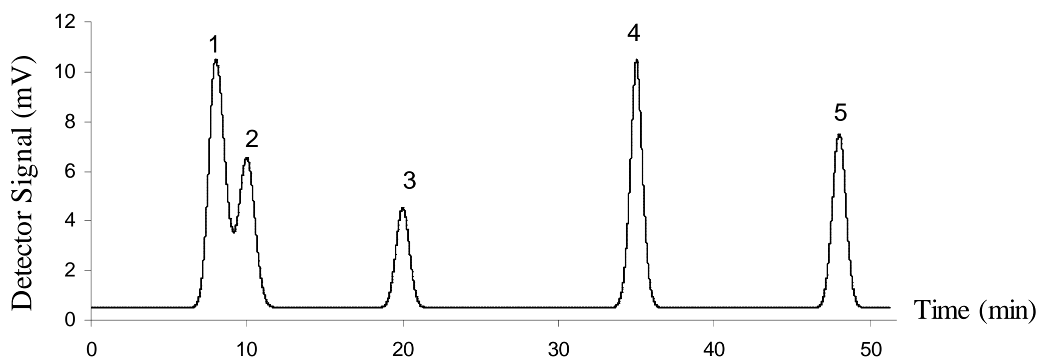

Figure 11. Lower quality chromatogram.

In conclusion, the ideal chromatographic separation has its number of peaks equal to the total number of components in the substance, all resolutions equal to 1.5, and requires minimal run time.

Analysis

The quality score of the ith chromatogram, Q_i, is a function of the form

$$Q_i = 1 - c_1 p_{1,i} - c_2 p_{2,i} - c_3 p_{3,i} . \tag{10}$$

Here, $p_{1,i}$, $p_{2,i}$ and $p_{3,i}$ are the punishment values for number of peaks, resolution, and run time assigned to the ith chromatogram by the punishment functions p_1, p_2, and p_3, respectively. All three punishment functions have range $[0,1]$, where zero corresponds to no punishment and one corresponds to maximum punishment. The numbers c_1, c_2 and c_3 are user-specified prioritization constants must satisfy

$$c_1 + c_2 + c_3 = 1 . \tag{11}$$

to ensure that Q_i has range [0,1]. The values of c_1, c_2 and c_3 are chosen to reflect the relative importance of the three quality attributes. For example, choosing $c_1 = 0.5$, $c_2 = 0.3$, and $c_3 = 0.2$ would be appropriate if the number of peaks is the most important quality attribute and the total run time is the least important attribute.

The peak punishment function value of the ith chromatogram $p_{1,i}$ is defined as

$$p_{1,i}(n_i) = \begin{cases} \dfrac{n_{\max} - n_i}{n_{\max} - n_{\min}} & \text{if } n_{\max} \neq n_{\min} \\ 0 & \text{if } n_{\max} = n_{\min} , \end{cases} \tag{12}$$

where n_i is the number of peaks in the ith chromatogram, and $n_{\max}$ and $n_{\min}$ are the maximum and minimum number of peaks among all chromatograms. That is, if there are N_{chr} total chromatograms in the set, then

$$n_{\min} = \min\{n_1, n_2, \ldots, n_{N_{\text{chr}}}\} \tag{13}$$

and

$$n_{\max} = \max\{n_1, n_2, \ldots, n_{N_{\text{chr}}}\} . \tag{14}$$

If $n_{\max} = n_{\min}$ then all chromatograms have the same number of peaks, so the appropriate punishment value is zero, as indicated by equation (12). Figure 12 shows the graph of the punishment function p_1.

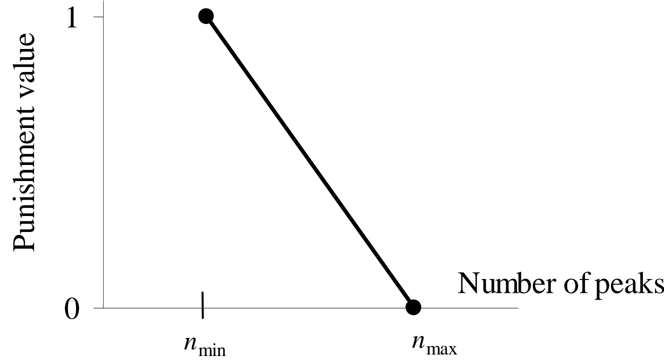

Figure 12. Peak punishment function.

The resolution punishment function p_2 is analyzed next. The *i*th chromatogram has $n_i - 1$ pairs of adjacent peaks, each of which is punished for resolution. So that the user may choose how to do this, a customizable function y was designed. The function y_k assigns to the *k*th peak a punishment value $y_k(R_k)$ based on its resolution R_k. The graph of this function is shown in Figure 13.

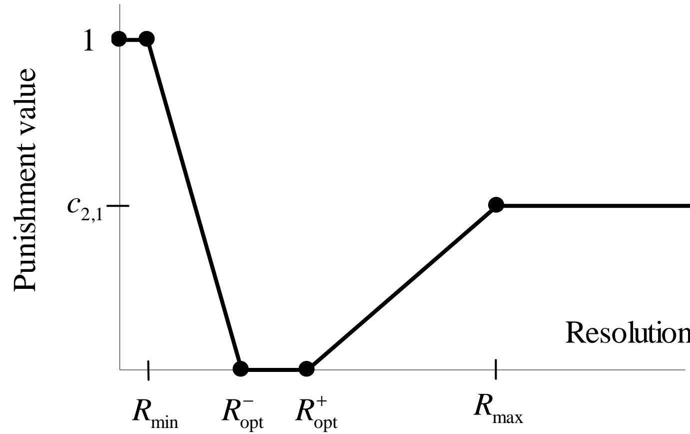

Figure 13. Punishment function for resolution of each adjacent peak pair.

The values $R_{\min}$, $R_{\max}$, R_{opt}^{-}, R_{opt}^{+}, and $c_{2,1}$ are adjustable so the user can modify the way resolutions are punished. The parameters R_{opt}^{-} and R_{opt}^{+} are used to specify a "neighborhood" around the optimal resolution so that $y_k(R_k) = 0$ if R_k is between R_{opt}^{-} and R_{opt}^{+}. Therefore the user should select R_{opt}^{-} and R_{opt}^{+} so that $R_{\text{opt}}^{-} \le 1.5 \le R_{\text{opt}}^{+}$. For example, if $R_{\text{opt}}^{-} = 1.45$ and $R_{\text{opt}}^{+} = 1.6$, peaks with resolutions between $R_{\text{opt}}^{-} = 1.45$ and

$R_{\mathrm{opt}}^{+} = 1.6$ will receive zero punishment. The parameters $R_{\min}$ and $R_{\max}$ are adjustable so that the user may decide to give maximum punishment to peaks with R_k sufficiently far away from 1.5. Finally, the user may adjust $c_{2,1}$ to punish resolutions that are too small differently than resolutions that are too large. The equation of y_k is

$$y_k(R_k) = \begin{cases} 1 & \text{if } R_k \le R_{\min} \\ \dfrac{R_{opt}^{-} - R_k}{R_{opt}^{-} - R_{\min}} & \text{if } R_{\min} \le R_k < R_{\mathrm{opt}}^{-} \\ 0 & \text{if } R_{\mathrm{opt}}^{-} \le R_k < R_{\mathrm{opt}}^{+} \\ c_{2,1}\dfrac{R_k - R_{opt}^{+}}{R_{\max} - R_{opt}^{+}} & \text{if } R_{\mathrm{opt}}^{+} \le R_k < R_{\max} \\ c_{2,1} & \text{if } R_k > R_{\max}. \end{cases} \tag{15}$$

After each adjacent peak pair has been assigned a punishment value according to the function y_k, the average resolution punishment of the *i*th chromatogram

$$\bar{y}_i = \frac{\sum_{k=1}^{n_i - 1} y_k}{n_i - 1} \tag{16}$$

is calculated. After this has been done for all the chromatograms, each chromatogram receives a resolution punishment according to the function p_2, given by

$$p_{2,i} = \frac{\bar{y}_{\max} - \bar{y}_i}{\bar{y}_{\max} - \bar{y}_{\min}}, \tag{17}$$

where $\bar{y}_{\min}$ and $\bar{y}_{\max}$ are the smallest and largest average resolution punishments among all chromatograms in the set. In other words,

$$\bar{y}_{\min} = \min\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{N_{\mathrm{chr}}}\} \tag{18}$$

and

$$\bar{y}_{\max} = \max\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{N_{\mathrm{chr}}}\}. \tag{19}$$

Thus, the chromatogram with the worst resolution profile receives a $p_{2,i}$ value of one, and the chromatogram with the best resolution profile receives a $p_{2,i}$ value of zero.

Finally, each chromatogram is punished for experiment run time. However, the run time is related to the number of peaks—more peaks should require more run time. Therefore the total run time punishment function is designed based on time per peak,

$$\tau_i = t_i / n_i \tag{20}$$

where t_i is the total run time of the *i*th chromatogram, calculated by adding the retention time of the last peak and half the baseline width of the last peak. Once τ_i is calculated for all chromatograms, the run time punishment values are assigned by

$$p_{3,i} = \begin{cases} \dfrac{\tau_i - \tau_{\min}}{\tau_{\max} - \tau_{\min}} & \text{if } \tau_{\max} \neq \tau_{\min} \\ 0 & \text{if } \tau_{\max} = \tau_{\min}. \end{cases} \tag{21}$$

In the unlikely event that $\tau_{\max} = \tau_{\min}$, all chromatograms have the same total run time, so a time punishment of zero is assigned to each of them.

In order to validate the numerical metric presented, the results were compared to those given by the method developed by Schlabach and Excoffier, which uses a value called the Chromatographic Resolution Statistic,

$$\text{CRS} = \left(\sum_{j=1}^{n-1} \left(\frac{R_j - R_{\text{opt}}}{R_j - R_{\min}} \right)^2 \frac{1}{R_j} + \sum_{j=1}^{n-1} \frac{(R_j)^2}{(n-1)(R_{\text{avg}})^2} \right) \frac{t_{R,n}}{n}. \tag{22}$$

Here, $R_{\min}$ is the smallest acceptable resolution value defined by the user, R_{opt} is the desired resolution (also specified by the user), and

$$R_{avg} = \frac{1}{n-1} \sum_{j=1}^{n-1} R_j . \tag{23}$$

These equations are presented in Chapter 2 of [2]. The CRS function assigns a positive score to each chromatogram. The smallest value corresponds to the best chromatogram in the set. The optimal value occurs when all individual resolutions approach the optimum resolution, are close to the average resolution, and the experiment requires the minimum analysis time.

Results

Quality Metric Results

In order to test the validity of the quality metric, the simulation developed in part 1 was used to generate twelve test chromatograms. In each case, the peaks are symmetric. These chromatograms are presented in Appendix B.

Since the user can specify which criteria are most important, it is necessary to test the quality metric for different scenarios. The scenarios tested are presented in Table 1. For brevity, only the results of scenario (d) are presented in detail.

Table 1. Scenarios tested.

Scenario	Description	c_1	c_2	c_3
(a)	Care only about peaks	1	0	0
(b)	Care only about resolution	0	1	0
(c)	Care only about total run time	0	0	1
(d)	Possible practical scenario #1	0.5	0.3	0.2
(e)	Possible practical scenario #2	0.65	0.35	0
(f)	Matching with CRS	0.1	0.7	0.2

The settings used for the y_k function in scenario (d) are given in Table 2 and shown in Figure 14. Since $c_{2,1}$ is set to zero, the value of $R_{\max}$ is unimportant.

Table 2. Resolution punishment parameters used in scenario (d).

Parameter	Value
$R_{\min}$	0.30
R_{opt}^{-}	1.5
R_{opt}^{+}	1.5
$R_{\max}$	N/A
$c_{2,1}$	0

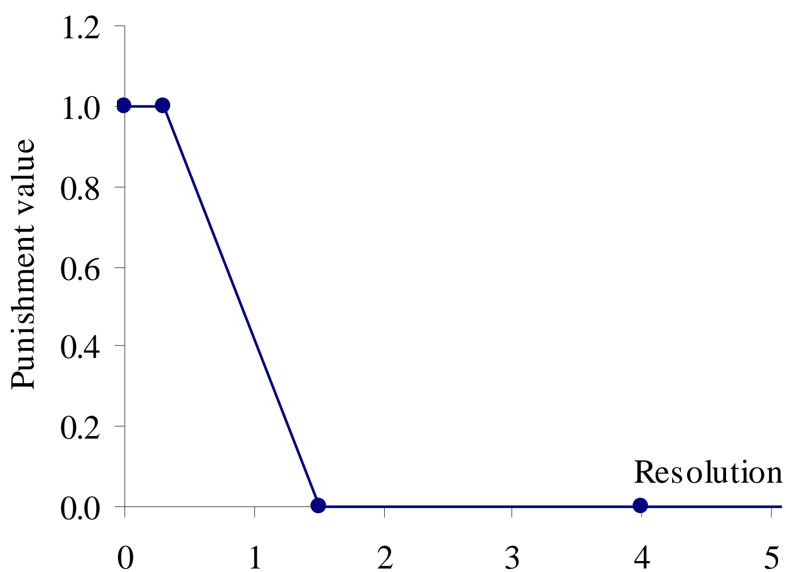

Figure 14. Punishment resolution for scenario (d).

The application was used to rank the 12 test case chromatograms using the settings given above. The results are given below in Table 3.

Table 3. Results for scenario (d).

Chromatogram	n_i	t_i	$y_{2,i}$	$p_{1,i}$	$p_{2,i}$	$p_{3,i}$	Quality
Chr 12	12	2.604	0.155	0.00	0.695	0.000	0.791
Chr 10	12	3.771	0.155	0.00	0.695	0.646	0.662
Chr 7	12	3.375	0.222	0.00	1.000	0.427	0.615
Chr 3	10	3.850	0.004	0.50	0.008	0.690	0.610
Chr 1	10	3.850	0.033	0.50	0.140	0.690	0.570
Chr 11	9	3.167	0.002	0.75	0.000	0.312	0.563
Chr 9	11	4.409	0.191	0.25	0.860	1.000	0.417
Chr 8	10	4.050	0.151	0.50	0.678	0.801	0.386
Chr 2	10	3.750	0.180	0.50	0.809	0.635	0.380
Chr 6	8	4.188	0.005	1.00	0.011	0.877	0.321
Chr 4	8	4.188	0.005	1.00	0.014	0.877	0.320
Chr 5	8	4.275	0.071	1.00	0.315	0.926	0.220

Since chromatograms 4, 5 and 6 have the fewest number of peaks, their peak punishment function values are one. Conversely, the peak punishment function values for chromatograms 7, 10 and 12 are zero, since these chromatograms have the greatest

number of peaks. By visual inspection, chromatograms 11 and 4 are the best in terms of resolution. Thus, these chromatograms should have the smallest punishment values for resolution. This agrees with the results shown in Table 3.

Chromatographic Resolution Statistic (CRS) Results

The results obtained using the CRS are shown in Table 4. These results used $R_{\min} = 0.3$ and $R_{\text{opt}} = 1.5$ in equation (22). Table 4 also shows the results obtained with the quality metric from scenario (f). The settings used are shown in Table 5. The results are quite similar, and in fact the top six chromatograms are the same.

Table 4. Comparison of the CRS (left) and quality metric (right) rankings.

Rank	Chromatogram	CRS
1	Chr 5	4.325
2	Chr 11	5.829
3	Chr 12	6.556
4	Chr 1	7.379
5	Chr 10	9.497
6	Chr 7	10.004
7	Chr 8	10.323
8	Chr 2	11.206
9	Chr 6	13.901
10	Chr 3	14.071
11	Chr 4	14.213
12	Chr 9	25.780

Rank	Chromatogram	Quality
1	Chr 5	0.715
2	Chr 11	0.646
3	Chr 12	0.549
4	Chr 1	0.547
5	Chr 10	0.420
6	Chr 7	0.312
7	Chr 3	0.310
8	Chr 8	0.281
9	Chr 2	0.188
10	Chr 9	0.075
11	Chr 4	0.072
12	Chr 6	0.038

Table 5. Resolution punishment parameters used in scenario (f).

Parameter	Value
$R_{\min}$	0.30
R_{opt}^{-}	1.50
R_{opt}^{+}	1.50
$R_{\max}$	30.00
$c_{2,1}$	1.00

Discussion

In this part of the project, a numerical metric was developed to indicate the best chromatogram in a set without the need of visual inspection. It is difficult to define what a good chromatogram is, as it will depend on the purpose of the experiment. The quality metric was developed with this principle in mind. The metric cannot determine the best chromatogram in the set without first knowing what parameters are most important for the particular experiment. The prioritization constants c_1, c_2 and c_3 allow the user to specify the importance of each criterion.

To illustrate this point, consider that in scenario (f), the best chromatogram is chromatogram 5, while in scenario (d), chromatogram 5 is the worst. Chromatogram 5 has the least number of peaks, all the peaks are completely resolved, and there is no waste of time between peaks. In scenario (d), the most important criterion is the number of peaks. In scenario (f), resolution is most important. Therefore, the difference is justified.

The scenarios shown in Table 4 were tested and the rankings corresponded to the results obtained by visual inspection. However, for scenario (c), when total run time is the only criterion under consideration, the quality metric is not as accurate. Ideally, if the user chooses to rank the chromatograms based on run time alone, the application should simply rank the chromatograms in order of increasing total run time (the chromatograms with the shortest and longest total run times should the best and the worst ranks). This could be accomplished by assigning the time punishment function values in terms of the *absolute* total run time, rather than the total run time *per peak* as done in equation 21. Unfortunately, when the algorithm using absolute run time was tested on scenario (d), in which all three quality criteria are important, the performance was far inferior to the algorithm in which run time per peak (equation 21) was used. If p_3 is based on absolute total run time, then chromatograms that have a longer run time because they have more peaks are punished rather severely for time. This is likely undesirable in many practical situations. Therefore a trade-off exists: the final algorithm can be selected so that either

- performance is good when all three quality factors are considered and imperfect when only total run time is considered, or
- performance is good when only total run time is considered, but worse when all three quality factors are considered.

The decision was to choose the first of the above two options, and sacrifice performance in the situation when only run time is important to gain performance in situations when all three factors are important. This choice was made for two reasons. First, it seems that scenarios when all three criteria are important will arise much more often than scenarios in which only run time is important. Second, even in the scenario when only run time is important, it is not very difficult to rank chromatograms manually. Visual inspection is far more difficult in situations when all three criteria are considered, so in that sense, it makes sense to design the application to function best in those situations.

Scenario (f) was developed in attempts to match the quality rankings with those given by the CRS method of Schlabach and Excoffier [2]. The CRS function evaluates a given chromatogram based on

- the extent of separation between each pair of adjacent peaks,
- the uniformity of spacing between peaks,
- the minimization of the analysis time.

The weights of these three criteria can the approximated by the algorithm from this work by setting $c_1 = 0.1$, $c_2 = 0.7$, and $c_3 = 0.2$ in equation 10. This points out that the CRS function is most useful when resolution is the most important quality attribute. If other attributes are more important, the CRS method does not handle this as well. The quality metric from this work is thus an improvement over the CRS method.

Future Work

The quality metric was developed assuming the peaks are well-modeled by Gaussian functions, and therefore the metric results are best when the chromatograms to be ranked have a minimal amount of tailing. To account for asymmetric peaks, the analysis in this work could be extended by modifying the resolution punishment function. If this were done, one possible approach is to use two Gaussian functions patched together at the mean, as was done in the simulation application developed in this project. The advantage to this method is that exact relationships for the peak widths and resolution can be developed. However, it may be difficult to correlate these relationships with the parameters that are given by the Empower software.

As mentioned in the Discussion section, the algorithm does not give ideal results when total run time is the only quality attribute considered, yet the algorithm was chosen because the alternatives performed worse in situations when all three attributes are taken into account. The run time punishment function p_3 is the crux of this issue. If p_3 is based on absolute total run time, then chromatograms that take longer because they have more peaks are unduly punished. Basing p_3 on run time per peak helps resolve this problem, but doing so gives poorer results in scenarios when only total run time is considered. It might be possible to merge the two concepts by weighting the two punishments based on absolute total run time and total run time per peak. This approach seems promising, but it was not attempted in this project due to schedule constraints.

Conclusions

- Two Excel applications were developed in this project—a chromatogram simulator, and a quality metric application that ranks chromatograms in a set based on their quality.

- The chromatogram simulator application can be used to generate realistic chromatograms that are well-suited for use in technical communications.

- Results obtained by the quality metric function correlate well with the results obtained by visual inspection.

- The metric application enables the user to choose the best chromatogram in a set of chromatograms with a much smaller need for visual inspection.

- The quality metric developed in this project can reproduce the results given by the CRS method. Since the quality metric developed in this work enables the user to specify the importance of each quality criterion, the algorithm from this project is an improvement over the CRS.

References

[1] Charles R. MacCluer, *Industrial Mathematics, Modeling in Industry, Science and Government*. New Jersey: Prentice Hall, 1999.

[2] Patrick H. S. Lukulay, *Parametric Modulation, A Novel Computer-Assisted Optimization Method for Liquid Chromatography*. East Lansing: Michigan State University, 1995.

[3] Empower System Suitability Quick Reference Guide, Revision A. http://www.waters.com.

[4] "Chromatography." Wikipedia, The Free Encyclopedia. 4 February 2007. Wikimedia Foundation, Inc. 5 February 2007. http://en.wikipedia.org/wiki/Chromatography.

Appendix A – Terminology

- Chromatography is a separation technique in which molecules or ions are distributed between two immiscible phases. One which is stationary and the other is mobile [2].

- Resolution is the degree of separation between any two solutes and is governed by the two independent processes that consist of solute retention and band broadening [2].

- The analyte is the substance which is to be purified or isolated during chromatography [4].

- A chromatogram is the visual output of the chromatographic separation. Different peaks or patterns on the chromatogram correspond to different components of the separated mixture [4].

- The mobile phase is the analyte and solvent mixture which travels through the stationary phase [4].

- The retention time is the characteristic time it takes for a particular molecule to pass through the system under set conditions [4].

- The stationary phase is the substance which is fixed in place for the chromatography procedure and is the phase to which solvents and the analyte travels through or binds to [4].

Appendix B – Test Case Chromatograms

In order to evaluate the quality metric, twelve chromatograms were generated using the simulation software. These chromatograms are shown in Figures B1-B12.

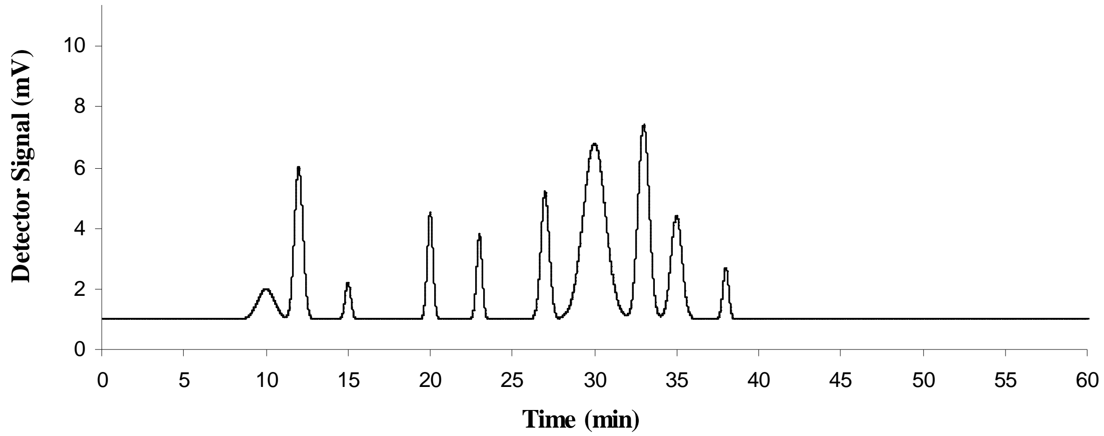

Figure B1. Chromatogram 1.

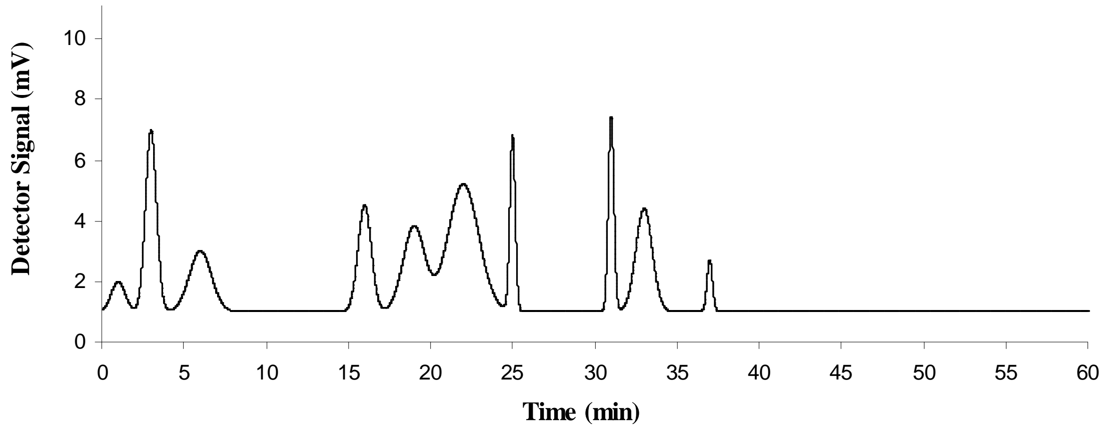

Figure B2. Chromatogram 2.

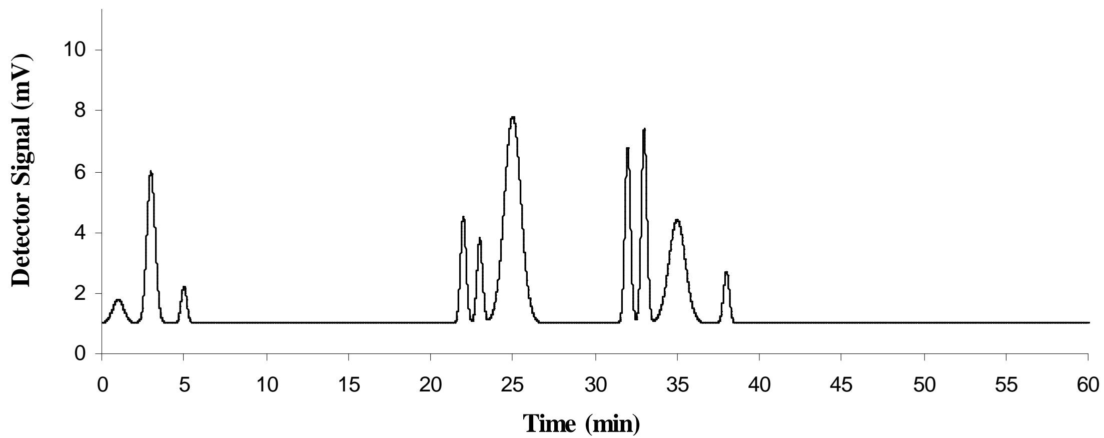

Figure B3. Chromatogram 3.

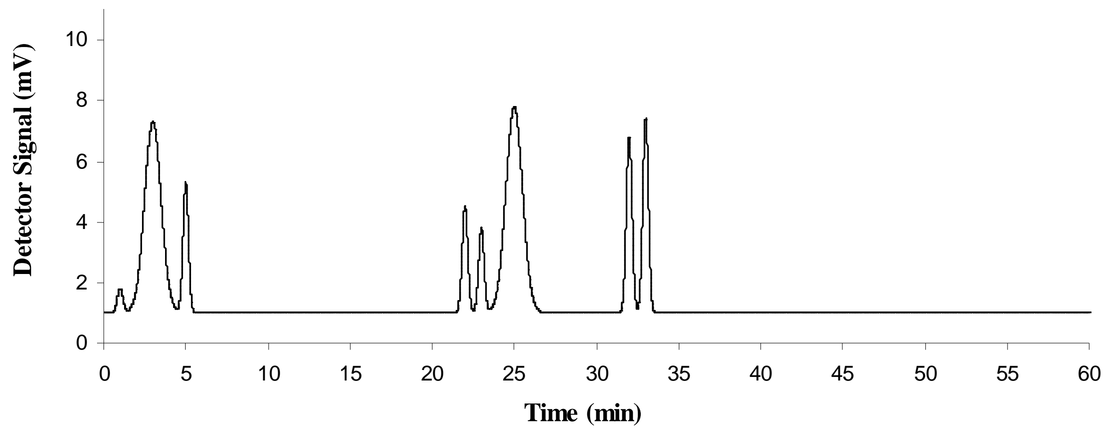

Figure B4. Chromatogram 4.

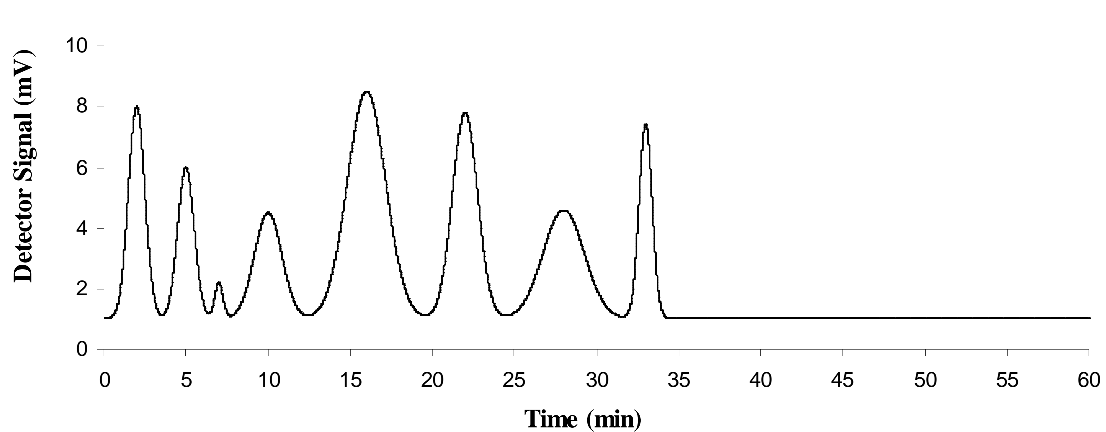

Figure B5. Chromatogram 5.

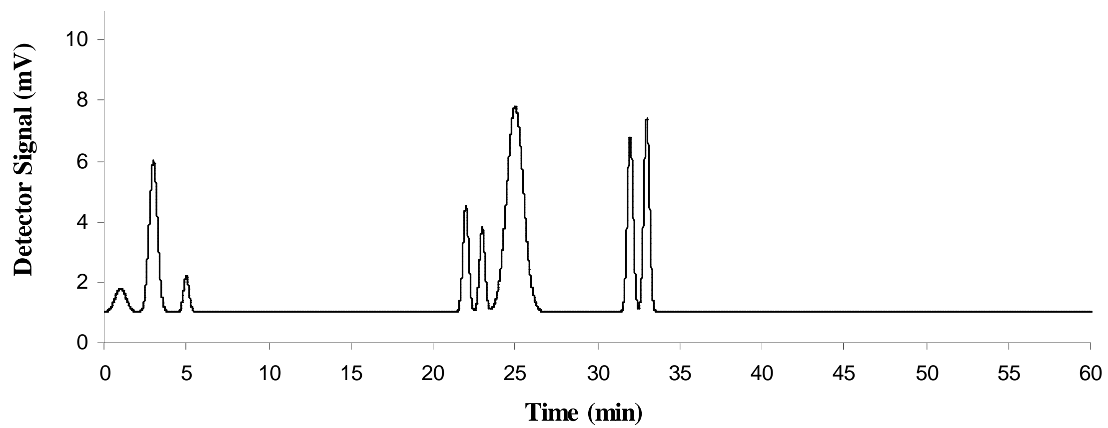

Figure B6. Chromatogram 6.

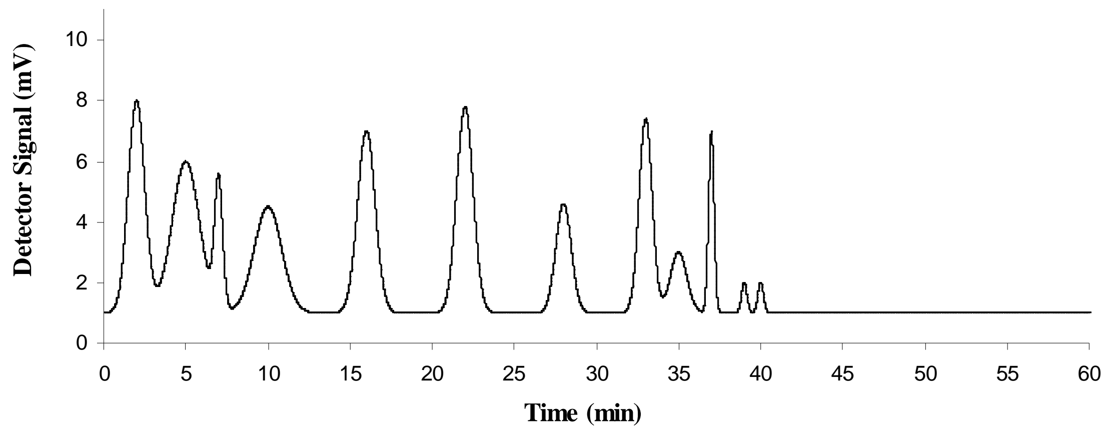

Figure B7. Chromatogram 7.

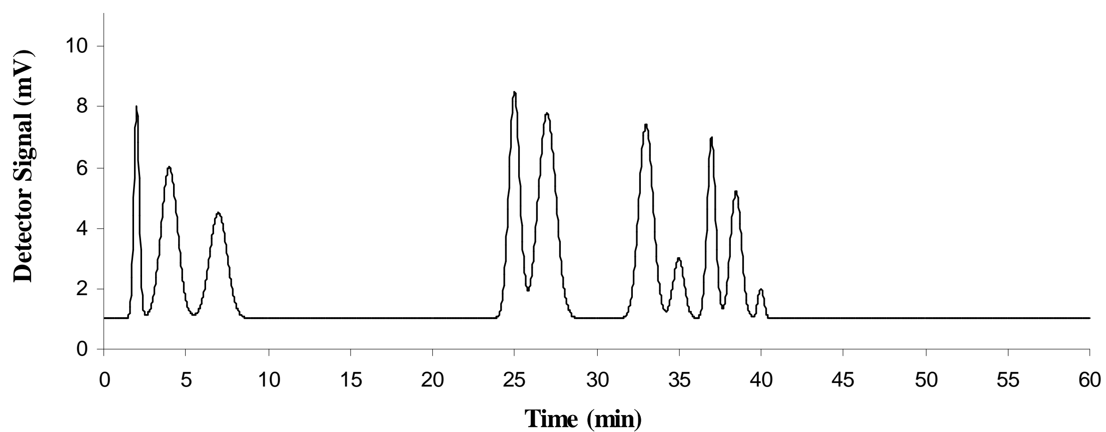

Figure B8. Chromatogram 8.

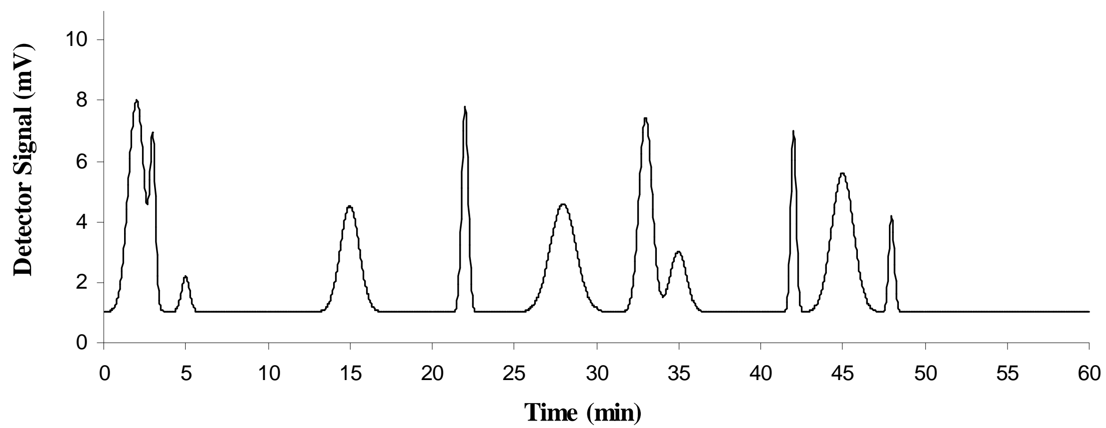

Figure B9. Chromatogram 9.

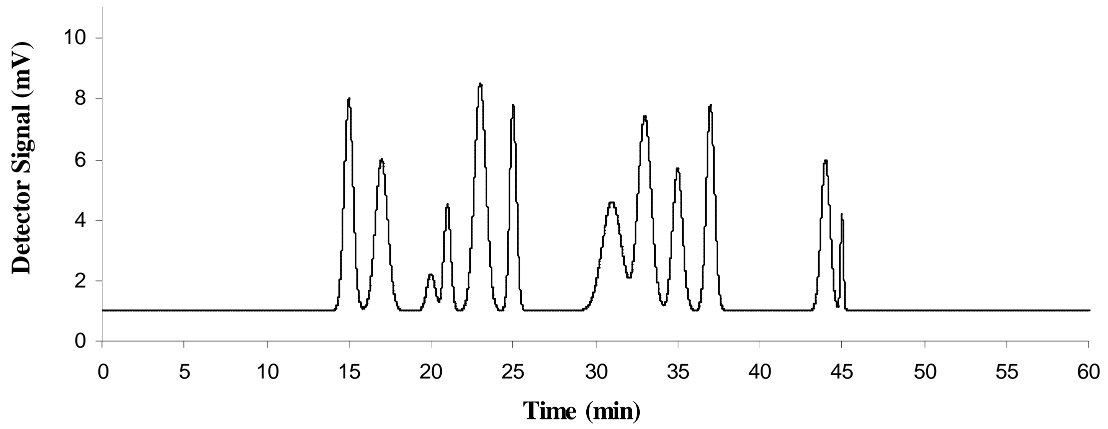

Figure B10. Chromatogram 10.

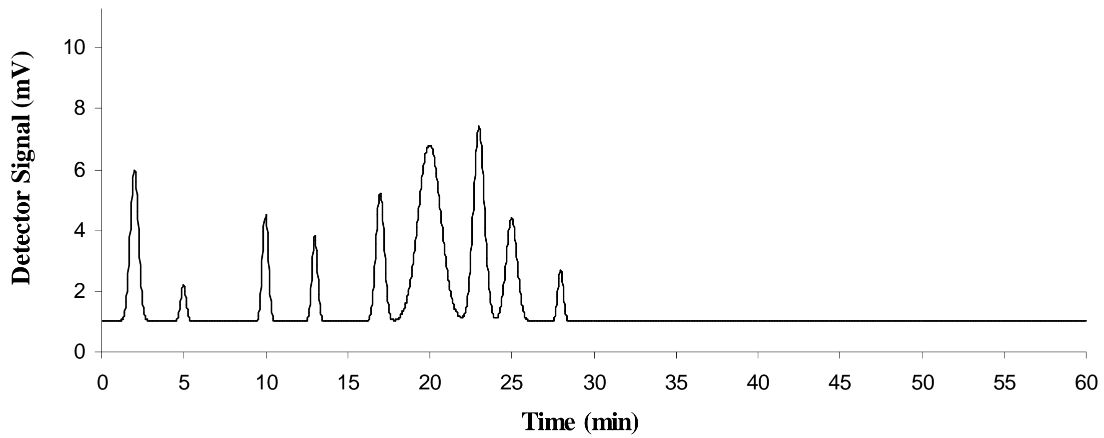

Figure B11. Chromatogram 11.

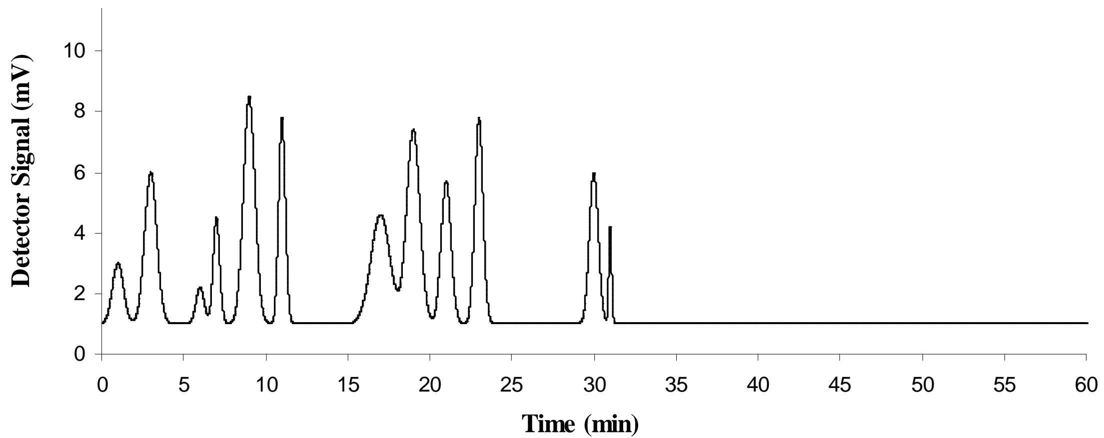

Figure B12. Chromatogram 12.

Appendix C – Derivations

The normal distribution density function is

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{C-1}$$

which has a height

$$h=f(\mu;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}, \tag{C-2}$$

and the area under the peak is 1.

For a given mean and standard deviation it is possible to graph a bell shaped curve. However, for the simulation application, the desired parameters are

- t_R - retention time;
- $w_{50\%}$ - peak width at half height;
- As_{10} - peak asymmetry; and
- h or $Area$ - peak height or peak area.

C.1 – Derive the expression that simulates an asymmetric peak with a specific height

First it is necessary to calculate the width at z% of the height, where z is a number between 0 and 100. The points $x_{z\%}$ that correspond to z% of the height can be calculated using (C-1) and (C-2). It follows that

$$\frac{z}{100}=\exp\left(-\frac{(x_{z\%}-\mu)^2}{2\sigma^2}\right), \tag{C-3}$$

and thus

$$x_{z\%}=\mu\pm\sigma\sqrt{2\ln(100/z)}\,. \tag{C-4}$$

For an asymmetric peak the width at 10% is given by

$$w_{10\%}=\sigma_B\sqrt{2\ln 10}+\sigma_A\sqrt{2\ln 10}\,, \tag{C-5}$$

where σ_B and σ_A denote the standard deviation to the left and to the right of the retention time (mean), respectively. It follows that a peak has an asymmetry of

$$As_{10} = \frac{A}{B} = \frac{\sigma_A \sqrt{-2\ln 0.1}}{\sigma_B \sqrt{-2\ln 0.1}} = \frac{\sigma_A}{\sigma_B}. \qquad \text{(C-6)}$$

From (C-4), the peak width at 50% is

$$w_{50\%} = \sigma_B \sqrt{2\ln 2} + \sigma_A \sqrt{2\ln 2}. \qquad \text{(C-7)}$$

From (C-6) and (C-7), the standard deviations to the left and to the right are

$$\sigma_A = \frac{w_{50\%} \cdot As_{10}}{(1 + As_{10})\sqrt{2\ln 2}} \qquad \text{(C-8a)}$$

and

$$\sigma_B = \frac{w_{50\%}}{(1 + As_{10})\sqrt{2\ln 2}}. \qquad \text{(C-8b)}$$

The last step is to incorporate the peak height. To accomplish this task the function (C-1) must be divided by the normal distribution height (C-2), and then multiplied by the specified height *h*. The function that simulates a peak with a retention time t_R, peak width at half-height $w_{50\%}$, asymmetry As_{10} and a given height *h* is

$$c_h(t) = \begin{cases} f(t; t_R, \sigma_B)\sigma_B \sqrt{2\pi} \cdot h & \text{if} \quad t \le t_R \\ f(t; t_R, \sigma_A)\sigma_A \sqrt{2\pi} \cdot h & \text{if} \quad t > t_R, \end{cases} \qquad \text{(C-9)}$$

with σ_A and σ_B given by equations (C-8a) and (C-8b).

C.2 – Derive the expression that simulates an asymmetric peak with a specific area

The area under an asymmetric peak is given by

$$Area = Area^B + Area^A, \qquad \text{(C-10)}$$

where $Area^B$ and $Area^A$ are the area under the peak to the left and to the right of the retention time, respectively. Since the area under a normal peak is 1, it follows that

$$Area^A = \frac{1}{2} \cdot \frac{h}{h_A} = \frac{h \cdot \sigma_A \sqrt{2\pi}}{2} \qquad \text{(C-11a)}$$

and

$$Area^{B} = \frac{1}{2} \cdot \frac{h}{h_{B}} = \frac{h \cdot \sigma_{B} \sqrt{2\pi}}{2}, \qquad \text{(C-11b)}$$

where h_A and h_B are the peak heights given by equation (C-2). The total area is then

$$Area = \frac{h\sqrt{2\pi}}{2}(\sigma_{A} + \sigma_{B}). \qquad \text{(C-12)}$$

From (C-12), (C-11a), and (C-11b) it follows that the areas to the left and to the right of the retention time as a function of the total area are given by

$$Area^{A} = \frac{Area \cdot \sigma_{A}}{(\sigma_{A} + \sigma_{B})} \qquad \text{(C-13a)}$$

and

$$Area^{B} = \frac{Area \cdot \sigma_{B}}{(\sigma_{A} + \sigma_{B})}. \qquad \text{(C-13a)}$$

With this result it is possible to generate a Gaussian peak with a given retention time, width at half-height, asymmetry, and area. The function that simulates the peak is

$$c_{A}(t) = \begin{cases} 2 \cdot f(t; t_{R}, \sigma_{B}) \cdot Area^{B} & \text{if} \quad t \leq t_{R} \\ 2 \cdot f(t; t_{R}, \sigma_{A}) \cdot Area^{A} & \text{if} \quad t > t_{R}. \end{cases} \qquad \text{(C-14)}$$

The factor 2 arises from the fact that to generate the left hand side and the right hand side of the peak it is necessary to assume that each side corresponds to a Gaussian curve. Since a complete Gaussian curve is symmetric, the area under the complete peak is twice the area calculated in equation (C-13).

C.3 – Derivation of the optimal resolution

The USP resolution between two symmetric peaks is given by

$$R = \frac{2(t_{R,2} - t_{R,1})}{1.7(w_{50\%}^{(1)} + w_{50\%}^{(2)})}. \qquad \text{(C-15)}$$

For a normally distributed peak the peak width is relatively close to six times the standard deviation. From this fact, in order to have totally resolved peaks it is necessary that

$$t_{R,2} = t_{R,1} + 3(\sigma_{1} + \sigma_{2}), \qquad \text{(C-16)}$$

where σ_1 and σ_2 are the standard deviations of peak 1 and peak 2, respectively. Using equation (C-7) to rewrite equation (C-16) in terms of peak width at half-height,

$$t_{R,2} = t_{R,1} + \frac{3}{2\sqrt{2\ln(2)}} \cdot (w_{50\%}^{(1)} + w_{50\%}^{(2)}) \,. \tag{C-17}$$

Substituting equation (C-17) in equation (C-15) shows that the optimal resolution is

$$R_{\text{opt}} = \frac{2\left(t_{R,1} + \dfrac{3(w_{50\%}^{(1)} + w_{50\%}^{(2)})}{2\sqrt{2\ln(2)}} - t_{R,1} \right)}{1.7(w_{50\%}^{(1)} + w_{50\%}^{(2)})} = \frac{3}{1.7\sqrt{2\ln(2)}} \approx 1.5. \tag{C-18}$$

C.4 – Simulating baseline noise

To best simulate baseline noise, it is necessary to use a function defined in terms of random variables. Several such functions were investigated, and the one that gave the most realistic results was selected. The noise function is based on two numbers,

$$u_1 \sim U[-1,1] \tag{C-19}$$

and

$$u_2 \sim U[-1,1]\,, \tag{C-20}$$

both of which are selected from the uniform distribution on $[-1,1]$. Then to each data point on the baseline is added the noise value

$$\phi = d \cdot \left| u_2 \sqrt{-2\ln(x)/x} \right|, \tag{C-21}$$

where

$$x = \frac{u_1^2 + u_2^2}{2} \tag{C-22}$$

and d is an adjustable parameter that controls the “amplitude” of the noise (increasing d increases the intensity of the noise).

Appendix D – User Guide – Simulation

Application Organization

The application consists of one worksheet, which contains the user options to simulate a chromatogram, the chromatographic inputs and the respective outputs. This worksheet is shown in Figure D1. Throughout, blue font indicates cells that the user should adjust.

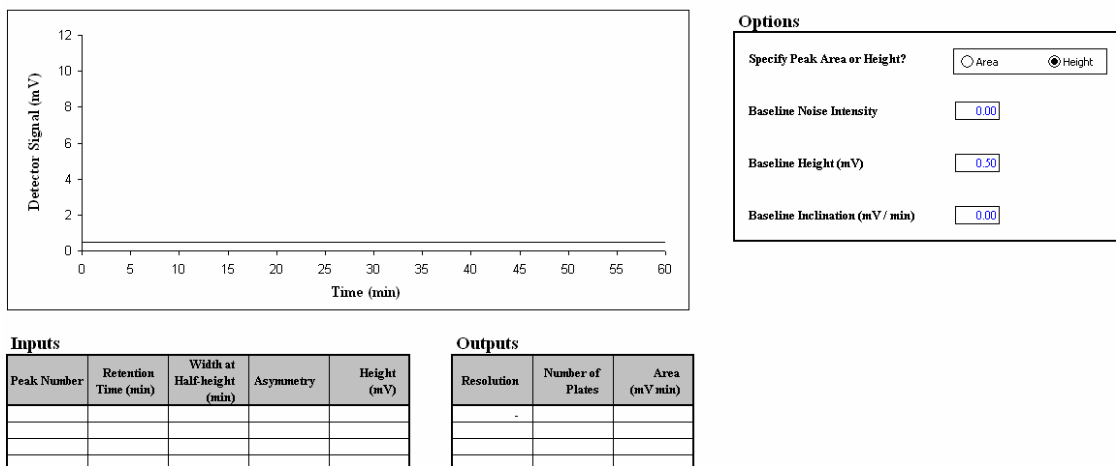

Figure D1. Screenshot of the simulation application.

Options Table

The Options table is enlarged in Figure D2. In this table, set the options as follows.

- For each peak in the simulated chromatogram, either the height or the area must be specified (the other is calculated automatically). To specify the area of each peak, click the Area radio button. To specify the height of each peak, click the Height radio button.

- The application has the capability to simulate baseline noise. To include baseline noise, change the Baseline Noise Intensity to a positive number (for example, setting the intensity to 0.05 gives light noise, and 0.20 gives heavier noise). To turn noise off, change the Baseline Noise Intensity to zero.

- To adjust the baseline height up or down, change the value in the Baseline Height box (this value must be a positive number).

- The application can simulate chromatograms with a sloped baseline. To adjust the inclination of the baseline, change the value in the Baseline Inclination box (this value must be non-negative).

Options

Specify Peak Area or Height?	○ Area ◉ Height
Baseline Noise Intensity	0.00
Baseline Height (mV)	0.50
Baseline Inclination (mV / min)	0.00

Figure D2. Screenshot of the Options table.

Inputs Table

The Inputs table with two peaks already entered is shown in Figure D3. This section allows the user to input the values for the retention time, width at half-height, asymmetry ratio, area or height of each peak.

- Do NOT enter the peak numbers. Peak numbers appear automatically once the four required input parameters are entered.

- Enter the retention time, width at half height, asymmetry ratio, and height or area (whichever was chosen as the input parameter in the Options table) for the first desired peak. When all four parameters are entered, the specified peak appears, and the peak number appears along the left edge of the Inputs table.

- Repeat the above step for each additional peak. Peaks must be entered in order of increasing retention times. If this is not done, an error message appears and the graph vanishes until the problem is corrected.

Inputs

Peak Number	Retention Time (min)	Width at Half-height (min)	Asymmetry	Height (mV)
1	5.0	0.6	1.0	4.0
2	7.2	0.5	1.9	3.0

Figure D3. Screenshot of the Inputs table with two peaks entered.

Outputs Table

The Outputs table screenshot is shown below in Figure D4. This section provides the resolution between adjacent peaks, the number of plates, and the area under the peak or the peak height (whichever was not chosen as the input parameter in the Options table). The first row for resolution is empty, and the second row contains the resolution between peak 1 and peak 2.

Outputs

Resolution	Number of Plates	Area (mV min)
-	384.20	2.55
2.35	1147.21	1.60

Figure D4. Screenshot of the Outputs table corresponding to the input values from Figure D3.

Appendix E – User Guide – Metric

Configuring Excel

The file Pfizer_QualityMetric.xls contains macros, so Excel must be configured to enable macros to run before the file is opened. This can be done using the following steps:

1. Start Microsoft Excel (but not by opening the Pfizer_ QualityMetric.xls file). If a new workbook called Book1 is opened automatically, close its window (but not the entire Excel application window) by clicking the Close Window box, or by selecting Close from the File menu.

2. Pull down the Tools menu, point to Macro, and select Security. The Security dialog box appears. Click the Security Level tab.

3. To use the application, the Security Level must be set to Medium or Low. Click the radio button for the desired setting, and then click OK.

4. Open the file Pfizer_Simulation.xls. If the Medium security setting was selected, a Security Warning dialog box similar to the one shown in Figure E1 will appear (this is normal). If this occurs, click the Enable Macros button.

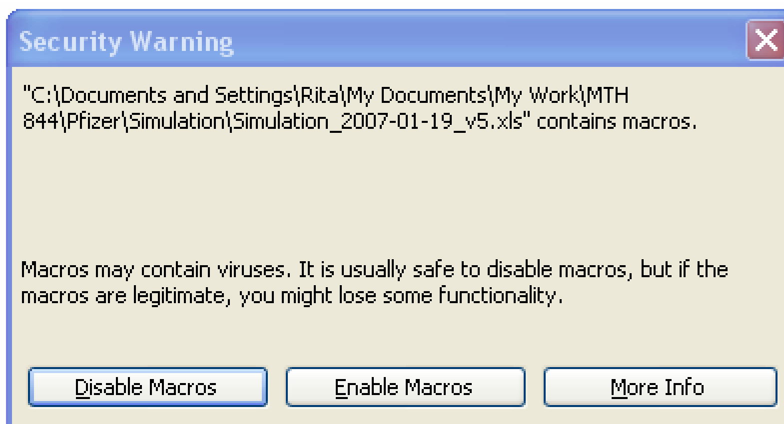

Figure E1. This Security Warning dialog box is normal.

Ranking Chromatograms By Quality

1. Click the Select Data File Button (see Figure E2). A dialog box opens. Browse until the desired data file is found, select it, and click Open. The name of the data file is then displayed on the screen.

2. Choose the desired threshold area. All peaks in the data file with an area below the threshold area will be automatically deleted. (By default, the threshold area is set to zero, so that if any peaks in the data file were not integrated and do not have data available, they are deleted automatically.)

Settings	
Input data file	Change Input Data File
Input data file is	WonderSubstanceData.xls
Thresholding	
Threshold area A_T	0

Figure E2. Use this part of the interface to configure the input data file.

NOTE: The data file must have the appropriate format or the application will not work (runtime errors may be generated, or the rankings may not be correct). The data file must consist of at least two worksheets, one worksheet for each chromatogram in the set. Each worksheet must have the format indicated in Figure E3. In particular, the following conditions are important.

- The name of each chromatogram (or set of experimental conditions) should appear in cell A1.
- The data for the first peak should be found in row 5.
- The peaks should be numbered sequentially beginning in cell A5 and continuing downward in column A.
- Columns B through K should contain the data indicated in Figure E3 for each peak.
- Peaks that are not integrated (such as peaks 8 and 9 in Figure E3) will be automatically deleted when the rankings are calculated. This is done by checking the area of each peak, and deleting peaks that have an area less than the specified threshold area (zero by default).

	A	B	C	D	E	F	G	H	I	J	K
1	Pfizer-Compound#7 Oxidation										
2											
3			Retention	Alpha	Peak	Peak	Peak	Width	USP	USP	USP
4		Name	Time	(RRT)	Area	Height	Width	@ 50%	Resolution	Tailing	Plates
5	1	base line disturbance 1	10.002	0.75	3866	316	29.352	0.241		1.04	19093
6	2	base line disturbance 2	10.887	0.81	7089	593	38.310	0.314	3.51	0.78	41382
7	3	Peak1	11.385	0.85	17842	2716	21.486	0.085	2.57	1.90	68628
8	4		12.214	0.91	2239	323	16.169	0.145	4.60	0.93	68688
9	5	Peak2	12.438	0.93	6703	1231	14.858	0.088	1.35	2.08	117648
10	6	Peak3	13.057	0.98	7490	902	23.379	0.222	3.43	0.80	58300
11	7	PF-043537209	13.389	1.00	2425167	435756	21.413	0.084	1.83	1.14	133810
12	8	Peak4	13.881	1.04	7743	1703	10.706	0.075	3.66	1.46	203815
13	9	Peak5	14.143	1.06							
14	10	base line disturbance 3	14.603	1.09							
15	11	base line disturbance 4	14.725	1.10	1069	251	9.541	0.083	6.91	0.98	236646
16	12	Peak6	15.063	1.13	19745	3829	15.878	0.076	2.67	1.70	208601
17	13	diluent peak	15.448	1.15	7618	1211	16.387	0.103	2.69	2.11	161262
18	14	Peak7	16.127	1.20	6487	675	24.108	0.205	3.50	0.90	75609
19	15	base line disturbance 5	16.421	1.23	4306	596	15.149	0.144	1.38	1.10	116148
20	16	base line disturbance 6	16.616	1.24							
21	17	Peak8	17.064	1.27	9123	1933	13.037	0.069	4.02	1.61	286956
22	18		17.918	1.34	1676	307	13.693	0.111	6.27	1.21	243667
23											
24											

Figure E3. Each worksheet in the data file must have the format shown here.

3. Set the prioritization constants and resolution punishment function parameters (see Figure E4) by modifying the default parameters as desired. The graph of the resolution punishment function is automatically updated to reflect the changes. The user should read the Analysis section of Part 2 of this report to understand how these parameters can be used to prioritize the quality attributes.

Quality criteria prioritization constants	
c_1 (peaks)	0.50
c_2 (resolution)	0.30
c_3 (time)	0.20
Total	1.00

Resolution punishment function parameters	
R_{min}	0.30
R_{opt}^{-}	1.30
R_{opt}^{+}	2.00
R_{max}	4.00
$c_{2,1}$	0.00

Figure E4. Use this part of the interface to configure the punishment parameters.

4. Click the Rank button. The following steps are performed automatically.

I. The specified data file is opened if it is not already open.
II. The first worksheet in the data file is checked to see whether the area of each peak is greater than or equal to the threshold area. Peaks whose areas are below the specified threshold are deleted.
III. The values from the first worksheet are copied to the Calculations worksheet of the Pfizer_QualityMetric workbook. The necessary calculations are done.
IV. The results of the first chromatogram calculations are copied to the Summary worksheet of the Pfizer_QualityMetric workbook.
V. Steps II through IV are repeated until each worksheet in the data file has been processed.
VI. The data on the Summary sheet are sorted according to the quality scores.
VII. The ranking and quality score of each chromatogram is inserted in the Results table on the Input-Output worksheet. The chromatogram with the best ranking (and the largest quality score) is displayed first (see Figure E5).

Results | Rank

Rank	Chromatogram	Score
1	Chr 12	0.83
2	Chr 10	0.70
3	Chr 7	0.64
4	Chr 1	0.61
5	Chr 3	0.61
6	Chr 11	0.56
7	Chr 8	0.52
8	Chr 2	0.46
9	Chr 9	0.38
10	Chr 4	0.32
11	Chr 6	0.32
12	Chr 5	0.30

Figure E5. After clicking the Rank button, the rankings of the chromatograms are listed.

5. To redo the rankings with different input parameters, simply adjust the parameters as desired and click the rank button again.

6. To rank another set of chromatograms, simply change the data file and repeat the same steps. It is possible to add or remove chromatograms from the set by adding or deleting worksheets in the data file, but the changes in the rankings will not be reflected until the Rank button is clicked again.