[About the GUI 2](#_Toc320860221)

[Settings 4](#_Toc320860222)

[Noise window 4](#_Toc320860223)

[Thres 4](#_Toc320860224)

[V1/V2 5](#_Toc320860225)

[Shift 5](#_Toc320860226)

[Min peak height 5](#_Toc320860227)

[From … min to … min 5](#_Toc320860228)

[Scan frequency 5](#_Toc320860229)

[Smoothing window 5](#_Toc320860230)

[Misc. features 6](#_Toc320860231)

[Load .xls 6](#_Toc320860232)

[Load 7](#_Toc320860233)

[Change dir. 7](#_Toc320860234)

[Begin calculations 7](#_Toc320860235)

[Export to Excel 7](#_Toc320860236)

[Plot peak width 7](#_Toc320860237)

[Plot retention time 7](#_Toc320860238)

[Close 8](#_Toc320860239)

[PCA 8](#_Toc320860240)

[GC data 8](#_Toc320860241)

[Compound levels 8](#_Toc320860242)

[Rt index 8](#_Toc320860243)

# About the GUI

The GUI is developed in Matlab R2008a with the purpose to detect all peaks present in a chromatogram. For further documentation we refer to the article “An automated method for baseline correction, peak finding and peak grouping in chromatographic data”, by L. Johnsen, T. Skov, U. Houldberg and R. Bro.

The GUI was originally developed for GC-FID data, but it can be used for all types of chromatographic data (eg. TIC from GC-MS). The GUI has been made so it can load data in four different formats: \*.RAX (csv format from Perking Elmer software), \*.cdf (gc-ms data), \*.mat and \*.xls.

If data is loaded in \*.mat or \*.xls formats it is important that the first column contains the time scale for the runs and the following columns contain the intensities obtained over time.

The layout of the GUI is shown in Figure 1. The functionality of the different parts will be reviewed in the following sections.

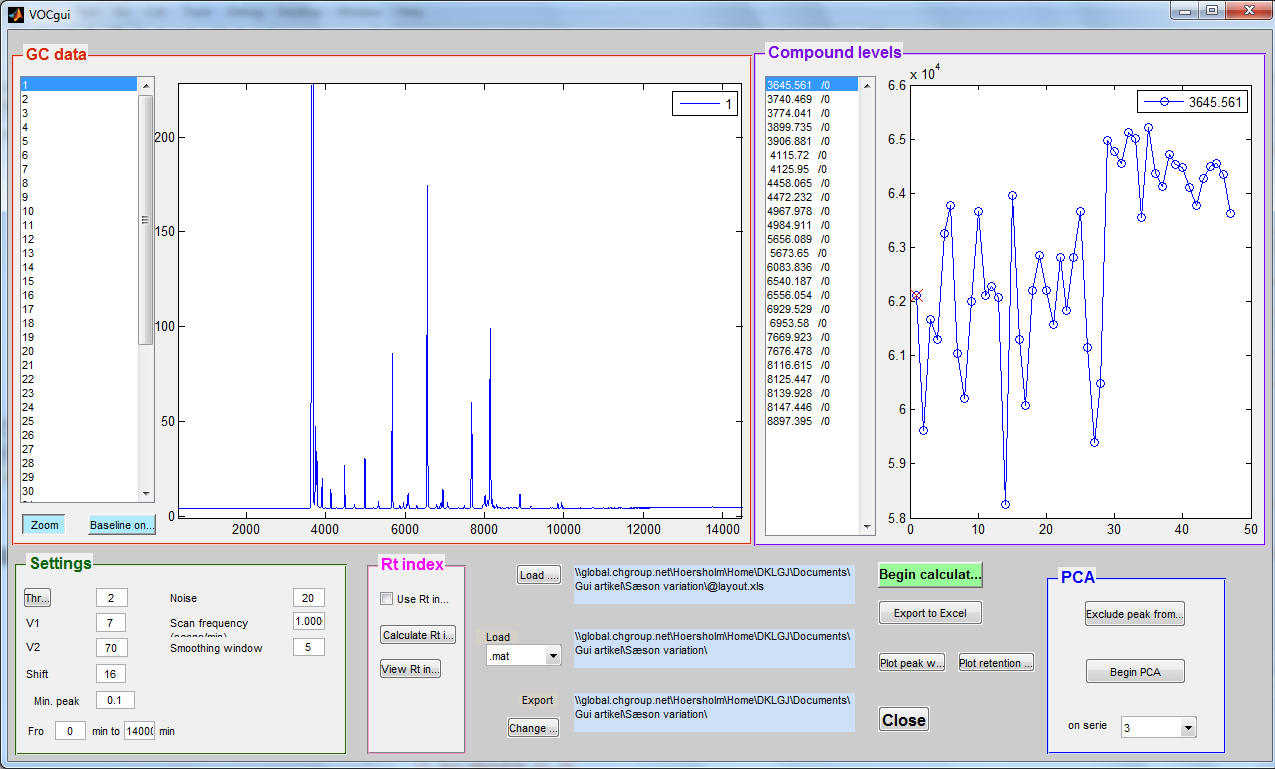


Figure 1. Screen dump of the GUI. The different elements is described in the sections below

# Settings

In this section the different settings for the algorithm should be adjusted to fit the relevant application.

## Noise window

This is the width of the window used to calculate the moving standard deviation. It is recommended that the window is set to approximately the peak width at half height (in data points).

## Thres

This is the threshold for the separation between regions with peaks and without peaks. The separation into the two groups is based on a moving standard deviation. The thres should be determined to fit the actual application but does not need to be recalculated every time the GUI is used. The determination of the Thres is fairly easy. Just push the button “Thres” and a window will appear as shown in Figure 4, the figure illustrates the chromatogram (in green) representing the sample marked in the “sample list” plotted together with the calculated std deviation (in blue). The red line illustrates a Thres on 3300. In this way it is easy to determine the effect of different threshold levels. The regions where the blue line is below the Thres are considered as non-informative and is used in the estimation of the baseline.

This function can only be used if an initial calculation has been performed.

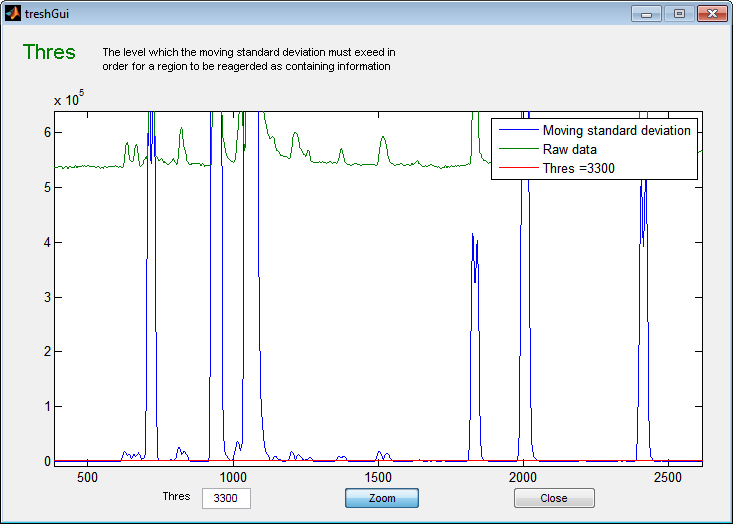


Figure 2. Illustration of how to determine the Thres. The red line illustrates a thres on 3300, regions where the blue line are below the red are considered noise and will be used in the estimation of the baseline.

## V1/V2

All peaks have to comply too a number of demands in order to be regarded as real peaks. V1 is the minimum peak width at half height and V2 is the maximum peak width.

## Shift

In order for peaks to be grouped they must not shift more than this number of scans. The shift can be determined by using the “peak” sheet in the excel sheet with the exported data (it will be necessary to perform an initial calculation and export the result to excel first). This sheet shows the distributions of the peaks before any alignment. If conditional formatting is used to colour all cells with a value greater than zero it is easy to determine the variation, by using peaks which are present in many (preferably all) the samples. Figure 3 shows an example with two very distinct peak groups. The shift should be odd and a bit higher than the deviation. In the example shown a shift on 13 would be appropriate.

Scan # 3541

Scan # 3511

Scan # 3520

Scan # 3530

Samples

RT

Scan # 3500

Scan # 3504

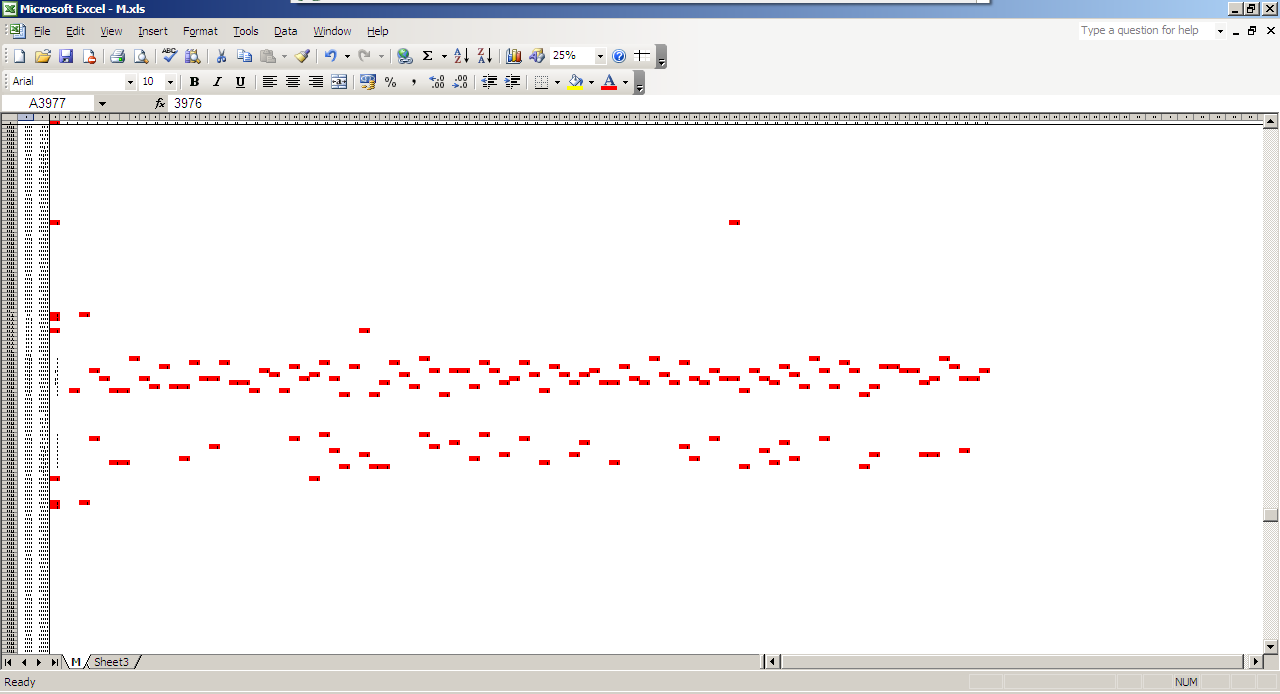


Figure 3. Illustration of the variation of peaks across samples. In the illustrated example it is clear that there are at least two peak groups represented. The variability of the two groups is at respectively 10 and 12 scans.

## Min peak height

Only peaks higher than this will be evaluated.

## From … min to … min

Only peaks eluting in this interval will be evaluated.

## Scan frequency

If the data is loaded as .RAX the scan frequency must be stated in scans/min, for other formats the GUI will calculate the actual scan frequency.

## Smoothing window

For very noisy data smoothing must be applied. The smoothing is simply done by calculating a moving average with a window size as stated here. If the smoothing window is set to 1 no smoothing will be applied.

# Misc. features

## Load .xls

This button is used to load the excel sheet containing the design of the experiment as well as the composition and index for possible Rt-index samples. The excel file must contain a sheet named “Samples” containing information about the experimental design, designation of RT standards (if included), control samples and experimental sample. If it is desired to export data to excel after processing the desired name of the excel sheet should be included in the layout sheet as well. In addition to this it is possible to give information about an optional number of classes; batch, good sample/bad sample etc. In the example shown in Figure 4 there are 3 different experiments (Exp\_1, Exp\_2 and Exp\_3). The results from the GUI will then be export to three different excel sheets (Exp\_1.xls, Exp\_2.xls and Exp\_3.xls) all containing control samples (indicated with 1 in the “Series” column) and Rt-index samples (indicated with 2 in the “Series” column).

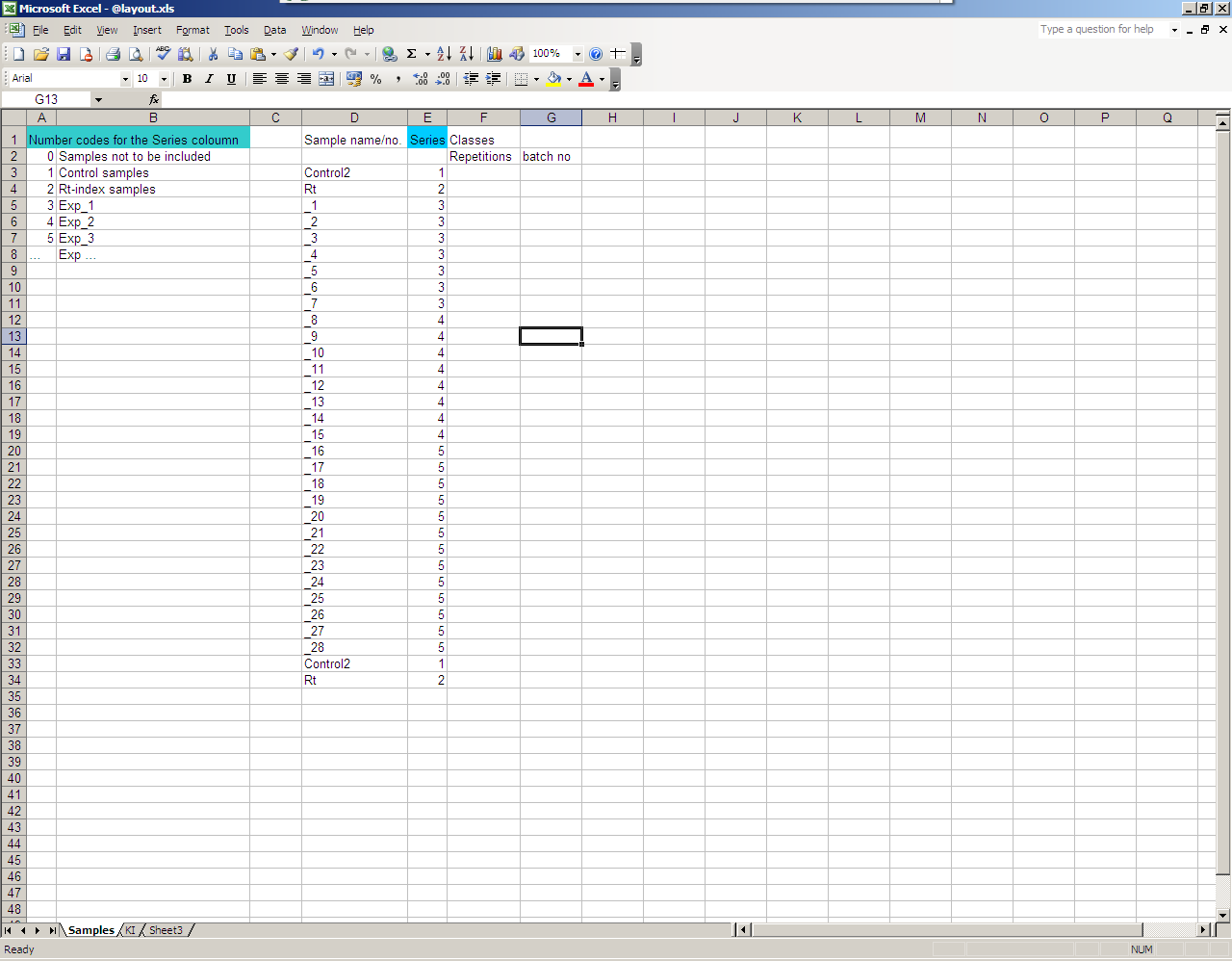


Figure 4. Example of the design of the “Samples” sheet

The samples name/number should always contain at least one character not being a number (eg. a letter or a symbol). The classes can be either numbers or names, and as many classes as desired can be added to the sheet.

If RT index is to be used also a sheet named “KI” with information about the rt-index compounds has to be included. The sheet must contain names of the compounds, their index, retention time and the window width where these peaks should be found. An example is shown in Figure 5

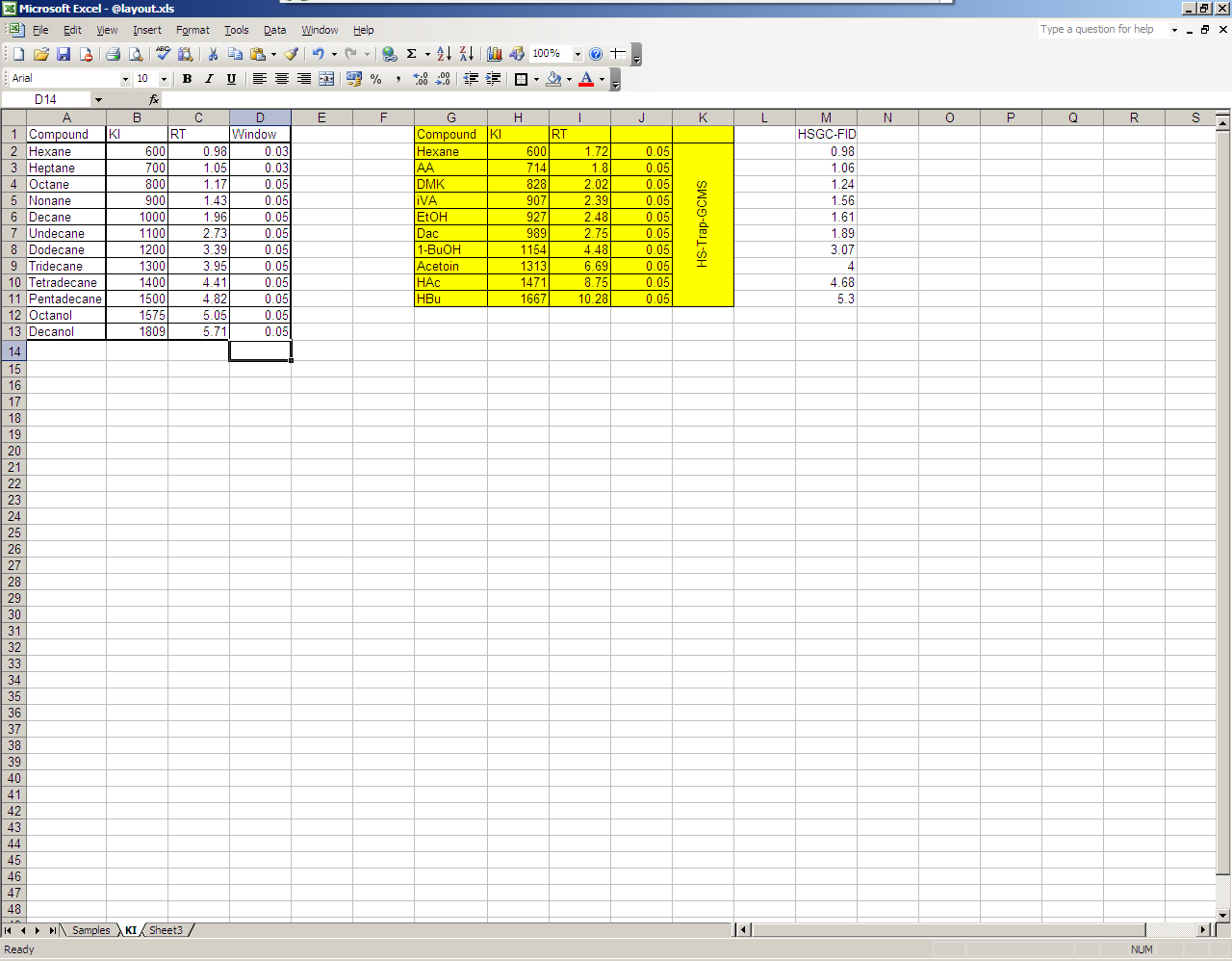


Figure 5 Example of the design of the “KI” sheet

## Load

Use the dropdown menu to choose the file-format and a window will appear where the location of the raw data should be chosen. If the data format is .cdf or .RAX the GUI will include all files in the specified format in the chosen folder in the analysis. If the file format is .mat the name of the matrix containing the data in the .mat file should be “x”. If the file format is .cdf the function iCDF is necessary (**Skov T and Bro R.** (2008) Solving fundamental problems in chromatographic analysis Analytical and Bioanalytical Chemistry, 390 (1): 281-285. http://www.models.life.ku.dk/iCDF)

## Change dir.

The directory indicates where the excel file with the exported data is saved. The directory containing the raw data is selected as default.

## Begin calculations

Before pressing this button settings should be adjusted to match the actual application, raw data and the excel sheet should be loaded and the Rt index should be calculated (if the evaluated peaks is to be assigned a Rt-index).

## Export to Excel

By pressing this button peak height, width, retention time and retention index is exported to an excel sheet (Import sheets). Also the settings used in the calculations are exported (“Settings” sheet) and the distribution of the peak before any alignment (“Peaks” sheet).

## Plot peak width

When pushing this button a window appears with a plot of the width of the peaks selected in the list box in the Compound levels section.

## Plot retention time

When pushing this button a window appears with a plot of the actual retention time of the peaks selected in the list box in the Compound levels section.

## Close

This button closes the GUI and clears the memory. If the GUI is closed in any other way data will be stored in the memory until Matlab is closed.

# PCA

This section prepares data to be used in PLS toolbox, by storing data as a dataset and assigning the classes from the layout excel sheet. The PLS\_toolbox (from eigenvector) needs to be installed before this section can be used. It is possible to excluded peaks selected in the Compound levels listbox from the dataset exported to the toolbox.

All samples with the series number chosen in the drop down menu as well as control samples will be exported.

When the “Begin PCA” button is pressed the PLS-toolbox will open in the PCA mode. Data needs to be loaded before the PCA can be made. The data is stored in a dataset called xp.

# GC data

In this window it is possible to inspect raw data. When samples are selected in the listbox to the left the raw chromatograms will be shown in the plotting window. By pressing the zoom button zooming becomes available both in the raw data and in the “compound level” plot.

Pressing the “baseline on/off” button makes it possible to see the estimated baseline for the selected samples.

Inspection of the raw data is only possible when calculations have been performed.

# Compound levels

When the calculations are done a list of peak-groups will be shown in the listbox to the left in this section. By selecting peak-groups the height of the peaks included in the group is shown in the plotting window.

# Rt index

If the experiment has Rt-index samples included, the box “Use Rt index” should be ticked off, and the button “Calculate Rt index” should be pressed before “Begin calculations”. Figure 6 shows the window which appears after the button “calculate Rt index” have been pressed. In the example there are two Rt-index samples included, each consisting of 12 compounds (hexane, heptane, octane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, octanol and decanol).

In the last row (“Use”) the number one can be replaced with a zero if a specific sample has been analysed wrongly, and not should be included in the calculation of the rt-indexes. Samples with zero in this row will not be used in the calculation of the Rt-index. It is also possible to manually change the retention time of the different compounds. If one compound in a sample not should be used the retention time can be replaced with NaN (Not A Number).

Pressing the zoom button allows zooming in the chromatogram shown (which shows the raw chromatograms of the Rt-index samples).

When the inspection and potential changes are done press the OK button (it is very important not to close the window in other ways since the index only is calculated by pressing OK.

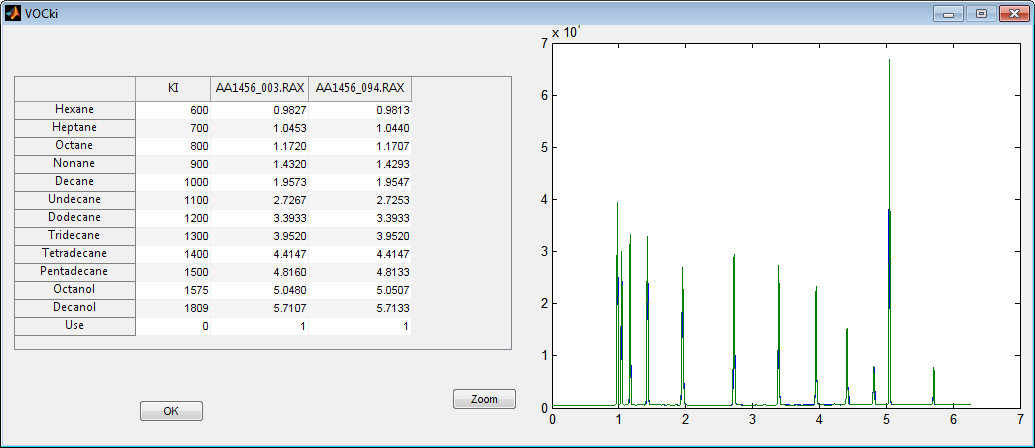


Figure 6. Illustration of the window used in calculation of the Rt-index.