# Supporting information

Selection of the appropriate optimization criterion is one of the most crucial factors when optimizing a chromatographic or electrophoretic separation, and it is well established that different criteria will produce different response surfaces and predict different 'optimum' conditions [1]. This largely results from the rather subjective nature of defining the 'optimum' or 'best' separation. When resolution is the most criticial parameter (for example in chiral separations), then the minimum resolution ($R_{min}$) defined as,

$$R_{si, i+1} = (2(t_{r_{i+1}}-t_{r_i})/(w_{i+1}+w_i)), \qquad (1)$$

is often the most useful. However in many cases it is desirable to include additional factors, such as analysis time or the distribution and even number of peaks. This can be achieved by creating a chromatographic response function (CRF) sometimes called an objective function. Many different CRFs have been developed, and are comprehensively covered in an excellent review by Siouffi and Phan-Tan-Luu [2].

In developing a buffer system to perform microchip DNA separations in an uncoated channel, we desired to optimize the separation conditions (field strength, buffer ionic strength, and polymer concentration) to give a rapid separation with sufficient resolution. In doing so, we constructed a CRF, which we called $Norm_{R,t}$ based on the product of the normalized time and resolution values for each separation. During the preparation of this manuscript, several questions were raised about the scientific basis of this criterion and on its suitability to find 'optimum' conditions. Here, we present a detailed discussion on the $Norm_{R,t}$ criterion and its application to optimizing two different data sets.

The criterion $Norm_{R,t}$ is defined as:

$$Norm_{R,t} = R_{norm} * t_{norm}, \qquad (2)$$

where $R_{norm}$ and $t_{norm}$ are defined as

$$R_{norm} = (R-R_{min})/(R_{max}-R_{min}) \qquad (3)$$

$$t_{norm} = (t_{max}-t)/(t_{max}-t_{min}) \qquad (4)$$

to give

$$Norm_{R,t} = (R-R_{min})/(R_{max}-R_{min}) * (t_{max}-t)/(t_{max}-t_{min}). \qquad (5)$$

In calculation of the value of $R_{norm}$, R is used to represent the resolution between the worst peak pair for each separation (more accurately defined as the minimum resolution), which was always between the 271 and 281 bp fragments. $R_{min}$ represents the minimum of this resolution over the experimental space, while $R_{max}$ represents the maximum of this resolution over this space (ie, when the resolution between the worst two peaks is at its greatest). Values for $R_{norm}$ will therefore take values between 0 and 1, corresponding to the separations with the worst and best resolution respectively. Similarly, t represents the migration time of the last DNA fragment, with $t_{max}$ being the longest time for migration of this fragment over the experimental space, while $t_{min}$ is the shortest migration time. Thus $t_{norm}$ values will likewise range between 0 and 1, with the shortest separation having a value of 1, while the longest has a value of 0. The product of $t_{norm}$ and $R_{norm}$ will have a value between 0 and 1, with a value of 1 occurring when the conditions have both the fastest time ($t_{norm} = 1$) and best resolution ($R_{norm} = 1$). It is important to note that values of 0 will be obtained when either the time is the longest ($t_{norm} = 0$) or when the resolution is the worst ($R_{norm} = 0$).

Upon careful consideration of the definitions of $Norm_{R,t}$ then the following simplification can be made. In many situations R will reduce to 0 – i.e. there is some point in the experimental space at which two peaks cannot be resolved. $R_{norm}$ is then reduced to the simple ratio $R/R_{max}$, and $Norm_{R,t}$ can be expressed as

$$Norm_{R,t} = R/R_{max} * (t_{max}-t)/(t_{max}-t_{min}).$$

$$= [R * (t_{max}-t)]/[R_{max} * (t_{max}-t_{min})] \tag{6}$$

Given that $t_{max}$, $R_{max}$ and $t_{min}$ are constants defined over the entire experimental space and do not vary for each point, then the term $R_{max} * (t_{max}-t_{min})$ can be eliminated as it is a common factor for every value of $Norm_{R,t}$, to give the following simplification:

$$Norm_{R,t} \propto R * (t_{max}-t) \tag{7}$$

The simplification in Equation (7) represents the core form of the $Norm_{R,t}$ function and there is no mathematical need to proceed through the normalization process. However, we prefer to normalize the data as it enables a quick determination of how close to the ideal best conditions the separation is (the ideal best conditions will have a value of 1). It is also straight forward to examine values of $t_{norm}$ and $R_{norm}$ to determine whether the criterion response is dominated by time or resolution.

Conditions with similar values of $Norm_{R,t}$ can be distinguished by their different values of $t_{norm}$ and $R_{norm}$ to allow further fine tuning in selecting the optimum. While this may dictate more user interaction than other CRFs, it is in line with the subjective nature of defining the optimal conditions.

In constructing this CRF, we became aware of a similar criterion introduced by Horvath and Lemander in which the resolution is divided by the time of the separation. In considering this application for our purposes, we used a slight modification in which the minimum resolution is divided by the migration time of the last fragment. However, upon closer examination of this CRF, it is apparent that the response is skewed towards quicker separations due to the hyperbolic nature of the $1/t$ function. This is represented in Figure 1, where low values of time receive a much higher weighting than higher values. For comparison, the time-weighted factor introduced in $Norm_R t$ depends on the factor ($t_{max} - t$) and is also shown in figure 1. In both cases, $t_{max}$ is defined as being 10. It is important to note that the y values of $1/t$ have been increased by a factor of 10 to aid in visualization, but will not effect the results as the aim is not to compare the magnitude of $1/t$ to $t_{max}$-t but the internal trend of each function. The function ($t_{max}$-t) is linear over the entire t values and as such there is no significant skew at shorter times, in contrast to the $1/t$ function. This has immense ramifications for optimization as the $1/t$ function will always predict more rapid separations to be more favorable. As such, we believe the product of R with ($t_{max}$-t) is a better approach for introducing a time component.

To verify the ability of $Norm_{R,T}$ to predict the optimum conditions, we have evaluated $Norm_{R,t}$, $R^*(t_{max}$-t), $R/t$ and R (the minimum resolution) using both the DNA migration data presented in the manuscript and an additional previously published dataset on the migration of metal complexes in MEKC [3].

Response surfaces for the metal complex MEKC data are shown in Figure 2 where it can be seen that all criteria select the same conditions as being the optimum. Minor differences can be seen throughout the remainder of the surface, but generally all the major features are apparent. The surfaces for $R^*(t_{max}$-t) and $Norm_{R,t}$ show the same features as expected from theory, with the only difference being the scale on the z axes. The similarity of all the surfaces support the use of $Norm_{R,t}$ as a criterion to find conditions under which satisfactory resolution in a reasonable time can be reached.

When applied to the DNA migration data collected for this manuscript, there are significant differences regarding the optimum conditions, displayed in Table 1. The conditions selected when optimizing using R are when the best resolution is obtained, hence high polymer concentrations and ionic strength is selected. The separation time at these conditions is nearly 700 seconds because there is no consideration of time when using this criterion. The conditions predicted by R/t provide separations in which the time is very rapid (<120 sec) due to the hyperbolic nature of 1/t and as such there is only marginal resolution between the 271 and 281 bp fragments. The best conditions selected where those by $Notm_{R,t}$ and $R^*(t_{max}-t)$. These conditions provided separations with sufficient resolution within a reasonable amount of time.

In conclusion, we have demonstrated that $Norm_{R,t}$ can be used to select conditions in which a compromise between the best resolution and the quickest time is reached. However, it is essential to recognize that it is truly a compromise. It does not necessarily select conditions under which peaks are baseline resolved, nor when the separation time is less than 5 minutes, or any other parameter that is absolute. It will select conditions under which 'nearly' the fastest separation is obtained and 'nearly' the best resolution. One of the downfalls of this criterion however is when the best resolution can only be obtained at the longest time. Nevertheless, we have demonstrated for two different systems that it can be used adequately as a tool in aiding selection of the optimum conditions.

# References

1.      Cela, R. *Quim.Anal.* **1999**, *18*, 49.

2.      Siouffi, A. M.; Phan-Tan-Luu, R. *J.Chromatogr.A* **2000**, *892*, 75-106.

3.      Breadmore, M. C.; Macka, M.; Haddad, P. R. *Anal.Chem.* **1999**, *71*, 1826-1833.

**Table 1: Optimum conditions selected with different criteria for the migration of DNA fragments through a sieving matrix for microchip CE.**

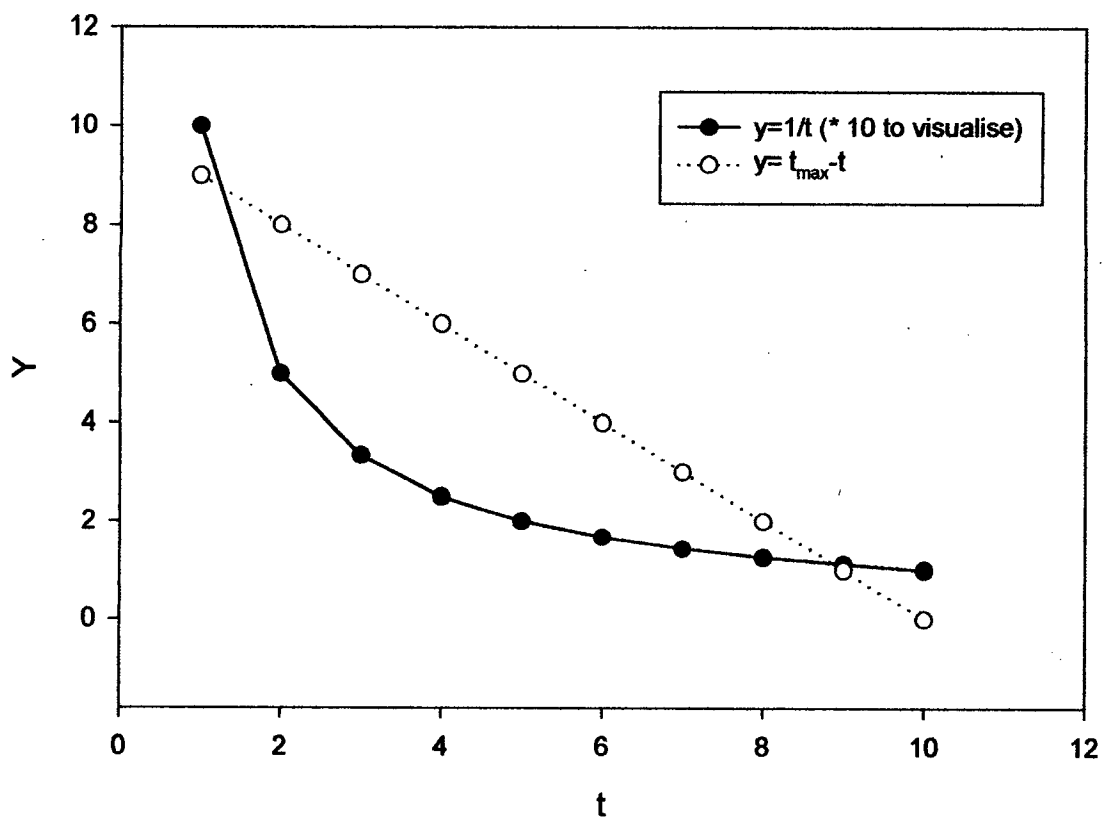| Criterion | V (V/cm) | % HPC (w/v) | [MES] (mM) |
|---|---|---|---|
| R | 295 | 5 | 105 |
| $Norm_{R,t}$ | 370 | 3.5 | 105 |
| $R*(t_{max}-t)$ | 370 | 3.5 | 105 |
| R/t | 520 | 5 | 30 |

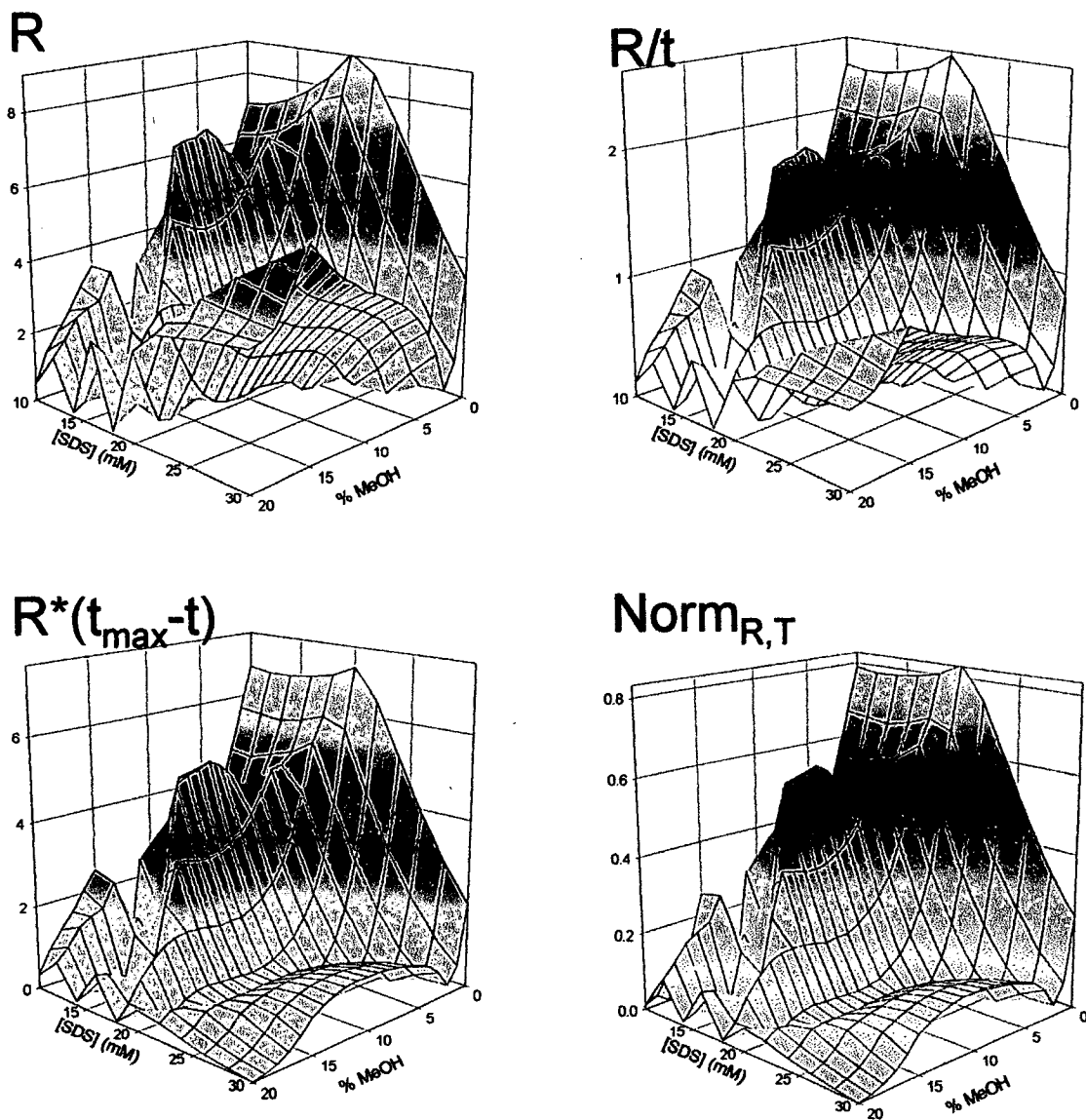**Figure 1: Variation of two possible functions for introducing a time component into the CRF.**

**Figure 2: Response surfaces for the optimization of the separation of 7 metal HEDTC complexes in MEKC using 4 different criterion. The optimum conditions predicted by all criterion are near 0% MeOH and 20 m M SDS.**