

Scoring function

To enhance the significance of the spot comparing feature we included a simple scoring function. The computed score rates the difference among two spots. It enables the user to efficiently compare and concentrate on specific protein spots, which exhibit significant differential expression within the performed proteomic experiments. Furthermore, it offers the possibility to compare all the proteins spot-by-spot.

The computation of the score was based on the two available (from PDQuest) spot-specific values, Normalized Quantity (NQ) and Peak Value (PV). These values enabled us to characterise the correlative protein amount of each spot. Thus, the score is based on the difference of Normalized Quantity and Peak Value of two spots. But due to basic limitations of the 2D gel electrophoresis these values are not reliable when applied to the faint spots. This limitation made it necessary to include a rating (in terms of scoring function) of the basic intensity and size of the spot. Using this, we adjusted the score accordingly. Furthermore, we normalised NQ, PV and the computed differences to values between 1 and 100 to receive a better comparability of the scores.

To compute the scoring function, first we calculated the highest measured Peak Value ($maxPVGel$) and the highest measured Normalized Quantity ($maxNQGel$) of the two gels G_1 and G_2 which we want to compare. These values enabled us to rate the relative size and intensity of a spot. They would be computed as follows:

$$maxPVGel_{G_1, G_2} = \max \left\{ \bigcup_{j=1}^n PV_{G_1(j)}, PV_{G_2(j)} \right\} \quad (1)$$

In this formula the variable n is related to the number of spots on a gel and the variable j is used to represent the corresponding spots on both gels. Below, only the computations of the PV-related values are shown to ensure a better overview. The computation of the NQ-related values was performed in an equivalent way.

Additionally, while normalisation we calculated the maximal PV-difference ($max\Delta PVGel$) and the maximal NQ-difference ($max\Delta NQGel$) of two corresponding spots on the gels which the user selects to compare. The related equation is showed here:

$$max\Delta PVGel_{G_1, G_2} = \max_{1 \leq j \leq n} \{ | PV_{G_1(j)} - PV_{G_2(j)} | \} \quad (2)$$

These gel and spot spanning values were combined with the specific values of each concrete spot comparison. Therefore it is necessary to compute the maximal PV-value ($maxPVSpot$) or the maximal NQ-value ($maxNQSpot$) of the viewed spot j respectively as follows:

$$maxPVSpot_{G_1(j), G_2(j)} = \max \{ PV_{G_1(j)}, PV_{G_2(j)} \} \quad \text{with } 1 \leq j \leq n \quad (3)$$

As discussed before, the differences between the measured values of two faint spots are not reliable. Hence we used the PV- and NQ-values to compute the score respectively. For computing the score where only one of the two spots of the comparison is faint are expected to have low PV- and NQ-values which are of highly significance and this type of

protein expression changes are not due to limitations of the 2D gel electrophoresis. These types of expression changes are of strong indication for significant changes in the proteome. Designating this, the score should represent these expression changes in a suitable way.

By using the maximal PV- and NQ-value of a spot comparison we could accomplish the requirement reliably. For example, a spot which is faint due to gel electrophoresis limitations would have similar low PV-values and NQ-values at both gels. Using the maximal PV- and NQ-value of these spots would certainly have a low score in our computation. In the case if a spot is faint on one gel but thick and dark on the other gel, the maximal PV- and NQ-value of these spots would certainly have a high score, which ignores the PV- and NQ-value of the faint spot. As a result, we obtain a score which are highly reliable based on the PV- and NQ-values of the thick and dark spots which is not influenced by the faint spots.

The following computation of the PV-difference ($\Delta PVSpot$) and the NQ-difference ($\Delta NQSpot$) of a pair of spots represents the base of the scoring function:

$$\Delta PVSpot_{G_1(j),G_2(j)} = | PV_{G_1(j)} - PV_{G_2(j)} | \quad \text{with } 1 \leq j \leq n \quad (4)$$

Utilizing the aforementioned derived values enabled us to compute intermediate scores for the concrete comparison of two spots. $PVNQScore$ rates the relative size and intensity, respectively. It is computed as follows:

$$PVNQScore_{G_1(j),G_2(j)} = \left(\frac{\max PVSpot_{G_1(j),G_2(j)}}{\max PVGel_{G_1,G_2}} + \frac{\max NQSpot_{G_1(j),G_2(j)}}{\max NQGel_{G_1,G_2}} \right) \cdot 50 \quad (5)$$

All scores was normalised to values between 1 and 100. PV- and NQ related values of a spot were treated as equal characterisations for the amount of protein expressed. The computation of the relative PV-difference score ($\Delta PVScore$) is shown in the following equation:

$$\Delta PVScore_{G_1(j),G_2(j)} = \frac{\Delta PVSpot_{G_1(j),G_2(j)}}{\max \Delta PVGel_{G_1,G_2}} \cdot 100 \quad (6)$$

The relative NQ-difference ($\Delta NQScore$) was determine in an equivalent way:

$$\Delta NQScore_{G_1(j),G_2(j)} = \frac{\Delta NQSpot_{G_1(j),G_2(j)}}{\max \Delta NQGel_{G_1,G_2}} \cdot 100 \quad (7)$$

Based on several empirical tests we concluded that $\Delta PVScore$ as well as $\Delta NQScore$ provides good indications to rate the differences between two spots. Mostly the obtained high score was related to significantly differential expressed protein spots. It was necessary to combine both the scores which would enhance the analysis of monitoring the differential expressed protein spots. This combination was necessary because $\Delta PVScore$ and $\Delta NQScore$ does not correlate with each other in every case.

The overall score ($Score$) to rate the difference of spot j on gel G_1 in comparison with its corresponding spot j on gel G_2 results from the average of the aforementioned intermediate scores. It was computed as follows:

$$Score_{G_1(j),G_2(j)} = \frac{PVNQScore_{G_1(j),G_2(j)} + \Delta PVScore_{G_1(j),G_2(j)} + \Delta NQScore_{G_1(j),G_2(j)}}{3} \quad (8)$$

By combining the spot-specific values (PV and NQ) along with the PVNQ intermediate score with the weighting ratio of 2:1 we were able to ensure two important things, on the one hand that the computation of the overall score is based on the differential expression of the protein spots. On the other hand, the score ensures that the faint spots attains lower score than the thick and dark spots (higher scores) harbouring similar relative differences between them.