

## Supplementary Note: Equations to calculate Flash Entropy Similarity in MS/MS queries

As defined before<sup>1</sup>, we calculate the unweighted entropy similarity as:

$$1 - \frac{2 \times S_{AB} - S_A - S_B}{\ln 4} \quad (1)$$

$$S_A = - \sum_i I_{A,i} \ln I_{A,i} \quad (2)$$

$$S_B = - \sum_j I_{B,j} \ln I_{B,j} \quad (3)$$

Here the  $S_{AB}$  is defined as the spectral entropy of a 1:1 mixed spectrum for spectra A and B. Therefore:

$$S_{AB} = - \sum_{i,j} \begin{cases} \frac{1}{2} I_{A,i} \ln \left( \frac{1}{2} I_{A,i} \right), \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ \frac{1}{2} I_{B,j} \ln \left( \frac{1}{2} I_{B,j} \right), \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ \left( \frac{1}{2} (I_{A,i} + I_{B,j}) \right) \ln \left( \frac{1}{2} (I_{A,i} + I_{B,j}) \right), \text{ where } m_{z_{A,i}} = m_{z_{B,j}} \end{cases} \quad (4)$$

After simplifying the formula (4), we obtain:

$$S_{AB} = - \frac{1}{2} \sum_{i,j} \begin{cases} I_{A,i} \ln I_{A,i} - (\ln 2) I_{A,i}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ I_{B,j} \ln I_{B,j} - (\ln 2) I_{B,j}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ (I_{A,i} + I_{B,j}) \ln(I_{A,i} + I_{B,j}) - (\ln 2)(I_{A,i} + I_{B,j}), \text{ where } m_{z_{A,i}} = m_{z_{B,j}} \end{cases} \quad (5)$$

Where we get:

$$2 \times S_{AB} = \ln 2 \sum_i I_{A,i} + \ln 2 \sum_j I_{B,j} - \sum_{i,j} \begin{cases} I_{A,i} \ln I_{A,i}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ I_{B,j} \ln I_{B,j}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ (I_{A,i} + I_{B,j}) \ln(I_{A,i} + I_{B,j}), \text{ where } m_{z_{A,i}} = m_{z_{B,j}} \end{cases} \quad (6)$$

We substitute formulas (2), (3), and (6) into formula (1), to obtain the unweighted entropy similarity as:

$$1 - \frac{1}{\ln 4} \left( \ln 2 \sum_i I_{A,i} + \ln 2 \sum_j I_{B,j} - \sum_{i,j} \begin{cases} I_{A,i} \ln I_{A,i}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ I_{B,j} \ln I_{B,j}, \text{ where } m_{z_{A,i}} \neq m_{z_{B,j}} \\ (I_{A,i} + I_{B,j}) \ln(I_{A,i} + I_{B,j}), \text{ where } m_{z_{A,i}} = m_{z_{B,j}} \end{cases} - S_A - S_B \right) \quad (7)$$

This formula equals:

$$1 - \frac{1}{\ln 4} \left( \ln 2 \sum_i I_{A,i} + \ln 2 \sum_j I_{B,j} \right) + Q \quad (8)$$

Where:

$$Q = \frac{1}{\ln 4} \left( \sum_{i,j} \begin{cases} I_{A,i} \ln I_{A,i}, \text{ where } m_{Z_{A,i}} \neq m_{Z_{B,i}} \\ I_{B,j} \ln I_{B,j}, \text{ where } m_{Z_{A,i}} \neq m_{Z_{B,i}} \\ (I_{A,i} + I_{B,j}) \ln(I_{A,i} + I_{B,j}), \text{ where } m_{Z_{A,i}} = m_{Z_{B,j}} \end{cases} - \sum_i I_{A,i} \ln I_{A,i} - \sum_j I_{B,j} \ln I_{B,j} \right) \quad (9)$$

The formula (8) equals:

$$1 - \frac{1}{2} \left( \sum_i I_{A,i} + \sum_j I_{B,j} \right) + Q \quad (9)$$

When the total intensity of spectrum A and B are normalized into 1, we will have unweighted entropy similarity equals  $Q$ . Therefore, we have unweighted entropy similarity equals:

$$\frac{1}{\ln 4} \left( \sum_{i,j} (I_{A,i} + I_{B,j}) \ln(I_{A,i} + I_{B,j}) - \sum_i I_{A,i} \ln I_{A,i} - \sum_j I_{B,j} \ln I_{B,j} \right), \text{ for all } m_{Z_{A,i}} = m_{Z_{B,j}} \quad (10)$$

This formula can be simplified to:

$$\frac{1}{2} \sum_{i,j} \left( (I_{A,i} + I_{B,j}) \log_2(I_{A,i} + I_{B,j}) - I_{A,i} \log_2 I_{A,i} - I_{B,j} \log_2 I_{B,j} \right), \text{ for all } m_{Z_{A,i}} = m_{Z_{B,j}} \quad (11)$$

If we define the function:

$$f(x) = x \log_2 x \quad (12)$$

we obtain the unweighted entropy similarity as:

$$\frac{1}{2} \sum_{i,j} \left( f(I_{A,i} + I_{B,j}) - f(I_{A,i}) - f(I_{B,j}) \right) \quad (13)$$

The formula (13) equals:

$$\sum_{i,j} \left( f\left(\frac{1}{2}I_{A,i} + \frac{1}{2}I_{B,j}\right) - f\left(\frac{1}{2}I_{A,i}\right) - f\left(\frac{1}{2}I_{B,j}\right) \right) \quad (14)$$

When normalizing the total spectral intensity to 0.5, we get an unweighted entropy similarity as:

$$\sum_{i,j} \left( f(I_{A,i} + I_{B,j}) - f(I_{A,i}) - f(I_{B,j}) \right) \quad (15)$$

As previously defined<sup>1</sup>, we apply the formula (16) to obtain a weighted entropy intensity:

$$I' = \begin{cases} I & (S \geq 3) \\ I^w, w = 0.25 + S * 0.25 & (S < 3) \end{cases} \quad (16)$$

We calculate the formula (15) to get the entropy similarity as published before<sup>1</sup>.

#### Reference:

1. Li, Y. et al. Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. *Nature methods* **18**, 1524-1531 (2021).