**Triple Motif Co-occurrence Score**

Our objective with this triple motif co-occurrence analysis was to find out to what extent any three motifs (from the discovered 129 peptide motifs) co-occur in droplet-promoting regions (DPRs) of Phase Separating Proteins (PhSePs). As such we created a scoring system using vectors meant to reflect both the extension and symmetry of co-occurrence, detailed in Fig. 1.

A diagram of a diagram of different colors

Description automatically generated

Figure - Conceptual approach to motif co-occurrence analysis.

To facilitate the explanation of our method, we use a practical example of the motif trio QQQQ, GRGG and GGYGG, as detailed below in row 10.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Motif List** | **Score** | **Possible Peptides** |
| **0** | ['QQQQ', 'GRGG', 'GGRG'] | 2.48 | ('QQQQGGRGG', 9) |
| **1** | ['QQQQ', 'GRGG', 'QQQQQ'] | 7.43 | ('QQQQQGRGG', 9) |
| **2** | ['QQQQ', 'GRGG', 'RGGG'] | 4.96 | ('QQQQGRGGG', 9) |
| **3** | ['QQQQ', 'GRGG', 'GGGGG'] | 4.96 | ('QQQQGRGGGGG', 11) |
| **4** | ['QQQQ', 'GRGG', 'GGGR'] | 2.48 | ('QQQQGRGGGR', 10) |
| **5** | ['QQQQ', 'GRGG', 'RGGR'] | 4.96 | ('QQQQGRGGR', 9) |
| **6** | ['QQQQ', 'GRGG', 'GGRGG'] | 2.48 | ('QQQQGGRGG', 9) |
| **7** | ['QQQQ', 'GRGG', 'QQQQQQ'] | 14.87 | ('QQQQQQGRGG', 10) |
| **8** | ['QQQQ', 'GRGG', 'GGGG'] | 12.39 | ('QQQQGRGGGG', 10) |
| **9** | ['QQQQ', 'GRGG', 'QQQP'] | 7.43 | ('QQQQPGRGG', 9) |
| **10** | ['QQQQ', 'GRGG', 'GGYGG'] | 12.39 | ('QQQQGRGGYGG', 11) |

We begin by computing in how many DPR sequences the motif QQQQ, co-occurs with the remaining two motifs (GRGG and GGYGG).

Each count can be seen as a Vector Tuple. For interactions between QQQQ,GGYGG and QQQQ,GRGG:

**GGYGG**

***Score Vector***

***Vector 2 Length***

***Vector 1 Length***

**GRGG**

**QQQQ**

The Sum of the two vectors is the **Score Vector.**

For **QQQQ:**

* **GRGG Interactions = 0**
* **GGYGG interactions = 0**
* **Score Vector Slope = 0**
* **Score Vector Length = 0**

For **GRGG:**

* **QQQQ Interactions = 0**
* **GGYGG interactions = 5**
* **Score Vector Slope = 0**
* **Score Vector Length = 5**

For **GGYGG:**

* **GRGG Interactions = 5**
* **QQQQ interactions = 0**
* **Score Vector Slope = 0**
* **Score Vector Length = 5**

To this data we apply the following equation, designated as **Vector Length Score:**

This Equation measures the extension of motif co-occurrences from 0 to 100, where 100 represents the maximum value of co-occurrence between motifs, and ***MAX Vector Length*** is the maximum Value of the **Score Vector** obtainedin our DPR dataset, which was **13.45**.

The following equation, designated as **Vector Slope Score**, measures the symmetry of motif pair co-occurrence:

In this calculation, if returns a division by 0 Error the **Vector Slope Score** is given the value of 0.

The following equation, **Vector Score**, finds the adjusted score for the sum of **Vector Length Score** and **Vector Slope Score** using **C1** and **C2** linear adjustment values.

For our analysis we used **C1** = **C2** = 0,5.

Finaly, we do the same operation for the two remaining motifs, with the list of the respective three **s** the final Equation is calculated.

For **QQQQ, GRGG, GGYGG:**

For our example comprising the motifs QQQQ, GRGG, GGYGG the value of the **Final** **Score** was 12.39, which indicates that these motifs have a low and unbalanced co-occurrence.