Given the data set, for each disease $d \in \{1, \ldots, 19\}$, we first compute the average value of protein j at region i, and we denote it by $m_{i,j}(d)$. Let $M(d) = \{m_{i,j}(d)\}_{i \in \{1,\ldots,24\}, j \in \{1,2,3,4\}}$ be the 24×4 matrix summarising all mean values related to disease d. Once we have the matrix for each disease, we compute the distance between disease d and disease d' by summing up the absolute distance for each entry in matrices $M(d)$ and $M(d')$, that is, we compute the L_0 norm between $M(d)$ and $M(d')$:

$$\delta(d, d') := L_0(M(d) - M(d')) = \frac{1}{24 \cdot 4} \left(\sum_{i=1}^{24} \sum_{j=1}^{4} (\texttt{abs}(m_{i,j}(d) - m_{i,j}(d')) \right).$$

The used distance definition is appropriate because we wanted the distance to reflect the total discrepancy between values at each region-protein combination. The approach can be generalized to other matrix norms.