

Calculate the similarity score between a 3-residues group (4-residues group) in test data and all the Zn3 templates (Zn4 templates) in training data.

For a 3-residues group [A, B, C] and a Zn3 template [D, E, F], the similarity score between [A, B, C] and [D, E, F] is calculated as follows:

- 1) Compute the permutations of (A, B, C), for a 3-residues group there are 6 permutations (i.e. [A, B, C], [A, C, B], [B, A, C], [B, C, A], [C, A, B], [C, B, A]).
- 2) For each permutations:

$$Score([A, B, C], [D, E, F]) = (S(A, D) + S(B, E) + S(C, F)) / 3 \quad (1)$$

$$S(a, b) = \sum_{i=1}^{15} \left(\sum_{j=1}^{20} (a_{ij} \log(\frac{b_{ij}}{f_j}) + b_{ij} \log(\frac{a_{ij}}{f_j})) \right) \quad (2)$$

$$a_{i,j} = e^{M_{i,j} \lambda_u} p_j \quad (3)$$

$$SimilarityScore([A, B, C], [D, E, F]) = \max \begin{cases} Score([A, B, C], [D, E, F]) \\ Score([A, C, B], [D, E, F]) \\ Score([B, A, C], [D, E, F]) \\ Score([B, C, A], [D, E, F]) \\ Score([C, A, B], [D, E, F]) \\ Score([C, B, A], [D, E, F]) \end{cases} \quad (4)$$

where a and b were profiles of two fragments in a sub group, M was the substitution of 20 amino acids at a specific position, λ_u was the Standard Ungapped Lambda value of a PSSM file and f was the background frequency for 20 amino acids. The form of the score function is the same as the PICASSO3 score mentioned in (Mittelman, et al., 2003; Shu, et al., 2008).