Definitions for Evaluation Metric

To formalize the evaluation metric, we introduce the following definitions: A classifier is a function that for a given doublet x and interaction class *Z* (base-base, stacking, base-phosphate, or base-ribose) returns an annotation from class Z, that is: $C_Z(x)$: doublet ↦ interaction annotation or UNDETECTED. The $ExpResult_Z(x)$ function returns the expected interaction type for doublets from the testing set, or returns UNDETECTED. For an interaction class Z, a confusion table can be defined as:

- true positives (TP_Z) are the doublets x such that:
  $C_Z(x) \neq$ UNDETECTED and $C_Z(x) = ExpResult_Z(x)$
- true negatives (TN_Z) are the doublets x such that:
  $C_Z(x) =$ UNDETECTED and $C_Z(x) = ExpResult_Z(x)$
- false positives (FP_Z) are the doublets x such that:
  $C_Z(x) \neq$ UNDETECTED and $C_Z(x) \neq ExpResult_Z(x)$
- false negatives (FN_Z) are the doublets x such that:
  $C_Z(x) =$ UNDETECTED and $C_Z(x) \neq ExpResult_Z(x)$

Formulas used to calculate doublet score using interatomic distances

Based on the analysis of the data set, the ClaRNA stores for each contact type, the distance matrix DM of n*m interatomic distance ranges computed from the doublets belonging to the interaction class.
Each cell of the matrix stores a triplet of values ($min_{i,j}$, $avg_{i,j}$, $max_{i,j}$) ($1 \leq i \leq n$, $1 \leq j \leq m$):

- the $min_{i,j}$ is the minimal observed distance between the i-th atom of the first residue and the j-th atom of the second residue,
- the $avg_{i,j}$, is the average observed distance between the i-th atom of the first residue and the j-th atom of the second residue,
- the $max_{i,j}$ is the maximal observed distance between the i-th atom of the first residue and the j-th atom of the second residue.

Finally for each class we choose scaling constant ϕ.

For analyzed doublet D with observed interatomic distances $d_{i,j}$ ($1 \leq i \leq n$, $1 \leq j \leq m$) the compatibility score is defined as:

$$\phi \cdot \frac{1}{n \cdot m} \sum_{i,j} e^{1-\max\left\{1,\left(\frac{d_{i,j}-\mathrm{avg}_{i,j}}{d_{i,j} \geq \mathrm{avg}_{i,j} \ ? \ \mathrm{max}_{i,j}-\mathrm{avg}_{i,j} \ : \ \mathrm{avg}_{i,j}-\mathrm{min}_{i,j}}\right)^2\right\}}$$