

# Constraints

## 1. Chiral Indices

The total and local n-linear indices, as defined in their “standard” form in the **QuBiLS-MIDAS** module do not codify information about the chiral properties of the molecular structure. Thus, a *correction factor* is introduced in the molecular vectors  $\bar{x}$ ,  $\bar{y}$ ,  $\bar{z}$  and  $\bar{w}$ , respectively, in order to take into account this criterion. In this way, chirality-based molecular vectors are computed ( $\bar{x}^*$ ,  $\bar{y}^*$ ,  $\bar{z}^*$  and  $\bar{w}^*$ ), where each coefficient is equal to the sum of the considered atomic property and the *correction factor*. This idea in several works reported in the literature has been used, but only employing opposed integer numbers as *correction factor*, *i.e.*: 1 when an atom is labeled as *rectus* (R), -1 when an atom is labeled as *sinister* (S) and 0 when an atom does not present a specific chiral environment according to the Cahn-Ingold-Prelog rules. However, inspired on the No Free Lunch Theorem that could be interpreted as no single *correction factor* (*i.e.* 1 and -1) to perform chirality-based studies yields superior results than others when its result is averaged over all possible chemical datasets, then it may be stated that the use of several integer and rational numbers as *correction factor* could contribute to obtain better results in cheminformatics applications (*e.g.* QSAR studies).

Therefore, values in the range  $[-3, 3]$  with step equal to 0.25 are accounted for. This range of values was chosen with the purpose of considering a suitable spectra of numbers as *correction factor*. It is important to highlight that those atoms labeled as R or S can take any value, which is rather than to the reported where 1 and -1 constitute the only values for atoms classified as R and S, respectively. In this way, the *correction factor* for atoms labeled as R or S can have assigned negative and positive values (*e.g.* -3 for R atom and 1 for S atom, or vice versa), opposed values (*e.g.* -3 for R atom and 3 for S atom, or vice versa), positive values (*e.g.* 3 for R atom and 1 for S atom) or negative values (*e.g.* -3 for R atom and -1 for S atom). Lastly, it is valid to clarify that this *correction factor* has essentially a mathematical means and must not be cause of any misinterpretation.

## 2. N-tuples

For **Duplex option (2)**, the index calculations are performed over vertex pairs  $i$  and  $j$ . Here, the  $k^{\text{th}}$  *spatial (dis-)similarity matrix* of the geometric molecular structure,  $\mathbb{G}^k$ , is used like matrix forms. For duplex relation, *bilinear*, *linear* and *quadratic indices* can be obtained. However, a generalization of these indices can be obtained using *n-linear maps*. That is to say, for **3-uples option**, the index calculations are performed over vertices  $i$ ,  $j$  and  $k$  using the  $k^{\text{th}}$  *three-tuples spatial (dis-)similarity matrix*,  $\mathbb{G}\mathbb{T}^k$ ; while for **4-uples option** the index calculations are performed over vertices  $i$ ,  $j$ ,  $k$  and  $l$  employing the  $k^{\text{th}}$  *four-tuples spatial (dis-)similarity matrix*,  $\mathbb{G}\mathbb{Q}^k$ . For ternary and quaternary matrices, previously mentioned, several matrices types are used, such as:

- **3C**: All the elements of the 3-tuples matrix, with coordinates  $(i, j, k)$ , are considered.
- **3nC**: Only elements of the triple matrix that satisfy the condition that all the 3 coordinates are different are selected.
- **4C**: All the elements of the quadruple matrix, with coordinates  $(i, j, k, z)$ , are considered.
- **4nC**: Only elements of the quadruple matrix that satisfy the condition that all the 4 coordinates are different are considered