Extended examples and description of various jCompoundMapper fingerprints in string format

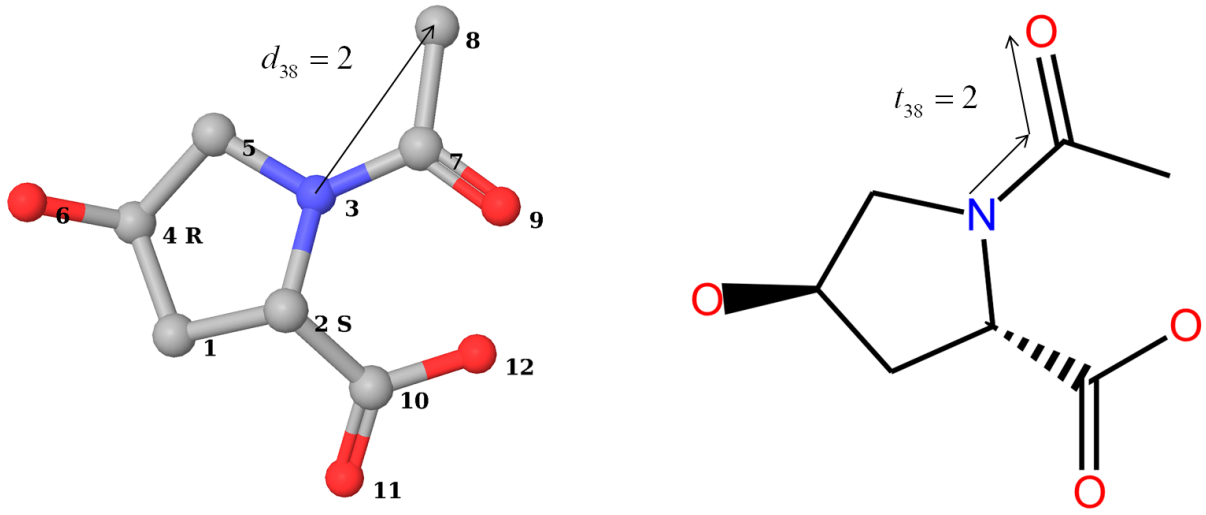


Figure 1: The geometry and topology of Oxaceprol. Pharmacophore types shown in the 3D structure are 1=[L], 3=[A - [#7H0]], 6=[D - [OH], A - [O]],8=[L], 9=[A - [O]],10=[N], 11=[A - [O]], 12=[D - [OH], A - [O]]. The geometry and topology of this compound is the basis for the exemplary fingerprints. The R and S tags in the three-dimensional representation refer to R/S stereochemistry.

For this extended explanation of the fingerprints we chose a search depth of 3 and the atom typing scheme was set to element plus neighbor. The representation shares the following properties:

- If pattern can be generated in two ways (chemical graphs are undirected), the pattern with the smaller hash code is stored.
- The patterns are stored as objects internally, therefore the labels are generated using a natural reading direction
- The count is incremented in the case of multiple occurrences. However, for hashed fingerprints, the features are regarded Boolean (so that a collision does not generate ambiguous counts).
- The topological search depth reduces either the search depth or the maximum allowed distance between two vertices in an extracted pattern
- An atom described in a feature of a fingerprint must be included only once per feature
- The atom types and pharmacophore types are generated exclusively by the CDK. Therefore some fingerprints may differ from their original implementation.

A main goal of fingerprints is to enable a quantitative comparison between chemical compounds. Therefore, the set of features must be comparable, which means the features must be described independently from the compounds they are derived of. In many cases, there will be several solutions to map a pattern back to the original structure.

The following example describes the fingerprints by string patterns. However, there are several alternative output formats possible:

1. LIBSVM feature format and similarity matrix format (kernel matrix) (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)

2. LIBSVM ARFF WEKA format (http://www.cs.waikato.ac.nz/ml/weka/)

Depth-First Search

The fingerprint consists of all possible unique paths that can be generated by visiting each atom and computing all paths up to length d. For instance „C.2-C.3-N.3-C.3=O.1“ means that a pattern can be generated which consists of a „carbon with two heavy atom neighbors, this atom is connected by a single bond to a carbon with three heavy atoms, this atom is connected by a single bond to a nitrogen with three heavy atoms, this atom is connected to a carbon atom with three neighbors, this atom is connected to an oxygen by a double bond“. In our example this can be, for example, the sequence 1,2,3,7,9.

```
C.2-C.3-N.3-C.3=O.1 C.2-C.3-C.2-C.3-C.3 C.2-C.3-C.2-C.3-N.3 C.3-C.2-C.3 C.2-C.3-C.3-O.1 C.1
C.2 C.3 C.3-N.3-C.3 C.2-C.3-N.3-C.2 C.2-C.3-N.3-C.3 C.1-C.3 C.3-N.3 C.3-N.3-C.3=O.1 C.1-C.3-
N.3-C.2 C.1-C.3-N.3-C.3 C.3-C.2-C.3-O.1 C.2-C.3-C.2 C.2-C.3-C.3 C.2-N.3-C.3 C.3-N.3-C.3-
C.3=O.1 C.3-C.2-C.3-C.2-N.3 C.2-C.3-N.3-C.2-C.3 C.3-C.3-O.1 C.2-C.3-C.2-C.3 C.3-N.3-C.2-C.3-
O.1 N.3 C.3-C.2-N.3-C.3=O.1 C.2-C.3 C.2-N.3 C.2-N.3-C.3-C.2-C.3 C.2-N.3-C.3-C.3-O.1 C.3-C.2-
C.3-C.3 C.3-C.2-C.3-N.3 N.3-C.3=O.1 C.2-C.3-N.3 C.1-C.3-N.3-C.2-C.3 C.1-C.3-N.3-C.3-C.2 C.1-
C.3-N.3-C.3-C.3 C.3-N.3-C.3-C.3-O.1 O.1-C.3=O.1 C.3-C.2-C.3-C.3-O.1 C.3-C.2-N.3 C.3-C.3-N.3
C.3-C.3=O.1 O.1 C.3-C.2-C.3-N.3-C.3 C.2-C.3-C.2-N.3-C.3 C.3-C.3-N.3-C.3=O.1 C.3-C.3-N.3-C.3
C.3-O.1 C.2-C.3-O.1 N.3-C.3-C.3-O.1 C.1-C.3=O.1 C.3-C.2-C.3-C.3=O.1 C.2-C.3-C.2-N.3 C.2-C.3-
C.3=O.1 C.2-N.3-C.3=O.1 C.1-C.3-N.3 C.3-C.3-C.2-C.3-O.1 C.3-C.2-N.3-C.3-C.3 C.2-N.3-C.3-
C.3=O.1 C.2-N.3-C.3-C.3 N.3-C.3-C.3=O.1 N.3-C.2-C.3-O.1 C.3-C.3 C.3=O.1 C.3-C.2-N.3-C.3 N.3-
C.3-C.2-C.3-O.1
```

All-Shortest Path

This fingerprint is similar to the depth-first search fingerprint with the constraint that only the shortest path between the first and the last atom is included in a pattern. The first example equals the first of the depth-first search fingerprint. Let us have a look at the second path computed by the depth-first search „C.2-C.3-C.2-C.3-C.3“

```
C.2-C.3-N.3-C.3=O.1 C.2-C.3-C.3-O.1 C.1-C.3 C.3-C.3 C.3-N.3 C.3-N.3-C.3=O.1 C.1-C.3-N.3-C.2
C.1-C.3-N.3-C.3 C.3-C.2-C.3-O.1 C.2-C.3-C.2 C.2-C.3-C.3 C.2-C.3-N.3 C.2-C.3-O.1 C.2-N.3-
C.3=O.1 C.3-C.2-N.3-C.3 C.3-N.3-C.3-C.3=O.1 N.3-C.3-C.3-O.1 C.1-C.3-N.3 C.3-C.2-C.3-C.3-O.1
C.3-C.2-C.3 C.3-C.3=O.1 C.2-C.3-C.3=O.1 C.3-N.3-C.2-C.3-O.1 C.3-C.3-N.3-C.3=O.1 C.3-C.3-N.3-
C.3 C.3-C.3-C.2-C.3-O.1 C.3-N.3-C.3 C.2-C.3-N.3-C.3 C.3-O.1 C.2-N.3-C.3-C.3-O.1 C.3-C.2-C.3-
C.3 C.2-N.3-C.3-C.3 C.1-C.3-N.3-C.2-C.3 C.2-N.3-C.3 C.3-N.3-C.3-C.3-O.1 O.1-C.3=O.1 C.1-
C.3=O.1 C.3-C.2-C.3-C.3=O.1 C.3-C.2-N.3 C.3-C.3-N.3 C.3-C.3-O.1 C.3-C.2-N.3-C.3=O.1 C.2-C.3
C.2-N.3 C.2-N.3-C.3-C.3=O.1 N.3-C.3=O.1 C.1-C.3-N.3-C.3-C.2 C.1-C.3-N.3-C.3-C.3 N.3-C.3-
C.3=O.1 N.3-C.2-C.3-O.1 C.3=O.1
```

Atom Pairs (Topological)

The patterns denote the distance between two atom types. For example, the feature „C.3-2-C.3“ describes two carbons connected to three heavy atoms which has the shortest-path distance 2 to another carbon with three neighboring heavy atoms. Because the patterns are symmetric, only the upper half of the distance matrix is considered.

```
C.3-1-C.3 C.3-2-C.2 C.3-2-C.3 C.3-3-C.1 C.3-3-C.2 C.3-3-C.3 C.3-4-C.1 N.3-1-C.2 N.3-1-C.3 N.3-
2-C.2 N.3-2-C.3
```

Atom Triplets (Topological)

The three-point patterns are extracted combinatorially. The patterns can be interpreted like the two-point patterns with an additional atom. The last number contained in a feature is the topological distance back to the first atom.

O.1-5-C.1-4-C.2-2 N.3-1-C.2-3-O.1-2 C.3-2-N.3-1-C.2-1 C.3-2-N.3-1-C.3-1 C.3-2-N.3-1-C.3-3 C.3-2-C.3-3-C.1-4 C.3-2-C.3-3-C.3-1 C.3-2-C.2-1-N.3-1 N.3-2-C.3-1-C.2-1 N.3-2-C.3-1-C.2-2 C.3-1-C.3-3-C.1-4 N.3-2-C.1-2-O.1-2 O.1-5-O.1-5-C.1-2 O.1-5-O.1-5-C.1-5 O.1-5-O.1-5-O.1-2 O.1-2-C.1-4-C.3-4 O.1-4-C.3-4-O.1-2 O.1-4-C.3-4-O.1-5 C.3-3-O.1-5-O.1-2 O.1-3-C.2-2-C.3-3 C.3-2-O.1-5-O.1-3 C.3-1-N.3-2-O.1-1 C.3-1-N.3-2-O.1-3 O.1-2-N.3-2-C.1-2 C.1-2-N.3-1-C.2-3 C.1-2-N.3-1-C.3-1 C.1-2-N.3-1-C.3-3 C.3-1-O.1-5-C.1-4 C.2-1-C.3-1-N.3-2 C.2-4-O.1-3-C.2-2 C.3-3-O.1-1-C.3-2 O.1-4-C.2-3-C.1-5 C.2-3-C.3-2-C.2-2 C.2-3-C.3-2-C.3-1 C.2-3-C.3-2-N.3-1 C.1-4-C.2-2-N.3-2 O.1-3-C.2-3-O.1-2 N.3-2-C.3-3-C.2-1 O.1-5-O.1-2-C.3-3 N.3-2-O.1-4-C.2-2 C.3-4-C.1-2-N.3-2 O.1-5-C.1-2-N.3-3 O.1-4-C.2-4-C.1-2 C.3-2-C.3-1-C.2-1 C.3-2-C.3-1-C.2-2 C.3-2-C.3-1-C.2-3 O.1-3-C.3-1-C.3-4 C.3-2-O.1-2-O.1-2 C.2-2-C.3-3-C.1-3 C.2-2-C.3-3-O.1-2 N.3-2-C.2-2-C.3-2 N.3-2-C.2-2-O.1-3 C.3-1-O.1-2-N.3-1 C.3-1-C.1-2-N.3-1 O.1-5-C.1-3-C.3-2 N.3-2-C.1-5-O.1-3 O.1-1-C.3-1-C.2-2 O.1-4-C.3-2-C.3-2 O.1-4-C.3-2-C.3-3 N.3-1-C.2-2-O.1-3 C.3-3-O.1-3-N.3-1 C.3-3-C.1-3-C.2-2 O.1-3-C.3-2-C.2-2 O.1-3-C.3-2-C.2-3 O.1-3-C.3-2-C.3-1 O.1-3-C.3-2-C.3-4 N.3-3-O.1-1-C.3-2 C.2-2-C.2-2-C.3-1 C.2-2-C.2-2-C.3-3 C.3-1-C.1-3-C.3-2 C.2-4-O.1-1-C.3-3 O.1-5-C.1-4-C.2-3 C.1-4-C.3-2-C.2-4 C.3-2-N.3-1-C.2-3 C.3-2-C.3-3-C.1-1 C.3-2-C.3-3-C.3-2 C.3-2-O.1-4-C.2-2 O.1-5-O.1-5-O.1-5 C.3-3-O.1-5-O.1-3 C.3-3-C.1-5-O.1-2 C.3-2-C.3-4-C.1-3 N.3-3-O.1-3-C.2-2 C.1-3-C.2-1-N.3-2 C.3-4-O.1-1-C.3-3 C.2-1-C.3-1-C.3-2 C.1-2-O.1-4-C.2-4 O.1-2-C.2-3-C.3-4 C.3-4-C.1-2-O.1-4 O.1-4-C.3-1-N.3-3 C.2-2-C.3-3-O.1-3 C.3-4-C.1-3-C.3-1 C.3-4-C.1-3-C.3-2 C.1-5-O.1-2-O.1-5 O.1-4-C.3-2-C.2-2 O.1-4-C.3-2-C.2-4 C.3-2-O.1-3-N.3-1 C.2-2-N.3-2-O.1-4 C.3-1-C.3-2-C.2-1 C.3-1-C.3-2-C.2-3 O.1-4-C.3-3-C.2-2 O.1-4-C.3-3-C.2-3 C.3-2-N.3-1-C.3-2 C.3-2-O.1-4-C.3-2 O.1-2-C.1-4-C.2-4 C.2-4-C.1-2-O.1-4 O.1-4-C.2-2-N.3-2 O.1-3-N.3-2-C.1-5 C.3-2-C.2-2-N.3-2 C.2-4-O.1-3-C.3-1 O.1-5-O.1-2-C.1-5 C.3-1-C.3-1-C.2-2 C.3-1-C.3-1-N.3-2 C.2-2-C.2-1-N.3-2 N.3-2-C.3-4-O.1-2 N.3-2-C.3-4-O.1-3 C.3-1-C.2-4-C.1-4 O.1-5-C.1-3-C.2-2 O.1-5-C.1-3-C.2-4 O.1-5-C.1-3-C.3-3 C.1-4-C.3-1-O.1-5 C.2-3-C.3-4-C.1-3 C.2-2-N.3-2-C.3-1 C.2-2-N.3-2-C.3-2 C.3-1-C.3-2-C.3-3 N.3-2-C.2-3-O.1-3 O.1-5-O.1-4-C.2-2 O.1-5-O.1-4-C.2-3 C.3-4-C.1-4-C.2-2 O.1-1-C.3-2-C.3-3 C.3-1-C.2-1-C.3-2 C.3-3-C.1-5-O.1-3 C.2-3-C.1-2-N.3-1 C.2-2-C.3-1-C.1-3 O.1-1-C.3-4-O.1-5 C.3-2-C.2-3-O.1-1 C.3-2-C.2-3-O.1-3 N.3-3-O.1-4-C.3-1 N.3-3-O.1-4-C.3-2 O.1-2-O.1-1-C.3-1 O.1-5-O.1-2-C.2-3 O.1-5-O.1-2-C.2-4 C.2-4-C.1-4-C.3-1 C.2-4-C.1-4-C.3-2 C.3-2-C.3-1-C.1-3 O.1-3-C.3-1-C.2-2 C.2-3-C.1-4-C.3-1 C.2-2-N.3-1-C.3-1 C.2-2-N.3-1-C.3-3 C.1-3-C.2-3-C.3-4 C.3-1-O.1-2-C.2-1 C.3-1-O.1-2-C.3-1 O.1-5-O.1-3-C.3-2 O.1-5-O.1-3-C.3-3 C.3-1-C.2-4-C.1-3 C.3-1-C.1-2-O.1-1 N.3-1-C.2-2-C.3-1 C.3-3-O.1-3-C.2-2 C.2-4-C.1-5-O.1-2 C.2-4-C.1-5-O.1-3 C.2-3-C.3-4-O.1-2 C.2-3-C.3-4-O.1-3 C.2-2-C.2-2-N.3-1 N.3-2-C.2-3-C.3-1 C.3-4-C.1-4-C.2-1 C.1-5-O.1-3-C.3-3 O.1-4-C.2-1-C.3-3 O.1-4-C.2-1-C.3-4 C.1-1-C.3-1-N.3-2 C.2-3-O.1-1-C.3-2 C.3-2-C.3-3-C.2-1 C.2-2-O.1-1-C.3-1 O.1-2-O.1-4-C.3-4 C.3-3-C.3-4-C.1-1 C.3-3-C.2-2-C.2-2 O.1-4-C.2-2-O.1-5 N.3-1-C.2-4-O.1-3 O.1-3-C.2-2-C.2-4 O.1-3-C.2-2-C.3-1 C.3-3-C.1-1-C.3-2 C.1-1-C.3-3-C.3-4 N.3-2-C.3-3-C.3-1 N.3-2-C.3-3-C.3-2 C.1-2-N.3-2-O.1-2 C.3-3-O.1-2-N.3-1 O.1-3-C.3-1-C.2-4 C.2-3-C.1-4-C.3-3 N.3-2-C.3-4-C.1-2 C.1-2-O.1-1-C.3-1 O.1-1-C.3-1-O.1-2 N.3-1-C.2-2-C.2-2 C.3-2-C.3-2-N.3-1 O.1-2-C.3-2-C.3-4 C.3-3-C.1-4-C.3-1 C.3-3-C.1-4-C.3-2 C.3-3-C.3-4-C.1-4 C.3-1-C.2-2-O.1-1 C.3-1-C.2-2-O.1-3 C.3-3-C.2-3-C.1-4 C.2-2-C.3-2-C.3-1 C.2-2-C.3-2-C.3-2 O.1-2-C.2-3-O.1-5 N.3-1-C.3-3-C.1-2 N.3-1-C.3-3-C.2-2 C.3-3-C.2-4-C.1-1 C.2-3-C.3-3-C.3-1 C.2-3-C.3-3-C.3-2 N.3-3-O.1-5-C.1-2 C.2-2-C.3-3-C.2-2 C.2-2-C.3-3-C.3-1 C.2-2-C.3-3-C.3-3 C.1-3-C.2-3-O.1-2 C.3-1-C.3-1-O.1-2 N.3-2-C.2-2-C.2-1 C.1-2-N.3-3-O.1-5 N.3-2-O.1-5-O.1-3 O.1-2-C.1-2-N.3-2 C.3-2-C.3-2-C.3-3 O.1-2-C.3-2-C.2-4 C.2-1-N.3-2-O.1-3 C.1-2-O.1-2-N.3-2 C.1-5-O.1-3-N.3-2 O.1-4-C.2-1-N.3-3 O.1-3-C.3-3-C.1-2 O.1-3-C.3-3-C.1-5 N.3-2-C.3-1-O.1-3 C.3-1-C.3-3-C.2-2 C.2-4-C.1-2-N.3-2 C.1-2-O.1-4-C.3-4 C.1-4-C.2-2-O.1-5 C.3-2-N.3-3-O.1-1 C.3-2-N.3-3-O.1-4 C.3-2-O.1-1-C.3-1 O.1-3-N.3-3-O.1-2 O.1-3-N.3-3-O.1-5 O.1-3-C.2-3-C.1-2 C.1-3-C.2-2-C.2-4 C.2-2-O.1-3-C.3-1 C.2-2-O.1-3-C.3-2 C.3-4-O.1-2-N.3-2 C.3-4-C.1-3-C.2-3 C.1-3-C.2-4-O.1-5 O.1-2-O.1-3-C.2-3 O.1-2-C.1-3-C.2-3 N.3-2-C.2-4-O.1-2 C.3-1-C.3-4-O.1-3 N.3-3-O.1-4-C.2-1 C.1-3-C.2-2-C.3-1 N.3-1-C.3-3-C.3-2 O.1-5-O.1-3-C.2-4 C.3-4-C.1-3-C.2-1 O.1-1-C.3-1-C.1-2 O.1-1-C.3-1-C.3-2 N.3-2-O.1-1-C.3-1 C.3-1-O.1-3-N.3-2 C.3-4-C.1-4-C.3-3 O.1-1-C.3-2-C.2-3 C.3-2-C.3-3-O.1-4 O.1-2-N.3-1-C.2-3 C.2-3-C.3-1-N.3-2 C.3-2-C.2-2-C.2-1 O.1-2-C.2-2-C.3-3 O.1-3-C.2-3-C.3-4 C.3-3-C.2-4-O.1-1 C.1-3-C.3-1-C.2-4 C.2-3-O.1-5-O.1-2 C.2-3-O.1-5-O.1-4 O.1-1-C.3-2-N.3-3 C.3-2-C.2-2-C.2-3 C.3-2-C.2-2-C.3-2 C.2-2-C.2-4-C.1-3 O.1-2-C.2-2-C.3-4 O.1-5-C.1-1-C.3-4 C.3-1-O.1-5-O.1-4 C.1-3-C.2-2-C.3-3 O.1-2-C.2-4-O.1-5 C.2-1-C.3-3-C.3-2 C.2-1-C.3-3-C.3-3 C.3-3-O.1-4-C.3-1 C.3-2-C.3-3-O.1-1 C.2-1-N.3-3-O.1-2 C.2-1-N.3-3-O.1-4 N.3-3-O.1-3-C.3-1 O.1-2-C.2-2-C.2-2 C.2-4-C.1-3-C.2-2 O.1-5-O.1-3-C.2-2 C.1-4-C.3-1-C.2-3 C.1-4-C.3-1-C.2-4 C.2-3-C.1-5-O.1-2 C.2-3-C.1-5-O.1-4 C.2-2-C.3-4-O.1-4 C.2-1-N.3-2-C.3-3 C.3-3-O.1-4-C.3-2 C.2-2-C.2-3-O.1-4 O.1-2-C.2-1-C.3-1 O.1-2-C.2-1-C.3-3 N.3-2-C.2-4-C.1-2 C.3-4-O.1-5-O.1-1 C.3-4-O.1-5-O.1-4 O.1-2-O.1-4-C.2-4 C.3-4-C.1-5-O.1-1 C.3-4-C.1-5-O.1-4 C.2-4-O.1-2-N.3-2 C.3-3-C.2-2-C.3-1 C.3-3-C.2-2-C.3-3 C.2-3-O.1-2-C.3-1 C.3-2-C.3-4-O.1-2 C.3-2-C.3-4-O.1-3 C.2-2-C.3-1-N.3-1 N.3-1-C.3-2-C.2-1 N.3-1-C.3-2-C.3-1 N.3-1-C.3-2-C.3-2 N.3-2-O.1-4-C.3-2 C.3-4-O.1-3-N.3-1 C.3-4-O.1-3-N.3-2 C.2-2-C.3-4-O.1-2 C.2-1-N.3-2-C.3-1 C.2-4-O.1-2-O.1-4 O.1-4-C.3-4-C.1-2 O.1-4-C.3-4-C.1-5 C.3-3-O.1-5-C.1-3 N.3-2-C.1-4-C.2-2 N.3-2-C.1-4-C.3-2 C.1-4-C.2-3-C.3-1 C.2-2-C.2-3-C.1-4 N.3-2-O.1-2-C.1-2 C.2-3-C.3-1-C.3-2 C.3-4-O.1-2-O.1-4 C.3-3-C.3-1-O.1-4 C.3-2-C.3-1-C.3-3 C.3-2-C.2-4-O.1-4 C.2-2-O.1-4-C.3-2 C.2-2-O.1-4-C.3-3 O.1-5-O.1-3-N.3-2 O.1-5-O.1-3-N.3-3 C.2-1-C.3-3-O.1-2 C.2-1-C.3-3-O.1-4 C.3-3-C.3-2-C.2-1 C.3-3-C.3-2-C.3-2 O.1-3-C.3-2-O.1-5 C.1-3-C.3-1-N.3-2 C.2-2-C.3-4-C.1-4 C.2-2-O.1-5-O.1-3 C.2-2-O.1-5-O.1-4 C.3-4-O.1-4-C.3-3 O.1-2-O.1-3-N.3-3 C.3-1-C.1-3-C.2-2 O.1-5-C.1-4-C.3-4 C.3-3-C.3-3-C.3-3 C.3-3-O.1-4-C.2-1 O.1-3-N.3-1-C.3-2 O.1-3-N.3-1-C.3-3 N.3-3-

O.1-2-C.2-1 O.1-3-C.3-3-O.1-5 O.1-3-C.2-1-C.3-2 C.3-1-C.3-3-O.1-4 C.3-1-C.2-1-N.3-2 C.1-2-O.1-
3-C.2-3 C.1-1-C.3-2-C.2-3 C.1-1-C.3-2-C.3-3 C.3-2-N.3-2-C.1-4 C.3-2-C.2-2-O.1-3 C.1-3-C.3-3-
O.1-2 C.3-2-O.1-5-C.1-3 C.2-2-C.3-1-O.1-3 C.3-1-N.3-2-C.3-1 C.3-1-N.3-2-C.3-2 C.3-1-N.3-2-C.3-
3 C.2-2-C.2-4-O.1-3 C.2-1-C.3-1-C.2-2 C.2-4-C.1-3-C.3-1 O.1-1-C.3-4-C.1-5 C.3-3-C.2-3-O.1-4
C.1-4-C.2-2-C.2-3 C.1-4-C.2-2-C.3-4 C.2-3-O.1-3-N.3-2 C.3-2-C.2-3-C.1-1 C.3-2-C.2-3-C.1-3 C.1-
3-C.2-2-O.1-5 O.1-2-N.3-3-O.1-5 C.3-1-C.2-3-C.3-2 C.3-1-C.2-3-C.3-3 C.1-5-O.1-1-C.3-4 O.1-2-
C.1-1-C.3-1 C.2-1-C.3-2-C.3-1 C.2-1-C.3-2-C.3-2 C.2-1-C.3-2-C.3-3 O.1-4-C.3-1-C.1-5 C.3-3-O.1-
2-C.2-1 C.3-3-O.1-2-C.2-2 C.3-2-C.3-1-O.1-3 C.2-3-C.1-4-C.2-2 C.3-2-C.2-4-O.1-2 C.3-1-O.1-2-
O.1-1 C.2-1-N.3-1-C.3-2 C.3-3-C.3-2-C.2-3 C.3-3-C.3-2-C.3-1 C.1-4-C.3-1-C.3-3 C.3-2-O.1-3-C.2-
1 C.2-2-N.3-2-C.1-4 O.1-2-C.3-2-O.1-2 C.2-2-O.1-5-C.1-4 C.3-1-O.1-3-C.3-2 C.2-1-N.3-2-C.1-3
O.1-5-C.1-4-C.3-1 C.2-4-C.1-1-C.3-3 O.1-3-N.3-1-C.3-4 N.3-3-O.1-2-C.2-2 O.1-3-C.2-1-C.3-4 O.1-
3-C.2-1-N.3-2 C.3-1-N.3-1-C.2-2 C.3-1-N.3-1-C.3-2 C.3-1-C.3-3-C.3-2 O.1-2-C.3-3-O.1-5 C.3-4-
O.1-5-C.1-1 C.3-4-O.1-5-C.1-4 N.3-1-C.3-1-O.1-2 C.2-4-O.1-2-C.3-2 C.3-3-C.3-4-O.1-1 C.3-3-C.3-
4-O.1-4 O.1-1-C.3-3-C.2-4 C.1-4-C.3-3-C.3-1 C.1-4-C.3-3-C.3-4 C.3-2-C.2-2-O.1-4 O.1-3-C.2-2-
N.3-3 C.1-3-C.3-3-O.1-5 O.1-2-N.3-2-C.2-4 O.1-2-N.3-2-C.3-4 O.1-5-O.1-1-C.3-4 C.3-1-C.2-2-N.3-
1 C.3-1-C.2-2-N.3-2 O.1-2-C.2-2-N.3-3 O.1-2-O.1-5-C.1-5 C.3-1-C.1-5-O.1-4 O.1-4-C.2-3-C.3-1
C.1-1-C.3-3-C.2-4 C.1-4-C.3-4-O.1-5 C.2-3-C.1-3-C.3-2 N.3-2-C.2-1-C.3-1 N.3-2-C.2-1-C.3-2 C.3-
4-O.1-2-C.2-3 C.1-2-N.3-2-C.2-4 O.1-5-O.1-2-N.3-3 O.1-4-C.3-1-C.2-3 O.1-4-C.3-1-C.2-4 O.1-4-
C.3-1-C.3-3 N.3-1-C.3-3-O.1-3 O.1-4-C.2-4-O.1-2 C.1-4-C.2-3-O.1-5 C.3-2-C.3-1-N.3-1 C.3-2-C.3-
1-N.3-2 C.2-3-O.1-4-C.2-2 O.1-2-C.2-4-C.1-5 C.3-3-C.3-2-N.3-1 C.3-3-C.3-2-N.3-2 C.3-2-C.3-2-
C.2-1 C.3-2-C.3-2-C.2-2 C.3-1-C.3-2-N.3-1 C.2-2-O.1-5-C.1-3 C.3-4-O.1-4-C.2-1 C.3-4-O.1-4-C.2-
2 C.2-1-C.3-4-C.1-4 C.2-1-C.3-4-O.1-3 C.2-1-C.3-4-O.1-4 C.1-4-C.3-2-C.3-3 C.2-2-C.2-3-C.3-2
C.3-1-O.1-4-C.2-3 C.3-1-O.1-4-C.3-3 C.3-1-C.1-4-C.2-3 C.3-1-C.1-4-C.3-3 O.1-1-C.3-3-C.3-4 C.2-
3-C.3-1-C.1-4 C.2-3-C.3-1-O.1-4 C.3-2-N.3-2-C.2-1 C.3-2-N.3-2-C.2-2 O.1-3-N.3-2-O.1-5 C.2-2-
C.3-1-C.2-2 C.2-2-C.3-1-C.3-1 C.2-2-C.3-1-C.3-3 C.2-1-C.3-1-O.1-2 O.1-2-O.1-5-O.1-5 C.3-2-C.2-
3-C.3-1 C.3-2-C.2-3-C.3-3 C.3-4-O.1-2-C.1-4 C.3-4-O.1-2-C.2-2 C.3-3-C.3-1-N.3-2 O.1-4-C.3-1-
O.1-5 N.3-1-C.3-3-O.1-2 C.3-2-C.2-4-C.1-4 N.3-2-C.1-1-C.3-1 C.1-4-C.2-1-C.3-3 C.3-1-C.2-2-C.2-
1 C.3-1-C.2-2-C.2-2 C.3-1-C.2-2-C.3-1 C.3-1-C.2-2-C.3-2 O.1-2-C.1-5-O.1-5 C.2-2-O.1-3-N.3-1
C.3-3-C.3-1-C.1-4 N.3-1-C.2-1-C.3-2 C.3-3-C.1-2-N.3-1 O.1-2-C.3-1-C.2-3 O.1-2-C.3-1-N.3-3 O.1-
2-O.1-2-C.3-2 C.3-1-C.2-4-O.1-3 C.3-1-C.2-4-O.1-4 C.1-5-O.1-2-C.2-3 C.1-5-O.1-2-C.2-4 C.3-1-
C.3-2-O.1-1 C.2-1-C.3-4-C.1-3 C.3-3-C.3-3-C.2-1 C.3-3-C.3-3-C.2-2 N.3-1-C.2-3-C.3-2 C.1-4-C.3-
2-N.3-2 C.2-3-C.1-1-C.3-2 N.3-3-O.1-2-O.1-3 O.1-2-C.2-1-N.3-3 O.1-5-C.1-5-O.1-5 C.3-3-C.2-2-
N.3-1 C.3-3-C.2-2-O.1-4 O.1-3-N.3-2-C.3-1 O.1-3-N.3-2-C.3-4 C.2-2-O.1-2-C.2-2 C.3-1-C.2-2-C.3-
3 N.3-2-O.1-3-C.2-1 C.3-4-C.1-1-C.3-3 C.1-5-O.1-5-O.1-5 C.2-4-O.1-3-N.3-1 C.1-4-C.3-4-O.1-2
C.2-2-C.3-2-N.3-2 C.2-2-C.3-2-O.1-4 C.3-1-N.3-3-O.1-2 C.3-1-N.3-3-O.1-3 C.3-1-N.3-3-O.1-4 C.3-
4-O.1-2-C.3-2 C.3-1-C.2-3-C.1-4 O.1-5-O.1-2-O.1-5 O.1-2-C.2-3-C.1-5 C.2-1-C.3-2-N.3-1 C.2-1-
C.3-2-N.3-2 C.3-3-O.1-2-C.1-3 C.3-3-C.1-2-O.1-3 O.1-3-C.3-1-N.3-2 O.1-3-C.3-1-N.3-3 C.2-3-O.1-
4-C.3-1 C.2-3-O.1-4-C.3-3 O.1-2-C.3-1-C.3-1 C.1-5-O.1-2-C.3-3 O.1-4-C.3-2-N.3-2 O.1-4-C.3-2-
N.3-3 C.3-2-C.3-2-O.1-4 C.3-1-O.1-3-C.2-2 O.1-2-C.1-3-C.3-3 N.3-1-C.2-3-C.1-2 C.2-2-N.3-3-O.1-
2 C.2-2-N.3-3-O.1-3 O.1-2-C.3-3-C.1-5 C.1-5-O.1-4-C.2-3 O.1-5-C.1-5-O.1-2 C.1-4-C.2-1-C.3-4
C.1-5-O.1-5-O.1-2 C.2-2-O.1-3-N.3-2 C.2-1-C.3-2-O.1-3 C.3-3-C.3-1-C.3-2 C.1-2-O.1-5-O.1-5 O.1-
3-C.2-4-C.1-5 C.3-4-O.1-3-C.3-1 C.3-4-O.1-3-C.3-2 C.2-1-N.3-2-C.2-2 C.1-3-C.3-2-C.3-1 C.1-3-
C.3-2-C.3-4 C.2-3-O.1-2-C.1-3 C.2-3-C.1-2-O.1-3 N.3-3-O.1-5-O.1-2 N.3-3-O.1-5-O.1-3 C.2-2-C.2-
1-C.3-1 C.2-2-C.2-1-C.3-2 C.1-3-C.3-1-C.3-4 C.1-5-O.1-3-C.2-4 C.3-3-C.2-1-C.3-3 O.1-4-C.3-3-
C.3-1 O.1-4-C.3-3-C.3-4 C.3-3-C.1-4-C.2-1 C.3-2-C.2-1-C.3-1 C.3-2-C.2-1-C.3-2 N.3-3-O.1-2-C.3-
1 C.1-3-C.3-2-C.2-3 O.1-2-N.3-1-C.3-1 O.1-2-N.3-1-C.3-3 N.3-1-C.3-1-C.1-2 N.3-1-C.3-1-C.2-2
N.3-1-C.3-1-C.3-2 C.2-4-O.1-2-C.1-4 C.1-4-C.3-3-C.2-3 C.3-2-N.3-2-C.3-3 N.3-2-O.1-3-C.3-1 O.1-
4-C.2-3-O.1-5 C.2-3-O.1-3-C.3-2 C.3-1-C.2-3-O.1-2 O.1-5-C.1-2-O.1-5 C.1-4-C.2-4-O.1-2 C.3-3-
C.2-1-C.3-2 C.3-2-C.2-1-C.3-3 C.2-3-O.1-2-N.3-1 N.3-2-C.1-3-C.3-1 N.3-1-C.3-2-O.1-3 C.3-1-C.2-
3-O.1-4 C.3-3-C.3-1-C.2-2 O.1-3-C.2-4-O.1-5 C.3-1-O.1-2-C.1-1 C.2-4-O.1-5-C.1-3 C.1-3-C.3-2-
O.1-5 C.3-3-C.3-1-C.2-3 C.2-2-N.3-1-C.2-2 C.2-2-C.2-2-O.1-2 C.3-3-C.2-1-N.3-2 O.1-4-C.2-2-C.3-
2 O.1-4-C.2-2-C.3-4 C.2-3-O.1-2-O.1-3 O.1-3-N.3-2-C.2-2 O.1-3-N.3-2-C.2-3 O.1-3-C.2-2-O.1-5
C.1-3-C.2-1-C.3-4 N.3-2-C.3-2-C.2-2 C.3-1-N.3-2-C.1-1 C.2-4-O.1-4-C.3-1 C.2-4-O.1-4-C.3-2 C.1-
1-C.3-4-O.1-5 C.3-4-O.1-3-C.2-1 C.3-4-O.1-3-C.2-3 C.2-1-C.3-3-C.1-4 O.1-1-C.3-1-N.3-2 N.3-1-
C.3-4-O.1-3 C.2-4-O.1-5-O.1-2 C.2-4-O.1-5-O.1-3 C.2-3-O.1-5-C.1-4 O.1-5-O.1-4-C.3-1 O.1-5-O.1-
4-C.3-4 C.1-1-C.3-1-O.1-2 O.1-3-N.3-1-C.2-4 N.3-2-C.3-1-C.3-1 C.1-5-O.1-4-C.3-1 C.3-2-N.3-2-
O.1-4 C.1-2-N.3-2-C.3-4 C.2-1-C.3-2-C.2-2 O.1-3-N.3-1-C.2-2 C.1-2-O.1-3-C.3-3 C.1-5-O.1-4-C.3-
4 O.1-4-C.2-2-C.2-3 N.3-2-C.3-2-C.3-1 C.3-1-N.3-2-C.1-3 C.3-1-N.3-2-C.2-1 C.3-1-N.3-2-C.2-3
C.3-1-C.3-4-C.1-3 N.3-2-C.1-3-C.2-1

Atom Pairs and Triplets (Geometrical)

These fingerprints can be interpreted like their topological versions. Nevertheless, the task is more difficult because the distances depend on the binning and the stretching factor and cannot be derived by simply regarding the chemical graph.

CATS2D fingerprints

A set of predefined pharmacophore types a matched against the atoms included in a chemical graph (15 unique combinations). After that, the count of features with a specific combination is assigned to fixed position in a vector. For example, consider entry "`CATS2D-2:2`". The first entry is indicates the position which corresponds to the pattern (AA; acceptor-acceptor; 0 (offset, defined in the paper) + 2 (distance)). To sum up, this means that two acceptor-acceptor combinations with distance 2 can be found in the compound. This corresponds to the pharmacophore points 3 and 9 and 11 and 12. Note that the bit positions can be shifted when the distances are varied.

```
CATS2D-10:4 CATS2D-11:0 CATS2D-12:1 CATS2D-13:2 CATS2D-14:0 CATS2D-15:5 CATS2D-16:0 CATS2D-
17:0 CATS2D-18:0 CATS2D-19:0 CATS2D-20:0 CATS2D-21:0 CATS2D-22:4 CATS2D-23:2 CATS2D-24:1
CATS2D-25:3 CATS2D-26:0 CATS2D-27:0 CATS2D-28:0 CATS2D-29:0 CATS2D-30:0 CATS2D-31:2 CATS2D-
32:1 CATS2D-33:0 CATS2D-34:2 CATS2D-35:0 CATS2D-36:0 CATS2D-37:0 CATS2D-38:0 CATS2D-39:0
CATS2D-40:0 CATS2D-41:0 CATS2D-42:0 CATS2D-43:0 CATS2D-44:0 CATS2D-45:0 CATS2D-46:0 CATS2D-
47:0 CATS2D-48:0 CATS2D-49:0 CATS2D-50:2 CATS2D-51:0 CATS2D-52:0 CATS2D-53:0 CATS2D-54:0
CATS2D-55 CATS2D-56:0 CATS2D-57:0 CATS2D-58:0 CATS2D-59:0 CATS2D-60:0 CATS2D-61:0 CATS2D-62
CATS2D-63 CATS2D-64:0 CATS2D-65:2 CATS2D-66:0 CATS2D-67:0 CATS2D-68:0 CATS2D-69:0 CATS2D-70:0
CATS2D-71 CATS2D-72:0 CATS2D-73:0 CATS2D-74 CATS2D-75:0 CATS2D-76:0 CATS2D-77:0 CATS2D-78:0
CATS2D-79:0 CATS2D-80:0 CATS2D-81:0 CATS2D-82:0 CATS2D-83:0 CATS2D-84:0 CATS2D-85:0 CATS2D-
86:0 CATS2D-87:0 CATS2D-88:0 CATS2D-89:0 CATS2D-90:2 CATS2D-91:0 CATS2D-92:0 CATS2D-93:0
CATS2D-94 CATS2D-95:0 CATS2D-96:0 CATS2D-97:0 CATS2D-98:0 CATS2D-99:0 CATS2D-100:0 CATS2D-
101:0 CATS2D-102:1 CATS2D-103:0 CATS2D-104:1 CATS2D-105:0 CATS2D-106:0 CATS2D-107:0 CATS2D-
108:0 CATS2D-109:0 CATS2D-110:0 CATS2D-111:0 CATS2D-112:0 CATS2D-113:0 CATS2D-114:0 CATS2D-
115:0 CATS2D-116:0 CATS2D-117:0 CATS2D-118:0 CATS2D-119:0 CATS2D-120:1 CATS2D-121:0 CATS2D-
122:0 CATS2D-123:0 CATS2D-124:0 CATS2D-125:0 CATS2D-126:0 CATS2D-127:0 CATS2D-128:0 CATS2D-
129:0 CATS2D-130:0 CATS2D-131:0 CATS2D-132:0 CATS2D-133:0 CATS2D-134:0 CATS2D-135:0 CATS2D-
136:0 CATS2D-137:0 CATS2D-138:0 CATS2D-139:0 CATS2D-140:0 CATS2D-141:0 CATS2D-142:0 CATS2D-
143:0 CATS2D-144:0 CATS2D-145:0 CATS2D-146:0 CATS2D-147:0 CATS2D-148:0 CATS2D-149:0 CATS2D-0:5
CATS2D-1:0 CATS2D-2:2 CATS2D-3:3 CATS2D-4:0 CATS2D-5:5 CATS2D-6:0 CATS2D-7:0 CATS2D-8:0
CATS2D-9:0
```

CATS3D fingerprints

Derived like the CATS2D fingerprint, with the exception that the binned geometrical distance matrix is used.

Pharmacophore Pairs

This fingerprint considers all pairs between PPPs. The patterns are extracted in the same fashion like the atom pair fingerprints.

```
A-2-A A-3-A A-5-A D-2-A D-3-A D-5-A D-5-D L-2-A L-2-D L-3-A L-3-D L-4-A L-4-L L-5-A L-5-D N-1-
A N-1-D N-2-A N-2-L N-4-A N-4-D N-4-L
```

Pharmacophore Triplets

Generated like the Atom triplets, but rely on the pharmacophore points. In case that the pharmacophore typing assigns multiple pharmacophore points to an atom, multiple pharmacophore triplets are extracted.

```
A-4-N-1-D-5 D-1-N-4-L-5 L-3-A-2-A-3 A-2-A-3-D-5 A-2-A-5-D-3 A-5-L-4-L-3 D-2-L-3-D-5 L-2-D-4-N-
2 L-2-D-5-A-3 L-2-D-5-A-4 L-2-D-5-L-4 D-5-D-3-A-3 A-1-N-1-A-2 A-3-D-2-L-2 A-3-D-4-N-2 A-3-D-5-
A-2 L-4-N-1-A-5 A-2-L-4-L-2 A-2-L-4-L-4 A-2-L-5-D-3 D-3-A-2-A-5 D-3-A-2-N-4 D-3-A-3-A-2 N-4-D-
2-L-2 N-4-D-5-L-4 A-5-D-2-A-5 L-2-A-2-A-2 L-2-A-2-N-2 L-2-A-2-N-4 L-2-A-3-A-2 L-2-A-3-A-3 L-2-
A-3-A-5 L-2-A-5-A-3 L-2-A-5-A-5 L-2-A-5-L-4 A-4-L-2-N-4 A-4-L-4-L-2 D-5-A-1-N-4 D-5-A-2-A-5 D-
5-A-3-A-3 D-5-A-5-A-2 D-5-A-5-A-5 L-4-A-2-A-2 L-4-A-5-D-2 L-4-A-5-D-3 N-2-A-2-A-4 N-2-A-2-L-4
N-2-A-3-A-1 N-2-A-3-A-4 A-3-A-2-L-2 A-3-A-2-L-3 A-3-A-2-N-1 A-3-A-2-N-4 A-3-A-3-A-2 A-3-A-4-N-
2 A-3-A-5-A-2 A-3-A-5-D-3 D-4-N-1-D-5 D-4-N-4-L-5 L-3-D-1-N-2 L-3-D-2-A-3 D-3-L-2-D-5 D-3-L-4-
L-5 L-3-D-5-A-4 L-3-D-5-D-2 L-3-D-5-L-4 N-1-D-2-A-1 N-1-D-3-L-2 A-2-D-1-N-1 A-2-D-3-L-3 A-2-D-
```

5-A-5 A-2-D-5-L-5 A-2-N-4-A-2 A-2-N-4-D-3 N-4-A-2-A-2 N-4-A-2-L-4 A-5-A-2-D-5 A-5-A-2-L-4 A-5-A-2-L-5 A-5-A-5-A-5 A-5-A-5-D-2 A-5-A-5-D-5 A-5-A-5-L-5 D-2-A-5-A-5 D-2-A-5-L-5 L-5-D-2-A-5 D-5-L-4-L-2 D-5-L-4-L-3 D-5-L-4-N-1 L-5-D-5-A-2 L-5-D-5-A-5 L-5-D-5-D-5 D-5-L-5-A-2 D-5-L-5-A-5 D-5-L-5-D-5 L-4-L-2-A-2 L-4-L-2-A-4 L-4-L-2-A-5 L-4-L-2-D-5 L-4-L-4-A-2 L-4-L-4-N-2 L-4-L-5-A-2 L-4-L-5-A-3 N-2-L-2-A-2 N-2-L-2-A-4 N-2-L-4-L-4 A-4-N-2-A-2 A-4-N-4-A-5 A-3-L-2-D-5 D-1-N-4-A-5 A-3-L-3-A-2 A-3-L-3-D-2 N-1-A-2-A-1 N-1-A-3-A-2 N-1-A-3-L-2 N-1-A-5-A-4 N-1-A-5-D-4 N-4-L-5-D-1 N-4-L-5-D-4 A-2-A-1-N-1 A-2-A-3-A-3 A-2-A-4-N-2 A-5-L-2-A-3 A-5-L-2-A-5 A-5-L-4-L-2 A-5-L-4-N-1 A-5-L-4-N-4 A-5-L-5-A-2 A-5-L-5-A-5 D-2-L-2-N-4 D-2-L-3-A-5 D-2-L-4-A-5 D-2-L-4-L-5 L-2-D-5-D-3 L-2-N-1-D-3 L-2-N-2-A-2 L-2-N-4-A-2 L-2-N-4-D-2 D-5-D-2-A-5 D-5-D-3-L-2 D-5-D-4-N-1 A-1-N-2-A-3 A-1-N-2-L-3 A-1-N-4-D-5 L-5-A-1-N-4 L-5-A-2-D-5 L-5-A-2-L-4 L-5-A-3-A-2 L-5-A-3-L-4 L-5-A-5-D-5 A-3-D-1-N-2 A-3-D-3-L-2 A-3-D-5-A-3 L-4-N-2-L-4 L-4-N-4-A-2 L-4-N-4-A-5 A-2-L-2-N-2 A-2-L-4-A-5 A-2-L-5-D-5 D-3-A-2-L-2 D-3-A-3-A-5 A-5-D-3-A-3 A-5-D-4-N-1 A-5-D-4-N-4 A-5-D-5-A-2 A-5-D-5-A-5 L-2-A-2-A-4 L-2-A-4-L-4 L-2-A-5-A-4 A-4-L-2-A-2 A-4-L-3-A-5 A-4-L-3-D-5 D-5-A-2-D-5 D-5-A-3-L-2 D-5-A-4-L-2 D-5-A-4-L-3 D-5-A-5-L-5 L-4-A-2-L-4 N-2-A-2-L-2 N-2-A-3-D-1 N-2-A-3-D-4 A-3-A-1-N-2 A-3-A-2-L-5 A-3-A-3-A-5 A-3-A-3-D-2 A-3-A-3-L-2 D-4-N-2-L-2 D-3-L-2-A-3 D-3-L-2-A-5 D-3-L-3-A-2 D-3-L-4-A-5 L-3-D-5-A-2 A-2-N-4-A-3 N-4-A-2-L-2 N-4-A-4-L-2 A-5-A-2-A-3 A-5-A-2-A-5 A-5-A-5-L-2 D-2-A-3-A-3 L-5-D-3-L-4 L-5-D-4-N-4 D-5-L-4-N-4 L-4-L-3-A-5 L-4-L-3-D-5 L-4-L-5-D-2 L-4-L-5-D-3 A-4-N-1-A-5 A-4-N-4-D-5 D-1-N-2-A-3 A-3-L-2-A-3 A-3-L-2-A-5 A-3-L-4-A-5 L-3-A-3-A-2 L-3-A-5-L-4 N-4-L-2-A-2 A-2-A-3-A-5 D-2-L-2-A-3 L-2-N-4-A-4 L-2-N-4-L-4 D-5-D-5-L-5 A-1-N-4-L-5 A-2-L-2-N-4 A-2-L-4-A-2 A-2-L-4-L-5 A-2-L-4-N-2 A-2-L-4-N-4 A-2-L-5-A-3 D-3-A-2-L-3 D-3-A-2-L-5 A-5-D-2-L-3 A-5-D-2-L-4 D-5-A-2-L-3 L-4-A-5-A-3 A-3-A-5-A-3 A-3-A-5-L-2 N-1-D-5-D-4 A-2-N-4-L-2 N-4-A-5-L-4 A-5-A-3-A-2 A-5-A-3-A-3 A-5-A-3-L-2 L-4-L-2-N-4 N-2-L-4-A-4 A-4-N-2-A-3 A-4-N-2-L-2 A-4-N-2-L-4 D-1-N-2-L-3 A-3-L-4-L-5 L-3-A-2-D-3 N-1-A-5-L-4 N-4-L-2-A-4 N-4-L-5-A-1 A-2-A-5-A-3 D-5-D-5-A-2 D-5-D-5-A-5 A-1-N-4-A-5 L-5-A-4-N-4 A-3-D-5-D-3 L-4-N-1-D-5 L-4-N-2-A-2 A-2-L-3-A-3 A-2-L-3-D-3 A-2-L-3-D-5 A-2-L-5-A-5 D-3-A-2-N-1 N-4-D-3-A-2 N-4-D-5-A-1 N-4-D-5-A-4 N-4-D-5-D-1 A-5-D-3-A-2 A-5-D-3-L-2 A-5-D-5-L-2 L-2-A-3-D-2 L-2-A-3-D-3 L-2-A-3-D-5 D-5-A-2-L-5 L-4-A-5-A-2 A-3-A-2-D-3 A-3-A-3-D-5 D-4-N-1-A-5 D-4-N-2-A-3 N-1-D-5-A-4 A-2-D-3-A-3 A-2-D-5-D-5 A-2-N-1-D-3 N-4-A-3-A-2 N-4-A-5-D-1 N-4-A-5-D-4 A-5-A-2-L-3 A-5-A-3-L-4 A-5-A-4-N-1 A-5-A-4-N-4 A-5-A-5-A-2 D-2-A-5-D-5 L-3-A-1-N-2 L-3-A-5-D-2 N-4-L-4-L-2 N-4-L-5-A-4 A-2-A-2-L-2 A-2-A-2-L-4 A-2-A-5-L-5 L-5-A-2-A-5 A-3-D-2-A-3 A-3-D-5-L-2 L-4-N-4-D-5 A-5-D-1-N-4 A-5-D-3-L-4 A-5-D-5-L-5 A-4-L-2-A-5 A-4-L-2-D-5 D-5-A-4-N-1 D-5-A-4-N-4 L-3-D-3-A-2 N-4-A-5-A-1 N-4-A-5-A-4 D-2-A-3-L-3 L-5-D-2-L-4 L-5-D-3-A-2 A-4-N-4-L-2 A-4-N-4-L-5 D-1-N-1-A-2 D-1-N-4-D-5 A-2-A-4-L-2 A-2-A-5-A-5 L-2-D-3-A-2 D-5-D-2-L-3 A-1-N-1-D-2 L-5-A-5-A-2 A-2-L-2-D-3 A-2-L-3-A-5 D-3-A-3-D-5 L-2-A-2-L-4 L-2-A-4-N-4 D-5-A-2-A-3 D-5-A-5-D-5 L-4-A-4-N-2 A-3-A-2-A-3 D-4-N-4-A-5 N-1-D-5-L-4 A-5-A-1-N-4 A-5-A-4-L-2 A-5-A-4-L-3 D-2-A-1-N-1 L-5-D-1-N-4 A-3-L-2-N-1 L-3-A-5-A-2 A-2-A-5-D-5 A-5-L-5-D-2 D-5-D-1-N-4 A-5-D-5-D-5 L-2-A-4-N-2 A-3-A-2-A-5 D-3-L-2-N-1 A-2-N-1-A-3 A-2-N-2-L-2 N-2-L-2-D-4 N-2-L-3-D-1 A-2-A-2-N-4 A-5-L-5-D-5 A-2-L-2-A-3 A-5-D-5-D-2 D-5-L-2-A-5 L-3-A-5-A-4 N-1-A-2-D-1 A-2-A-3-L-3 L-5-A-5-A-5 A-2-L-2-A-2 L-2-A-5-D-3 L-2-A-5-D-5 N-1-D-3-A-2 D-5-L-2-A-3 N-2-L-3-A-1 L-2-N-1-A-3:1

Pharmacophore Pairs and Triplets (Geometrical)

Generated like the topological versions but using the binned geometrical distance between PPPs.

Extended Connectivity Fingerprint

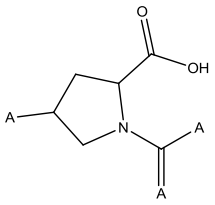
[*]C [*]=O [*]C([*])O [*]N([*])C(=O)C [*]=C([*])N1CC(O)CC1([*]) [*]=C([*])N(C[*])C([*])[*] [*]N([*])CC([*])[*] O=C(O)C1N(C(=O)C)CC(O)C1 [*]=C([*])C [*]C([*])=O [*]C(=[*])C1N(C(=O)C)CC(O)C1 [*]O [*]C([*])[*] [*]CN(C(=O)C)C([*])[*] [*]CC(O)C[*] [*]C(=[*])O [*]C([*])C(=O)O [*]N1CC(O)CC1([*]) [*]C([*])CC([*])[*] [*]N([*])C(C(=O)O)C[*] [*]C[*] [*]C(=[*])C1N(C(=O)C)CC([*])C1 [*]=C([*])N1CC(O)CC1(C([*])=[*]) [*]=C([*])N1CC(O)CC1(C(=O)O) [*]N([*])[*] [*]C([*])=[*] [*]C(=[*])C(N([*])[*])C[*] [*]C(=[*])C1N([*])CC(O)C1 [*]=C([*])N1CC([*])CC1(C(=O)O)

Examples

(1) SMILES

[*]=C([*])N1CC([*])CC1(C(=O)O)

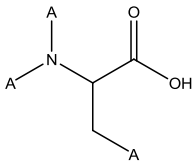
equals



(2) SMILES

[*]N([*])C(C(=O)O)C[*]

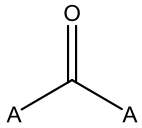
equals



(2) SMILES

[*]C([*])=O

Equals



Local Star Fingerprint

This fingerprint describes atom environments by simply generating all paths originating from an atom and generating a canonical string representation from it. Each shell is stored. For example, the pattern `[N.3-C.2,N.3-C.3,N.3-C.3]` encodes the shell of depth 1 originating from the nitrogen. It is similar to the Molprint2D fingerprint but includes the bond information.

[O.1] [C.3-C.3-C.2,C.3-C.3-N.3] [C.3-C.2-C.3-C.2-N.3-C.3-C.1,C.3-C.2-C.3-C.2-N.3-C.3=O.1] [C.3-C.2-C.3-C.3-O.1,C.3-C.2-C.3-C.3=O.1,C.3-C.2-C.3-N.3-C.2,C.3-C.2-C.3-N.3-C.3,C.3-C.2-N.3-C.3-C.1,C.3-C.2-N.3-C.3-C.2,C.3-C.2-N.3-C.3-C.3,C.3-C.2-N.3-C.3=O.1] [C.3-C.2,C.3-C.3,C.3-N.3] [C.3-C.2,C.3-C.2,C.3-O.1] [C.3-C.3-C.2-C.3-C.2-N.3,C.3-C.3-N.3-C.2-C.3-C.2,C.3-C.3-N.3-C.2-C.3-O.1] [O.1=C.3-C.1,O.1=C.3-N.3] [O.1-C.3-C.3-C.2,O.1-C.3-C.3-N.3] [C.3-C.3-C.2-C.3-C.2,C.3-C.3-C.2-C.3-O.1,C.3-C.3-N.3-C.2-C.3,C.3-C.3-N.3-C.3-C.1,C.3-C.3-N.3-C.3=O.1] [O.1-C.3-C.2-C.3-C.3-O.1,O.1-C.3-C.2-C.3-C.3=O.1,O.1-C.3-C.2-C.3-N.3-C.2,O.1-C.3-C.2-C.3-N.3-C.3,O.1-C.3-C.2-N.3-C.3-C.1,O.1-C.3-C.2-N.3-C.3-C.2,O.1-C.3-C.2-N.3-C.3-C.3,O.1-C.3-C.2-N.3-C.3=O.1] [C.2-C.3-C.2-N.3-C.3-C.1,C.2-C.3-C.2-N.3-C.3-C.3,C.2-C.3-C.2-N.3-C.3=O.1,C.2-C.3-N.3-C.2-C.3-O.1] [O.1-C.3-C.2-C.3,O.1-C.3-C.2-N.3] [N.3-C.2-C.3-C.2-C.3,N.3-C.3-C.2-C.3-C.2,N.3-C.3-C.2-C.3-O.1] [C.3-C.2-C.3-C.2-N.3,C.3-N.3-C.2-C.3-C.2,C.3-N.3-C.2-C.3-O.1] [N.3] [C.2-C.3-C.2-N.3-C.3-C.3-O.1,C.2-C.3-C.2-N.3-C.3-C.3=O.1] [C.3-N.3-C.2-C.3-C.2,C.3-N.3-C.2-C.3-O.1,C.3-N.3-C.3-C.2-C.3,C.3-N.3-C.3-C.3-O.1,C.3-N.3-C.3-C.3=O.1] [C.3-C.2-C.3-C.2-N.3-C.3] [C.3-C.1,C.3-N.3,C.3=O.1] [N.3-C.2-C.3-C.2-C.3-C.3-O.1,N.3-C.2-C.3-C.2-C.3-C.3=O.1] [O.1-C.3-C.3-C.2-C.3-C.2,O.1-C.3-C.3-C.2-C.3-O.1,O.1-C.3-C.3-N.3-C.2-C.3,O.1-C.3-C.3-N.3-C.3-C.1,O.1-C.3-C.3-N.3-C.3=O.1] [O.1=C.3-C.3-C.2,O.1=C.3-C.3-N.3] [C.3] [C.3-C.2-C.3-C.3,C.3-C.2-C.3-N.3,C.3-C.2-N.3-C.3,C.3-C.2-N.3-C.3] [C.1-C.3-N.3-C.2-C.3,C.1-C.3-N.3-C.3-C.2,C.1-C.3-N.3-C.3-C.3] [N.3-C.2-C.3-C.2,N.3-C.2-C.3-O.1,N.3-C.3-C.2-C.3,N.3-C.3-C.3-O.1,N.3-C.3-C.3=O.1] [C.3-C.2-C.3-N.3-C.3-

```
C.1,C.3-C.2-C.3-N.3-C.3=O.1,C.3-C.2-N.3-C.3-C.3-O.1,C.3-C.2-N.3-C.3-C.3=O.1] [C.2-C.3-C.2-C.3-
N.3-C.3-C.1,C.2-C.3-C.2-C.3-N.3-C.3=O.1] [C.2] [O.1=C.3-N.3-C.2,O.1=C.3-N.3-C.3] [C.2-C.3-C.2-
C.3-C.3,C.2-C.3-C.2-C.3-N.3,C.2-N.3-C.3-C.2-C.3,C.2-N.3-C.3-C.3-O.1,C.2-N.3-C.3-C.3=O.1] [C.1-
C.3-N.3-C.2-C.3-C.2-C.3,C.1-C.3-N.3-C.3-C.2-C.3-C.2,C.1-C.3-N.3-C.3-C.2-C.3-O.1] [C.2-C.3-C.2-
N.3-C.3,C.2-C.3-C.2-N.3-C.3,C.2-C.3-N.3-C.2-C.3,C.2-C.3-N.3-C.3-C.1,C.2-C.3-N.3-C.3=O.1] [C.2-
C.3-C.2,C.2-C.3-C.3,C.2-C.3-N.3,C.2-C.3-O.1] [C.2-C.3-C.2-N.3,C.2-C.3-C.3-O.1,C.2-C.3-
C.3=O.1,C.2-C.3-N.3-C.2,C.2-C.3-N.3-C.3] [C.2-C.3,C.2-C.3] [C.2-C.3,C.2-N.3] [O.1=C.3-N.3-C.2-
C.3-C.2,O.1=C.3-N.3-C.2-C.3-O.1,O.1=C.3-N.3-C.3-C.2-C.3,O.1=C.3-N.3-C.3-C.3-O.1,O.1=C.3-N.3-
C.3-C.3=O.1] [C.1] [C.3-N.3-C.2-C.3-C.2-C.3,C.3-N.3-C.3-C.2-C.3-C.2,C.3-N.3-C.3-C.2-C.3-O.1]
[O.1=C.3-N.3-C.2-C.3-C.2-C.3,O.1=C.3-N.3-C.3-C.2-C.3-C.2,O.1=C.3-N.3-C.3-C.2-C.3-O.1] [O.1-
C.3-C.3,O.1-C.3=O.1] [O.1=C.3-C.3-C.2-C.3-C.2-N.3,O.1=C.3-C.3-N.3-C.2-C.3-C.2,O.1=C.3-C.3-N.3-
C.2-C.3-O.1] [O.1=C.3-C.3,O.1=C.3-O.1] [O.1=C.3-C.3-C.2-C.3,O.1=C.3-C.3-N.3-C.2,O.1=C.3-C.3-
N.3-C.3] [N.3-C.2-C.3,N.3-C.3-C.1,N.3-C.3-C.2,N.3-C.3-C.3,N.3-C.3=O.1] [C.3-C.3-C.2-C.3,C.3-
C.3-N.3-C.2,C.3-C.3-N.3-C.3] [C.3-N.3-C.2-C.3-C.2-C.3-C.3] [O.1=C.3] [O.1-C.3-C.3-C.2-C.3,O.1-
C.3-C.3-N.3-C.2,O.1-C.3-C.3-N.3-C.3] [C.2-C.3-C.2-C.3-C.3-O.1,C.2-C.3-C.2-C.3-C.3=O.1,C.2-C.3-
C.2-C.3-N.3-C.3,C.2-N.3-C.3-C.2-C.3-O.1] [O.1-C.3-C.2-C.3-N.3-C.3-C.1,O.1-C.3-C.2-C.3-N.3-
C.3=O.1,O.1-C.3-C.2-N.3-C.3-C.3-O.1,O.1-C.3-C.2-N.3-C.3-C.3=O.1] [C.1-C.3-N.3-C.2,C.1-C.3-N.3-
C.3] [N.3-C.2-C.3-C.2-C.3-C.3] [C.3-C.2-C.3,C.3-C.2-N.3] [C.3-C.3-C.2-C.3-C.2-N.3-C.3] [C.3-
N.3-C.2-C.3,C.3-N.3-C.3-C.2,C.3-N.3-C.3-C.3] [C.2-C.3-C.2-C.3,C.2-N.3-C.3-C.1,C.2-N.3-C.3-
C.2,C.2-N.3-C.3-C.3,C.2-N.3-C.3=O.1] [C.3-C.3,C.3-O.1,C.3=O.1] [C.3-N.3-C.2,C.3-N.3-C.3] [C.2-
C.3-C.2,C.2-C.3-O.1,C.2-N.3-C.3,C.2-N.3-C.3] [O.1-C.3-C.2-C.3-C.3,O.1-C.3-C.2-C.3-N.3,O.1-C.3-
C.2-N.3-C.3,O.1-C.3-C.2-N.3-C.3] [O.1=C.3-N.3-C.2-C.3,O.1=C.3-N.3-C.3-C.2,O.1=C.3-N.3-C.3-C.3]
[O.1=C.3-C.3-C.2-C.3-C.2,O.1=C.3-C.3-C.2-C.3-O.1,O.1=C.3-C.3-N.3-C.2-C.3,O.1=C.3-C.3-N.3-C.3-
C.1,O.1=C.3-C.3-N.3-C.3=O.1] [C.1-C.3-N.3,C.1-C.3=O.1] [C.1-C.3-N.3-C.2-C.3-C.2,C.1-C.3-N.3-
C.2-C.3-O.1,C.1-C.3-N.3-C.3-C.2-C.3,C.1-C.3-N.3-C.3-C.3-O.1,C.1-C.3-N.3-C.3-C.3=O.1] [C.1-C.3]
[C.3-C.2-C.3,C.3-C.3-O.1,C.3-C.3=O.1,C.3-N.3-C.2,C.3-N.3-C.3] [C.3-C.2-C.3-C.2,C.3-C.2-C.3-
O.1,C.3-N.3-C.2-C.3,C.3-N.3-C.3-C.1,C.3-N.3-C.3=O.1] [O.1-C.3-C.3-C.2-C.3-C.2-N.3,O.1-C.3-C.3-
N.3-C.2-C.3-C.2,O.1-C.3-C.3-N.3-C.2-C.3-O.1] [N.3-C.2,N.3-C.3,N.3-C.3] [O.1-C.3-C.2,O.1-C.3-
C.2] [O.1-C.3]:1
```

SHED Key

This fingerprint describes the entropy of all possible pairs of PPPs regarding the distribution of the corresponding distances. The atom type is restricted to the PPPs.

```
AA:2.8 AD:2.46 AL:3.596 AN:2.872 AP:0 DD DL:2.828 DN:2 DP:0 LL LN:2 LP:0 NN NP:0 PP:0
```

Molprint2D

This fingerprint extends an atom by the atoms up to the defined maximum search depth. `0[N.3]1[C.2 C.3 C.3]2[C.1 C.2 C.3 C.3 O.1]` encodes the topological environment of depth 2 of the nitrogen.

```
0[N.3]1[C.2 C.3 C.3]2[C.1 C.2 C.3 C.3 O.1] 0[C.3]1[C.1 N.3 O.1]2[C.2 C.3] 0[C.1]1[C.3]2[N.3
O.1] 0[C.3]1[C.3 O.1 O.1] 0[O.1]1[C.3]2[C.2 C.2]3[C.3 N.3] 0[C.3]1[C.2 C.3 N.3]2[C.2 C.3 C.3
O.1 O.1]3[C.1 O.1 O.1] 0[C.2]1[C.3 N.3]2[C.2 C.3 C.3 O.1] 0[O.1]1[C.3]2[C.3 O.1] 0[C.3]1[C.2
C.2 O.1]2[C.3 N.3]3[C.3 C.3] 0[C.2]1[C.3 C.3] 0[C.2]1[C.3 N.3] 0[C.3]1[C.2 C.3 N.3]2[C.2 C.3
C.3 O.1 O.1] 0[O.1]1[C.3]2[C.3 O.1]3[C.2 N.3] 0[C.2]1[C.3 C.3]2[C.2 C.3 N.3 O.1]3[C.3 O.1 O.1]
0[C.3]1[C.1 N.3 O.1] 0[O.1]1[C.3] 0[C.3]1[C.3 O.1 O.1]2[C.2 N.3]3[C.2 C.3 C.3]
0[O.1]1[C.3]2[C.1 N.3] 0[O.1]1[C.3]2[C.2 C.2] 0[C.3]1[C.2 C.2 O.1]2[C.3 N.3] 0[C.1]1[C.3]
0[C.3]1[C.1 N.3 O.1]2[C.2 C.3]3[C.2 C.3 C.3] 0[N.3]1[C.2 C.3 C.3] 0[C.1]1[C.3]2[N.3 O.1]3[C.2
C.3] 0[O.1]1[C.3]2[C.1 N.3]3[C.2 C.3] 0[N.3]1[C.2 C.3 C.3]2[C.1 C.2 C.3 C.3 O.1]3[O.1 O.1 O.1]
0[C.3]1[C.2 C.2 O.1] 0[C.3]1[C.3 O.1 O.1]2[C.2 N.3] 0[C.2]1[C.3 C.3]2[C.2 C.3 N.3 O.1]
0[C.2]1[C.3 N.3]2[C.2 C.3 C.3 O.1]3[C.1 C.3 O.1] 0[C.3]1[C.2 C.3 N.3]
```

Molprint3D

Similar to Molprint2D, but using the binned geometrical distance matrix.