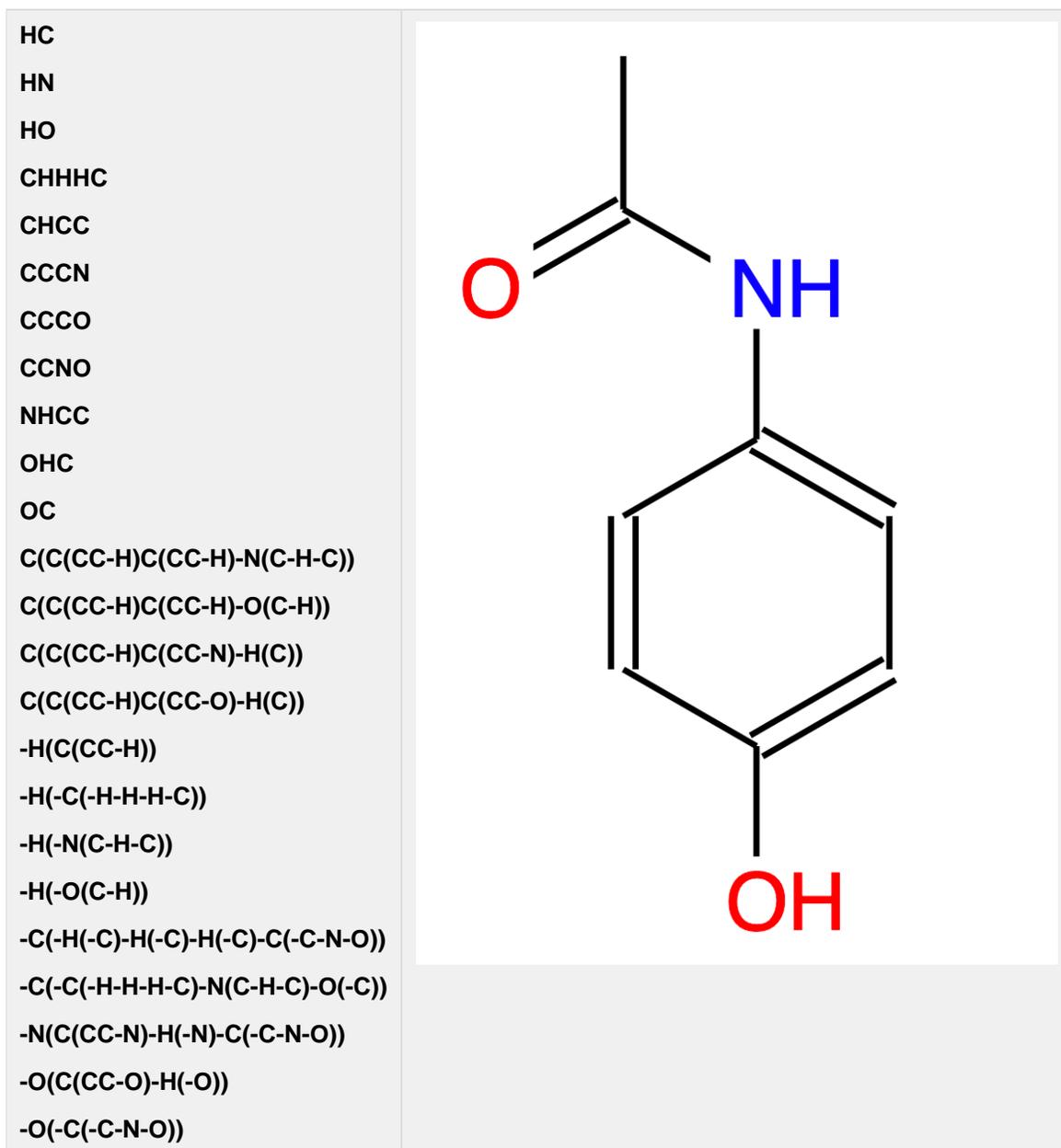


MNA

The **Multilevel Neighbourhoods of Atoms (MNA)** descriptors represent a molecule as a set of character strings. An example of MNA descriptors is given below for **Paracetamol**:



An important feature of the MNA descriptors is that they are constructed directly using the structural formula rather than a prescribed list of structural fragments. Yet another feature of these descriptors consists in that they retain the integrity of structural fragments in the sense that for each MNA descriptor the researcher can draw the corresponding structural fragment provided some skill.

The 2D structural formulae of compounds were chosen as the basis for *description of chemical structure* because this is the only information available in the early stage of research. The MNA descriptors are based on the molecular

structure representation, which includes the hydrogens according to the valences and partial charges of other atoms and does not specify the types of bonds.

The MNA descriptors are generated as recursively defined sequence:

- zero-level MNA descriptor for each atom is the mark **A** of the atom itself;
- any next-level MNA descriptor for the atom is the sub-structure notation $A(D_1D_2...D_i...)$,

where D_i is the previous-level MNA descriptor for i -th immediate neighbour's of the atom **A**.

The mark of atom may include not only the atomic type but also any additional information about the atom. In particular, if the atom is not included into the ring, it is marked by "-". The neighbour descriptors $D_1D_2...D_i...$ are arranged in unique manner, e.g., in lexicographic order. Iterative process of MNA descriptors generation can be continued covering first, second, etc. neighbourhoods of each atom.

The molecular structure is represented by the set of unique MNA descriptors of the 1st and 2nd levels. Since MNA descriptors do not represent the stereochemical peculiarities of a molecule, the substances whose structures differ only stereochemically, are formally considered as equivalent.

The MNA descriptors (for prediction of activity spectra or for adding substances to SAR Base) are generated only if structure corresponds to the following criteria:

- each of the atoms in a molecule must be presented by atom symbol from the periodic table. Symbols of unspecified atom A, Q, *, or R group labels are not allowed;
- each of the bonds in a molecule must be covalent bond presented by single, double or triple bond types only;
- structure must include three or more carbon atoms;
- structure must include only one component. Single atom parts like HCl, Cl⁻, OH⁻, Na⁺, etc., (hydrogen atoms do not take into account) are excluded from MNA descriptors generation;
- structure (main part, see the previous sentence) must be uncharged;
- the absolute molecular weight of a substance must be less than 1250.

REFERENCES

1. V. V. Poroikov, D. A. Filimonov, T. A. Glorizova, A. A. Lagunin, D. S. Druzhilovskiy, A. V. Rudik, L. A. Stolbov, A. V. Dmitriev, O. A. Tarasova, S. M. Ivanov, P. V. Pogodin. Computer-aided prediction of biological activity spectra for organic compounds: the possibilities and limitations. *Russ. Chem. Bull., Int. Ed.*, 2019, **68**, 2143–2154. <https://doi.org/10.1007/s11172-019-2683-0>
2. D. A. Filimonov, A. A. Lagunin, T. A. Glorizova, A. V. Rudik, D. S. Druzhilovskii, P. V. Pogodin, V. V. Poroikov. Prediction of the Biological Activity Spectra of Organic Compounds Using the Pass Online Web Resource. *Chemistry of Heterocyclic Compounds*, 2014, **50**, 444-457. <https://doi.org/10.1007/s10593-014-1496-1>
3. Filimonov D.A., Poroikov V.V. Probabilistic approach in activity prediction. In: *Chemoinformatics Approaches to Virtual Screening*. Eds. Alexandre Varnek and Alexander Tropsha. Cambridge (UK): RSC Publishing, 2008, 182-216.
4. Lagunin A., Stepanchikova A., Filimonov D., Poroikov V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*, 2000, **16**, 747-748. <https://doi.org/10.1093/bioinformatics/16.8.747>
5. Poroikov V. V., Filimonov D. A., Borodina Yu. V., Lagunin A. A., Kos A. Robustness of Biological Activity Spectra Predicting by Computer Program PASS for Noncongeneric Sets of Chemical Compounds. *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 1349-1355. <https://doi.org/10.1021/ci000383k>
6. Filimonov D., Poroikov V., Borodina Yu., Glorizova T. Chemical Similarity Assessment through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 666-670. <https://doi.org/10.1021/ci980335o>