View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Kazan Federal University Digital Repository

V. Solov'ev [1]
A. Solovev [2]

3D MOLECULAR FRAGMENT DESCRIPTORS FOR STRUCTURE-PROPERTY MODELING

[1] Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Leninskiy prosp., 31, 119071 Moscow, Russia;
[2] School of Software Engineering, Faculty of Computer Science, National Research University Higher School of Economics, Moscow, Russia;
solovev-vp@mail.ru

2D molecular fragment descriptors are one of the most important types of descriptors used for design of substances with the desired properties by Quantitative Structure Property Relationship (QSPR) modeling [1]. Despite many advantages of 2D fragment descriptors, there is a problem of modeling of molecular properties, which depend on stereochemical molecular structure [1]. In this work for the QSPR modeling of various physical and chemical properties, as well as biological activity, the problem is overcome by developing of new 3D molecular fragment descriptors taking into account spatial arrangement of fragment atoms.

Initially similar to 2D fragment descriptors [2], 3D fragment descriptors were generated as subgraphs of molecular graph of 3D molecular structure. For every pair of atoms in molecule, the shortest chains of bonded atoms were chosen as subgraphs. Then bond and dihedral angles of the generated fragments of molecule were evaluated. For the limitation of the number of generated 3D fragments, both of bond and dihedral angles were categorized into basic groups. For the identification of 3D fragment descriptors, the unique codes were generated, which for every chain include data concerning of atoms, bonds, and bond and dihedral angles. 3D molecular fragment descriptors recognize quite clearly different stereoisomeric molecular forms and conformers. Generation and coding of new 3D fragment descriptors were realized in the computer program *mfSpace* (*M*olecular *F*ragments *Space*) which uses the Singular Value Decomposition for Multiple Linear Regression analysis as machine learning method.

New 3D fragment descriptors have been applied to discriminate between stereoisomers in predictive QSPR modeling of the standard free energy ($\Delta G°$) for the 1:1 inclusion complexation of 76 chiral guests with β-cyclodextrin and 40 chiral guests with 6-amino-6-deoxy-β-cyclodextrin in water at 298 K. QSPR predictions of the free energies for the inclusion complexation were performed considering different modeling strategies: the chiral guest structures were represented with explicit and implicit hydrogen atoms; both pure 3D fragments and different combinations of the 2D and 3D fragments as descriptors were tested; predictive performance of consensus models was estimated both without and with the models applicability domain. The models based on the compositions of the 2D and 3D fragment descriptors discriminate between the two stereoisomers predicting the different free energies for antipodal guests. The compositions of the 2D and 3D fragment descriptors with explicit hydrogen atoms provided the best predictions in external 5-fold cross-validation: *RMSE* = 1.1 kJ/mol and R_{det}^2 = 0.918 (β-cyclodextrin), *RMSE* = 0.89 kJ/mol and R_{det}^2 = 0.910 (6-amino-6-deoxy-β-cyclodextrin).

1. Baskin I. et al. *In Chemoinformatics Approaches to Virtual Screening*, Cambridge: RCS Publishing, 2008: 1 – 43.
2. Solov'ev V. et al. *Chem. Inf. Comput. Sci.*, 2000, **40:** 847-858.