Kazan Summer School on Chemoinformatics

PROGRAM
&
ABSTRACTS

KSSCI KAZAN 2013

August 26-29, 2013
KAZAN, RUSSIA

KAZAN (VOLGA REGION) FEDERAL KFU 1804 UNIVERSITY

GENERAL INFORMATION

ORGANIZERS

Kazan (Volga region) Federal University
Russian Foundation for Basic Research
Russian Section of the International Cheminformatics and QSAR Society
D.I. Mendeleev Chemical Society of Republic of Tatarstan

SPONSOR

Chemical Abstract Service (https://www.cas.org/)

ORGANIZING COMMITTEE

Chairmen of the committee:
Prof. Alexandre Varnek (Strasbourg, France)
Prof. Igor S. Antipin (Kazan, Russia)
Prof. Vladimir V. Poroikov (Moscow, Russia)

Scientific Secretary: **T.I. Madzhidov** (Kazan, Russia)

Members of the Committee:

N.I. Ivanova	M.A. Kazymova	V.A. Burilov
M.D. Misarov	P.O. Sidorov	A.R. Kurbangalieva
D.R. Chubukaeva	A.V. Bodrov	R.I. Nougmanov
T.R. Gimadiev	A.S. Petrovsky	

SCIENTIFIC PROGRAMME

The programme of the First Kazan Summer School on Chemoinformatics includes 10 lectures, 2 plenary and 6 oral reports, 4 tutorials and 24 poster presentations.

OFFICIAL LANGUAGE

The official School language is English. No translation will be provided.

OFFICE OF THE ORGANIZING COMMITTEE

The office of the organizing committee is located at **A.M. Butlerov Chemical Institute** building, **Auditorium No. 218**. Participants have the ability to use the telephone and the Internet there.

VENUE

The event will be held mainly in the **A.M. Butlerov Chemical Institute** building (Lobachevskogo St. 1). Lectures, plenary and oral presentations will take placeinthe **Hall No. 319**. Tutorials will be held in **Hall No. 401**. Only a few stand-alone computers will be installed for the usage by participants. However Wi-Fi connection for personal laptops of the participants is provided.

Substructural molecular fragments in consensus QSPR modeling

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

[1] Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Leninskiy avenue, 31/4, 119071 Moscow, Russia
[2] Laboratoire d'Infochimie, UMR 7177 CNRS, Université de Strasbourg, 4, rue B. Pascal, Strasbourg, 67000, France
solovev-vp@mail.ru

Here we report about Substructural Molecular Fragments (SMF) as descriptors for QSPR modeling, their application for ensemble QSPR modeling with combined applicability domain approach, and some tools for compound design and property optimization on the basis of SMF and their contributions in individual models.

QSPR modeling with SMF descriptors is performed using computer programs and tools, which were developed for Windows operating systems: the ISIDA/QSPR program for Ensemble Multiple Linear Regression Analysis for QSPR modeling, data manager EdiSDF, which is editor of MDL Structure Data Files and 2D sketcher EdChemS, which is editor of 2D chemical structures.

SMF are subgraphs of molecular graphs [1]. The counts of SMF of the graphs are descriptor values. Two principal classes of SMF are generated: sequences or topological paths and augmented atoms or atoms with nearest neighbors. The minimal and maximal lengths of sequences are varied from 2 to 15. They represent shortest paths or all paths, those with explicit representation of atoms and bonds or terminal groups as paths defined by length and explicit identification of terminal atoms and bonds. In SMF, atom has different attributes: it can be presented by element symbol only, or atomic hybridization can be taken into account, or Bensons' scheme can be used, where some atomic groups are presented as extended atoms, or atom can be labeled for indication of reaction center and a property of selected atom. Similar molecular fragments belonging to two different regents are considered as different. Hereby, labeled units can be fragments and atoms. Chemical bonds have also several attributes: types for compounds, complexes and reactions; order (single, double, triple and aromatic etc.); cyclic or acyclic for taking into account of topology of 2D structure.

The ISIDA/QSPR program with combined forward and backward stepwise variable selection techniques can generate thousands of linear relationships (individual models) between dependent variable (here, property, activity) and independent variables (here, SMF descriptors) using hundreds of SMF types and several forward variable selection techniques. The property is reliable predicted as an arithmetic mean of values obtained by individual models excluding those leading to outlying values and being outside applicability domains of individual models.

SMF and their contributions in the QSPR models are convenient building blocks for compound design and property optimization. For this aim, ISIDA Predictor (*http://infochim.u-strasbg.fr/cgi-bin/predictor.cgi*), generator of virtual combinatorial libraries, interactive compound designer [2], and a tool for coloring of atoms of 2D structure according fragment contributions of individual models have been elaborated.

The programs have been developed in the framework of the ISIDA project. It is collaborative project between the Laboratory of Chemoinformatics under the direction of Prof. A. Varnek and the Laboratory of New Physico-Chemical Problems under the direction of Academician Aslan Tsivadze. The programs are available for the end users on the web sites via the internet: http://infochim.u-strasbg.fr/spip.php?rubrique53 or http://vpsolovev.ru/programs/.

[1] Varnek A., Fourches D., Hoonakker F., Solov'ev V. *J. Comp.-Aided Mol. Design,* 2005, 19, 693-703.
[2] Solov'ev V., Sukhno I., Buzko V., Polushin A., Marcou G., Tsivadze A., Varnek A. *J. Incl. Phenom. Macrocycl. Chem.*, 2012, 72, 309-321.

VS thanks Organizers of Kazan Summer School on Chemoinformatics for support.