References: [1] Nasal, A.; Siluk, D.; Kaliszan, R. Chromatographic retention parameters in medicinal chemistry and molecular pharmacology. *Curr. Med. Chem.* **2003**, *10,* 381–426; [2] Head, G.A.; Mayorov, D.N. Imidazoline receptors, novel agents and therapeutic potential. *Cardiovasc. Hematol. Agents Med. Chem.* **2006**, *4,* 17-32.

## P50

## STRUCTURAL CONNECTIVITY FINGERPRINTS – A NEW WAY TO REPRESENT AND CLASSIFY COMPOUNDS

**Krzysztof Rataj**[1], Wojciech Czarnecki[2], Sabina Podlewska[1,3], Andrzej J. Bojarski[1]

*[1]Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences,12 Smętna Street, 31-343 Cracow, Poland; [2]Faculty of Mathematics and Computer Science, Jagiellonian University, 6 Łojasiewicza Street, 30-348 Kraków, Poland; [3]Faculty of Chemistry Jagiellonian University 3 Ingardena Street 30-060 Krakow*

The drug design community is still struggling with the problem of proper representation of chemical compounds. One of such approaches is substructure key based fingerprints, which depict existence and number of pre-defined chemical groups within the target compound. There are multiple fingerprints using various sets of such keys, e.g. MACCS, Substructure FP, Klekota-Roth FP, CACTVS FP. As the information content of these fingerprints differs, they all share a major disadvantage: the relative positions of the substructures are not encoded by any means, which can lead to two significantly different compounds sharing almost identical fingerprint. To address this issue, we have designed a new method of compound representation: the Substructural Connectivity Fingerprint (SCFP). This new approach uses substructure keys definitions from some of the well-established methods mentioned before, however it adds additional information about the internal connectivity of those groups. In this way, the compound is represented more accurately, which in turn enables more efficient classification of screened chemicals. The SCFP excelled at machine learning classification tests, with balanced accuracy measure by a couple percentage points higher than state of the art fingerprints. However, we have also adapted a novel machine learning methodology called Extreme Entropy Machines (EEM)[1] to further increase the screening efficacy. This has also increased the classification score in numerous cases, while behaving more consistently between different targets and different substructure key sets. We believe, that the SCFP combined with EEM may have a huge impact on the process of drug discovery.

References: Czarnecki WM, Tabor J: Extreme Entropy Machines: Robust information theoretic classification. 2015, Pattern Anal. Appl.

Acknowledgments: The study was partially supported by the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project PLATFORMex (Pol-Nor/198887/73/2013).

## P51

## PROTEIN BINDING SITES DETERMINATION BY USING VARIOUS MACHINE LEARNING METHODS

**Georgina Mirceva**[1], Andreja Naumoski[1], Andrea Kulakov[1]

*Faculty of computer science and engineering, Ss. Cyril and Methodius University in Skopje, Skopje, Macedonia*

We present an approach for determination of the protein binding sites, so latter these predictions could be used to perform functional annotations of the protein structures based on the characteristics of the