Journal of **Chem**informatics

POSTER PRESENTATION

Open Access

A large scale classification of molecular fingerprints for the chemical space representation and SAR analysis

Fabian López-Vallejo[*], Jacob Waddell, Austin B Yongye, Richard A Houghten, José L Medina-Franco

Fingerprint-based structure representation has a broad range of applications including, but not limited to, diversity analysis, compound classification, chemical space visualization [1], activity landscape modelling and similarity searching. It has been shown that depending on the particular fingerprints used, the outcome of similarity searching [2] or activity landscapes [3] can be very different. Combining structure representations is a common practice to increase the performance of similarity searching [4]. Also, combining representations for activity landscape modelling has been proposed to generate robust descriptive SAR models [5]. However, the selection of fingerprints to be combined is not an easy task. As part of our efforts to select fingerprint representations to generate consensus representations of chemical space and activity landscapes [5,6] herein we discuss the results of a systematic comparison of more than 10 2D and 3D fingerprint representations in terms of performance in diversity analysis (as opposed to similarity searching). We employed more than 20 data sets from different sources relevant to drug discovery. In this work the widely used Tanimoto coefficient was employed. The approach presented here can be easily extended to other similarity measures, additional fingerprints and molecular databases. We also discuss the typical mean/median similarity values of selected fingerprints across databases from different sources.

Published: 1 May 2012

References
1. Medina-Franco JL, Martínez-Mayorga K, Giulianotti MA, Houghten RA, Pinilla C: **Visualization of the Chemical Space in Drug Discovery.** *Curr Comput-Aided Drug Des* 2008, **4**:322-333.
2. Bender A: **How Similar Are Those Molecules after All? Use Two Descriptors and You Will Have Three Different Answers.** *Expert Opin Drug Discovery* 2010, **5**:1141-1151.
3. Wassermann AM, Wawer M, Bajorath J: **Activity Landscape Representations for Structure-Activity Relationship Analysis.** *J Med Chem* 2010, **53**:8209-8223.
4. Chen B, Mueller C, Willett P: **Combination Rules for Group Fusion in Similarity-Based Virtual Screening.** *Mol Inf* 2010, **29**:533-541.
5. Yongye A, Byler K, Santos R, Martínez-Mayorga K, Maggiora GM, Medina-Franco JL: **Consensus Models of Activity Landscapes with Multiple Chemical, Conformer and Property Representations.** *J Chem Inf Model* 2011, **51**:1259-1270.
6. López-Vallejo F, Nefzi A, Bender A, Owen JR, Nabney IT, Houghten RA, Medina-Franco JL: **Increased Diversity of Libraries from Libraries: Chemoinformatic Analysis of Bis-Diazacyclic Libraries.** *Chem Biol Drug Des* 2011, **77**:328-342.

Torrey Pines Institute for Molecular Studies, Port St. Lucie, Florida, 34987, USA

ChemistryCentral