Accepted Manuscript

Interaction prediction in structure-based virtual screening using deep learning

Adam Gonczarek, Jakub M. Tomczak, Szymon Zaręba, Joanna Kaczmar, Piotr Dąbrowski, Michał J. Walczak

PII: S0010-4825(17)30297-4

DOI: 10.1016/j.compbiomed.2017.09.007

Reference: CBM 2778

To appear in: *Computers in Biology and Medicine*

Received Date: 31 January 2017

Revised Date: 22 August 2017

Accepted Date: 8 September 2017

Please cite this article as: A. Gonczarek, J.M. Tomczak, S. Zaręba, J. Kaczmar, P. Dąbrowski, Michał.J. Walczak, Interaction prediction in structure-based virtual screening using deep learning, *Computers in Biology and Medicine* (2017), doi: 10.1016/j.compbiomed.2017.09.007.

Interaction Prediction in Structure-based Virtual Screening using Deep Learning

Adam Gonczarek[a,c,*], Jakub M. Tomczak[a,*], Szymon Zaręba[a,c,*], Joanna Kaczmar[a,*], Piotr Dąbrowski[a,b,*], Michał J. Walczak[c,*]

[a] *Department of Computer Science, Wrocław University of Science and Technology*
[b] *Indata SA, Wrocław, Poland*
[c] *Alphamoon, Wrocław, Poland*

Abstract

We introduce a deep learning architecture for structure-based virtual screening that generates fixed-sized fingerprints of proteins and small molecules by applying learnable atom convolution and softmax operations to each molecule separately. These fingerprints are further non-linearly transformed, their inner product is calculated and used to predict the binding potential. Moreover, we show that widely used benchmark datasets may be insufficient for testing structure-based virtual screening methods that utilize machine learning. Therefore, we introduce a new benchmark dataset, which we constructed based on DUD-E, MUV and PDBBind databases.

Keywords: Virtual Screening, Neural fingerprint, Graph convolution, Deep Learning, PDBBind, DUD-E, MUV

1. Introduction

Virtual screening is one of the leading methods in computational drug discovery, which aims at identification of novel small molecules that are capable of binding a drug target, usually a protein. In general, there are two main approaches of virtual screening, ligand-based (LBVS, Ligand-Based Virtual Screening) and structure-based (SBVS, Structure-Based Virtual Screening). Ligand-based virtual screening relies on empirically established data, which

*Corresponding author, Tel.: (+48) 71 320 44 53.
Email address: adam.gonczarek@pwr.edu.pl (Adam Gonczarek)