

ORAL PRESENTATION

Open Access

Integrating logic-based machine learning and virtual screening to discover new drugs

Christopher R Reynolds*, Michael JE Sternberg

From 7th German Conference on Chemoinformatics: 25 CIC-Workshop
Goslar, Germany. 6-8 November 2011

Investigational Novel Drug Discovery by Example (INDDEx™) is a technology developed to guide hit to lead discovery by learning rules from existing active compounds that link activity to chemical substructure. INDDEx is based on Inductive Logic Programming [1], which learns easily interpretable *qualitative* logic rules from active ligands that give an insight into chemistry, relate molecular substructure to activity, and can be used to guide the next steps of drug design chemistry. Support Vector Machines weight the rules to produce a *quantitative* model of structure-activity relationships. Whereas earlier testing [2,3] was performed on single dataset examples, this talk presents the largest and fullest test of the method. The method was benchmarked

on the Directory of Useful Decoys (DUD) datasets [4], using the same methodology described in the paper on the assessment of LASSO [5] and DOCK. For each of the DUD datasets, the known active ligands were mixed with all the decoy compounds in DUD, and the retrieval rates of INDDEx and DUD were measured when they were trained on 2, 4, and 8 of the known active ligands (Figure 2). Early retrieved compounds showed high topological differences to molecules used as training data, showing the strength of this method for scaffold hopping. This work was supported by a BBSRC case studentship with Equinox Pharma Ltd (<http://www.equinoxpharma.com>).

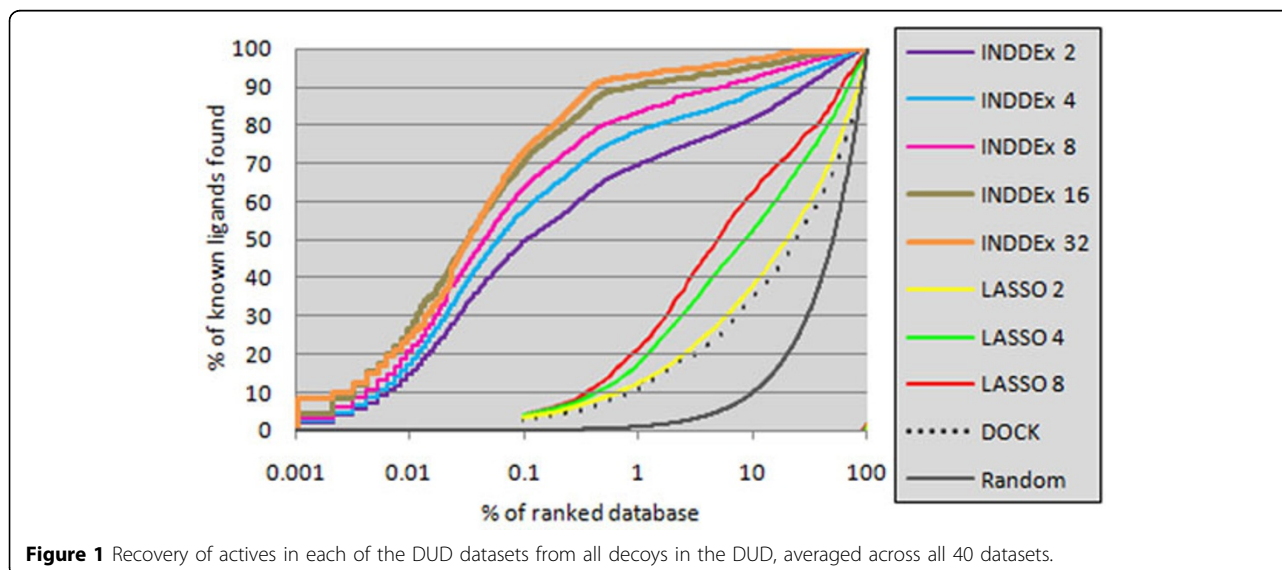


Figure 1 Recovery of actives in each of the DUD datasets from all decoys in the DUD, averaged across all 40 datasets.

* Correspondence: cr308@imperial.ac.uk
Imperial College, London, N19 3LA, UK

Published: 1 May 2012

References

1. Muggleton SH: **Inductive logic programming.** *New Gen Comp* 1995, **13**:245-286.
2. Amini A, et al: **A Novel Logic-Based Approach for Quantitative Toxicology Prediction.** *J Chem Inf Model* 2007, **47**:998-1006.
3. Cannon EO, et al: **Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds.** *J Comput Aided Mol Des* 2007, **21**:269-280.
4. Huang N, et al: **Benchmarking Sets for Molecular Docking.** *J Med Chem* 2006, **49**:6789-6801.
5. Reid D, et al: **LASSO-ligand activity by surface similarity order: a new tool for ligand based virtual screening.** *J Comput Aided Mol Des* 2008, **22**:479-487.

doi:10.1186/1758-2946-4-S1-O10

Cite this article as: Reynolds and Sternberg: Integrating logic-based machine learning and virtual screening to discover new drugs. *Journal of Cheminformatics* 2012 **4**(Suppl 1):O10.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

“Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge.”

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>


ChemistryCentral