In Silico Lead Discovery by Integrating Ligand Screening and Chemical Synthesis Rules

Christopher R Reynolds, Stephen H Muggleton and Michael JE Sternberg

Imperial College London

Department of Life Sciences, Imperial College,
and Department of Computing, Imperial College

INDDEx: Logic-based drug discovery

- Technology based on a ligand-based virtual screening method, INDDEx: **I**nvestigational **N**ovel **D**rug **D**iscovery by **Ex**ample.
- INDDEx uses a combination of Inductive Logic Programming and Support Vector Machines to derive a predictive activity model of ***qualitative*** logical rules with ***quantitative*** weighting.
- INDDEx takes as input a set of active and inactive molecules and learns a QSAR based on a series of logic-based rules relating activity to the presence of two sub-pharmacophores and their separation.
- The QSAR screens a library of molecules to find novel hits.
- INDDEx was benchmarked on 40 drug targets using the Database of Useful Decoys (DUD[4]). The added value of sampling the top x% of predicted hits over random selection of x% of database molecule is quantified by an enrichment factor E(x). In a cross-validated study[2], INDDEx (learning from 8 actives) obtained E(0.1%) of more than 600 fold.

INDDEx for Lead Discovery

- An approach was developed to predict active leads that could be synthesised via a one-step addition reaction (see Figure 1).
- Computational rules for 26 two-reactant addition reactions were taken from the ChemAxon Reactor[3] library.
- For each of the 40 DUD targets, INDDEx rules were learnt from 8 randomly-selected active molecules and used to screen for putative actives within the ZINC fragment database[4] (474,770 fragments) with any molecules similar to the DUD training sets (Tanimoto Coefficient (TC) > 0.5) removed.
- Virtual product libraries were generated around the putative actives, and their activity evaluated by the INDDEx QSAR This was done with and without a method to select putative actives based on rule fulfilment requirements (see Figure 2).
- From the 40 targets, there was at least one product molecule deemed to be active by a TC > 0.7 for 4 targets and by a TC > 0.6 for 7 targets.

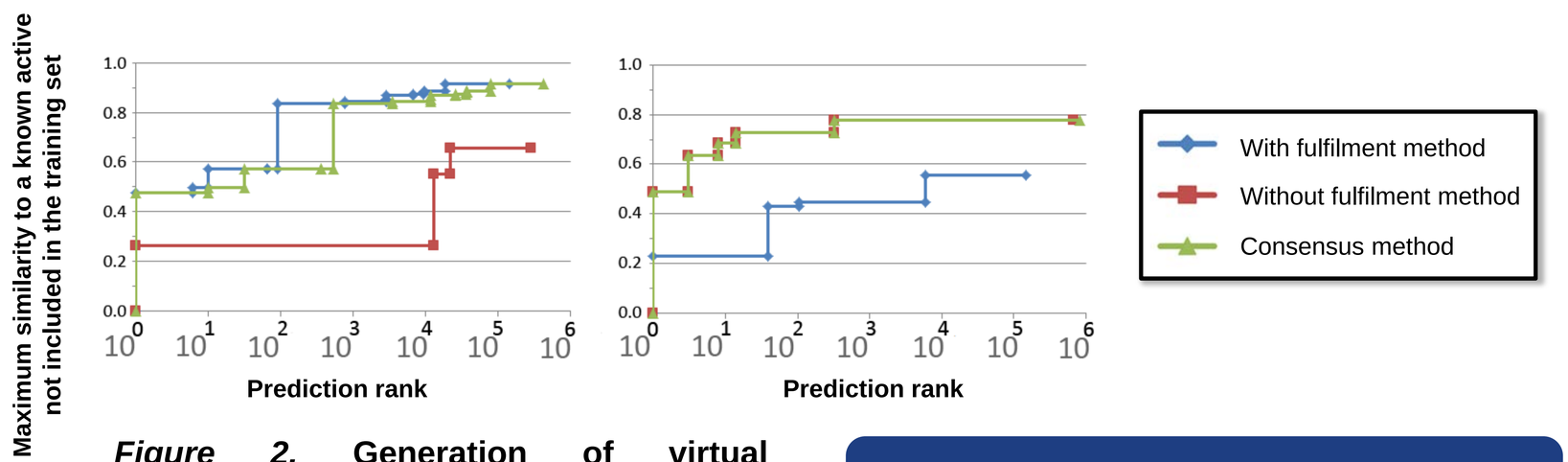

Figure 2. **Generation of virtual compounds similar to known actives for the COX-2 and PPAR γ targets.**

Figure 1. **Diagram showing the process of logic-based drug discovery.**

References

1. Huang, N, *et al. J Med Chem* **49**, 6789-6801 (2006).
2. Reynolds, CR, Amini, AC, Muggleton, SH & Sternberg, MJE. *J Phys Chem B* **116**, 6732-6739 (2012).
3. Pirok, G, *et al. J Chem Inf Model* **46**, 563-568, (2006).
4. Irwin, JJ & Shoichet, BK. *J Chem Inf Model* **45**, 177-182 (2005).
5. Reynolds, CR, Amini, AC, Muggleton, SH & Sternberg, MJE. *Mol Inform* **Online** (2015).