

Evaluation and expression of “drug-likeness” of small molecules using Semantic Web technologies

Joo-Young Choi¹
j.choi@imb.uq.edu.au

Melissa J. Davis¹
m.davis@imb.uq.edu.au

Mark A. Ragan¹
m.ragan@imb.uq.edu.au

¹ The University of Queensland, Institute for Molecular Bioscience, ARC Centre of Excellence in Bioinformatics, Brisbane, QLD 4072, Australia

Abstract

To be effective as an orally administered drug, a chemical compound must be both pharmacologically active, and exhibit an appropriate pharmacokinetic profile as described by measures of absorption, distribution, metabolism and excretion (ADME). Programs of drug design and development are therefore usually guided by criteria (*e.g.* the number and types of atoms, molecular weight, and relative solubility) that correlate with solubility and permeability. One such set of criteria, known as Lipinski’s Rule of Five, is widely used to assess the “drug-likeness” of a molecule.

We have developed an ontology for representing biologically relevant attributes of small molecules (including drugs and drug candidates) using the Semantic Web technologies Web Ontology Language (OWL) and Resource Description Framework (RDF). In this poster we describe the subset of our small molecule ontology that describes physical attributes of small molecules, and instance it with chemicals from publicly available small-molecule databases. We apply Lipinski’s Rule of Five in the form of a Simple Protocol and RDF Query Language (SPARQL) query to extract molecules with drug-like properties. We also demonstrate that Lipinski’s Rule can be constructed as a Semantic Web Rule Language (SWRL) rule.

We constructed known and unknown drug dataset from DrugBank and PubChem, and used Protégé as an ontology editor to run the SPARQL queries. Each dataset from 1000 small molecules separately was used to instance our ontology. The query implementing Lipinski’s Rule retrieved 787 and 853 of these small molecules from the true-drug and unknown-drug datasets respectively. Thus OWL, RDF, SPARQL and SWRL can effectively express complex concepts like “drug-likeness” and can be used to retrieve drug-like molecules from an ontology-based representation of publicly available data.

Keywords: Small molecules, Semantic Web, Ontology, SPARQL, drug-likeness

1 Introduction

Interactions between small molecules and their biological targets play key roles in cellular biology, and are of interest in the pharmaceutical industry. Our research aims to represent these data using Semantic Web technologies [1-3]. Lipinski’s Rule of Five specifies that poor absorption and permeation are more likely when a small molecule possesses more than five hydrogen-bond donors, has a molecular weight over 500 g/mol, has a LogP over 5, and contains more than 10 hydrogen-bond acceptors [6]. We implemented the rule as a SPARQL query and as a SWRL rule to identify drug-like molecules in our data [4-5].

2 Materials and Methods

We integrated data from DataBank and PubChem on 1,000 small molecules as our known-drug dataset, and

used these data to instance our Small-Molecule Ontology (SMO) [7-10]. We used an ontology management application, Protégé, as an ontology editor to create our SMO and to run the SPARQL queries. The query implementing Lipinski's rule was executed over these two datasets.

3 Results

We ran a SPARQL query implementing Lipinski's Rule of Five over 1000 small molecules in the true-drug dataset to instance our SM ontology. We consider small molecules retrieved by the query to be predicted drug-like molecules according to the rule. The query retrieved 787 small molecules from the 1000 in true-drug dataset. We also executed the query over the set of 1000 molecules of unknown drug status. Of these small molecules, 853 were predicted by the query to have drug-like properties.

We expressed Lipinski's Rule of Five as a SWRL rule. In our SMO, a small-molecule class is an OWL class, and has several properties relevant to its physical properties including *hasMolecularWeight*, *hasXLogP*, *hasHydrogenBondDonors*, and *hasHydrogenBondAcceptors*. These are the key values to assess drug-likeness of small molecules according to Lipinski's Rule.

4 Conclusion

The goal of this research has been to demonstrate the applicability of Semantic Web technologies for computational reasoning in drug discovery. We have applied Lipinski's Rule to find 'drug-like' molecules. The rule does not retrieve 'known' drug-like molecules with 100% accuracy, but its high success rate (nearly 80%) indicates that the rule can be a valuable tool. This work demonstrates the capacity of Semantic Web technologies to manipulate chemical data with potential application in pharmacogenomic research.

References

- [1] <http://www.w3c.org/2001/sw>
- [2] RDF Primer. W3C Recommendation 10 February 2004
- [3] OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004
- [4] SPARQL Query Language for RDF. W3C Recommendation 15 January 2008
- [5] SWRL: a Semantic Web Rule Language combining OWL and RuleML. W3C Member Submission 21 May 2004
- [6] Lipinski CA, Lombardo F, Dominy BW & Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* 46:3-26, 2001
- [7] Wishart DS, Knox C, Gou An Chi, Shrivastava S, Hassanali M, Stothard P, Chang Zhan & Woolsey J. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration, *Nucleic Acids Research* 34:D668-D672, 2006
- [8] Wishart DS, Knox C, Gou An Chi, Cheng D, Shrivastava S, Tzur D, Gautam B & Hassanali M. *Nucleic Acids Research* 36:D901-D906, 2008
- [9] O'Connor M, Knublauch H, Tu S, Grosz B, Dean M, Grosso W & Mususen M. Supporting rule system interoperability on the Semantic Web with SWRL. Fourth International Semantic Web Conference (ISWC2005), Galway, Ireland, 2005
- [10] Hur J & Wild DJ. PubChemSR: a search and retrieval tool for PubChem. *Chemistry Central Journal* 2:1-11, 2008