

# Predicting Chemical Properties of Small Molecules



Roman Eisner\*, Craig Knox,  
Russell Greiner, and David S. Wishart

eisner@cs.ualberta.ca

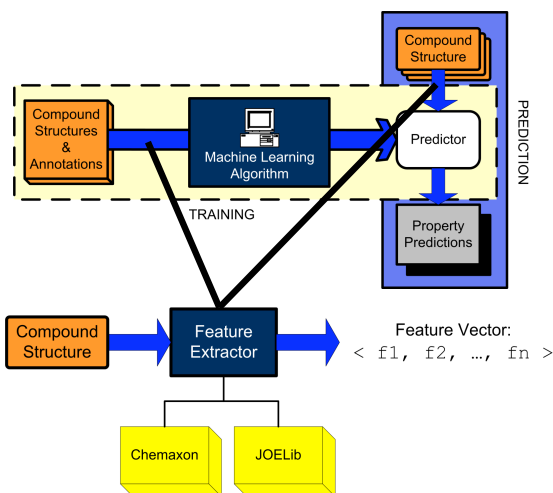
<http://hmdb.med.ualberta.ca/~eisner/PropPred/>



## Introduction

The chemical properties of molecular compounds (such as solubility, logP, and melting point) are important for both understanding, and finding uses for, these compounds. The Human Metabolome Project (<http://www.metabolomics.ca>) is therefore attempting to include as many of these properties as possible, for the entire catalogue all of the metabolites in the human body. To deal with situations where these properties have not been acquired experimentally, we provide analytic tools for estimates of their values. We present our general framework for learning functions that can predict the values of an arbitrary target property of chemical compounds. This framework involves first preprocessing each chemical in a training set (each labeled with the value of the target property) into a fixed set of features, based on the structure of the chemical, then using standard learning tools to produce a predictor that maps the features into the target property. We evaluate this system in the context of the water solubility of compounds, and other properties. We also present our online tool that can be used to predict chemical properties of compounds.

## General Predictive Framework



- General Framework
- Exhaustive Feature Set Allows for little or no human input
- Does not require domain expertise - Little or no tuning required
- Can train a predictor for properties where quality data is available
- All possible features from JOELib<sup>6</sup> and ChemAxon<sup>5</sup> tools are used
- Predictors trained using structures from PubChem<sup>4</sup>

## Acknowledgements

The Human Metabolome Project is supported by Genome Alberta, in part through Genome Canada.

- [1] N Jain and S H Yalkowsky. Estimation of the Aqueous Solubility I: Application to Organic Nonelectrolytes. *Journal of Pharmaceutical Science*, 90(2), 2001
- [2] E Rytting, K A Lentz, X Q Chen, F Qian, and S Venkatesh. Aqueous and Cosolvent Solubility Data for Drug-like Organic Compounds. *The AAPS Journal*, 7(1), 2005.
- [3] J K Wegner and A Zell. Aqueous Solubility and Partition Coefficient Optimized by a Genetic Algorithm Based Descriptor Selection Method. *J. Chem. Inf. Comput. Sci.*, 43:1077-1084, 2003.
- [4] Pubchem: (<http://pubchem.ncbi.nlm.nih.gov/>)
- [5] Marvin Beans, Version 4.0.1. ChemAxon, Budapest, Hungary: ([www.chemaxon.com/products.html](http://www.chemaxon.com/products.html))
- [6] JOELib2: (<http://www-ra.informatik.uni-tuebingen.de/software/joelib/>)
- [7] PhysProp: (<http://www.syrres.com/esc/physprop.htm>)

## Results of Property Prediction

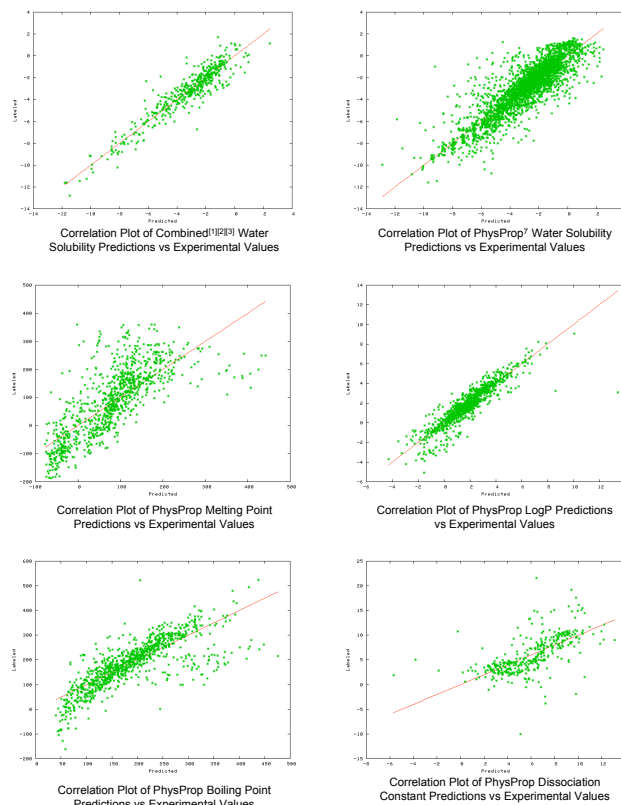


Table 1: 5-fold Cross Validation Results of Predictors

Predictor	r <sup>2</sup>	Average Error	n
H2O Solubility - Combined	0.88	0.62	463
H2O Solubility - PhysProp	0.75	0.86	2913
LogP - PhysProp	0.82	0.55	1000
pKa - PhysProp	0.43	1.98	369
Boiling Point - PhysProp	0.63	41.33	961
Melting Point - PhysProp	0.58	60.19	940

## Online Server

- Predictors are available via an online tool:  
[ <http://hmdb.med.ualberta.ca/~eisner/PropPred/> ]
- Only required input is the chemical compound's structure
- Compounds run through predictive framework
- Output is predicted target properties and calculated feature set