Bioinformatics, 2016, 1–3
doi: 10.1093/bioinformatics/btw491
Advance Access Publication Date: 2 August 2016
Applications Note

OXFORD

Systems biology

ReactPRED: a tool to predict and analyze biochemical reactions

Tadi Venkata Sivakumar[1], Varun Giri[1†], Jin Hwan Park[2], Tae Yong Kim[2,*] and Anirban Bhaduri[1,*]

[1]Bioinformatics Lab, Samsung Advanced Institute of Technology, Bangalore 560037, India and [2]Biomaterials Lab, Materials Center, Samsung Advanced Institute of Technology, Gyeonggi-Do 443803, Korea

*To whom correspondence should be addressed.
[†]Present address: Biological Experimental Physics, Saarland University, 66123 Saarbrücken, Germany
Associate Editor: Jonathan Wren

Received on April 21, 2016; revised on June 24, 2016; accepted on July 19, 2016

Abstract

Motivation: Biochemical pathways engineering is often used to synthesize or degrade target chemicals. *In silico* screening of the biochemical transformation space allows predicting feasible reactions, constituting these pathways. Current enabling tools are customized to predict reactions based on pre-defined biochemical transformations or reaction rule sets. Reaction rule sets are usually curated manually and tailored to specific applications. They are not exhaustive. In addition, current systems are incapable of regulating and refining data with an aim to tune specificity and sensitivity. A robust and flexible tool that allows automated reaction rule set creation along with regulated pathway prediction and analyses is a need. ReactPRED aims to address the same.
Results: ReactPRED is an open source flexible and customizable tool enabling users to predict biochemical reactions and pathways. The tool allows automated reaction rule creation from a user defined reaction set. Additionally, reaction rule degree and rule tolerance features allow refinement of predicted data. It is available as a flexible graphical user interface and a console application.
Availability and implementation: ReactPRED is available at: https://sourceforge.net/projects/reactpred/.
Contact: anirban.b@samsung.com or ty76.kim@samsung.com
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

In silico approaches are increasingly being used to predict biochemical pathways and reactions with an objective to synthesize or degrade chemicals (Araki *et al.*, 2015; Nakamura *et al.*, 2012; Medema *et al.*, 2012). The approach is often used to identify xenobiotic degradation pathways (Wicker *et al.*, 2016; Gao *et al.*, 2011; Dimitrov *et al.*, 2011; Finley *et al.*, 2009). Its applicability in predicting pathways for chemical synthesis is however, a maturing field (Carbonell *et al.*, 2012; Moura *et al.*, 2013; Moriya *et al.*, 2010). Its evolution is similar to 'click chemistry' (Kolb and Sharpless, 2003; Durrant and McCammon, 2012).

The current study presents ReactPRED a robust, flexible graphical user interface and a console application. The tool predicts reactions and pathways using biochemical transformation rules or reaction rules set. It is enabled to create customizable reaction rule set automatically from an input reaction set. Moreover, the tool is able to simulate pathways or reactions in a synthetic or retrosynthetic mode allowing a larger scope of applications for drug metabolism and biochemical engineering.

2 Methods

ReactPRED is a tool enabling reaction rule creation, biochemical reaction prediction, pathway prediction and analyses. It is a compilation of three independent systems namely, reaction rule creation, pathway prediction and pathway analysis. Detailed architecture is represented in the Supplementary Material S1.

Acting on an input reaction set, the reaction rule creation system extracts all bond rearrangements defining reaction or transformation center along with its neighborhood in SMIRKS notation (Warner *et al.*, 2010). Reaction rules are created from the extracted SMIRKS and stored in a reaction rule library. Identical reaction rules extracted from different reactions are merged into single rule (Supplementary Material S2). It should be noted that along with transformation information the reaction rule library records substrate mass range associated with reaction rules and commonly occurring co-substrates. Co-substrate selection for a reaction rule is based on maximum likelihood of molecules participating in reactions defining the reaction rule.

Pathway prediction system simulates and reports reactions and pathways for input molecules using a reaction rule library. The reaction library is customizable based on a user input reaction set. Alternatively a default reaction rule library derived from MetaCyc 19.1 (Caspi *et al.*, 2014) may be used. To detect potential reaction centers on an input molecule, substructures identical to regions defined in the reaction rule is identified using an in-house algorithm. On detection the input molecule is processed in accordance to the reaction rule (Supplementary Material S3).

Pathway analysis modules allow assessment of the simulation output. Feasibility of simulated pathways is assessed based on standard Gibbs free energy derived from group contribution method (Jankowski *et al.*, 2008; Noor *et al.*, 2012). For details please refer, Supplementary Material S4. Further, outputs are screened on user defined molecular substructures and mass range (Supplementary Material S5).

For brevity, discussion on the above mentioned systems are limited in the manuscript and are extended in the Supplementary Material.

3 Features and results

ReactPRED is implemented in Java 1.7 and uses chemical development kit (CDK) (Steinbeck *et al.*, 2006) and Ambit-SMARTS (Jeliazkova and Kochev, 2011) libraries for processing chemical notations.

A key challenge for reaction prediction system is reporting high coverage and data enrichment. ReactPRED is enabled by two unique features, reaction rule degree and prediction tolerance to achieve the same.

3.1 Reaction rule degree

Reaction rule degree feature describes the neighborhood around the transformation center detailed within a reaction rule. Zero degree reaction rule is constructed from only the bonds undergoing rearrangement in a reaction. An *n*th degree rule includes information from atoms at a distance of *n* atoms from transformation region (Supplementary Material S2). Increasing neighborhood information restricts search space and improves prediction specificity (or data enrichment). Lower reaction rule degree allows exploring a larger simulation space enhancing prediction sensitivity or coverage.

Effectiveness of the feature was elucidated on a data set of 1381 reactions clustered in 435 reaction groups. Each reaction group comprised of multiple reactions sharing identical EC number. Assuming transformation conservation across reactions sharing identical EC number (Finley *et al.*, 2009) predictability with change of reaction rule degree was assessed. For each reaction group a leave one out experiment was performed by constructing a reaction rule from one reaction and predicting the others. This is performed with reaction rule degrees 0–5.

Reaction rule degree influence on predictions were assessed based on coverage ($\mathbf{C} = \sum\left(\frac{r_p}{r_t}\right)/\boldsymbol{N}$) and enrichment factor ($\mathbf{EF} = \sum\left(\frac{r_p}{r_{pt}}\right)/\boldsymbol{N}$). For predictions with a reaction rule, r_t, r_{pt} and $r_p$are total reaction numbers in a group, total predicted reactions and number of reactions predicted from the group, respectively. Total number of reaction groups is represented by $\boldsymbol{N}$. It was observed that change in reaction rule degree influenced enrichment factor and coverage inversely (Fig. 1A). This enables sensitivity and specificity customization for a simulation (recommended degree is 1, degree beyond size of molecule is not advised). Manual inspection suggested that prediction failure with some specific 0 degree reaction rules was due to diverse transformations within a group.

3.2 Prediction tolerance

Prediction tolerance (Supplementary Material S6) defines the allowed substrate range for a reaction rule. Prediction tolerance for

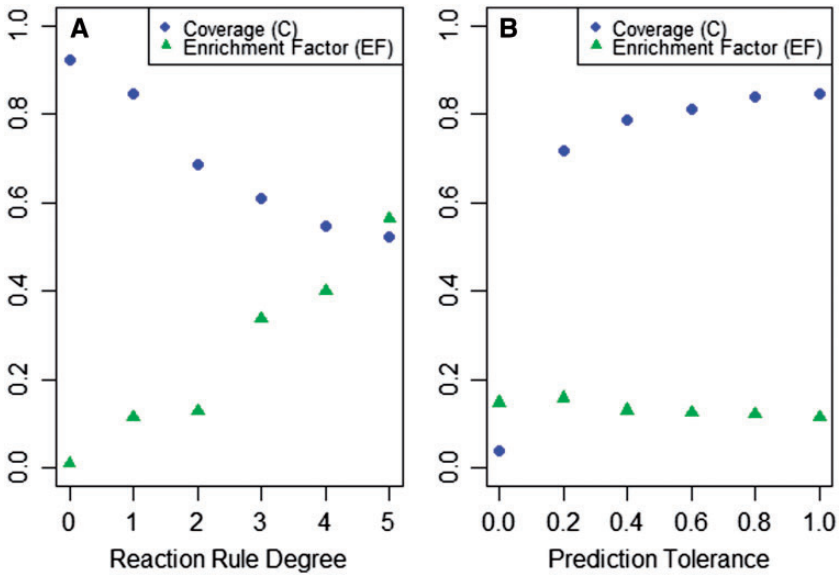

Fig. 1. [A] Influence of reaction rule degree evaluated through enrichment factor (EF) and coverage (C), tolerance was set to 1. [B] Influence of prediction tolerance evaluated through enrichment factor (EF) and coverage (C), reaction rule degree was set to 1

a reaction rule is computed and if the input molecule is within a user defined allowed threshold, it is processed.

Impact of prediction tolerance on coverage and enrichment was assessed considering the data used to analyze reaction rule degree. Results are reported in Figure 1B. It was observed that below 0.2, coverage is significantly compromised (owing to stringent threshold). Default and recommended score is 1. Influence on enrichment factor is less significant.

3.3 Click biochemistry

Similar to 'Click chemistry' systems, ReactPRED is structured in a user friendly interface to predict biochemical pathways for chemical synthesis/degradation. It may be tuned to predict pathways through interactive or automated approaches. The tool is enabled to perform systematic searches or explorative searches based on a user requirement.

Acknowledgement

Authors acknowledge support from Samsung Advanced Institute of Technology.

Conflict of Interest: none declared.

References

Araki,M. *et al*. (2015) M-path: a compass for navigating potential metabolic pathways. *Bioinformatics*, **31**, 905–911.

Carbonell,P. *et al*. (2012) Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Syst. Biol*., **6**, 10.

Caspi,R. *et al*. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*., **42**, D459–D471.

Dimitrov,S. *et al*. (2011) Simulation of chemical metabolism for fate and hazard assessment. II CATALOGIC simulation of abiotic and microbial degradation. *SAR QSAR Environ. Res*., **22**, 719–755.

Durrant,J. and McCammon,J. (2012) AutoClickChem: click chemistry in silico. *PLoS Comput. Biol*., **8**, e1002397.

Finley,S. *et al*. (2009) Computational framework for predictive biodegradation. *Biotechnol. Bioeng*., **104**, 1086–1097.

Gao,J. *et al*. (2011) The university of Minnesota pathway prediction system: multi-level prediction and visualization. *Nucleic Acids Res*., **39**, W406–W411.

Jankowski,M. *et al*. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J*., **95**, 1487–1499.

Jeliazkova,N. and Kochev,N. (2011) Ambit-SMARTS: efficient searching of chemical structures and fragments. *Mol. Inform*., **30**, 707–720.

Kolb,H. and Sharpless,K. (2003) The growing impact of click chemistry on drug discovery. *Drug Discov. Today*, **8**, 1128–1137.

Medema,M. *et al*. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol*., **10**, 192–202.

Moriya,Y. *et al*. (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res*., **38**, W138–W143.

Moura,M. *et al*. (2013) Computational tools for guided discovery and engineering of metabolic pathways. *Methods Mol. Biol*., **985**, 123–147.

Nakamura,M. *et al*. (2012) An efficient algorithm for de novo predictions of biochemical pathways between chemical compounds. *BMC Bioinformatics*, **13**, S8.

Noor,E. *et al*. (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics*, **28**, 2037–2044.

Steinbeck,C. *et al*. (2006) Recent developments of the chemistry development kit (CDK) - an open-source Java library for chemo- and bioinformatics. *Curr. Pharm. Des*., **12**, 2111–2120.

Warner,D. *et al*. (2010) WizePairZ: a novel algorithm to identify, encode, and exploit matched molecular pairs with unspecified cores in medicinal chemistry. *J. Chem. Inf. Model*., **50**, 1350–1357.

Wicker,J. *et al*. (2016) enviPath—the environmental contaminant biotransformation pathway resource. *Nucleic Acids Res*., **44**, D502–D508.