MetRxn: A Knowledgebase of Metabolites and Reactions Spanning Metabolic Models and Databases

Akhil Kumar[1,*] (azk172@psu.edu), Patrick F. Suthers[2] and Costas D. Maranas[2]

[1]Bioinformatics & Genomics, [2]Department of Chemical Engineering

The Pennsylvania State University, University Park, PA 16802

Project Goals: The goal of this project is to develop a comprehensive knowledgebase that standardizes metabolite and reaction entries consolidated from a wide range of databases and existing genome-scale metabolic models.

Increasingly, metabolite and reaction information is organized in the form of community, organism, or even tissue-specific genome-scale metabolic reconstructions. These reconstructions account for reaction stoichiometry and directionality, gene to protein to reaction associations, organelle reaction localization, transporter information, transcriptional regulation and biomass composition. A key bottleneck in the pace of reconstruction of new high quality metabolic models is our inability to directly make use of metabolite/reaction information from biological databases (e.g., BRENDA, KEGG, MetaCyc, EcoCyc, BioCyc, BKM-react, UM-BBD, Reactome.org, Rhea, PubChem, ChEBI etc.) or other models due to incompatibilities of representation, duplications and errors.

A major impediment is the presence of metabolites with multiple names across databases and models, and in some cases within the same resource, which significantly slows down the pooling of information from multiple sources. Therefore, the almost unavoidable inclusion of multiple replicates of the same metabolite can lead to missed opportunities to reveal (synthetic) lethal gene deletions, repair network gaps and quantify metabolic flows. Moreover, most data sources inadvertently include some reactions that may be stoichiometrically inconsistent and/or elementally / charge unbalanced, which can adversely affect the prediction quality of the resulting models if used directly. Finally, a large number of metabolites in reactions are partly specified with respect to structural information and may contain generic side groups (e.g., alkyl groups -R), varying degree of a repeat unit participation in oligomers, or even just compound class identification such as “an amino acid” or “electron acceptor”.

MetRxn is a knowledgebase that includes standardized metabolite and reaction descriptions by integrating information from BRENDA, KEGG, MetaCyc, Reactome.org and 44 metabolic

models into a single unified data set. All metabolite entries have matched synonyms, resolved protonation states and are linked to unique structures. All reaction entries are elementally and charge balanced. This is accomplished through the use of a workflow of lexicographic, phonetic, and structural comparison algorithms (see Figure 1). MetRxn allows for the download of standardized versions of existing genome-scale metabolic models and the use of metabolic information for the rapid reconstruction of new ones.

We describe the development and highlight applications of the web-based resource MetRxn that integrates, using internally consistent descriptions, metabolite and reaction information from 8 databases and 44 metabolic models. The MetRxn knowledgebase contains over 76,000 metabolites and 72,000 reactions (including unresolved entries) that are charge and elementally balanced. By conforming to standardized metabolite and reaction descriptions, MetRxn enables users to efficiently perform queries and comparisons across models and/or databases. For example, common metabolites and/or reactions between models and databases can rapidly be generated along with connected paths that link source to target metabolites. MetRxn supports export of models in SBML format. New models are being added as they are published or made available to us. MetRxn uses relational database models (MySQL) and is available as a web-based resource at http://metrxn.che.psu.edu.

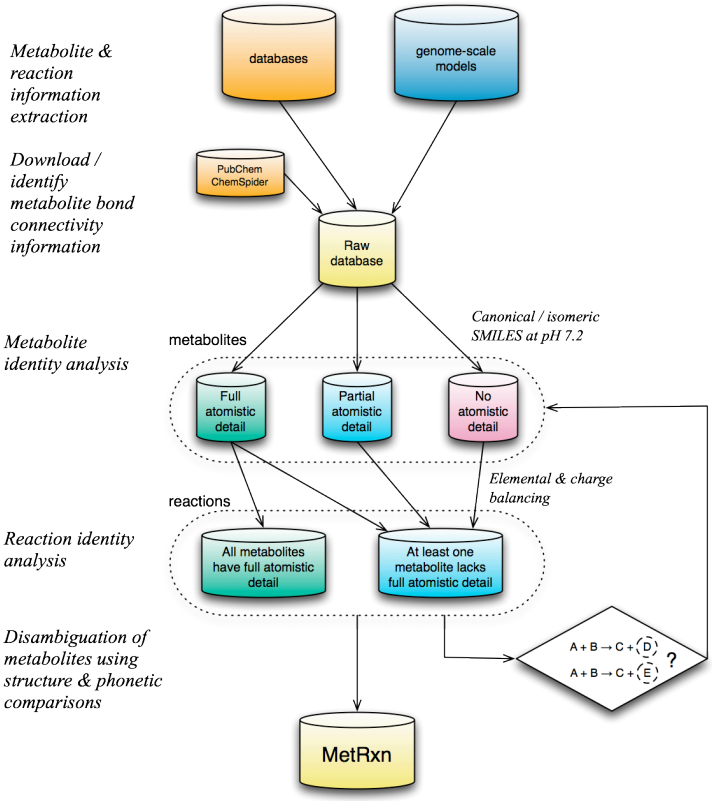

Figure 2: Flowchart outlining the construction of MetRxn.