Reaction Decoder Tool (RTD): Extracting Chemical Features from Chemical Reactions

Syed Asad Rahman[+#], Gilliean Torrance[+], Lorenzo Baldacci[+, 1], Sergio M. Cuesta[+, 2], Nimish Gopal[+], Saket Choudhary[+], John W. May[+, 3], Gemma L. Holliday[+, 4], Christoph Steinbeck[+] and Janet M. Thornton[+]

[+]European Molecular Biology Laboratory - European Bioinformatics Institute EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

Current affiliation:

1: Dipartimento di Informatica - Scienza e Ingegneria, Mura Anteo Zamboni 7, Bologna, Italy

2: Balasubramanian laboratories, Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK

3: NextMove Software Ltd, *Innovation Centre (Unit 23)*, Cambridge Science Park, Cambridge, CB4 0EY, UK

4: Bioengineering, UCSF School of Pharmacy, 1700 4th Street, San Francisco CA 94158, USA

[#] Corresponding Author: Syed Asad Rahman {asad@ebi.ac.uk, s9asad@gmail.com}

Abstract

Summary: Extracting chemical features like Atom-Atom Mapping (AAM), Bond changes (BC) and Reaction Centres (RC) from biochemical reactions helps us understand the chemical composition of enzymatic reactions. Reaction Decoder is a robust command line tool, which performs this task with high accuracy. It supports standard chemical input/output exchange formats i.e. RXN/SMILES, computes AAM, highlights bond changes and creates images of the mapped reaction. This aids in the analysis of metabolic pathways and the ability to perform comparative studies of chemical reactions based on these features.

Availability and Implementation: This software is implemented in Java, supported on Windows, Linux and Mac OSX, and freely available at https://github.com/asad/ReactionDecoder

Contact: asad@ebi.ac.uk

Introduction:

Large-scale chemical reaction databases such as KEGG (Kanehisa *et al.*, 2013), BRENDA (Chang *et al.*, 2015), Rhea (Alcántara *et al.*, 2012) and MetaCyc (Latendresse *et al.*, 2012) link reactions to enzymes and provide data-mining opportunities for novel pathways (Hatzimanikatis *et al.*, 2004; Rahman *et al.*, 2005), and the discovery of drugs (Rydberg *et al.*, 2010), natural products and green chemistry. One of the primary bottlenecks for automated analyses of these chemical reactions comes from the realizations of the imperfect quality of data, such as unmapped or unbalanced reactions. Accurate Atom-Atom Mapping (AAM) - the one-to-one correspondence between the substrate and product atoms (Gasteiger, 2003), will lead to correct prediction of bond changes (Gasteiger, 2003; Rahman *et al.*, 2014) and the ability to locate the fate of interesting atoms or substructure across the metabolic networks (Faulon and Bender, 2010), etc. Linking novel pathways or optimising pathways of biological/commercial relevance demands better understanding of metabolic routes (Rahman, 2007) and pathway annotation (May *et al.*, 2013).

Bond changes in chemical reactions refers to the cleavage and formation of chemical bonds, changes in bond order and stereo changes, which are due to chemical processes such as chiral inversions or cis-trans isomerisation(s). Cleaved and formed bonds are indicated as lines connecting atoms, for instance **C–C** means a single carbon-carbon bond that is cleaved or formed in the reaction. Bond order changes are represented as double arrows connecting bonds, for example **C–C** ↔ **C=C** means a single carbon-carbon bond turning into double carbon-carbon bond or vice versa. Stereo changes are represented as atoms that change their absolute configuration, for instance **C(R/S)** means a carbon atom that changes from R to S configuration. A reaction centre is the collection of atoms and bonds that are changed during the reaction (Warr, 2014), also known as the local atomic environment around the atoms involved in bond changes.

The primary computational approaches to measure enzyme similarity on the basis of their catalysed reactions traditionally relied upon comparing their ligands directly (Chiang *et al.*, 2008; Izrailev and Farnum, 2004; Nobeli *et al.*, 2005). Therefore methods had to be extended in order to account for the transformation between two or more molecules. Comparing and finding similar reactions and linking them to their corresponding enzymes and chemical changes (similar bond changes, reaction centres or substructures) helps enrich knowledge of enzyme superfamilies (Akiva *et al.*, 2014) and understanding the evolution of enzymes (Martínez Cuesta *et al.*, 2015; Brown and Babbitt, 2002; Holliday *et al.*, 2014). Finding similar enzyme reactions is a valuable tool for biochemists working across areas as diverse as chemical synthesis, enzyme reaction databases, enzyme design, metabolic network reconstruction and overall, discerning evolutionary relationships between enzyme sequence, structure and function.

Features:

The key features of this tool are:

a) Ability to perform Atom-Atom Mapping (AAM) on chemical reactions catalysed by enzymes (Figure 1).
b) Reaction Decoder works on both chemically balanced and unbalanced reactions. Semantically speaking, in a balanced reaction, the total number of atoms on the left (Reactant) of the equation, equal the total number of atoms on the right (Product).
c) Generates images of the mapped reactions where matching substructures are highlighted.
d) Generates reaction patterns and bond changes for input reactions.
e) The input format can be SMILES or RXN file (Gasteiger, 2003).
f) The mapped reaction can in an RXN file or SMILES with AAM information.
g) This is built upon the SMSD (Rahman *et al.*, 2009) and CDK (Steinbeck *et al.*, 2006), hence it is pure Java (7.0+) and platform independent.

Usage and applications

The EC-Blast (Rahman et al., 2014) - a novel tool to map and compare biochemical reactions, uses "Reaction Decoder" in the background to mine and extract chemical information from thousands of reactions. The success rate of mapping is more than 99% when compared to manual AAM mapping (Rahman et al., 2014). Databses like FunTree (Furnham *et al.*, 2011), MACiE (Holliday *et al.*, 2012) and Catalytic Site Atlas (CSA) (Furnham *et al.*, 2013) use EC-Blast. While the former two uses mapped chemical information to understand evolution of enzyme sequences, the later uses mapped reaction images to demonstrate chemical changes in an enzymatic reaction.

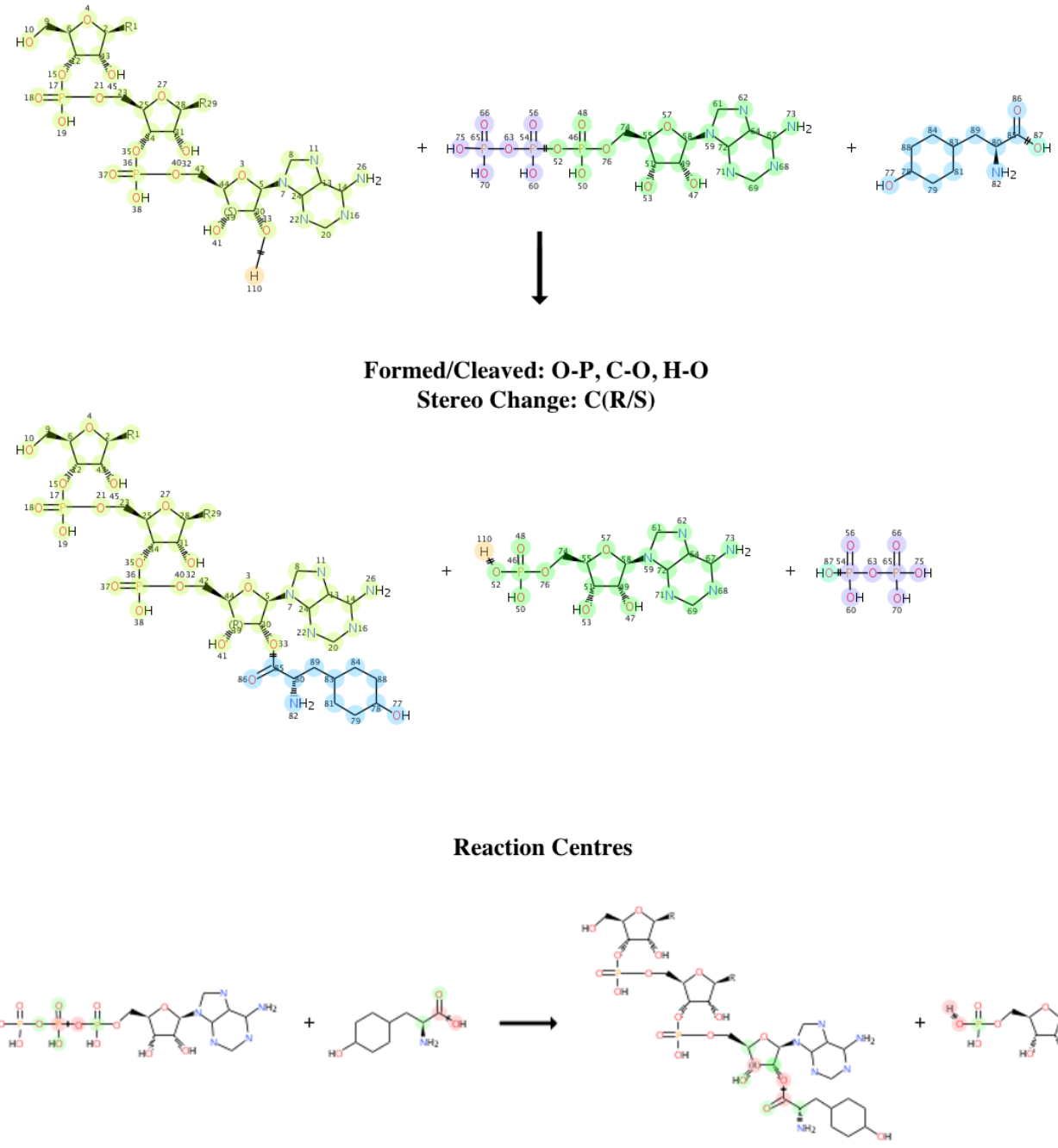

Figure 1: Atom Atom Mapping performed by the Reaction Decoder. The bond changes (Formed/Cleaved: O-P, C-O, H-O, Stereo change: C(R/S)) and reaction centres are highlighted in the images.

Conclusion

Reaction Decoder is a fast and robust tool to compute AAM and extract bond changes in chemical reactions. This is coded in java and optimised to run as a computationally asynchronous process.

Funding

Authors would like to thank EMBL for funding this project. LB was partially funded by Marie Curie fellowship.

Acknowledgements

We would like to thank the CDK team and anonymous users for their valuable feedback.

Conflict of interest

NONE

References

Akiva,E. *et al.* (2014) The Structure-Function Linkage Database. *Nucleic Acids Res.*, **42**, D521–30.

Alcántara,R. *et al.* (2012) Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **40**, D754–D760.

Brown,S. and Babbitt,P. (2002) Using the Structure-Function Linkage Database to Characterize Functional Domains in Enzymes John Wiley & Sons, Inc., Hoboken, NJ, USA.

Chang,A. *et al.* (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res.*, **43**, D439–D446.

Chiang,R.A. *et al.* (2008) Evolutionarily Conserved Substrate Substructures for Automated Annotation of Enzyme Superfamilies. *PLoS Comput Biol*, **4**, e1000142.

Faulon,J.-L. and Bender,A. (2010) Handbook of Chemoinformatics Algorithms 1st ed. Bender,A. and Faulon,J.-L. (eds) Chapman and Hall/CRC.

Furnham,N. *et al.* (2011) FunTree: a resource for exploring the functional evolution of structurally defined enzyme superfamilies. *Nucleic Acids Res.*, **40**, D776–D782.

Furnham,N. *et al.* (2013) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.*, **42**, D485–D489.

Gasteiger,J. (2003) Handbook of chemoinformatics Vch Verlagsgesellschaft Mbh.

Hatzimanikatis,V. *et al.* (2004) Metabolic networks: enzyme function and metabolite structure.

Current Opinion in Structural Biology, **14**, 300–306.
Holliday,G.L. *et al.* (2014) Exploring the biological and chemical complexity of the ligases. *J. Mol. Biol.*, **426**, 2098–2111.
Holliday,G.L. *et al.* (2012) MACiE: exploring the diversity of biochemical reactions. *Nucleic Acids Res.*, **40**, D783–9.
Izrailev,S. and Farnum,M.A. (2004) Enzyme classification by ligand binding. *Proteins*, **57**, 711–724.
Kanehisa,M. *et al.* (2013) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
Latendresse,M. *et al.* (2012) Accurate Atom-Mapping Computation for Biochemical Reactions. *J. Chem. Inf. Model.*, **52**, 2970–2982.
Martínez Cuesta,S. *et al.* (2015) The Classification and Evolution of Enzyme Function. *Biophysical Journal*, **109**, 1082–1086.
May,J.W. *et al.* (2013) Metingear: a development environment for annotating genome-scale metabolic models. *Bioinformatics*, **29**, 2213–2215.
Nobeli,I. *et al.* (2005) A Ligand-centric Analysis of the Diversity and Evolution of Protein–Ligand Relationships in E.coli. *J. Mol. Biol.*, **347**, 415–436.
Rahman,S.A. (2007) Pathway Hunter Tool (PHT)- A Platform for Metabolic Network Analysis and Potential Drug Targeting. *Ph.D Thesis, University of Cologne, Cologne, Germany*.
Rahman,S.A. *et al.* (2014) EC-BLAST: a tool to automatically search and compare enzyme reactions. *Nat Meth*, **11**, 171–174.
Rahman,S.A. *et al.* (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, **21**, 1189–1193.
Rahman,S.A. *et al.* (2009) Small Molecule Subgraph Detector (SMSD) toolkit. *J Cheminf*, **1**, 12–12.
Steinbeck,C. *et al.* (2006) Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Current pharmaceutical design*, **12**, 2111–2120.
Warr,W.A. (2014) A Short Review of Chemical Reaction Database Systems, Computer-Aided Synthesis Design, Reaction Prediction and Synthetic Feasibility. *Mol. Inf.*, **33**, 469–476.