Creating Atom-Level Chemical Reaction Networks from Biochemical Models

Casper Asbjørn Eriksen
casbjorn@imada.sdu.dk

40th TBI Winterseminar in Bled • February 2025

SDU MATOMIC novo nordisk foundation

Why?

- Recall Richard's talk

Why?

- Recall Richard's talk

Existing atom-atom mapped networks:

Paper	**# reactions**	**Curated**	**Organism**
Suthers et al., 2007	238	Yes	E. coli

Why?

- Recall Richard's talk

Existing atom-atom mapped networks:

Paper	**# reactions**	**Curated**	**Organism**
Suthers et al., 2007	238	Yes	E. coli
Antoniewicz et al., 2007	75	Yes	E. coli

Why?

- Recall Richard's talk

Existing atom-atom mapped networks:

Paper	**# reactions**	**Curated**	**Organism**
Suthers et al., 2007	238	Yes	E. coli
Antoniewicz et al., 2007	75	Yes	E. coli
Ravikirthi et al., 2011	2.077	Nope	E. coli

Why?

- Recall Richard's talk

Existing atom-atom mapped networks:

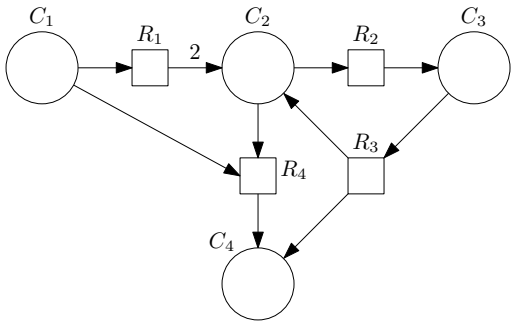
Paper	**# reactions**	**Curated**	**Organism**
Suthers et al., 2007	238	Yes	E. coli
Antoniewicz et al., 2007	75	Yes	E. coli
Ravikirthi et al., 2011	2.077	Nope	E. coli

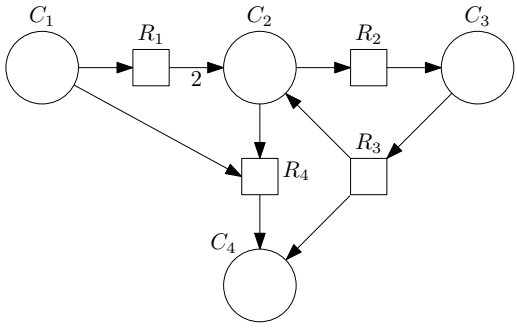
Goal: Create a method for constructing atom-atom mapped networks, which:

- Is generic - takes any genome-scale model as input
- Uses manually curated atom-atom maps whenever possible

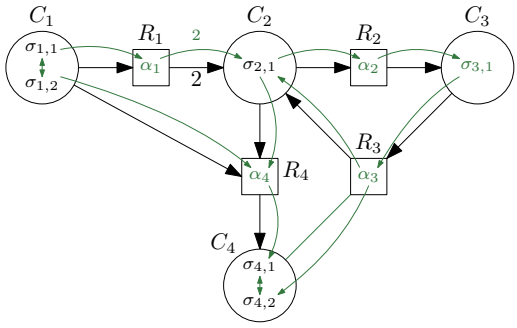
Definition 1 (Chemical Reaction Network). A chemical reaction network (CRN), is a directed multi-hypergraph **CRN** $= (\mathbb{C}, \mathbb{R})$, in which $\mathbb{C}$ is a set of vertices (compounds), and $\mathbb{R}$ is a set of hyperedges (reactions). Each hyperedge is a pair $(\mathcal{E}, \mathcal{P}), \mathcal{E}, \mathcal{P} \in \mathbb{C}$ of multisubsets of compounds, with $\mathcal{E}$ representing the educts and $\mathcal{P}$ representing the products of a reaction.

$$
\begin{aligned}
R_1: &\quad C_1 \to 2C_2 \\
R_2: &\quad C_2 \to C_3 \\
R_3: &\quad C_3 \to C_2 + C_4 \\
R_4: &\quad C_1 + C_2 \to C_4
\end{aligned}
$$





Definition 2 (Atom-level chemical reaction network). An atom-level Chemical Reaction Network (ACRN) is a directed multi-hypergraph $\mathbf{ACRN} = (\mathbb{C}, \mathbb{R}, \mathbf{\Sigma}, \mathbf{A})$, where $\mathbb{C}$ and $\mathbb{R}$ are the sets of compounds and reactions, as in the CRN, which define the network structure; $\mathbf{\Sigma}$ and $\mathbf{A}$ are sets of molecular structures and atom-atom maps. Define by $\mathbf{\Sigma}_C : \mathbb{C} \to \mathcal{P}(\mathbf{\Sigma})$ the set of molecular structures associated with compound C, and by $\mathbf{A}_R : \mathbb{R} \to \mathbf{A}$ the atom-atom map associated with reaction R.



Definition 2 (Atom-level chemical reaction network). An atom-level Chemical Reaction Network (ACRN) is a directed multi-hypergraph $\mathbf{ACRN} = (\mathbb{C}, \mathbb{R}, \mathbf{\Sigma}, \mathbf{A})$, where $\mathbb{C}$ and $\mathbb{R}$ are the sets of compounds and reactions, as in the CRN, which define the network structure; $\mathbf{\Sigma}$ and $\mathbf{A}$ are sets of molecular structures and atom-atom maps. Define by $\mathbf{\Sigma}_C : \mathbb{C} \to \mathcal{P}(\mathbf{\Sigma})$ the set of molecular structures associated with compound C, and by $\mathbf{A}_R : \mathbb{R} \to \mathbf{A}$ the atom-atom map associated with reaction R.

How to determine the structure of a compound based on a database id?

Traverse cross-database references to construct the **'identifier graph'**.

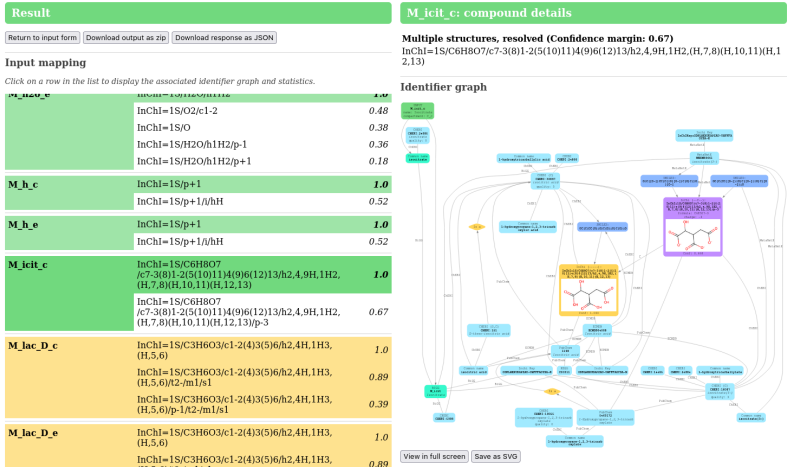
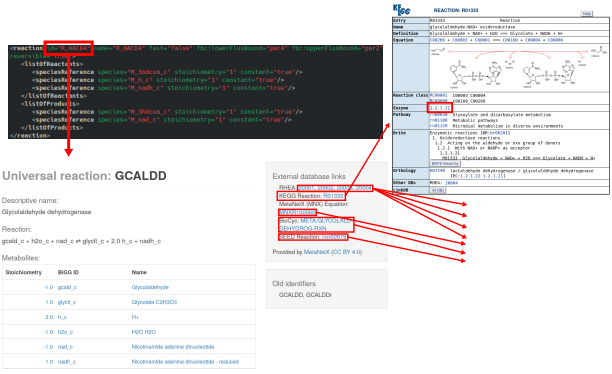


Figure: Eriksen et al., 2024

Identifiers in the model are used as a starting point to traverse the database graph.



- Can be used to get EC numbers and associated reactions with AAMs.

The StructRecon database contains reactions from:

- BiGG
- (KEGG)
- **M-CSA**
- (MetaCyc)
- **MetAMDB**
- MetaNetX
- RHEA
- SABIO-RK
- SEED

MetAMDB - Metabolic Atom Mapping Database

- Approx. 70.000 reactions, **1.003** manually curated.

MetAMDB - Metabolic Atom Mapping Database

- Approx. 70.000 reactions, **1.003** manually curated.

M-CSA - Mechanism and Catalytic Site Atlas

- Reaction mechanisms given as Marvin files representing arrow-pushing diagrams
- Tools developed by Juri Kolčák, Christophe Laurent, Nicolai Nøjgaard for cleaning up and extracting atom maps. (Andersen et al., 2022)
- 1.005 reactions, **430** survive the clean-up and extraction procedures.

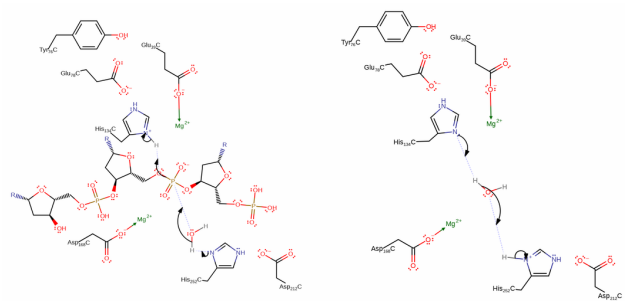


Figure: M-CSA

In case an atom-atom map is not available:

- The third number (X.Y.**Z**.-) describes the type of reaction.
- Given X.Y.**Z**.**A** with a known atom-atom map is similar, we may be able to extract the 'reaction centre' and apply it to X.Y.**Z**.**B**.

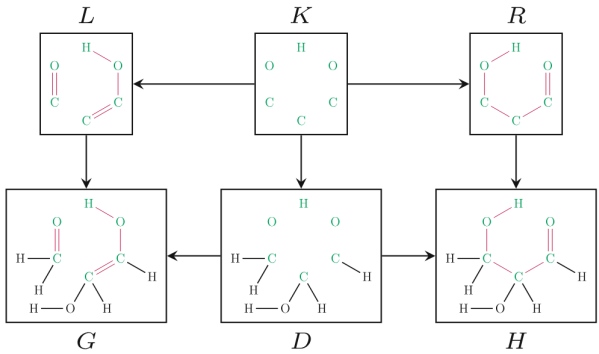
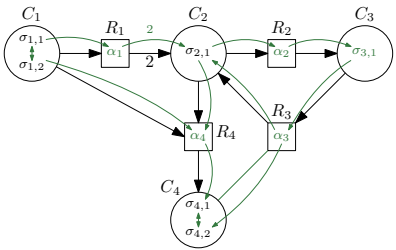


Figure: Andersen et al., 2019

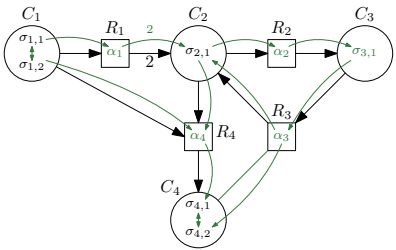
- For each reaction R, attempt to find atom-atom mapped reactions through the *identifier graph*.
- If this is not possible, find an atom-atom mapped reaction up to the third EC digit.
- For some subset $\mathbb{R}' \subseteq \mathbb{R}$, we now have atom maps α_R.
- Each compound $C \in \mathbb{C}$ will have (possibly several) structures inferred from the atom map.
- We create 'pseudo-reactions' with simple atom-atom maps between structures of the same compound.
- For compounds not yet associated with a reaction, generate the consensus structure using the previous StructRecon workflow.
- All remaining unmapped reactions are mapped via RDT.



Priority of atom-atom map sources:

1. MetAMDB map
2. M-CSA map

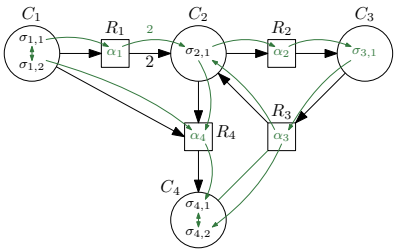
- For each reaction R, attempt to find atom-atom mapped reactions through the *identifier graph*.
- If this is not possible, find an atom-atom mapped reaction up to the third EC digit.
- For some subset $\mathbb{R}' \subseteq \mathbb{R}$, we now have atom maps α_R.
- Each compound $C \in \mathbb{C}$ will have (possibly several) structures inferred from the atom map.
- We create 'pseudo-reactions' with simple atom-atom maps between structures of the same compound.
- For compounds not yet associated with a reaction, generate the consensus structure using the previous StructRecon workflow.
- All remaining unmapped reactions are mapped via RDT.



Priority of atom-atom map sources:

1. MetAMDB map
2. M-CSA map
3. Apply level 3 match from MetAMDB as graph transformation
4. Apply level 3 match from M-CSA as graph transformation

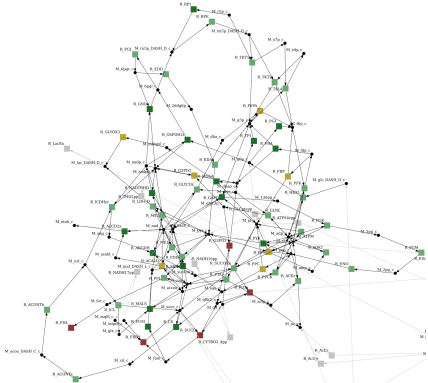
- For each reaction R, attempt to find atom-atom mapped reactions through the *identifier graph*.
- If this is not possible, find an atom-atom mapped reaction up to the third EC digit.
- For some subset $\mathbb{R}' \subseteq \mathbb{R}$, we now have atom maps α_R.
- Each compound $C \in \mathbb{C}$ will have (possibly several) structures inferred from the atom map.
- We create 'pseudo-reactions' with simple atom-atom maps between structures of the same compound.
- For compounds not yet associated with a reaction, generate the consensus structure using the previous StructRecon workflow.
- All remaining unmapped reactions are mapped via RDT.



Priority of atom-atom map sources:

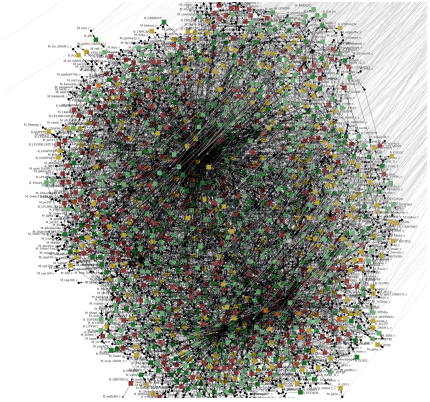
1. MetAMDB map
2. M-CSA map
3. Apply level 3 match from MetAMDB as graph transformation
4. Apply level 3 match from M-CSA as graph transformation
5. Produce map via RDT ($\sim 90\%$ accuracy, depends on enzyme class)

EColiCore2 (compressed)



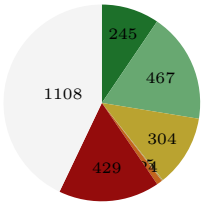
- MetAMDB
- M-CSA
- MetAMDB (L3)
- M-CSA (L3)
- M-CSA (map failed)
- No EC number
- Transport

EColiCore2 (genome scale)

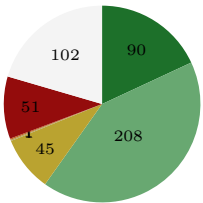


- MetAMDB
- M-CSA
- MetAMDB (L3)
- M-CSA (L3)
- M-CSA (map failed)
- No EC number
- Transport

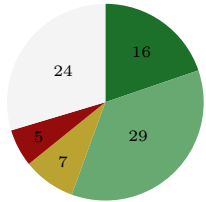
EColiCore2 (genome scale)
2582 reactions



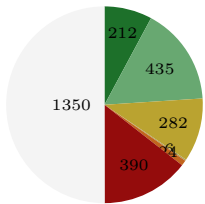
EColiCore2
498 reactions

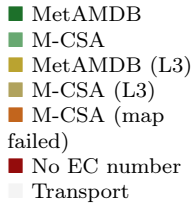


EColiCore2 (compressed)
81 reactions



E. coli via CarveMe
2699 reactions





- Most reactions seems to be associated with manually curated AAMs.
- Especially the ‘core’ pathways.

But,

- Work in progress!
- Success rate and accuracy of level-3 matching not known yet.

- Most reactions seems to be associated with manually curated AAMs.
- Especially the ‘core’ pathways.

But,

- Work in progress!
- Success rate and accuracy of level-3 matching not known yet.

Planned availability:

- Web application
- Python library
- Consolidated database (SQL query)

Thank you!

SDU

MATOMIC

novo nordisk
foundation