Atom Mapping in MetaCyc and Pathway Tools

Mario Latendresse

March 2013

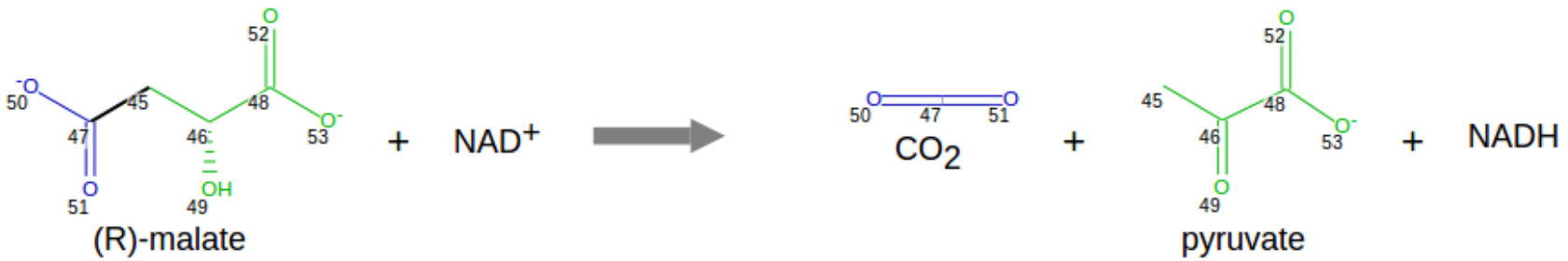
Outline

What is an Atom Mapping?

Definition of a valid atom mapping

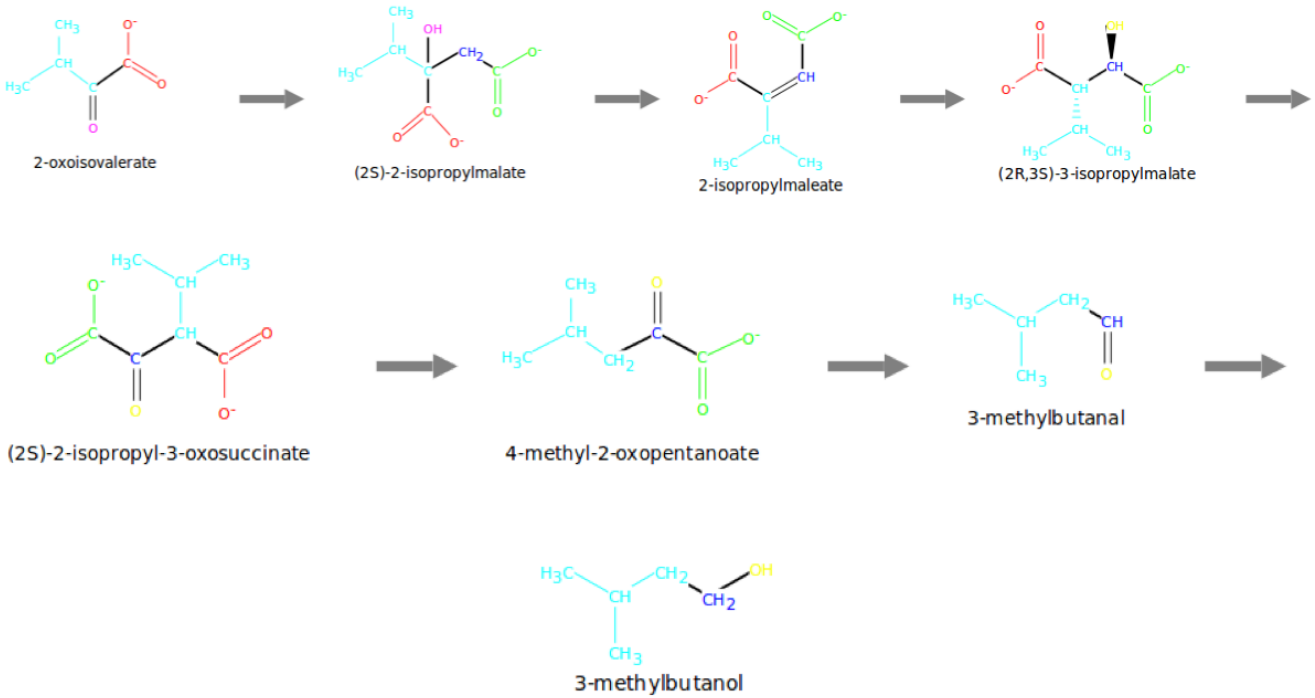
A bijection of the reactant atoms to the product atoms of a (bio)chemical reaction such that atom species are conserved.

Reaction EC 1.1.1.83



Applications

- Bioengineering: computing the conserved atoms from a source to a target compound in pathways



- Computing fluxes of reactions based on atom tracing (atom labeling)
- Better understanding of the reaction mechanisms (e.g., teaching)

Details in Published Paper

The details about how the atom mappings were computed and validated were published in

JCIM

Mario Latendresse, Jeremiah P. Malerich, Mike Travers, and Peter D. Karp, **Accurate Atom-Mapping Computation for Biochemical Reactions**, Journal of Chemical Information and Modeling, September 2012

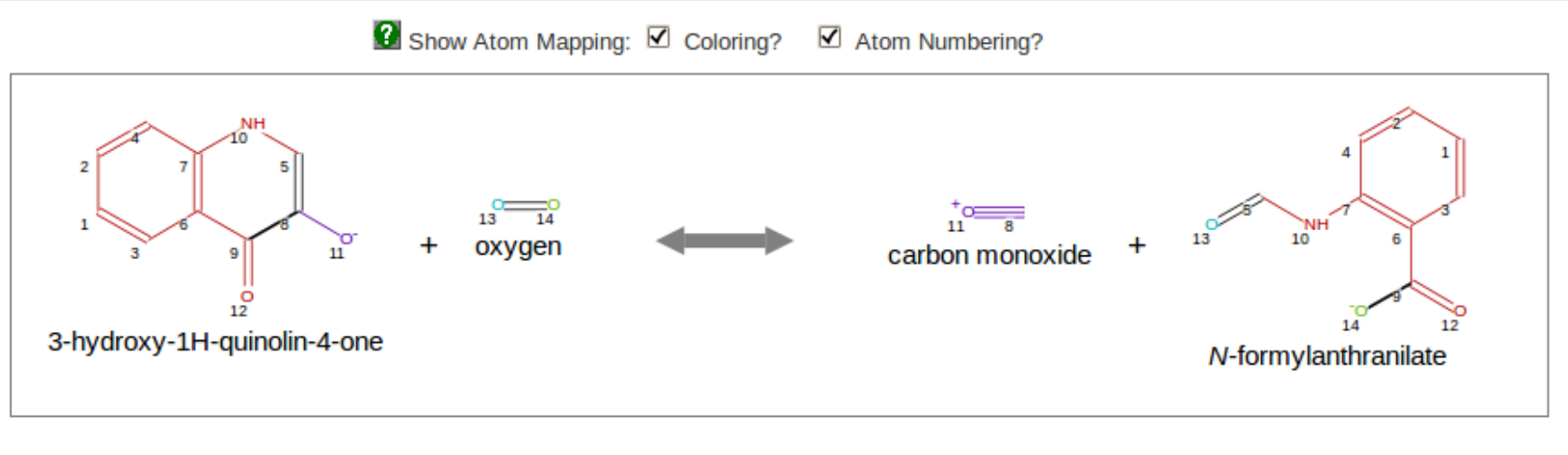
Atom Mappings in MetaCyc

- MetaCyc is a **manually curated** multi-organism database of biochemical reactions and pathways (main curators: Ron Caspi & Carol Fulcher)
- It is the main database of BioCyc and Pathway Tools
- Atom mappings were **computed** for 9,387 of its reactions
- Version 17.0 (March 2013) has 11,362 reactions (enzymatic and spontaneous)
- Some reactions do not have compound structures or are generic and not mass balanced: no atom mapping for them
- The computation took too long (> 30 minutes) or had too many equivalent atom mappings (> 1000) for about 150 reactions

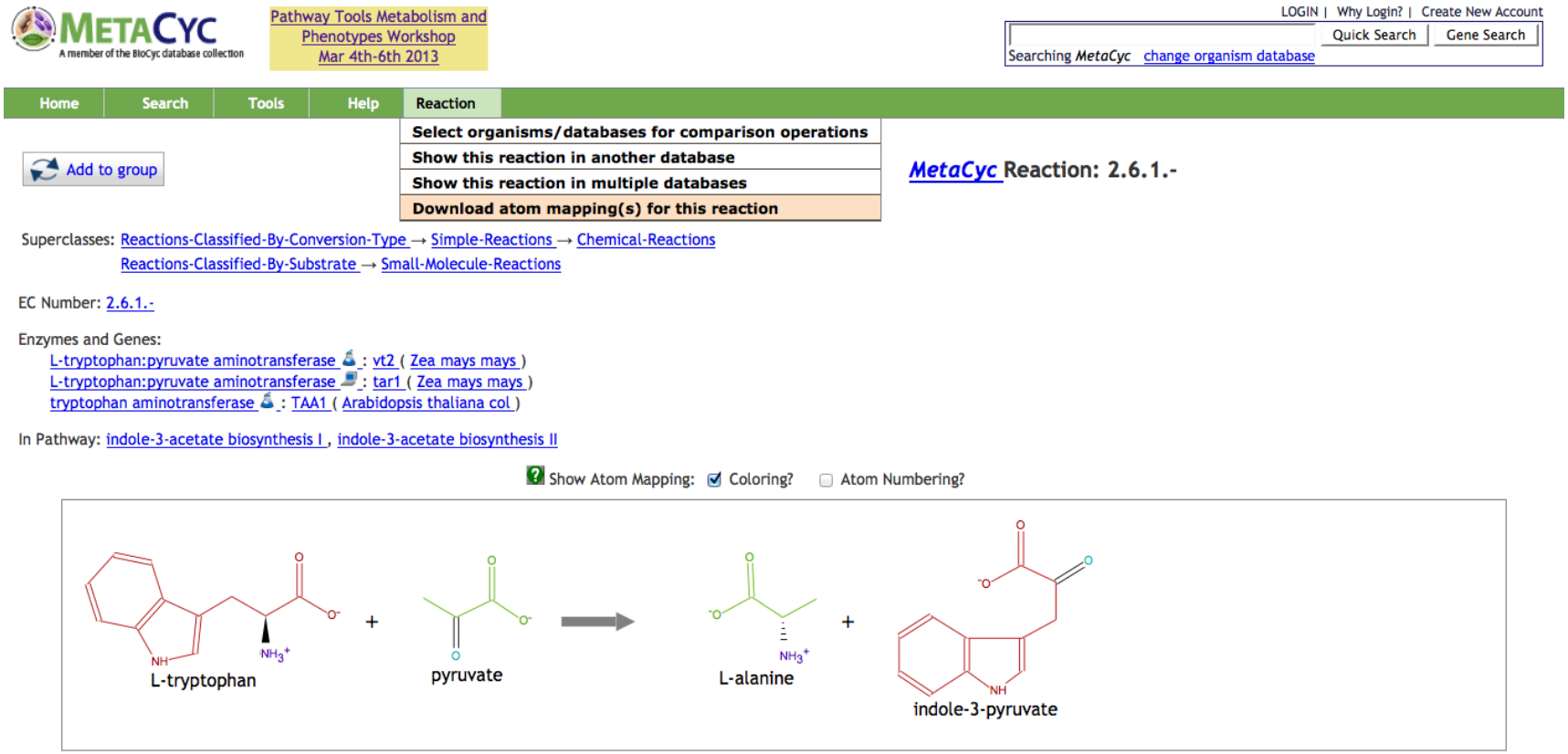
Atom mappings at BioCyc (Web)

- Atom mappings are displayed at BioCyc.org (Web) for all databases (over 2000)
- But, atom mappings are mostly stored in MetaCyc
- When displaying a reaction, show the atom mappings stored in the database, if any, otherwise show the atom mappings stored in MetaCyc, if any

Reaction EC 1.13.11.47



Downloading Atom Mappings



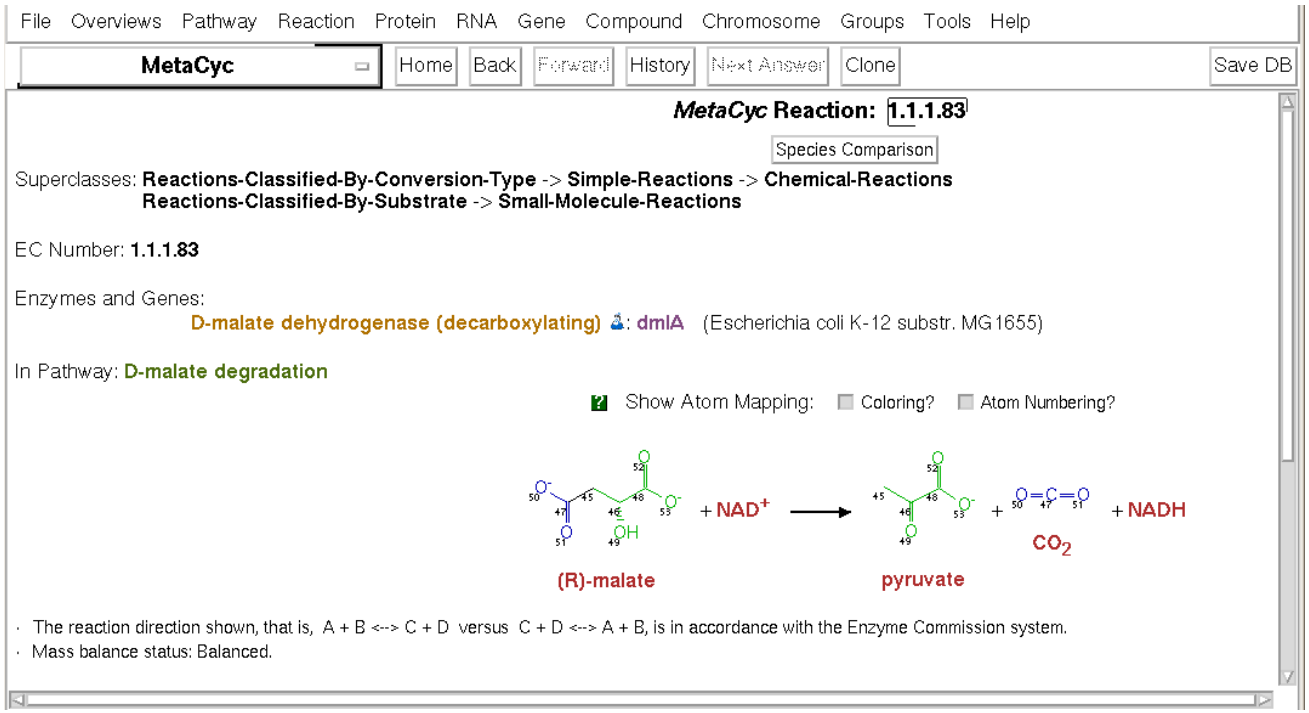
Text Representation of Atom Mappings

```
# Exported atom mapping(s) for reaction RXN-10139 in PGDB META done on
 05-Mar-2013.
# There is 1 atom mapping.
# Please consult Help->PGDB Concepts Guide, Section Atom Mapping, to
  interpret the following encoding.

REACTION - RXN-10139
NTH-ATOM-MAPPING - 1
MAPPING-TYPE - NO-HYDROGEN-ENCODING
FROM-SIDE - (TRP 0 14) (PYRUVATE 15 20)
TO-SIDE - (INDOLE_PYRUVATE 0 14) (L-ALPHA-ALANINE 15 20)
INDICES - 0 1 2 3 4 5 6 7 9 8 10 12 18 13 14 15 16 17 11 20 19
//
```

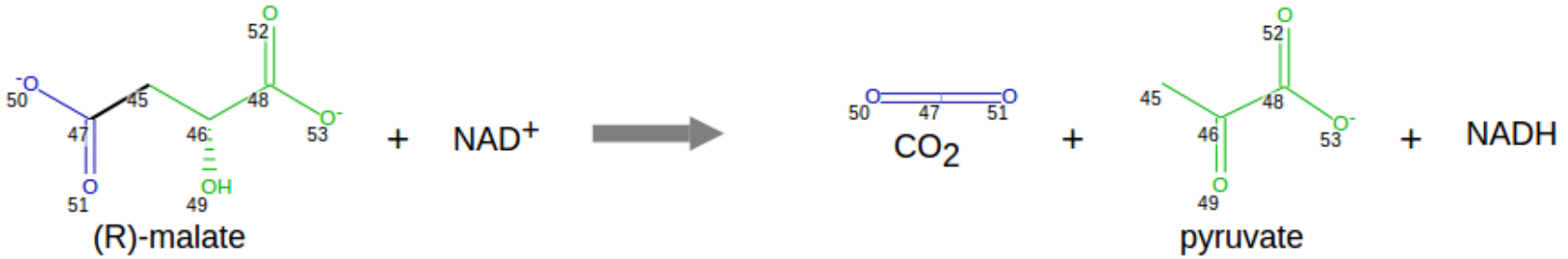
Atom mappings in Pathway Tools (Desktop)

- Atom mappings can be computed for your **own** database
- Creating or modifying a reaction using Pathway Tools (Desktop) computes its atom mappings
- Atom mappings are displayed as on the Web
- Currently it is **not** possible to (manually) edit the atom mappings



Maximize Keeping the Right Bonds Intact

Compute atom mappings as an optimization problem



- For all possible valid atom mappings ...
- Keep intact the bonds that are not likely to break
- Do not make bonds that are not likely to form
- For example, C-C bonds do not break or form as often as P-O bonds
- Assign appropriate **propensity** values to bonds to break or form

Basic Bond Propensity Values (Magic Table)

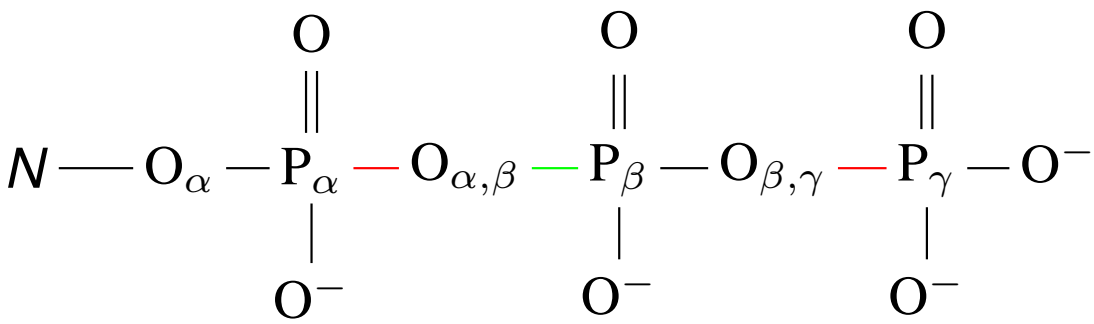
The larger the value, the less likely the bond breaks or forms.

	C	O	N	P	H	S
C	400|24	48*|8	56*|8	48	72	48*
O	48*|8	16|8	8|72	8*|72	4	8|72
N	56*|8	8|72	16	8	8	24
P	48	8*|72	8	na	na	8
H	72	4	8	na	na	8
S	48*	8|72	24	8	8	16

The numbers marked by * are tuned for special cases.

Special Bond Values (Example)

Special propensity values for compounds with a triphosphate group (e.g., ATP).



***N*TP**

The bonds P_α—$O_{\alpha,\beta}$ and $O_{\beta,\gamma}$—P_γ are more likely to break compared with the other P—O bonds; except for compounds dGTP, dCTP, dTTP, and dUTP, where only $O_{\alpha,\beta}$—P_β is more likely to break.

Basic MILP Variables

Use a linear solver (e.g., SCIP, CPLEX, Gurobi).

Basic Sets and Symbols

A_r: set of atoms on the reactant side
A_p: set of atoms on the product side
$s(x)$: the species of atom x

Variables Directly Controlling the Atom Mapping

$\forall a \in A_r, x \in A_p, s(a) = s(x)$, define binary (0,1) variable m_{ax}
The solver will say $m_{ax} = 1$only if a is mapped to$x$

Variables Controlling Bonds

For all bonds broken (a, b) or made (x, y), define variable e_{abxy}
The solver will say $e_{abxy} = 1$ only if $m_{ax} = 1$ and $m_{by} = 1$

Basic Constraints and Minimization

Injection Constraints

$$\forall a \in A_r, \sum_{x \in A_p, s(x)=s(a)} m_{ax} = 1 \quad (1)$$

Surjection Constraints

$$\forall x \in A_p, \sum_{a \in A_r, s(x)=s(a)} m_{ax} = 1 \quad (2)$$

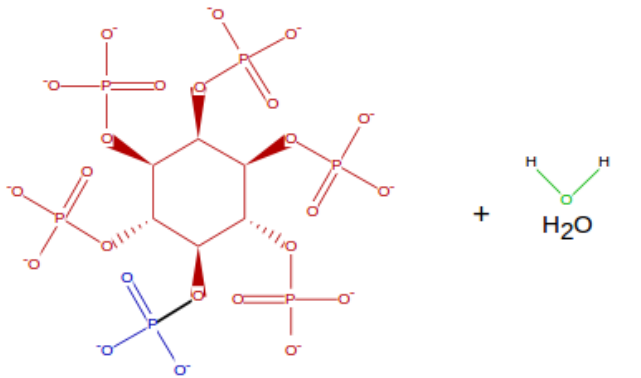
Minimize

$$\sum_{(a,b)(x,y)} P_{(a,b)} e_{axby} + P_{(x,y)} e_{axby} \quad (3)$$

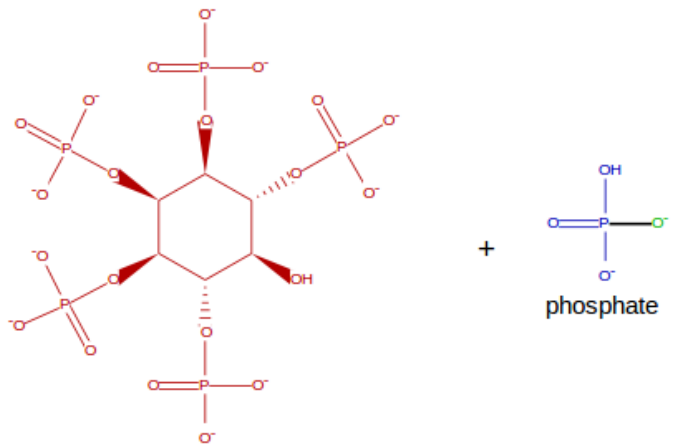
Removing Multiple Equivalent Atom Mappings

- We try to keep (i.e., store and display) only the **non equivalent** atom mappings
- Two atom mappings **are equivalent** if the same bonds are broken/made taking into account indistinguishable atoms and symmetries of compounds
- Equivalent atom mappings are (tentatively) detected **after** the linear solver has found all the optimal atom mappings
- Sometimes, due to the complexity of detecting symmetries, some equivalent atom mappings are not detected

380 Equivalent Atom Mappings for EC 3.1.3.72



1D-*myo*-inositol 1,2,3,4,5,6-hexakisphosphate



1D-*myo*-inositol (1,2,3,4,6)-penta*kis*phosphate

Multiple Atom Mappings Stored in MetaCyc 17.0

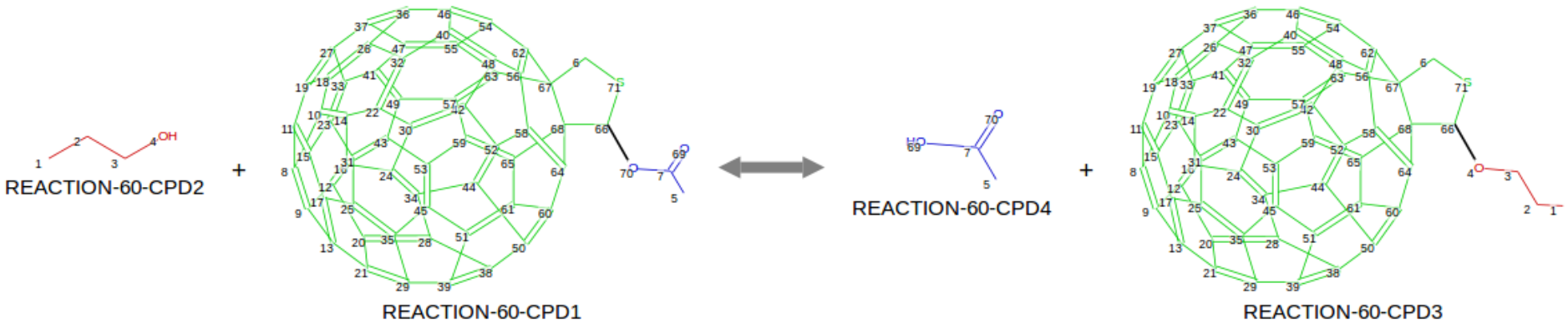
- There are 9,387 reactions with atom mappings in MetaCyc 17.0 (11,362 reactions)
- Many multiple atom mappings are actually equivalent but were not automatically detected as equivalent

	Number of Atom Mappings				
	1	**2**	**3-4**	**5-8**	**9-24**
reactions	94%	5%	1%	0.2%	0.09%

Rings are Common in Compounds

- Many biochemical reactions have at least one compound with rings
- Rings do not often form or break
- When similar rings can be potentially mapped, a model is created to tentatively mapped them directly, bypassing the direct mapping of every atom in the similar rings
- This ring mapping helps the MILP solver to find the atom mappings faster
- If the model is infeasible (as detected by the MILP solver), the modeling of rings is removed and a basic model is solved

A (synthetic) Reaction with Lots of Rings



One atom mapping found in 5 seconds

Speed of Execution

- Highly depends on the solver (**SCIP**, CPLEX, Gurobi) used
- The following numbers applied to version MetaCyc 16.0

	Solved Under a Time Limit, Seconds				
	$<$ **0.1s**	$<$ **1s**	$<$ **10s**	$<$ **60s**	$<$ **1800s**
1	51%	73%	91%	96%	98%
n	47%	72%	87%	93%	98%

Correctness of MetaCyc Atom Mappings

- An error rate of 0.9% for MetaCyc
- KEGG RPAIR is a manually curated atom mapping database
- Programmatically compared 2,446 reaction atom mappings from the KEGG RPAIR database with the corresponding atom mappings of MetaCyc 16.0
- 22 reaction atom mappings were found incorrect for MetaCyc
- 2 reaction atom mappings were found incorrect in KEGG RPAIR (verified by a literature search)
- The exact correctness of the atom mappings in MetaCyc is not known

Future Work

- Modeling compound symmetries, and stoichiometry to reduce the number of equivalent atom mappings
- Better modeling of stereochemistry
- Compute tracing of atoms in pathways taking into account compound symmetries, indistinguishable atoms, and stoichiometry
- More precise modeling to help the solver execute faster

The End

Thank You

Questions?

Comments?