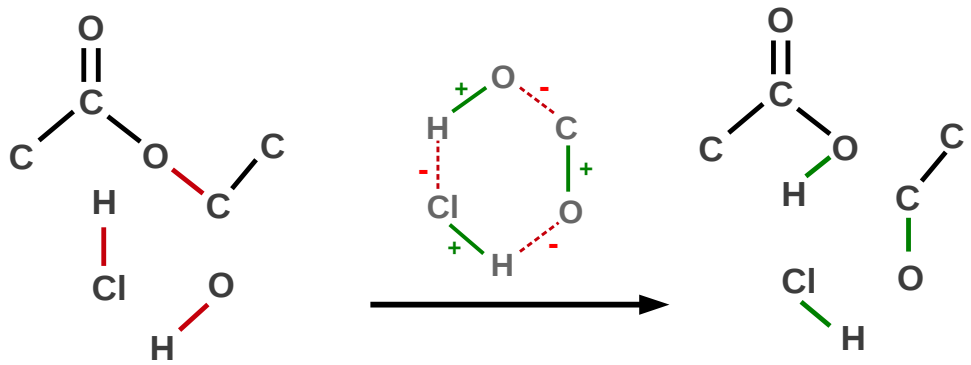
Atom Mapping with Constraint Programming

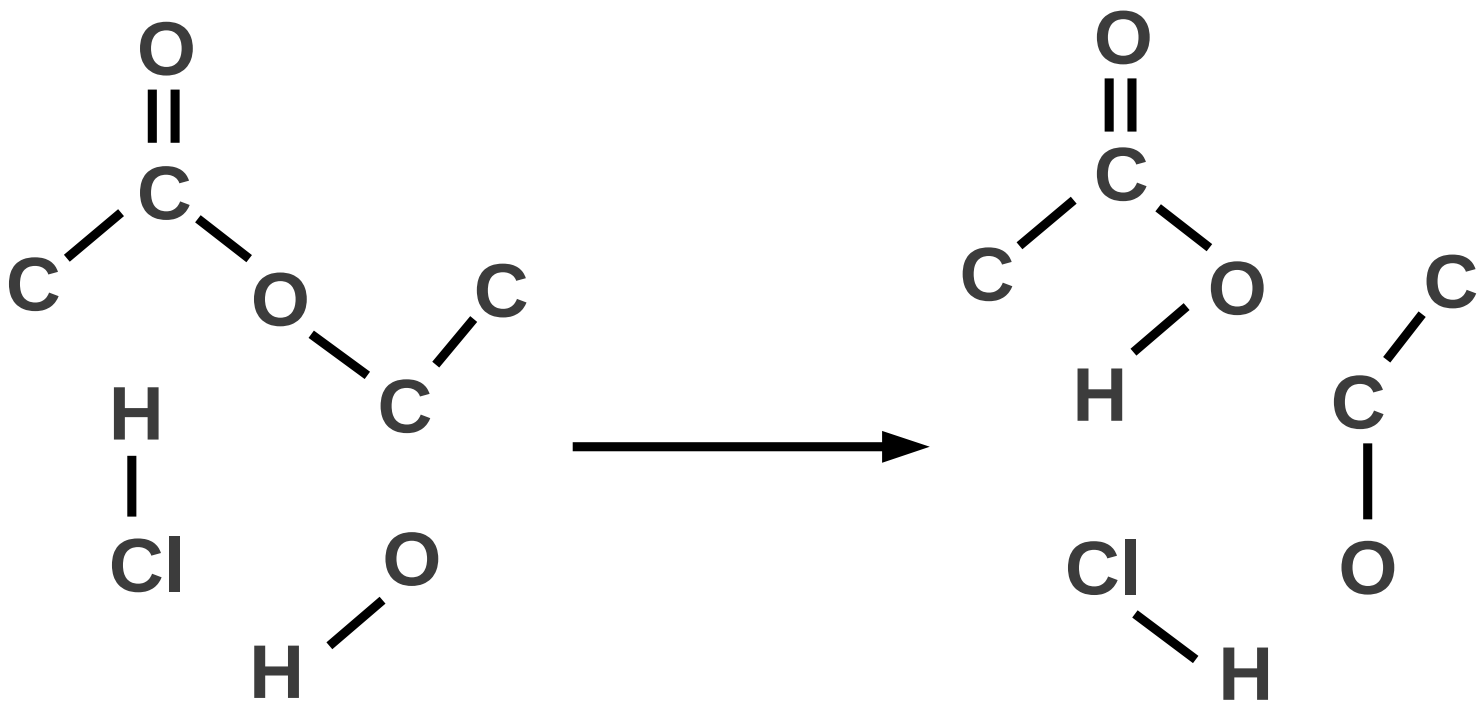
Martin Mann, Feras Nahar, Heinz Ekker, Rolf Backofen,
Peter F. Stadler, and Christoph Flamm

Bioinformatics Group
University of Freiburg



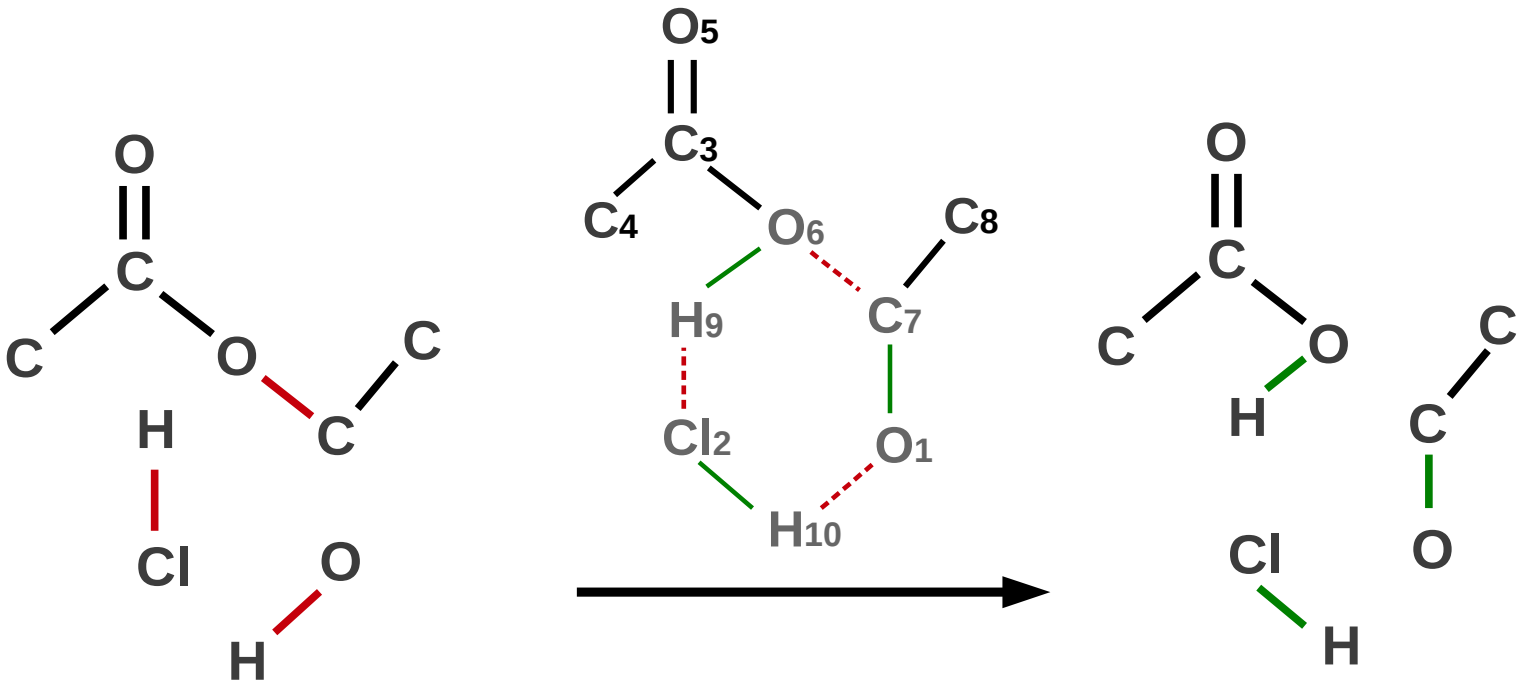
Problem

A chemical reaction is a transformation of educts molecules into products molecules.



Problem

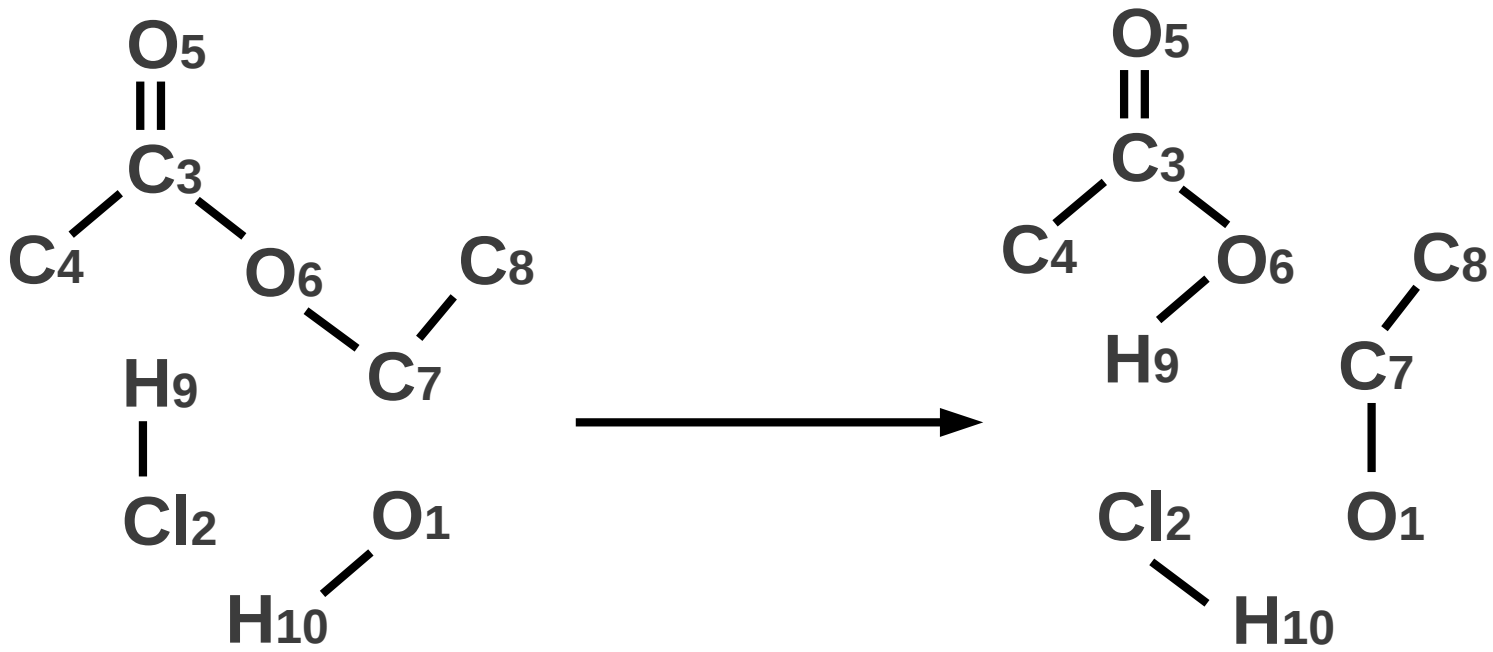
A chemical reaction is a transformation of educts molecules into products molecules.



- In this process, chemical bonds are redistributed.
- Redistribution process (reaction mechanism) is unknown.
- Experimental identification tricky and expensive.

Atom Mapping = Reaction Mechanism Encoding

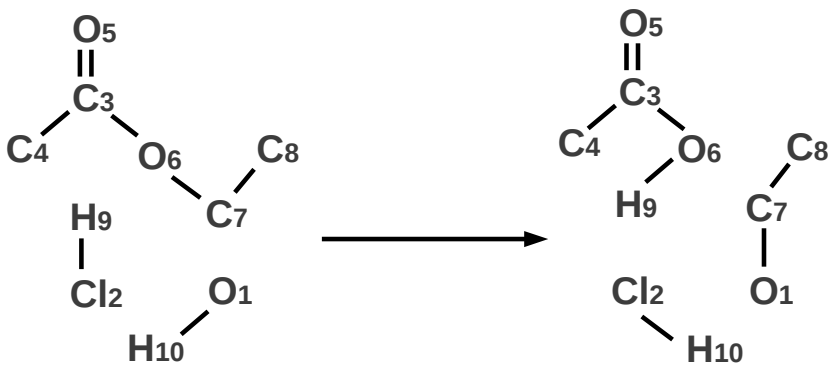
Each atom in an educt molecule appears in a specific position of a reaction product.



Atom Mapping

one-to-one correspondence between atoms in educts and products

Computational Approaches Problem

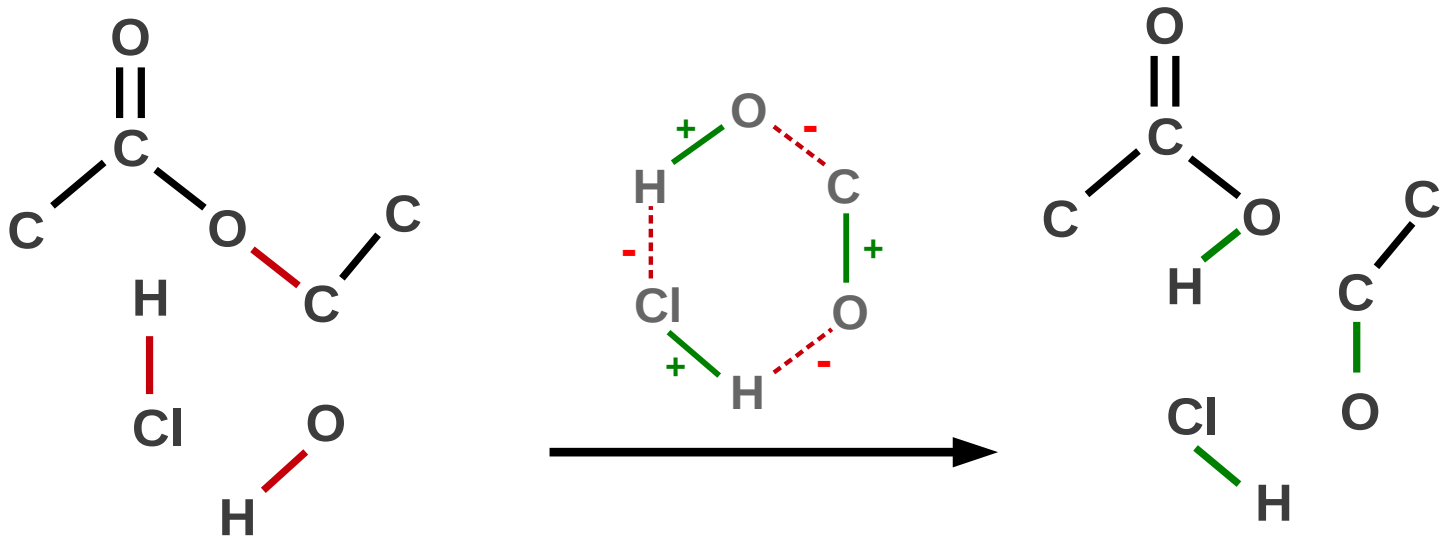


Previous approaches are based on:

1. Maximum common subgraph (McGregor 1981, 1982).
2. Maximum common edge subgraph (Raymond 2002).
3. Graph partitioning and graph isomorphism (Akutsu 2004).
4. Weighted common edge subgraph (Körner, Apostolakis 2008).
5. Canonical graph naming (Crabtree & Mehta 2009, 2010).
6. Integer linear programming (First 2011).

The Imaginary Transition State (ITS)

- The changing bonds (broken, formed) are described by an intermediate transition state (ITS).



- The ITS shows cyclic/pseudo-cyclic topology.
- Different ITS sizes and layouts.
- Once the ITS is identified, the rest is graph isomorphism.

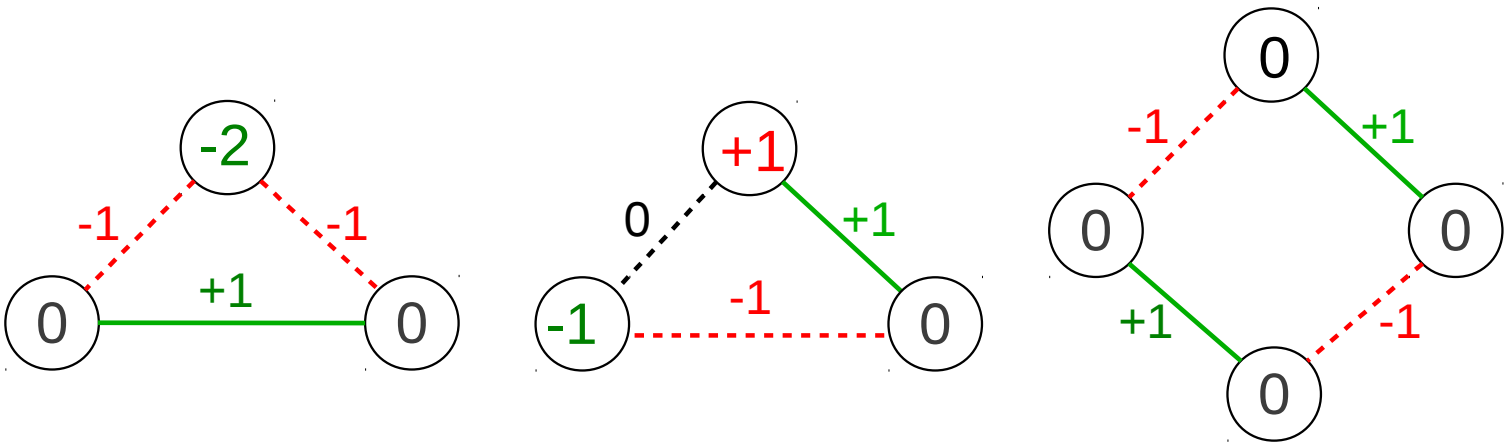
ITS Encoding

Observation:

- Finite number ITS layouts
- Can be encoded by graphs

ITS graph encoding C describes:

- Charge changes of ITS-participating atoms.
- Bond valence changes between them.
- Important: Every bond breaking is compensated by bond formation or charge change (and vice versa).



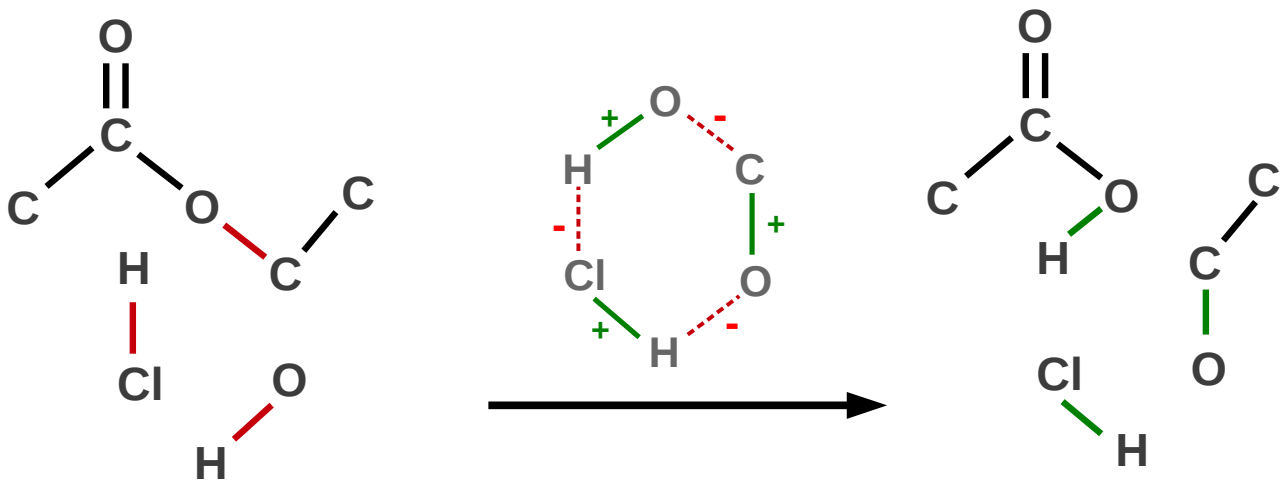
Atom Mapping - Formally

Given:

1. Two molecule graphs $I = (V_I, E_I)$ and $O = (V_O, E_O)$.
2. Bond numbers via adjacency matrices $\mathcal{I}$ and $\mathcal{O}$.

A *chemically correct* atom map is a bijective map $m : V_I \to V_O$ such that:

1. Atom labels are preserved: $\forall_{x \in V_I} : l(x) = l(m(x))$.
2. Total bond orders are preserved: $(\mathcal{O} - (\mathcal{I} \circ m))\,\overrightarrow{1} = 0$.
3. Bond changes form a valid cyclic ITS C.



Constraint Programming Approach

Consider the **ITS identification** as **Constraint Satisfaction Problem** (CSP) for a **given ITS** of size k.

Given:

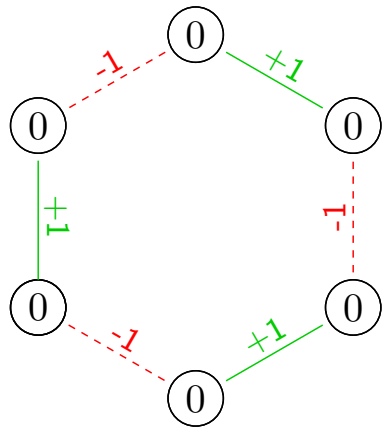
- $I = (V_I, E_I)$ and $O = (V_O, E_O)$.
- Adjacency matrices $\mathcal{I}$ and $\mathcal{O}$.
- ITS matrix C such that for $1 \leq i \leq j \leq k$:
  - $C_{i,i} = \textit{charge change}$
  - $C_{i,j} = \textit{bond change}$

We encode **2k ITS variables**:

- in the educts $\{X_1^I, \ldots, X_k^I\}$ with the domain $D_i^I = V_I$.
- in the products $\{X_1^O, \ldots, X_k^O\}$ with the domain $D_i^O = V_O$.

Constraints

1. **Bijective Mapping:** $\forall i \neq j : X_i^I \neq X_j^I$ and $\forall i \neq j : X_i^O \neq X_j^O$
2. **Atom Label Preservation:** $l(X_i^I) = l(X_i^O)$
3. **Atom Charge Change:** $\mathcal{I}_{X_i^I, X_i^I} - \mathcal{O}_{X_i^O, X_i^O} = C_{i,i}$
4. **Edge Degree Change:** $|degree(X_i^I) - degree(X_i^O)| \leq max(1, C_{i,i})$
5. **Ring Bonding:** $\mathcal{O}_{X_i^O, X_{i+1}^O} - \mathcal{I}_{X_i^I, X_{i+1}^I} = C_{i,i+1}$
6. **Coverage of Connected Components**
7. **Symmetry Breaking**

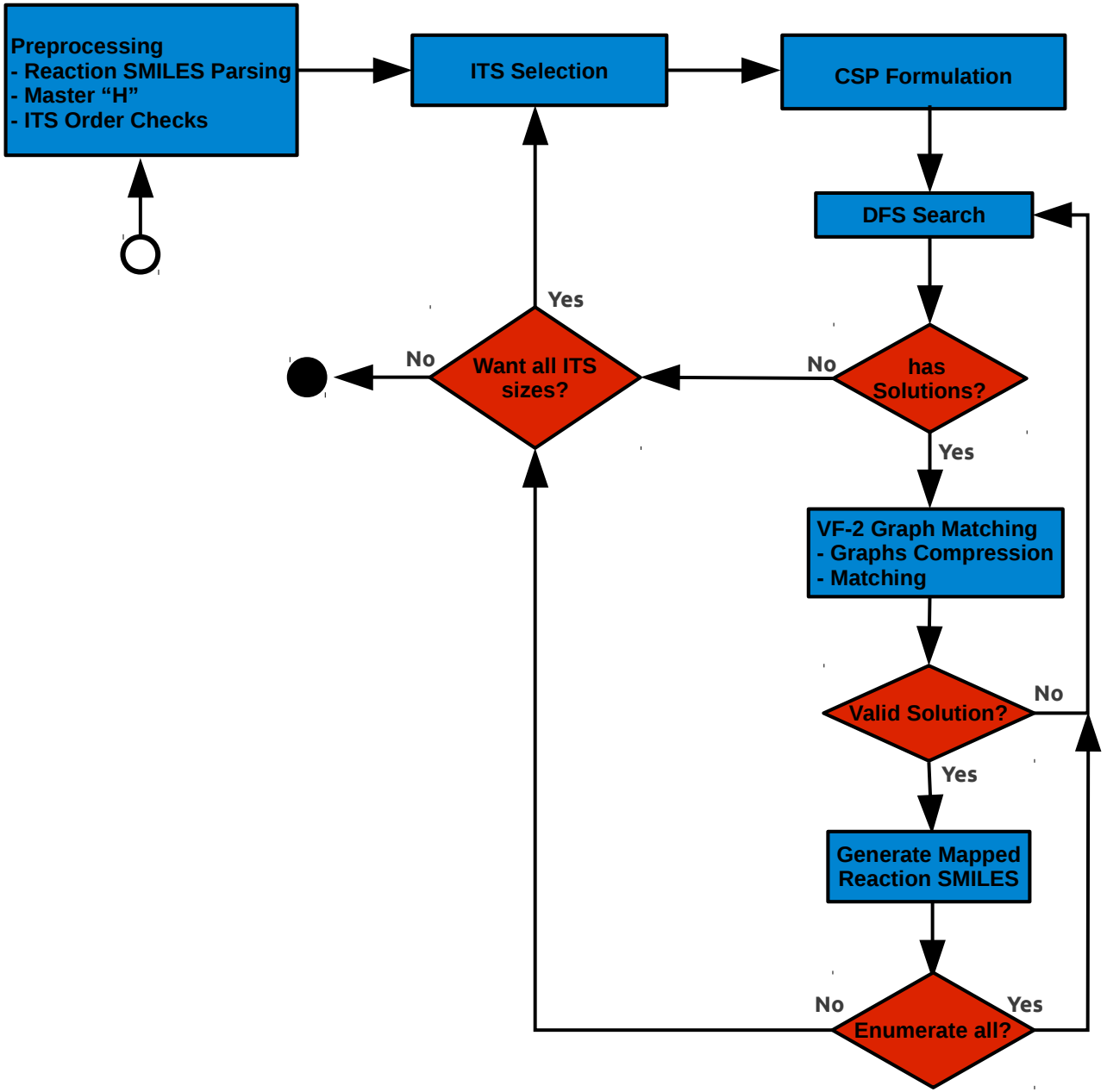


Overall Atom Mapping

- Small CSP size (only ITS atoms).
- Depth-First-Search (DFS) to find CSP solutions.
- CSP solutions that fulfil the constraints are ITS candidates.
- Non-ITS atoms still have to be mapped.
- Extend the ITS mapping to global mapping using graph matching procedure VF-2.

The CSP filters the ITS candidates for the subsequent graph matching.

Atom Mapping Framework



Experimental Results

KEGG ID	Atoms	k	ITS CSP Sol.	Valid Sol.	Time (s)
R00009	8	6	1	1	**0**
R00013	14	6	19	1	**0.13**
		8	41	1	**0.6**
R00018	36	4	2	1	**0.15**
		6	1438	13	**4.5**
R00048	30	4	8	2	**0.24**
		6	1792	20	**5.93**
R00059	44	4	1	1	**1.05**
R00120	66	6	1	1	**3.13**
R00207	20	8	1	1	**1.33**

Evaluation for some reactions from KEGG LIGAND database.

Current Work and Plans

- Identification of all occuring ITS layouts.
- Extensive atom mapping evaluation using known reactions.
- Atom mapping computation for available reaction databases.
- Webserver for ad hoc usage.
- Deduction of chemical graph grammar rules and application.

Thanks for your attention!