Disconnection Rules are Complete for Chemical Reactions

Towards functorial chemistry

Leo Lobski

joint work with Fabio Zanasi and Ella Gale

University College London

21st International Colloquium on Theoretical Aspects of Computing, Bangkok, Thailand

27 November 2024

Outline

- Two perspectives on chemical processes:

Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change

Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change
  2. Disconnection rules: hypothetical breaking of *local* bonds

Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change
  2. Disconnection rules: hypothetical breaking of *local* bonds
- We define the following categories, which share the same objects (chemical graphs):

Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change
  2. Disconnection rules: hypothetical breaking of *local* bonds
- We define the following categories, which share the same objects (chemical graphs):
  1. **React –** partial bijections that encode any physically feasible reactions (atoms and charge are preserved),

Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change
  2. Disconnection rules: hypothetical breaking of *local* bonds
- We define the following categories, which share the same objects (chemical graphs):
  1. **React** – partial bijections that encode any physically feasible reactions (atoms and charge are preserved),
  2. **Disc** – local graph rewrites of chemical bonds

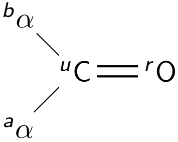
Outline

- Two perspectives on chemical processes:
  1. Reactions: predictive model of *global* chemical change
  2. Disconnection rules: hypothetical breaking of *local* bonds
- We define the following categories, which share the same objects (chemical graphs):
  1. **React** – partial bijections that encode any physically feasible reactions (atoms and charge are preserved),
  2. **Disc** – local graph rewrites of chemical bonds
- There is a functor R : **Disc** $\rightarrow$ **React** which is faithful, and full up to an isomorphism

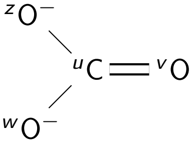
Chemical graphs

Molecular entities are represented by labelled graphs:

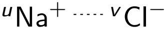
A



B

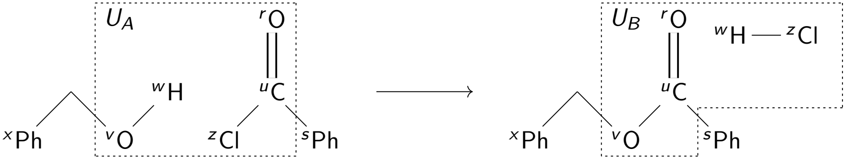


C

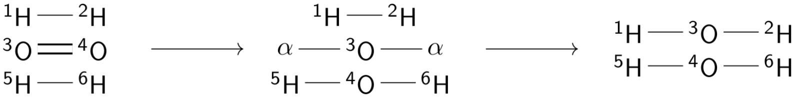


Reactions: Example

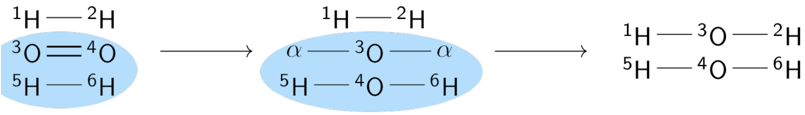
Formation of benzyl benzoate from benzoyl chloride and benzyl alcohol



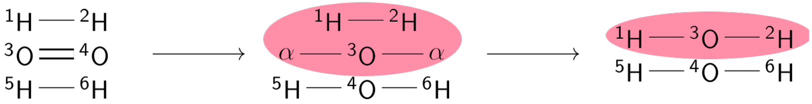
Composition in **React**



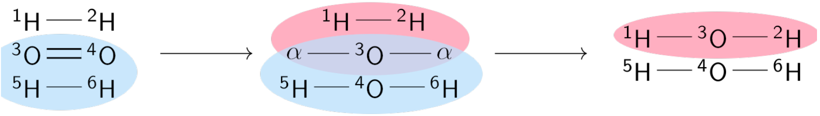
Composition in **React**



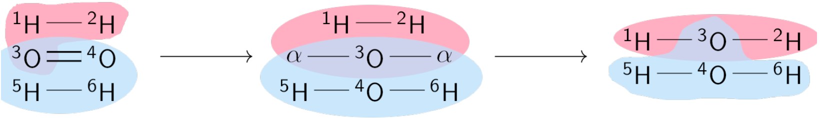
Composition in **React**



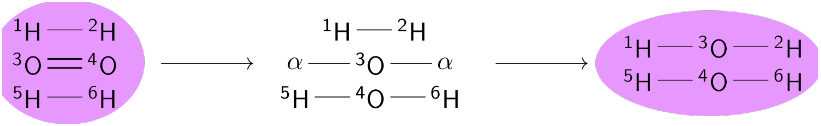
Composition in **React**



Composition in **React**

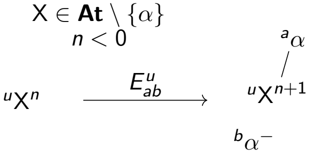


Composition in **React**

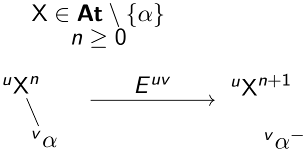


Disconnection rules

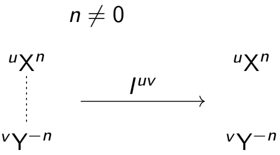
Motto: Chemical processes are movements of electrons



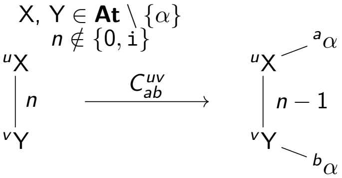
Electron detachment (negative charge)



Electron detachment (nonnegative charge)



Ionic bond breaking



Covalent bond breaking

Disconnection category: Terms

We define the set of terms with types:

- $\mathrm{id} : A \to A$,
- if $u \in V_A$, let $S^u : A \to A$,
- if $u \in \alpha(A)$ and $v \notin V_A \setminus \{u\}$, let $R^{u \mapsto v} : A \to A[v/u]$,
- $d^{uv}_{ab} : A \to d^{uv}_{ab}(A)$ and $\bar{d}^{uv}_{ab} : d^{uv}_{ab}(A) \to A$,
- if $\mathtt{t} : A \to B$ and $\mathtt{s} : B \to C$, then $\mathtt{t};\mathtt{s} : A \to C$.

Disconnection category: Equations

$$R^{u\mapsto z}; R^{z\mapsto w} \lesssim R^{u\mapsto w} \tag{1}$$
$$R^{u\mapsto z}; R^{v\mapsto w} \approx R^{v\mapsto w}; R^{u\mapsto z} \tag{2}$$
$$R^{u\mapsto u} \lesssim S^{u} \tag{3}$$
$$R^{b\mapsto z}; R^{a\mapsto b} \approx S^{b}; R^{a\mapsto z} \tag{4}$$
$$R^{u\mapsto v}; S^{w} \approx S^{w}; R^{u\mapsto v} \tag{5}$$
$$R^{u\mapsto v}; S^{v} \simeq S^{u}; R^{u\mapsto v} \simeq R^{u\mapsto v} \tag{6}$$

$$R^{u\mapsto v}; d_D^U \approx d_D^U; R^{u\mapsto v} \tag{7}$$
$$R^{u\mapsto v}; E^{wv} \simeq E^{wu}; R^{u\mapsto v} \tag{8}$$
$$d^U_{D[u]}; R^{u\mapsto v} \simeq d^U_{D[v/u]} \tag{9}$$
$$d^{U'}_{ij}; \bar{h}^U_{ab}; R^{i\mapsto c}; R^{j\mapsto d} \lesssim \bar{h}^U_{ab}; d^{U'}_{cd} \tag{10}$$

$$d^U_{ab}; \bar{d}^U_{cd} \approx S^U; R^{c\mapsto a}; R^{d\mapsto b} \tag{11}$$
$$d^U_{ab}; \bar{d}^U_{cb} \approx S^U; R^{c\mapsto a} \tag{12}$$
$$d^U_{ab}; \bar{d}^U_{ad} \approx S^U; R^{d\mapsto b} \tag{13}$$
$$d^U_D; \bar{d}^U_D \lesssim S^U \tag{14}$$
$$\bar{d}^U_D; d^U_D \lesssim S^U; S^D \tag{15}$$
$$E^{ua}; \bar{E}^{ub} \approx S^u; R^{a\mapsto z}; R^{b\mapsto a}; R^{z\mapsto b} \tag{16}$$
$$\bar{d}^{uv}; d^{wz} \approx d^{wz}; \bar{d}^{uv} \tag{17}$$

$$S^u; S^v \simeq S^v; S^u \tag{18}$$
$$S^u; S^u \simeq S^u \tag{19}$$
$$S^u; d^U_D \lesssim d^U_D; S^u \tag{20}$$
$$d^{U[v]}_D; S^v \simeq d^{U[v]}_D \tag{21}$$
$$C^{uv}_{ab} \simeq C^{vu}_{ba} \tag{22}$$

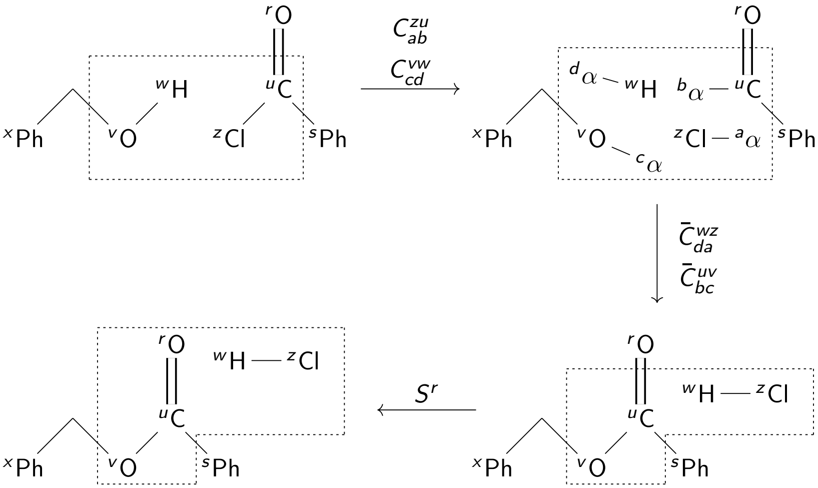
$$d^U_D; d^{U'}_{D'} \simeq d^{U'}_{D'}; d^U_D \tag{23}$$
$$C^{uv}_{ab}; I^{wz} \simeq I^{wz}; C^{uv}_{ab} \tag{24}$$
$$E^u_{ab}; I^{wz} \lesssim I^{wz}; E^u_{ab} \tag{25}$$
$$E^{uv}; I^{wz} \lesssim I^{wz}; E^{uv} \tag{26}$$
$$\bar{E}^{uv}; I^{wz} \lesssim I^{wz}; \bar{E}^{uv} \tag{27}$$
$$\bar{E}^u_{ab}; I^{wz} \lesssim I^{wz}; \bar{E}^u_{ab} \tag{28}$$
$$\bar{C}^{uv}_{ab}; I^{wz} \lesssim I^{wz}; \bar{C}^{uv}_{ab} \tag{29}$$
$$E^u_{ab}; C^{wz}_{cd} \simeq C^{wz}_{cd}; E^u_{ab} \tag{30}$$
$$E^{uv}; C^{wz}_{cd} \lesssim C^{wz}_{cd}; E^{uv} \tag{31}$$
$$\bar{E}^{uv}; C^{wz}_{cd} \simeq C^{wz}_{cd}; \bar{E}^{uv} \tag{32}$$
$$E^{uv}; E^w_{cd} \lesssim E^w_{cd}; E^{uv} \tag{33}$$
$$\bar{E}^{uv}; E^w_{cd} \simeq E^w_{cd}; \bar{E}^{uv} \tag{34}$$

From disconnections to reactions

Define the translation $R : \textbf{Disc} \to \textbf{React}$ by

- $R(\mathrm{id}_A) := (\varnothing, \varnothing)$,
- $R\left(S^u\right) := (\{u\}, \{u\})$,
- $R\left(R^{u \mapsto v}\right) := (\{u\}, \{v\})$,
- $R(d^U_D) := (U, U \cup D)$,
- $R\left(\bar{d}^{uv}_{ab}\right) := \overline{R\left(d^{uv}_{ab}\right)}$,
- $R(\mathtt{t}; \mathtt{s}) := R(\mathtt{t}); R(\mathtt{s})$.

From disconnections to reactions: Example



Soundness, completeness, universality

Proposition (Soundness)

R : **Disc** $\to$ **React** *is a dagger functor.*

Soundness, completeness, universality

Proposition (Soundness)

*R : **Disc** $\to$ **React** is a dagger functor.*

Theorem (Completeness)

*R : **Disc** $\to$ **React** is faithful: for any terms $\mathtt{t}, \mathtt{s}$ we have $R(\mathtt{t}) = R(\mathtt{s})$ in **React** if and only if $\mathtt{t} = \mathtt{s}$ in **Disc**.*

Soundness, completeness, universality

Proposition (Soundness)

*R : **Disc** $\to$ **React** is a dagger functor.*

Theorem (Completeness)

*R : **Disc** $\to$ **React** is faithful: for any terms $\mathtt{t}, \mathtt{s}$ we have $R(\mathtt{t}) = R(\mathtt{s})$ in **React** if and only if $\mathtt{t} = \mathtt{s}$ in **Disc**.*

Theorem (Universality)

*Given a reaction $r : A \to C$ in **React**, there is a term $\mathtt{t} : A \to B$ in **Disc** and an isomorphism $\iota : B \xrightarrow{\sim} C$ in **React** such that $R(\mathtt{t}); \iota = r$.*

Example

The reaction we saw (formation of benzyl benzoate from benzoyl chloride and benzyl alcohol) decomposes into:

$$C^{zu}_{ab}; C^{vw}_{cd}; C^{ru}_{ij}; C^{ru}_{nm}; E^{vc}; E^{wd}; E^{za}; E^{ub}; E^{ri}; E^{uj}; E^{rn}; E^{um};$$
$$\bar{E}^{vc}; \bar{E}^{wd}; \bar{E}^{za}; \bar{E}^{ub}; \bar{E}^{ri}; \bar{E}^{uj}; \bar{E}^{rn}; \bar{E}^{um}; \bar{C}^{ru}_{ij}; \bar{C}^{ru}_{nm}; \bar{C}^{wz}_{da}; \bar{C}^{uv}_{bc};$$
$$S^{z}; S^{u}; S^{v}; S^{w}; S^{r},$$

Example

The reaction we saw (formation of benzyl benzoate from benzoyl chloride and benzyl alcohol) decomposes into:

$$C^{zu}_{ab}; C^{vw}_{cd}; C^{ru}_{ij}; C^{ru}_{nm}; E^{vc}; E^{wd}; E^{za}; E^{ub}; E^{ri}; E^{uj}; E^{rn}; E^{um};$$
$$\bar{E}^{vc}; \bar{E}^{wd}; \bar{E}^{za}; \bar{E}^{ub}; \bar{E}^{ri}; \bar{E}^{uj}; \bar{E}^{rn}; \bar{E}^{um}; \bar{C}^{ru}_{ij}; \bar{C}^{ru}_{nm}; \bar{C}^{wz}_{da}; \bar{C}^{uv}_{bc};$$
$$S^{z}; S^{u}; S^{v}; S^{w}; S^{r},$$

which is equal to:

$$C^{zu}_{ab}; C^{vw}_{cd}; \bar{C}^{wz}_{da}; \bar{C}^{uv}_{bc}; S^{r}.$$

Future work

Chemical questions:

Future work

Chemical questions:

- Incorporate stereochemistry?

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

- Implement the algorithm that turns an arbitrary reaction into a sequence of disconnection rules

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

- Implement the algorithm that turns an arbitrary reaction into a sequence of disconnection rules
- Implement the normalisation procedure

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

- Implement the algorithm that turns an arbitrary reaction into a sequence of disconnection rules
- Implement the normalisation procedure

Mathematical questions:

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

- Implement the algorithm that turns an arbitrary reaction into a sequence of disconnection rules
- Implement the normalisation procedure

Mathematical questions:

- What is the categorical structure of **React** and **Disc**?

Future work

Chemical questions:

- Incorporate stereochemistry?
- Can we talk about dynamics/energy?
- Is the notion of composition appropriate?

Computational questions:

- Implement the algorithm that turns an arbitrary reaction into a sequence of disconnection rules
- Implement the normalisation procedure

Mathematical questions:

- What is the categorical structure of **React** and **Disc**?
- Monoidal structure?

References

- Andersen, Flamm, Merkle, Stadler. *Inferring chemical reaction patterns using rule composition in graph grammars*. Journal of Systems Chemistry. 2013.
- Clayden, Greeves, Warren. *Organic chemistry*. Oxford University Press. 2012.
- Gale, Lobski, Zanasi. *A categorical approach to synthetic chemistry*. International Colloquium on Theoretical Aspects of Computing. 2023.
- Gale, Lobski, Zanasi. *A categorical model for organic chemistry*. Theoretical Computer Science. Elsevier. To appear.
- Habel, Plump. *$\mathcal{M},\mathcal{N}$-Adhesive Transformation Systems*. International Conference on Graph Transformation. 2012.
- Warren, Wyatt. *Organic synthesis: the disconnection approach*. Wiley. 2008.

Thank you for your attention!

Reactions

The category **React** is defined as:

- objects: chemical graphs
- morphisms $A \to B$: tuples (U_A, U_B, b, i), where
  - $U_A \subseteq V_A$ and $U_B \subseteq V_B$ with $\mathtt{Net}\,(U_A) = \mathtt{Net}\,(U_B)$
  - $b : \mathtt{Chem}\,(U_A) \to \mathtt{Chem}\,(U_B)$ is a bijection preserving the atoms
  - $i : V_A \setminus U_A \to V_B \setminus U_B$ is an isomorphism

  such that for all $u \in \mathtt{Chem}\,(U_A)$ and $a \in V_A \setminus U_A$ we have

$$m_A(u, a) = m_B(bu, ia)$$

- the composition of $(U_A, U_B, b, i) : A \to B$ and $(W_B, W_C, c, j) : B \to C$ is

$$(Z_A, Z_C, (c + j)(b + i), ji) : A \to C$$

  where $Z_A := U_A \cup i^{-1}(W_B \setminus U_B)$ and $Z_C := W_C \cup j(U_B \setminus W_B)$
- the identity on A: $(\varnothing, \varnothing, !, \mathrm{id}_A)$

Proof ideas

Completeness

1. Show that every term is equal to the form

$$\mathtt{I}; \mathtt{C}; \mathtt{E}^{<0}; \mathtt{E}^{\geq 0}; \bar{\mathtt{E}}^{\geq 0}; \bar{\mathtt{E}}^{<0}; \bar{\mathtt{C}}; \bar{\mathtt{I}}; \mathtt{R}; \mathtt{S}$$

2. Under certain conditions, such normal form is unique
3. Show that if $R(\mathtt{t}) = R(\mathtt{s})$, then $\mathtt{t}$ and $\mathtt{s}$ have the same normal form

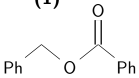
Universality

1. Every reaction $r : A \to B$ factorises as $(\varnothing, \varnothing, !, \iota) \circ (A, B, \mathrm{id}, \mathrm{id})$
2. Keep applying disconnections to A until there is nothing to disconnect
3. Apply connections to obtain B: preservation of atoms and charge guarantees that this can always be done

Retrosynthetic analysis

(1) Start with the target molecule(s)

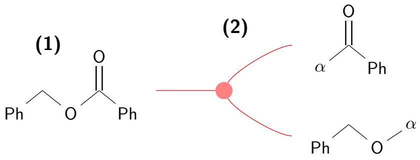
(1)



Retrosynthetic analysis

(1) Start with the target molecule(s)

(2) Cut the target along some bond, creating *synthons*

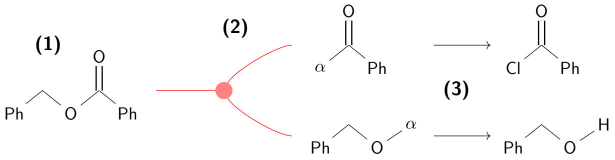


Retrosynthetic analysis

(1) Start with the target molecule(s)

(2) Cut the target along some bond, creating *synthons*

(3) Search for *synthetic equivalents*



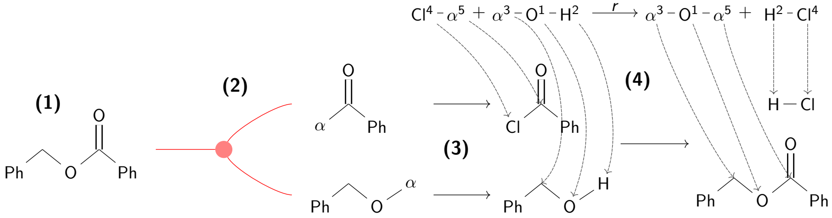
Retrosynthetic analysis

(1) Start with the target molecule(s)

(2) Cut the target along some bond, creating *synthons*

(3) Search for *synthetic equivalents*

(4) Search for a reaction whose reactants contain the synthetic equivalents, and whose products contain the target



Retrosynthetic analysis

(1) Start with the target molecule(s)

(2) Cut the target along some bond, creating *synthons*

(3) Search for *synthetic equivalents*

(4) Search for a reaction whose reactants contain the synthetic equivalents, and whose products contain the target

(5) Check whether the synthetic equivalents are known molecules: if yes, terminate, if no, return to **(1)** taking them as the target