A categorical approach to synthetic chemistry

Leo Lobski
joint work with Fabio Zanasi (UCL) and Ella Gale (Bristol)

University College London, UK

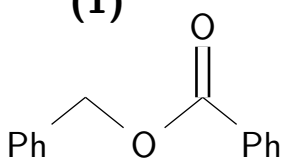
International Colloquium on Theoretical Aspects of Computing
8 December 2023, Lima, Peru

Outline

Retrosynthetic analysis

(1) Start with the target molecule(s)

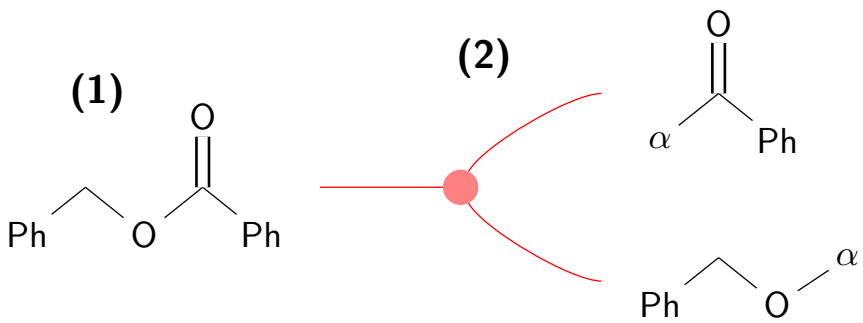
(1)



Retrosynthetic analysis

(1) Start with the target molecule(s)

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

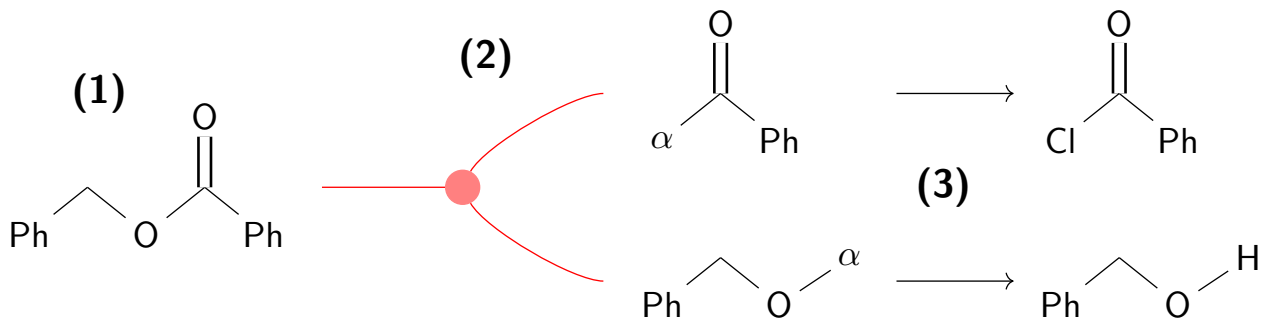


Retrosynthetic analysis

(1) Start with the target molecule(s)

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

(3) Search for *synthetic equivalents*



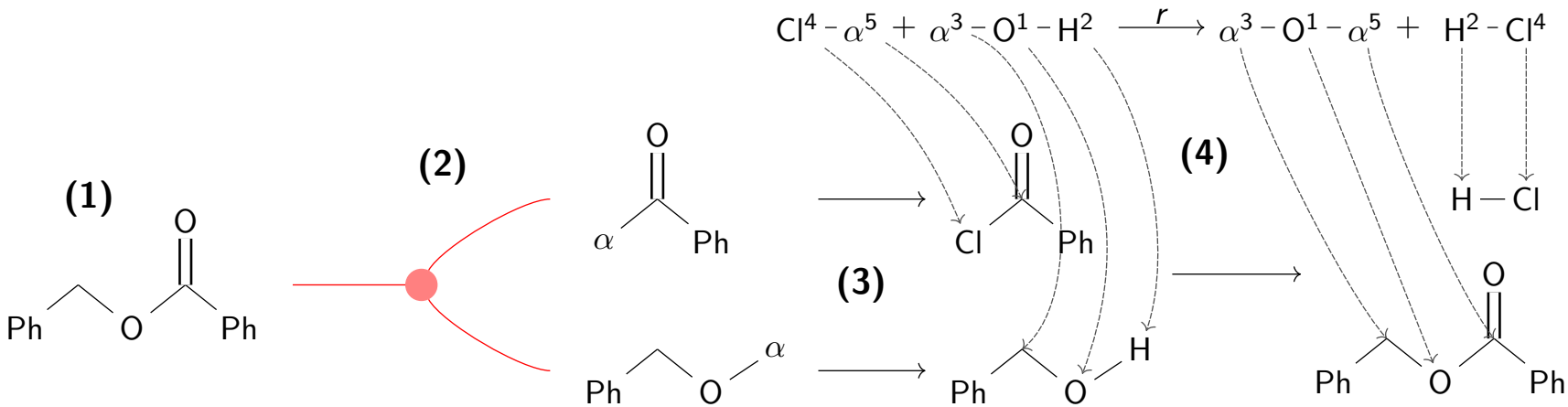
Retrosynthetic analysis

(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

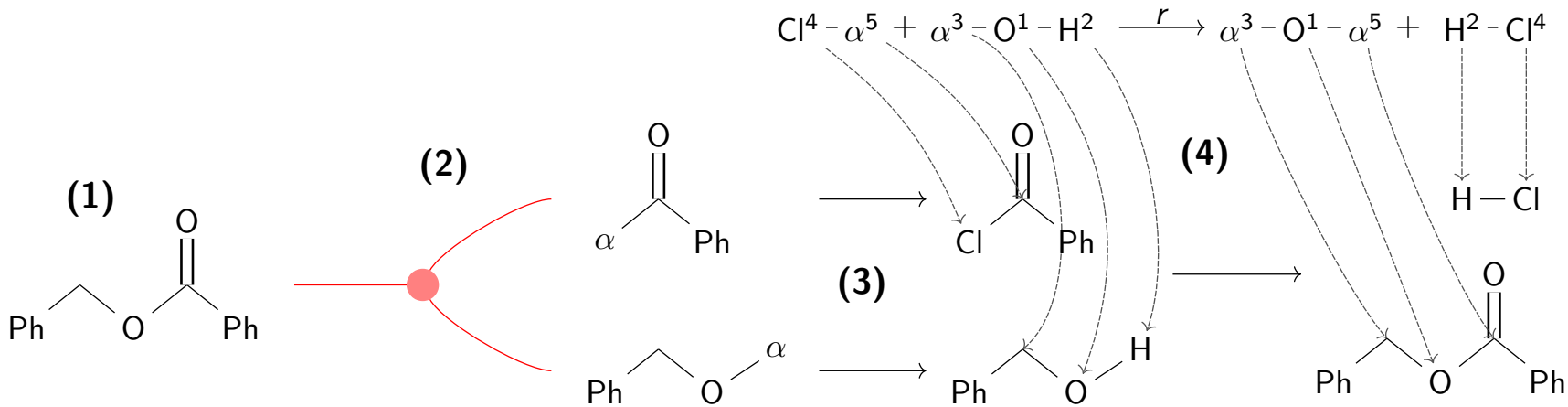
(1) Start with the target molecule(s)

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

(3) Search for *synthetic equivalents*

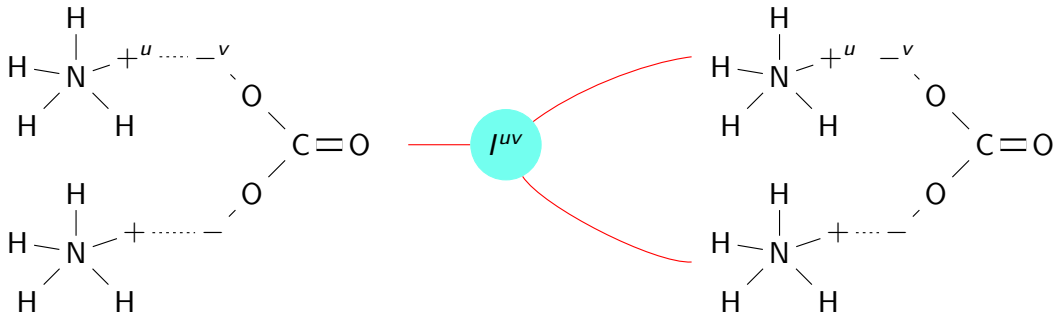
(4) Search for a reaction whose reactants contain the synthetic equivalents, and whose products contain the target

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



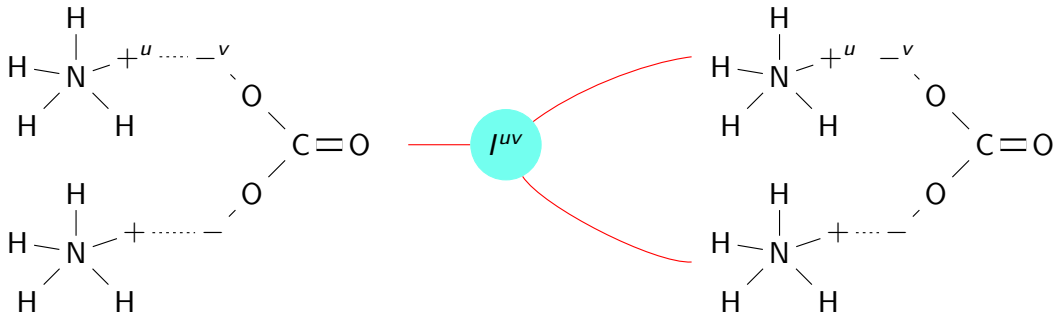
Disconnection rules

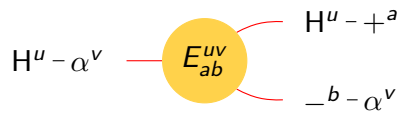
Motto: Chemical reactions are movements of electrons



Disconnection rules

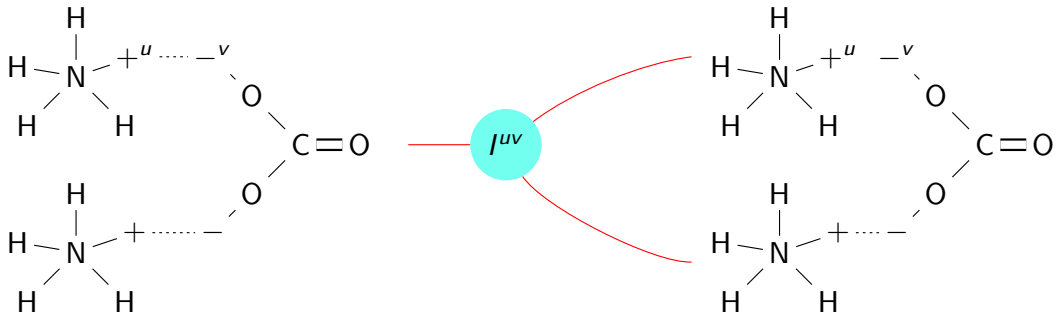
Motto: Chemical reactions are movements of electrons

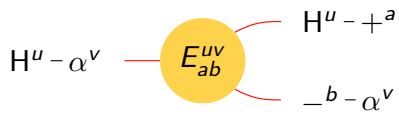


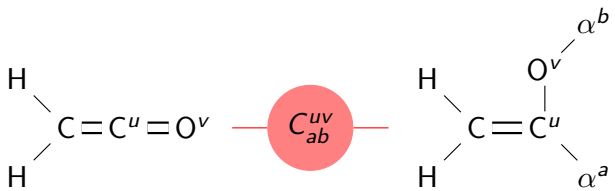


Disconnection rules

Motto: Chemical reactions are movements of electrons

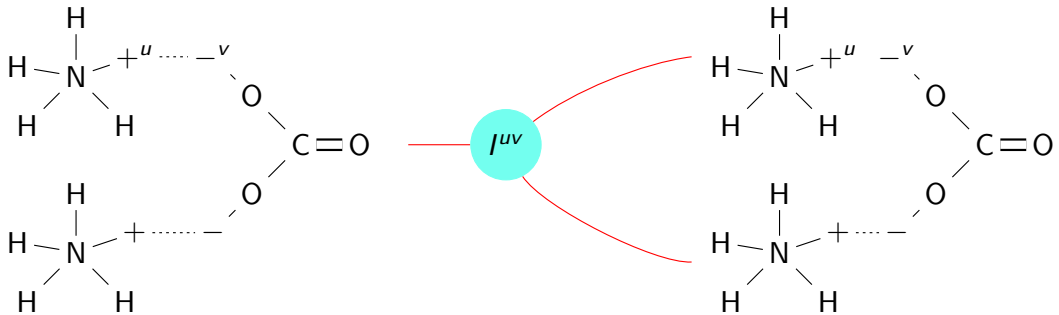


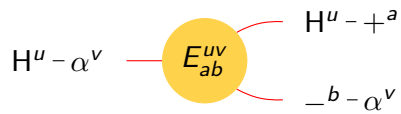


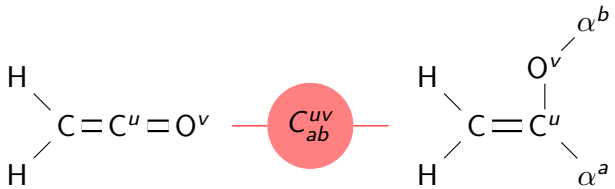


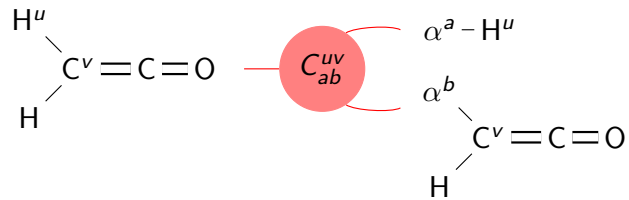
Disconnection rules

Motto: Chemical reactions are movements of electrons



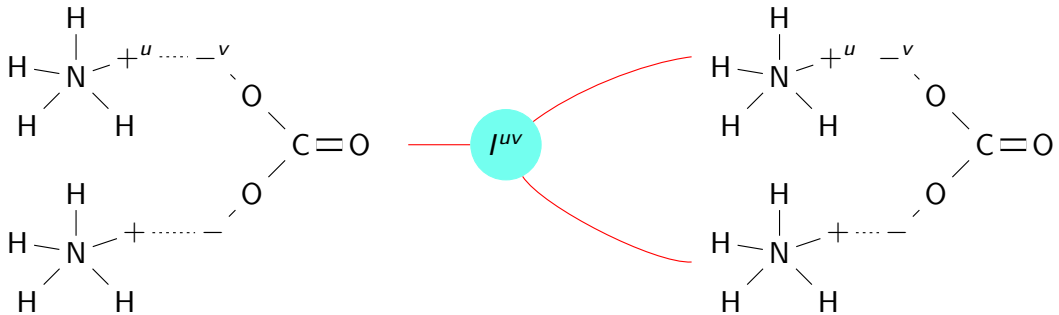


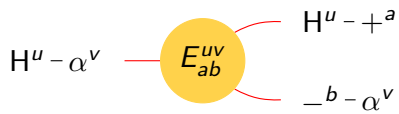


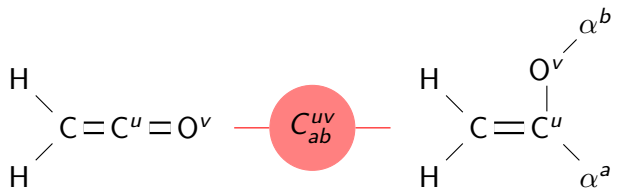


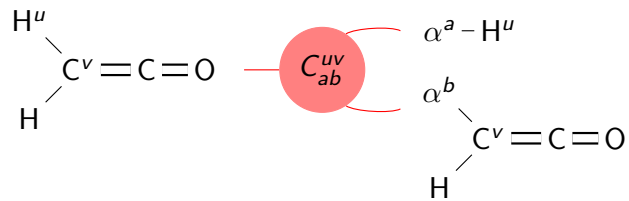
Disconnection rules

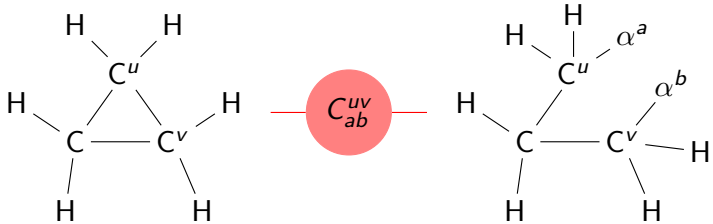
Motto: Chemical reactions are movements of electrons





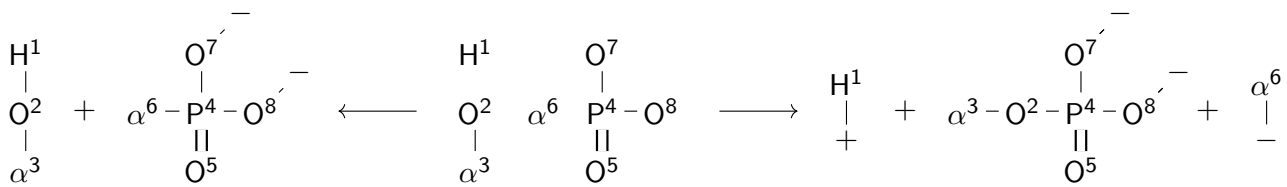






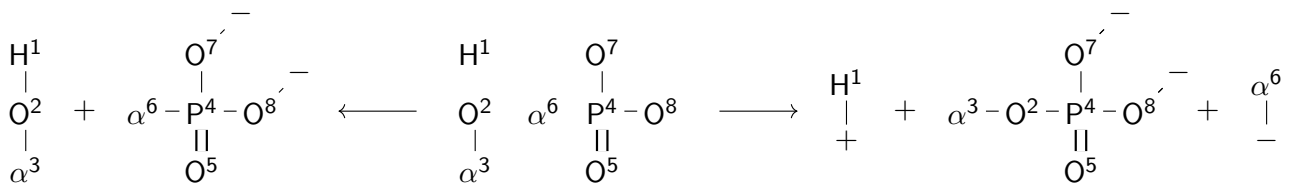
Reactions

Reactions are graph rewrites generated using *reaction schemes*:

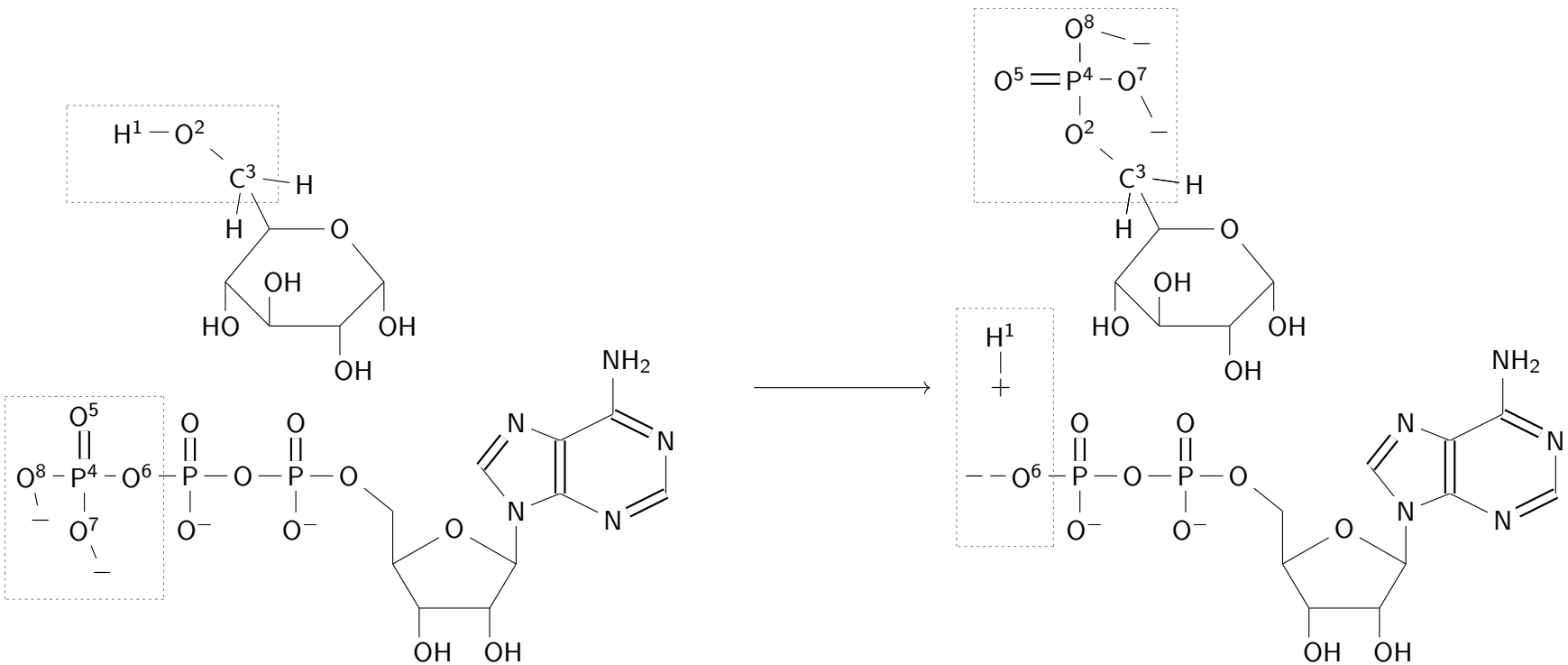


Reactions

Reactions are graph rewrites generated using *reaction schemes*:

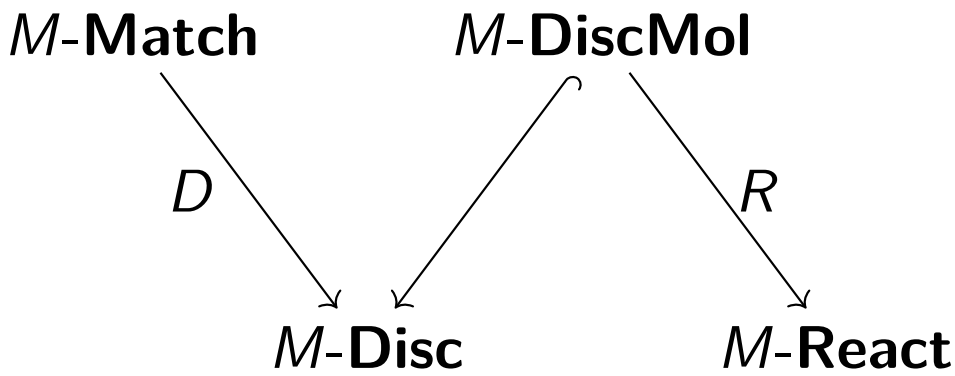


by finding a *matching* for the left-hand side in a larger entity:



Retrosynthesis, formalised

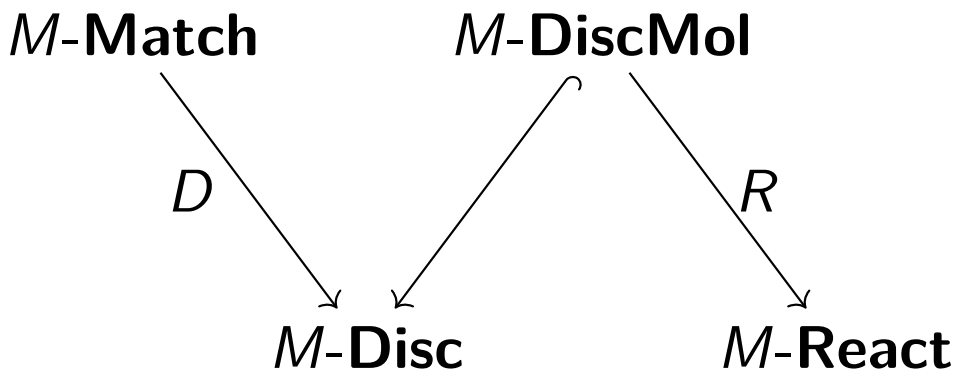
There are the following functors:



between categories of matchings, disconnection rules and reactions,

Retrosynthesis, formalised

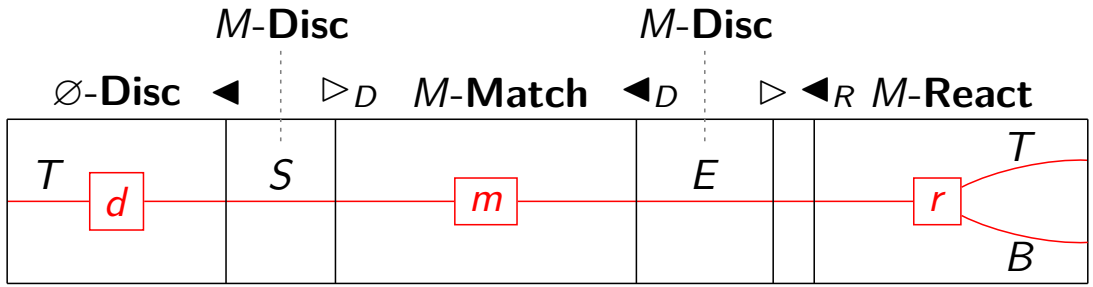
There are the following functors:



between categories of matchings, disconnection rules and reactions, all parameterised by the environmental molecules (solvent, catalyst, reagent). This is represented by a finite set of molecules M.

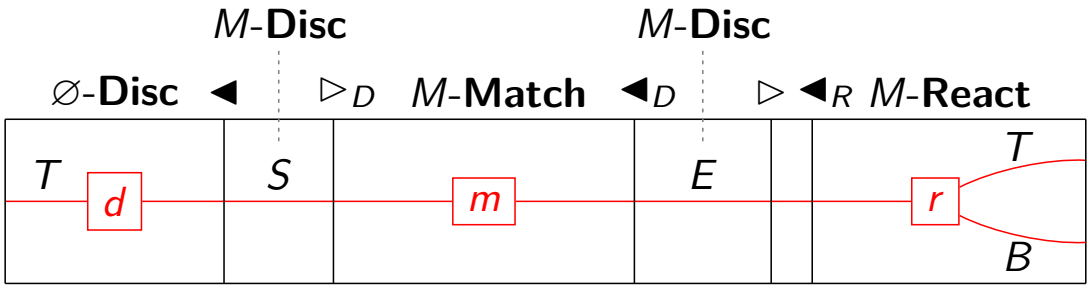
Retrosynthesis, formalised

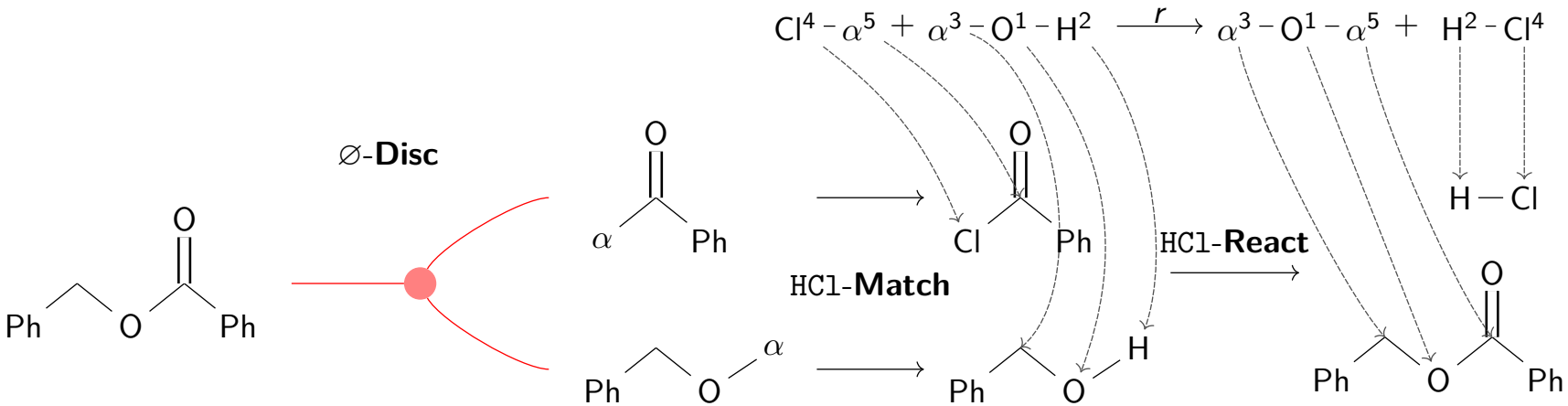
This allows for the following definition of a retrosynthetic step:



Retrosynthesis, formalised

This allows for the following definition of a retrosynthetic step:

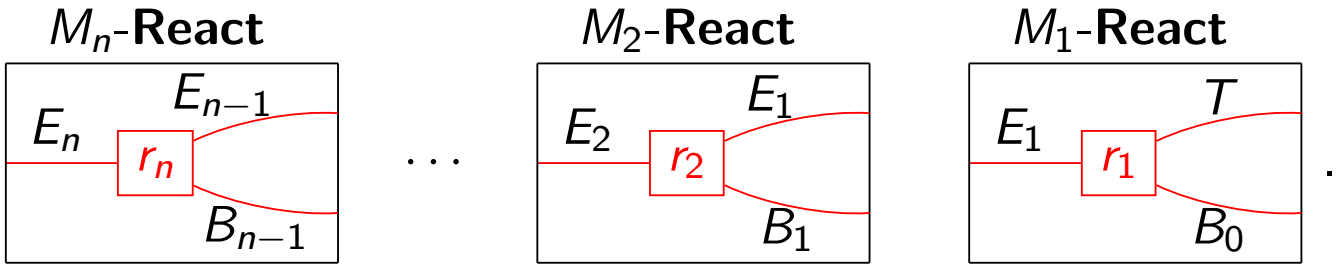




Retrosynthesis, formalised

Definition (Retrosynthetic sequence)

A *retrosynthetic sequence* for a target molecular entity T is a sequence of morphisms $r_1 \in M_1\text{-}\mathbf{React}(E_1, T + B_0)$, $r_2 \in M_2\text{-}\mathbf{React}(E_2, E_1 + B_1)$, $\ldots$, $r_n \in M_1\text{-}\mathbf{React}(E_n, E_{n-1} + B_{n-1})$ such that the domain of r_i is a connected subgraph of the codomain of r_{i+1}:



Retrosynthesis, formalised

Let T be some fixed molecular entity. We initialise by setting $i = 0$ and $E_0 := T$.

1. Choose a subset $\mathcal{D}$ of disconnection rules,
2. Provide at least one of the following:
   (a) a finite set of reaction schemes $\mathcal{S}$,
   (b) a function $\mathfrak{F}$ from molecular graphs to finite sets of molecular graphs,
3. Search for a retrosynthetic step with $d \in \varnothing\text{-}\mathbf{Disc}(E_i, S)$, $m \in M\text{-}\mathbf{Match}(S, E)$, and $r \in M\text{-}\mathbf{React}(E, E_i + B_i)$ such that all disconnection rules in d and $D(m)$ are in $\mathcal{D}$, and we have at least one of the following:
   (a) there is an $s \in \mathcal{S}$ such that the reaction r is an instance of s,
   (b) $E_i + B_i \in \mathfrak{F}(E)$;

   if successful, set $E_{i+1} := E$, $M_{i+1} := M$, $r_{i+1} := r$ and proceed to Step 4; if unsuccessful, stop,
4. Check if the molecular entities in E_{i+1} are known (commercially available): if yes, terminate; if no, increment $i \mapsto i + 1$ and return to Step 1.

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Future work:

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Future work:

- Wish to model the retrosynthetic steps and full retrosynthesis

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Future work:

- Wish to model the retrosynthetic steps and full retrosynthesis
- Prove that all reactions arise from the (dis)connection rules

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Future work:

- Wish to model the retrosynthetic steps and full retrosynthesis
- Prove that all reactions arise from the (dis)connection rules
- Encode existing retrosynthesis algorithms within framework

Conclusion

- Gave precise definitions of retrosynthetic steps and sequences
- Described how to model retrosynthetic algorithms
- Wrote down the retrosynthetic disconnection rules

Future work:

- Wish to model the retrosynthetic steps and full retrosynthesis
- Prove that all reactions arise from the (dis)connection rules
- Encode existing retrosynthesis algorithms within framework
- Suggest features for new algorithms: e.g. chirality, reaction environment, protection-deprotection steps

References

- Andersen, Flamm, Merkle, Stadler. *An intermediate level of abstraction for computational systems chemistry*. Philosophical Transactions of the Royal Society A. 2017.
- Clayden, Greeves, Warren. *Organic chemistry*. Oxford University Press. 2012.
- Lobski, Zanasi. *String diagrams for layered explanations*. Applied Category Theory. 2022.
- Sun, Sahinidis. *Computer-aided retrosynthetic design: fundamentals, tools, and outlook*. Current Opinion in Chemical Engineering. 2022.
- Warren, Wyatt. *Organic synthesis: the disconnection approach*. Wiley. 2008.

Thank you for your attention!

Layers of abstraction

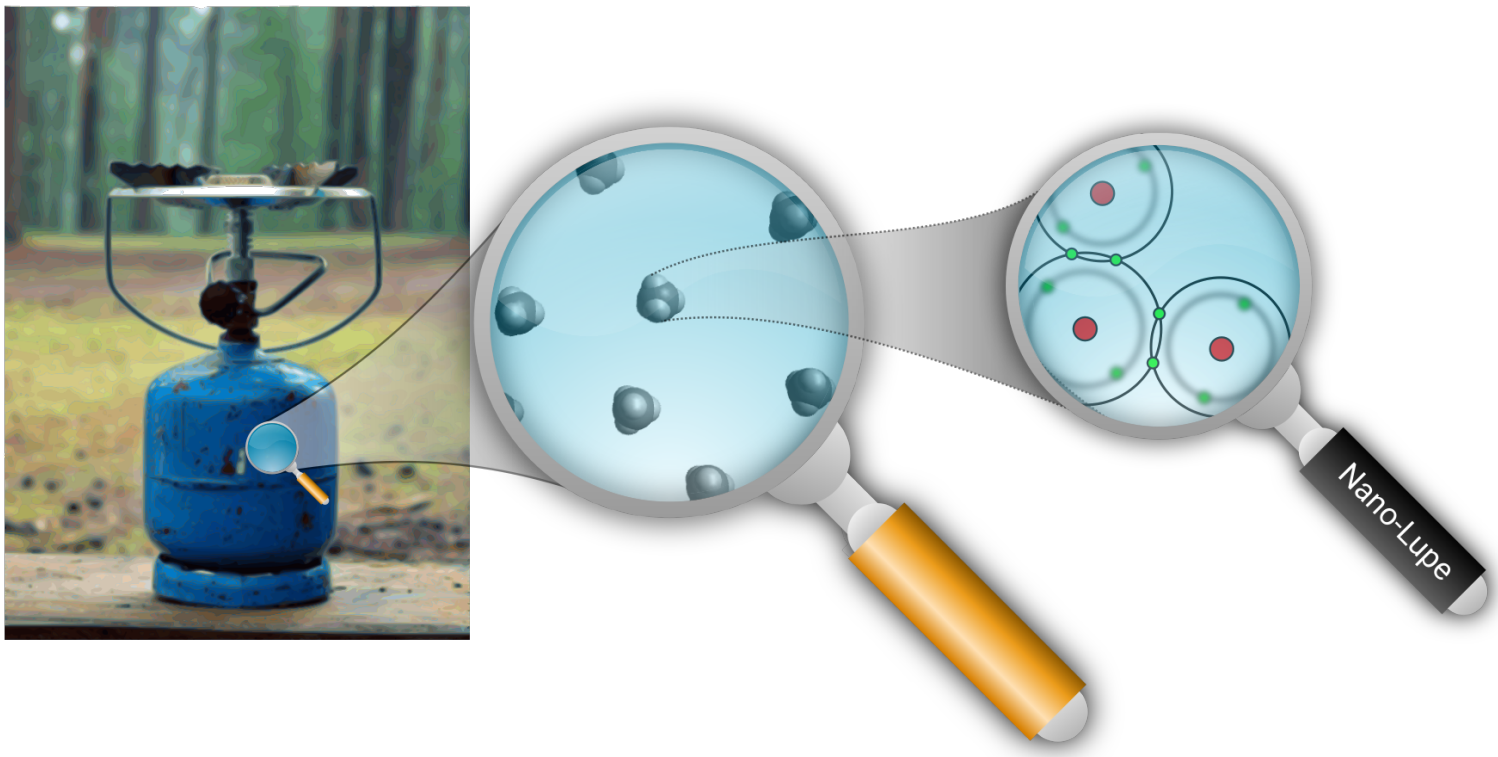


Image source: Openclipart

Layered props

Definition

A *layered signature* is a functor $\Omega : P \to \mathbf{StrMon}$ from a poset P which does not send non-isomorphic objects to isomorphic objects.

Layered props

Definition

A *layered signature* is a functor $\Omega : P \to \mathbf{StrMon}$ from a poset P which does not send non-isomorphic objects to isomorphic objects.

Definition

A *layered prop* generated by Ω is a 2-category whose 0-cells are lists $(\omega_1, a_1; \ldots; \omega_n, a_n)$ of pairs (ω, a), where $\omega \in P$ and $a \in \Omega(\omega)$,

Layered props

Definition

A *layered signature* is a functor $\Omega : P \to \mathbf{StrMon}$ from a poset P which does not send non-isomorphic objects to isomorphic objects.

Definition

A *layered prop* generated by Ω is a 2-category whose 0-cells are lists $(\omega_1, a_1; \ldots; \omega_n, a_n)$ of pairs (ω, a), where $\omega \in P$ and $a \in \Omega(\omega)$,and whose 1-cells are generated by the procedure below:

Layered props: 1-cells

$$\frac{\omega \le \tau \quad \alpha \in \Omega(\omega) \quad \Omega(\omega \le \tau) = f}{\omega \;\; \alpha \;\; f\alpha \;\; \tau : (\omega, \alpha \mid \tau, f\alpha)}$$

$$\frac{\omega \le \tau \quad \alpha \in \Omega(\omega) \quad \Omega(\omega \le \tau) = f}{\tau \;\; f\alpha \;\; \alpha \;\; \omega : (\tau, f\alpha \mid \omega, \alpha)}$$

$$\frac{\alpha, \beta \in \Omega(\omega) \quad \sigma : \alpha \to \beta}{\omega \;\; \alpha \;\; \sigma \;\; \beta \;\; \omega : (\omega, \alpha \mid \omega, \beta)}$$

$$\frac{\alpha \in \Omega(\omega) \quad \beta \in \Omega(\tau)}{\omega \;\; \alpha \;\; \beta \;\; \tau \;\; \tau \;\; \beta \;\; \alpha \;\; \omega : (\omega, \alpha; \tau, \beta \mid \tau, \beta; \omega, \alpha)}$$

$$\frac{}{: (\varepsilon \mid \varepsilon)}$$

$$\frac{}{\varepsilon \;\; \varepsilon \;\; \omega : (\varepsilon \mid \omega, I_\omega)}$$

$$\frac{}{\omega \;\; \varepsilon \;\; \varepsilon : (\omega, I_\omega \mid \varepsilon)}$$

$$\frac{\alpha, \beta \in \Omega(\omega)}{\omega \;\; \alpha \;\; \omega \;\; \beta \;\; \alpha \;\; \beta \;\; \omega : (\omega, \alpha; \omega, \beta \mid \omega, \alpha \otimes_\omega \beta)}$$

$$\frac{\alpha, \beta \in \Omega(\omega)}{\omega \;\; \alpha \;\; \beta \;\; \alpha \;\; \omega \;\; \beta \;\; \omega : (\omega, \alpha \otimes_\omega \beta \mid \omega, \alpha; \omega, \beta)}$$

$$\frac{x : (t \mid s) \quad y : (s \mid u)}{x ; y : (t, u)}$$

$$\frac{x : (t \mid s) \quad y : (u \mid w)}{x \otimes y : (t; u \mid s; w)}$$

Layered props: 2-cells

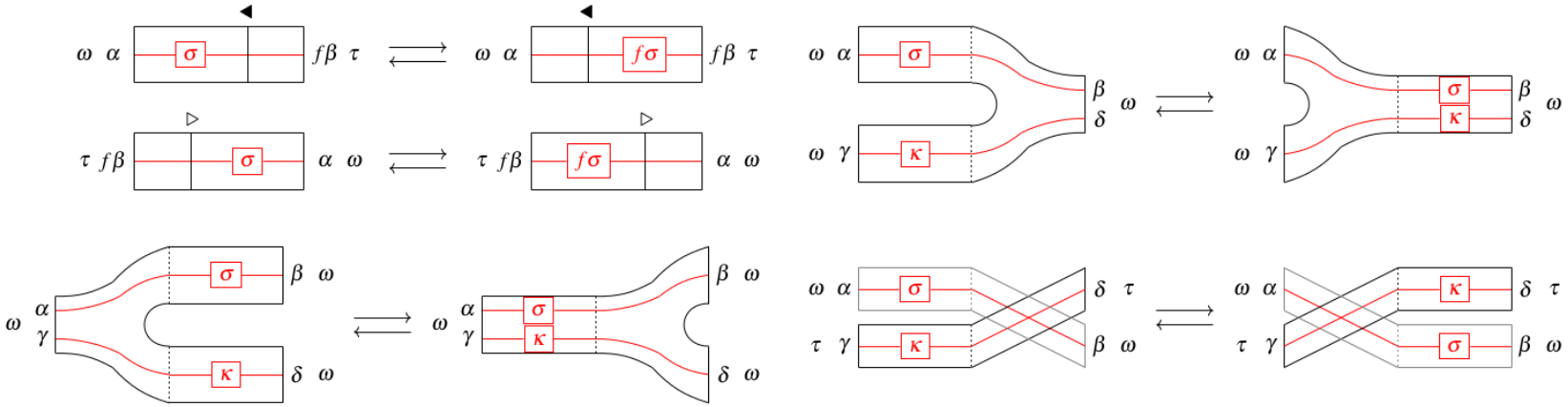


Figure 2: 2-cells of a layered prop expressing functoriality of refinement, coarsening, pants and copants.

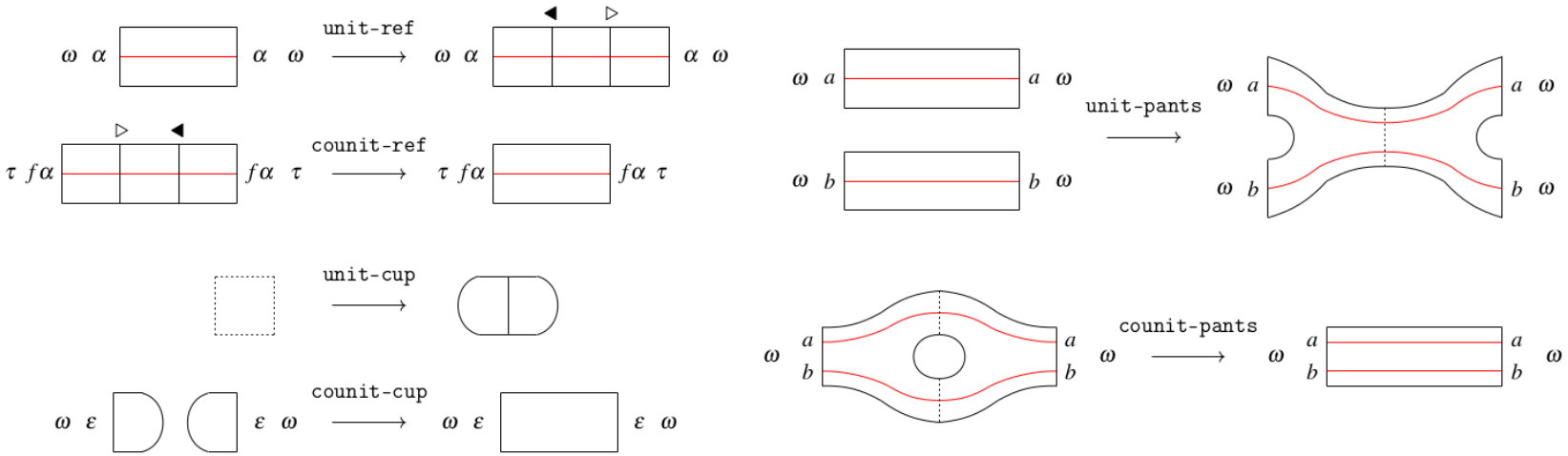


Figure 3: 2-cells of a layered prop that exhibit pants-copants and refinement-coarsening as two adjoint pairs.

Layered props: 2-cells

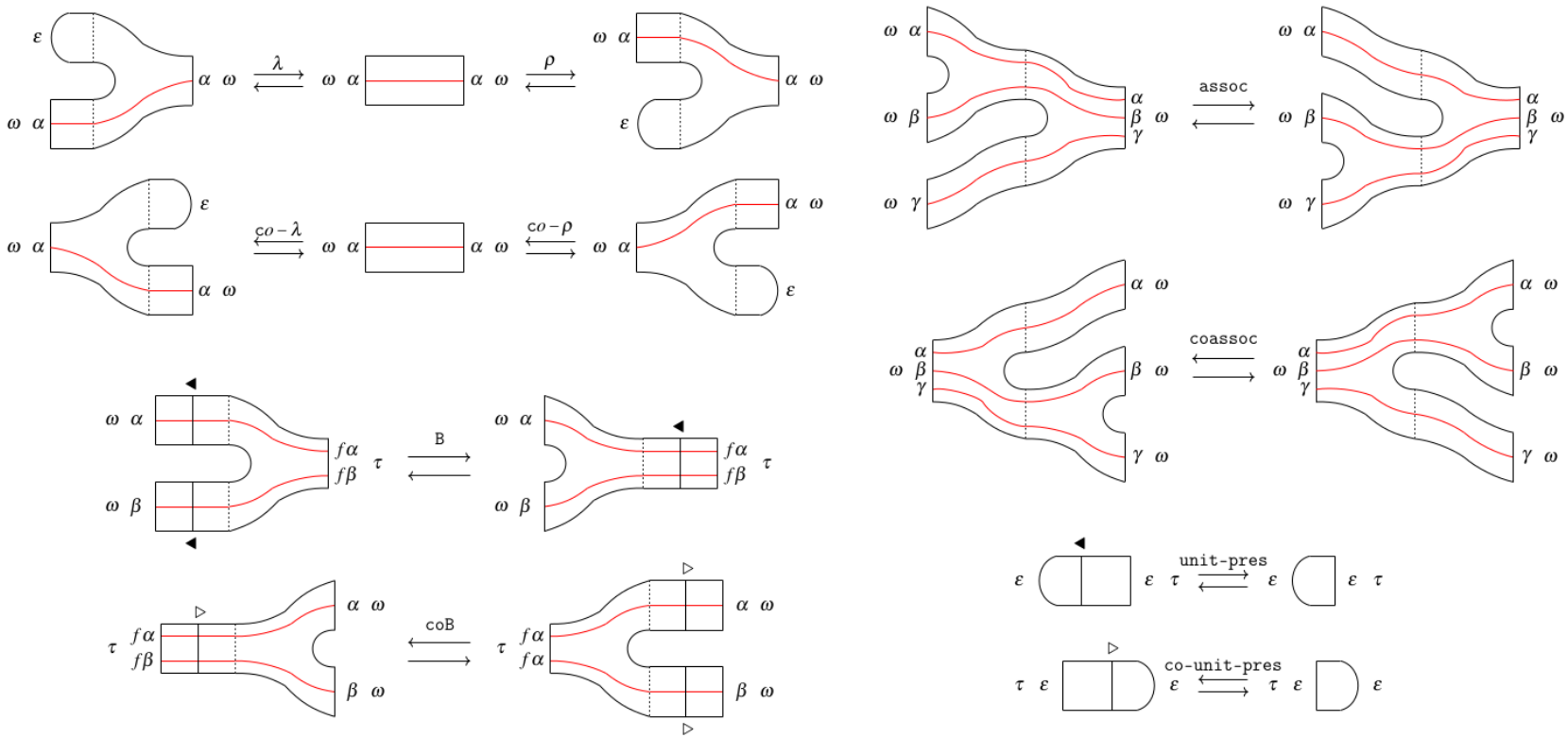


Figure 4: 2-cells of a layered prop that are motivated by monoidal categories and functors.