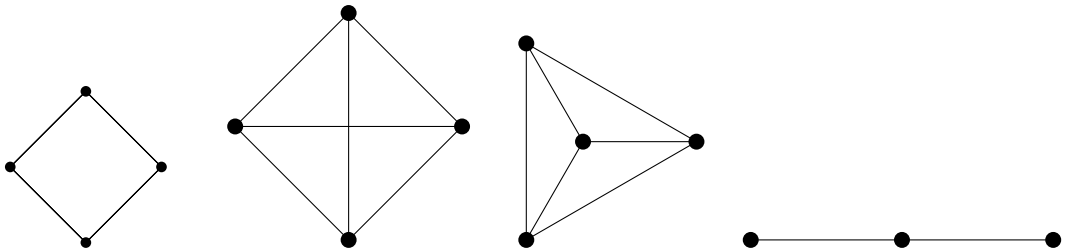
392013 Exercises Algorithmic Cheminformatics

Exercise 01. April 24, 2025
(no mandatory exercises)

1 Morphisms

Given the following 4 graphs $G_1, \ldots G_4$



G_1 G_2 G_3 G_4

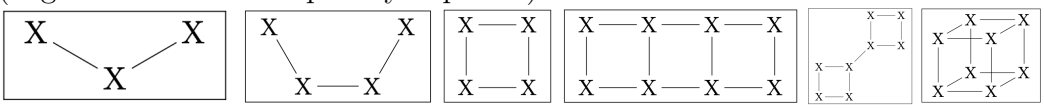
For all pairs of graphs (G_i, G_j), determine if

1. there exists a morphism from G_i to G_j.
2. there exists a monomorphism from G_i to G_j.
3. there exists a subgraph isomorphism from G_i to G_j.
4. there exists an isomorphism from G_i to G_j.

2 Double Pushout Approach (DPO)

1. Design a graph transformation rule $p = (L \xleftarrow{l} K \xrightarrow{r} R)$ that removes exactly one edge from a (not necessarily chemical) graph G. For the rule you can assume, that the label of any vertex in G is `X`, and that the label of any edge in G is `-`.

2. Answer all of the following questions for the following 5 graphs $G_1, \ldots, G_6$ (edge labels are not explicitly depcited):



   (a) How many different derivations $G_x \overset{p,m}{\Longrightarrow} H$ (resulting from the application of the rule p to the graph G_x) do exist? (Note: two derivations are different if the resulting multiset of graphs H is different). For each derivation specify the morphism m and the resulting multiset H.

   (b) Draw the DPO diagram for all different derivations $G_6 \overset{p,m}{\Longrightarrow} H$.

   (c) Let $G \overset{p,m}{\Longrightarrow} H$ be a derivation. A *one-to-one* derivation is a derivation for which $|G| = |H| = 1$. A *one-to-many* derivation is a derivation for which $|G| = 1$ and $|H| > 1$.

      i. What is the minimum number of consecutive *one-to-one* derivations that must be applied to G_x, before a *one-to-many* derivation can be applied.

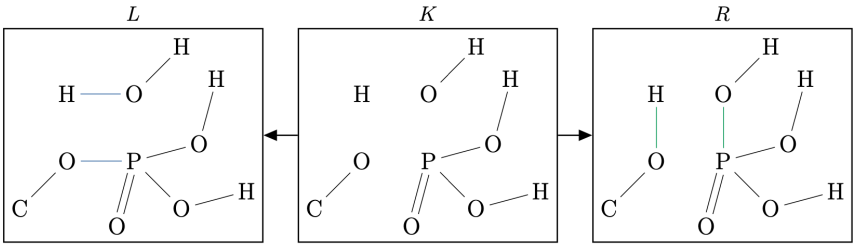
      ii. What is the maximum number of consecutive *one-to-one* derivations, that is applicable to G_x.

      iii. What is the minimum number of consecutive *one-to-one* derivations, such that a subsequent one-to-many derivation would result in a multiset $H = \{H_1, H_2\}$, where H_1 and H_2 contain the same number of vertices.

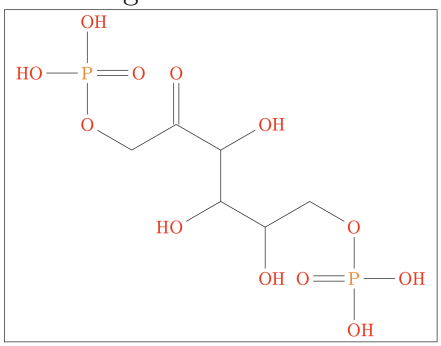
      iv. (*) Can you relate the questions i.) to iii.) to problems you (might) know of? Do you know the the computational complexity of answering those questions?

3 Biochemical DPO Rule Application

The following DPO rule models an enzymatic reaction (phophohydrolase).

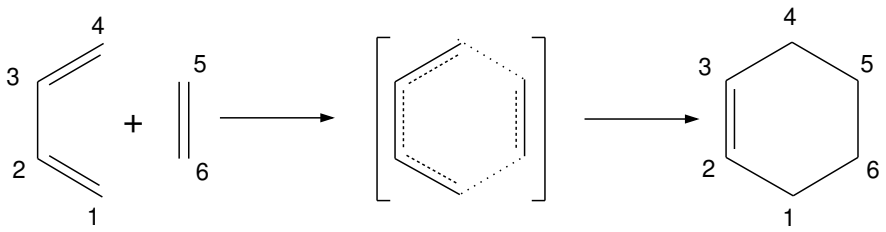


Apply the rule to the set consisting of a water molecule and the following molecule:



4 Writing a Graph Transformation Rule

The Diels-Alder reaction is a [4+2]-cycloaddition between a conjugated diene and an alkene, to form a (substituted) cyclohexene system. The reaction is thought to proceed *via* a 6-cyclic transition state (bracketed structure). The Diels-Alder reaction is used to illustrate how graph rewrite rules are formulated.




```
rule [
 ruleID "Diels-Alder reaction"
 left [
  edge [ source 1 target 2 label "=" ]
  edge [ source 2 target 3 label "-" ]
  edge [ source 3 target 4 label "=" ]
  edge [ source 5 target 6 label "=" ]
 ]
 context [
  node [ id 1 label "C" ]
  node [ id 2 label "C" ]
  node [ id 3 label "C" ]
  node [ id 4 label "C" ]
  node [ id 5 label "C" ]
  node [ id 6 label "C" ]
 ]
 right [
  edge [ source 1 target 2 label "-" ]
  edge [ source 2 target 3 label "=" ]
  edge [ source 3 target 4 label "-" ]
  edge [ source 4 target 5 label "-" ]
  edge [ source 5 target 6 label "-" ]
  edge [ source 6 target 1 label "-" ]
 ]
]
```

1. Create a rewrite rule as given above and name the corresponding file `DA.gml`.
2. Generate a file named `grammar-DA.py` with the following content

```
smiles('C=C(C)C=C', 'Isoprene')
ruleGML('DA.gml')
```

3. Generate a file named `doit.py` with the following content

```
# exploration strategy for chemical space defined by grammar-DA.py
strat = (addSubset(inputGraphs) >> repeat[1](inputRules))

# calculate and hypergraph
dg = dgRuleComp(inputGraphs, strat)
dg.calc()
dg.print()

# print DPOs of Rules
postSection("Rule")
for r in inputRules:
    r.print()
```

4. Expand chemical space for 1 step using the tool `mød`

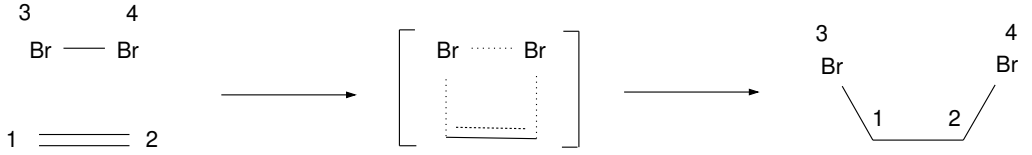
```
> mod -f grammar-DA.py -f doit.py
```

5. Look at the generated summary `summary/summary.pdf`.

6. Change the startegy statement in file `doit.py` such that the chemical space is expanded for 2 steps and start the calculation and inspect the summary.

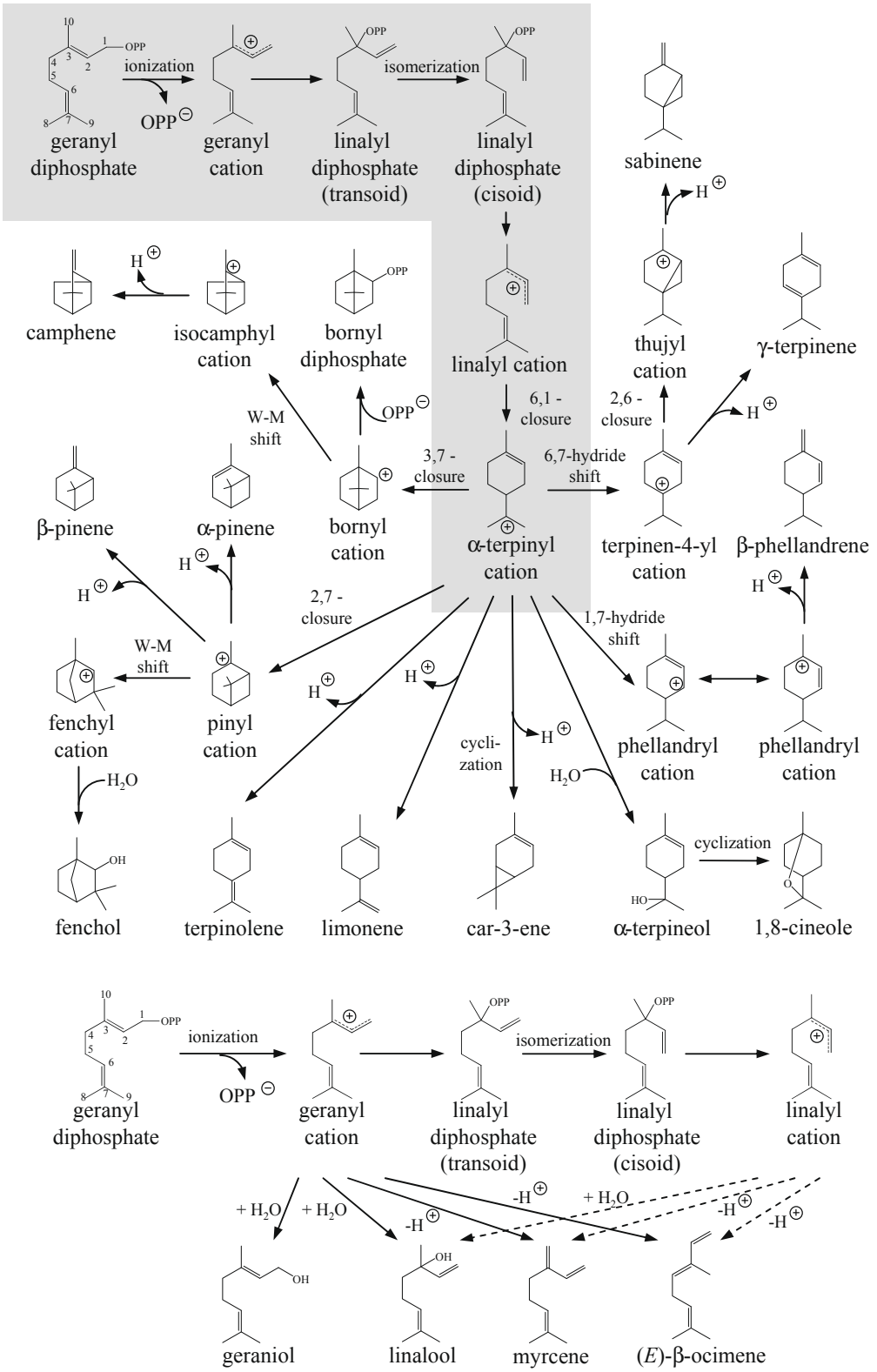
7. Use an editor of your choice to formulate the bromination depicted below in a file called `brom.gml`. In this reaction a Br_2 molecule is added to a C=C bond. The reaction is thought to proceed *via* a 4-cyclic transition state (bracketed structure).

8. Write a `grammar-Br.py` file and check, with `mød` that the rule works.



9. Mix the Diels-Alder and bromination grammars and look at resulting chemical space using 1 or 2 expansion steps.

5 Monoterpene Reaction Chemistry (voluntary)

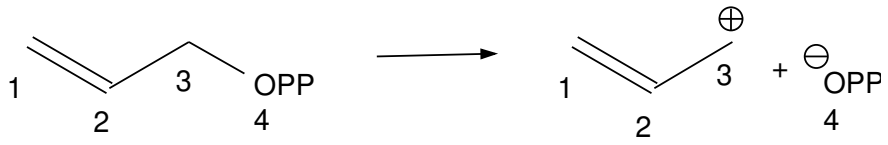


Monoterpene Synthetases The reaction mechanism starts with the ionization of the geranyl diphosphate substrate. The resulting carbocation can undergo a range of cyclizations, hydride shifts and rearrangements before reaction is terminated by deprotonation or water capture. Figure from Degenhardt J et al (2009), Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, Phytochem **70**:1621-1637 | doi:10.1016/j.phytochem.2009.07.030

Charges and complex groups such diphosphate (OPP) play an essential role in terpene chemistry. Molecules containing complex groups, which shall not be modeled explicitly but abbreviated as multi-character labels are defined in møl using the `SMILES`-like `graphDFS` syntax. For example the diphosphate group in dimethylallyl pyrophosphate DMAPP is abbreviated as OPP label

```
# DMAPP, dimethylallylpyrophosphate (OPP group treated as entity)
graphDFS("CC(C)=CC[OPP]");
```

which can than be used as label in graph rewrite rules. Charges are treated as label changes, hence atoms gaining/loosing charges show up as nodes in the `left` and `right` graph with the same `id`. The following rule illustrates these features, encoding the ionization reaction of geranyl diphosphatre.




```
rule [
 ruleID "Loss of OPP"
 left [
  node [id 4 label "OPP" ]
  node [id 3 label "C" ]
  edge [source 3 target 4 label "-" ]
 ]
 context [
  node [id 1 label "C" ]
  node [id 2 label "C" ]
  edge [source 1 target 2 label "=" ]
  edge [source 2 target 3 label "-" ]
 ]
 right [
  node [id 4 label "OPP-" ]
  node [id 3 label "C+" ]
 ]
]
```

10. Formulate some of the reaction arrows as graph-rewrite rules and test them out (Idealy choose a reaction sequence starting from geranyl diphosphate ending in one of the uncharged cyclic endproducts).