An Ontological Approach to Medicinal Chemistry

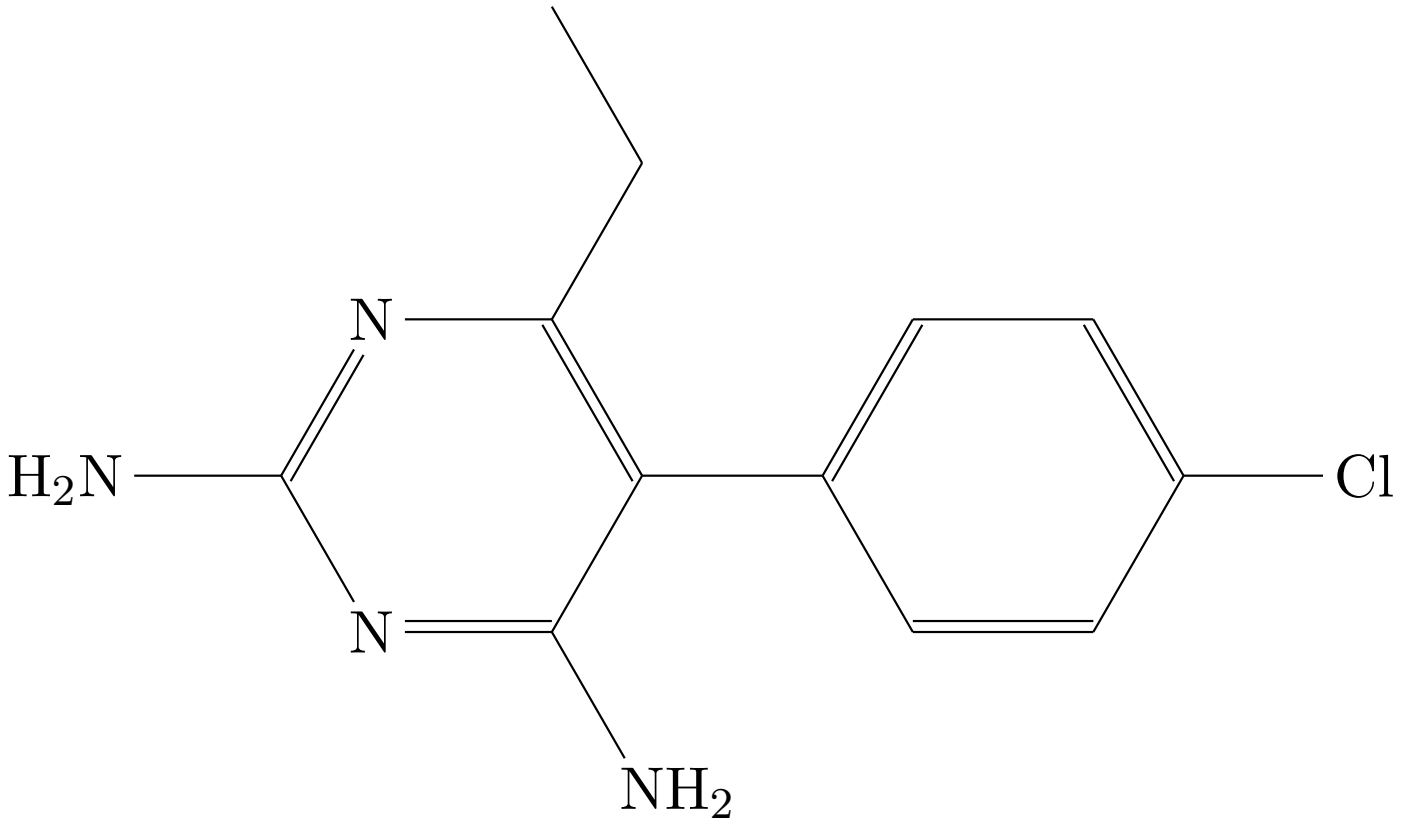
Carmen Chui and Michael Grüninger

Semantic Technologies Lab, University of Toronto

FOIS 2016

How would you describe and represent this molecule?

Pyrimethamine/Daraprim ($C_{12}H_{13}ClN_4$)



Approach #1: Cheminformatics (Identifiers)

- Names:
  - Pyrimethamine; 58-14-0; Daraprim; Chloridine; 5-(4-chlorophenyl)-6-ethylpyrimidine-2,4-diamine; Ethylpyrimidine;
- IUPAC name:

  `5-(4-chlorophenyl)-6-ethylpyrimidine-2,4-diamine`
- InChI identifier:

  `InChI=1S/C12H13ClN4/c1-2-9-10(11(14)17-12(15)16-9)7-3-5-8(13)6-4-7/h3-6H,2H2,1H3,(H4,14,15,16,17)`

Approach #2: Cheminformatics (molfile output)

```
 17 18  0  0  0  0            999 V2000
   18.1293  -10.4004    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.4147  -10.8129    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.1293   -9.5754    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.8437  -10.8129    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.4147  -11.6379    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.7003  -10.4004    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.8437   -9.1629    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   17.4147   -9.1629    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.5582  -10.4004    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   18.8437  -11.6379    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   16.7003  -12.0504    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   15.9858  -10.8129    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   19.5582   -9.5754    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.7003   -9.5754    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   15.9858  -11.6379    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.2727   -9.1629    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   15.2713  -12.0504    0.0000 Cl  0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  3  2  0  0  0  0
  1  4  1  0  0  0  0
  2  5  2  0  0  0  0
  2  6  1  0  0  0  0
            ...
 13 16  1  0  0  0  0
 15 17  1  0  0  0  0
  9 13  1  0  0  0  0
 12 15  1  0  0  0  0
M  END
```

Approach #3: ChEBI

(Excerpts extracted from ChEBI)

- pyrimethamine (CHEBI:8673) **has role** antimalarial (CHEBI:38068)
- pyrimethamine (CHEBI:8673) **has role** antiprotozoal drug (CHEBI:35820)
- pyrimethamine (CHEBI:8673) **has role** EC 1.5.1.3 (dihydrofolate reductase) inhibitor (CHEBI:50683)
- pyrimethamine (CHEBI:8673) **is a** aminopyrimidine (CHEBI:38338)
- pyrimethamine (CHEBI:8673) **is a** monochlorobenzenes (CHEBI:83403)

Approach #4: Molecular Graph Theory

- Represent a molecule as a graph
  - Vertices represent atoms
  - Edges represent the adjacency relation

What's the Problem?

Conventional Approaches

- No *semantics* for symbols used to represent molecules
- No *reasoning* capabilities about shape

Ontological Approaches

- Not enough *semantics* for shape representation
- No *reasoning* capabilities about shape

Competency Questions

Similarity

- Which molecules have common substructures with a given molecule?
- Which antibiotics contain a β-lactam ring?
- What are molecules that contain two fused rings?
- Which molecules contain a given functional group?

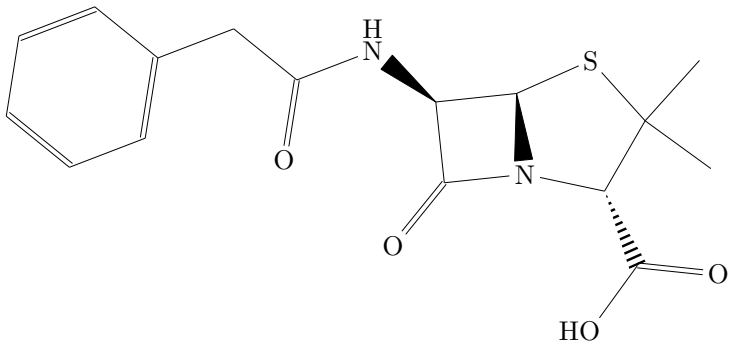


Figure: Penicillin ($C_9H_{11}N_2O_4S$)

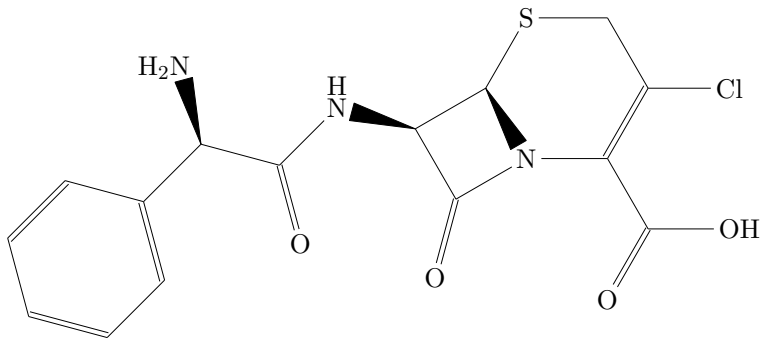


Figure: Cefaclor ($C_{15}H_{14}ClN_3O_4S$)

Competency Questions

Substitution

- What molecules are equivalent to molecule x after we substitute substructure y with substructure z?
  - e.g., dimethylmethylene group $C(CH_3)_2$ in Bisphenol A (BPA) can be replaced with a sulfone group SO_2 in Bisphenol S (BPS).
- What molecules have the same shape if you substitute one functional group with another?

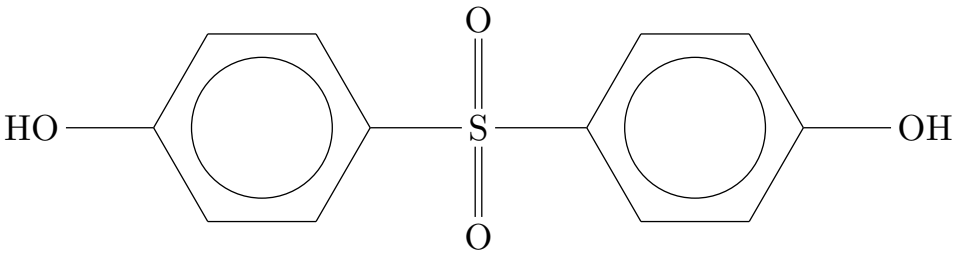


Figure: Bisphenol S (BPS)

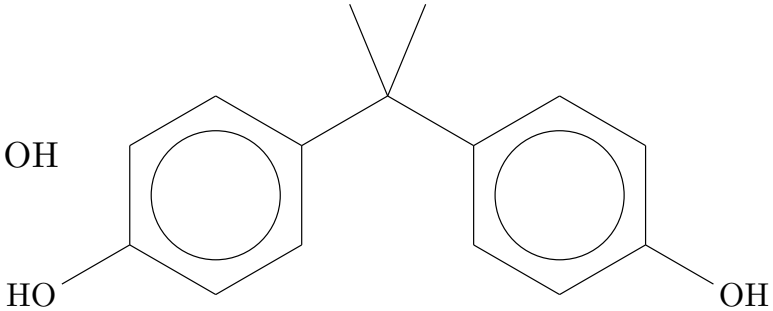


Figure: Bisphenol A (BPA)

Competency Questions

Synthesis

- What molecules contain the combination of elements/atoms x, y, z?
- What molecules contain functional groups x and y?

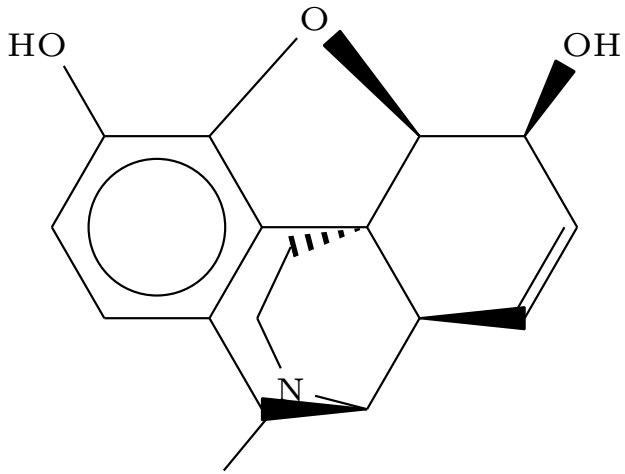


Figure: Morphine ($C_{17}H_{19}NO_3$)

Ontological Commitments as Requirements

We can extract requirements that drive the design of the ontology to ensure that we capture all of the information required to represent chemical structure:

R-1 The ontology must represent the properties of elements, functional groups, connections between functional groups and components of molecules, along with a classification of molecules with respect to their structures.

R-2 The ontology must represent molecules as graphs, such that molecules can be decomposed into their primitive functional groups.

Existing Ontological Approaches

- Existing ontological approaches found in Magka et al. (2014) are synonymous with molecular graph theory.
- Cohn (2001) is an extension of molecular graph theory with a mereology.
- Both ontologies satisfy Requirements R-1 and R-2.

Ontological Commitments as Requirements (cont.)

Atoms, Bonds, Functional Groups

R-3 Atoms, bonds, and functional groups are *primitives* in the ontology related by a notion of *composition*.

R-4 Primitive functional groups are *rings* and *chains*, which correspond to induced cyclic and path subgraphs, respectively.

R-5 The carbon backbone of organic molecules forms the essential structure of a given compound, which we call a '*skeleton*'. The skeleton consists of the various combinations of rings, chains, and atoms that are not in any functional groups.

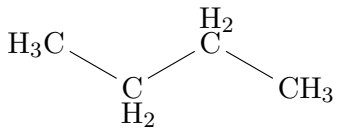


Figure: Butane

Figure: Butane (Carbon Backbone)

Ontological Commitments as Requirements (cont.)

Connections Between Functional Groups

Functional groups can be categorized according to how groups can share atoms or are bonded together; we categorize the various types of connections as requirements for the ontology:

R-6 The ontology must represent the *fusion* connection: there are two overlapping atoms that share a bond between the groups.

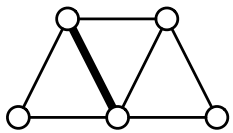


Figure: Fusion (Rings)

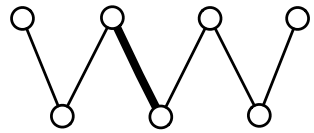


Figure: Fusion (Chain)

Ontological Commitments as Requirements (cont.)

Connections Between Functional Groups

R-7 The ontology must represent the *spiro* connection: rings or chains share an atom: they overlap at one atom.

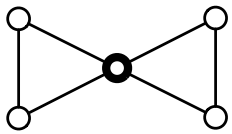


Figure: Spiro (Rings)

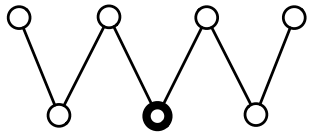


Figure: Spiro (Chain)

Ontological Commitments as Requirements (cont.)

Connections Between Functional Groups

R-8 The ontology must represent the *tether* connection: no atoms between the groups are shared, but any two groups are bonded together by a bond between each group.

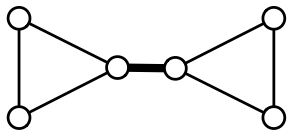


Figure: Tether (Rings)

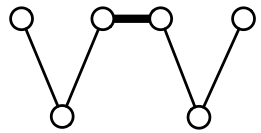


Figure: Tether (Chain)

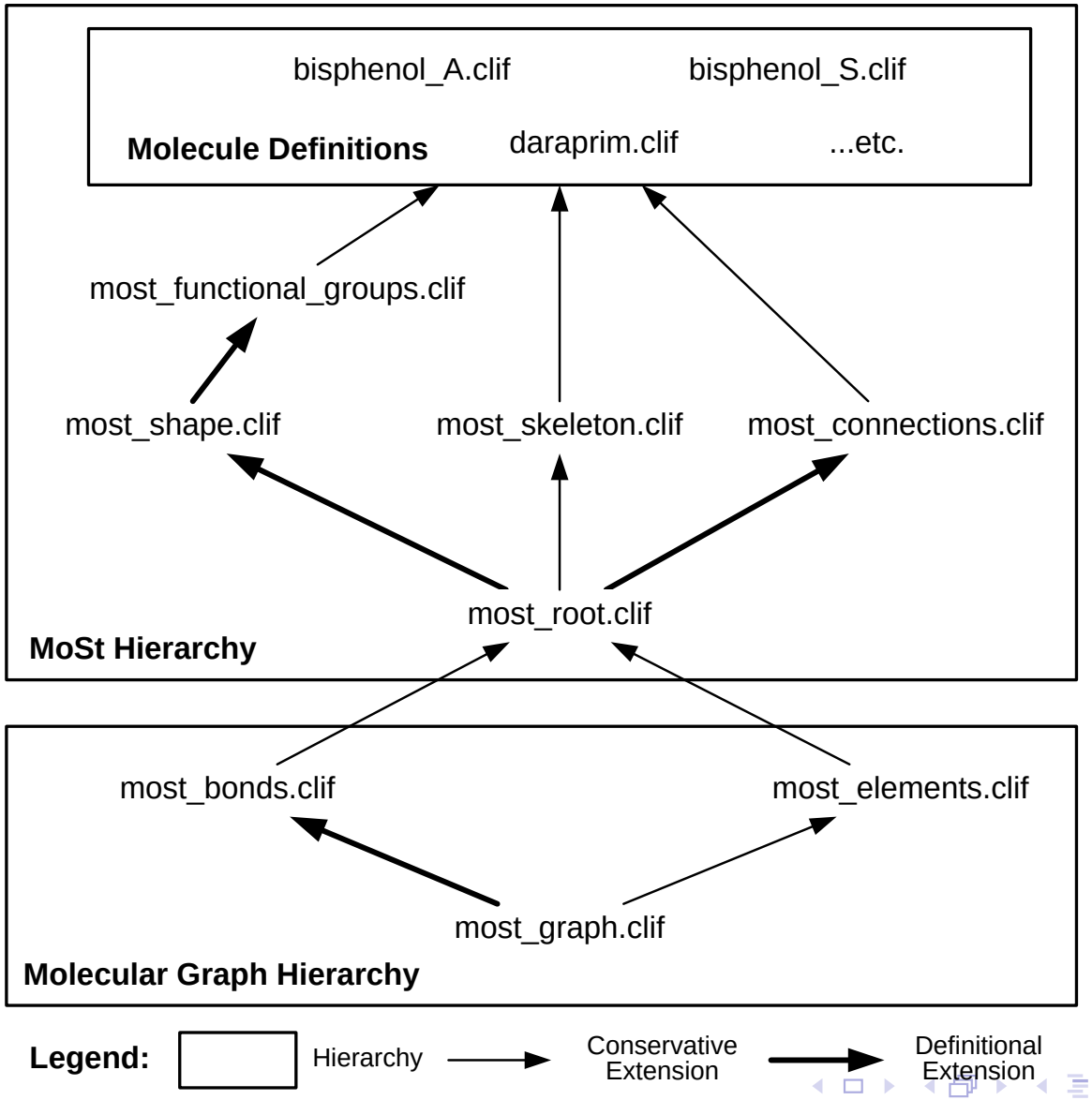
MOlecular Structure Ontology (MoSt)

We propose an ontology that combines both the conventional and ontological approaches to describing molecules, and adheres to the requirements.

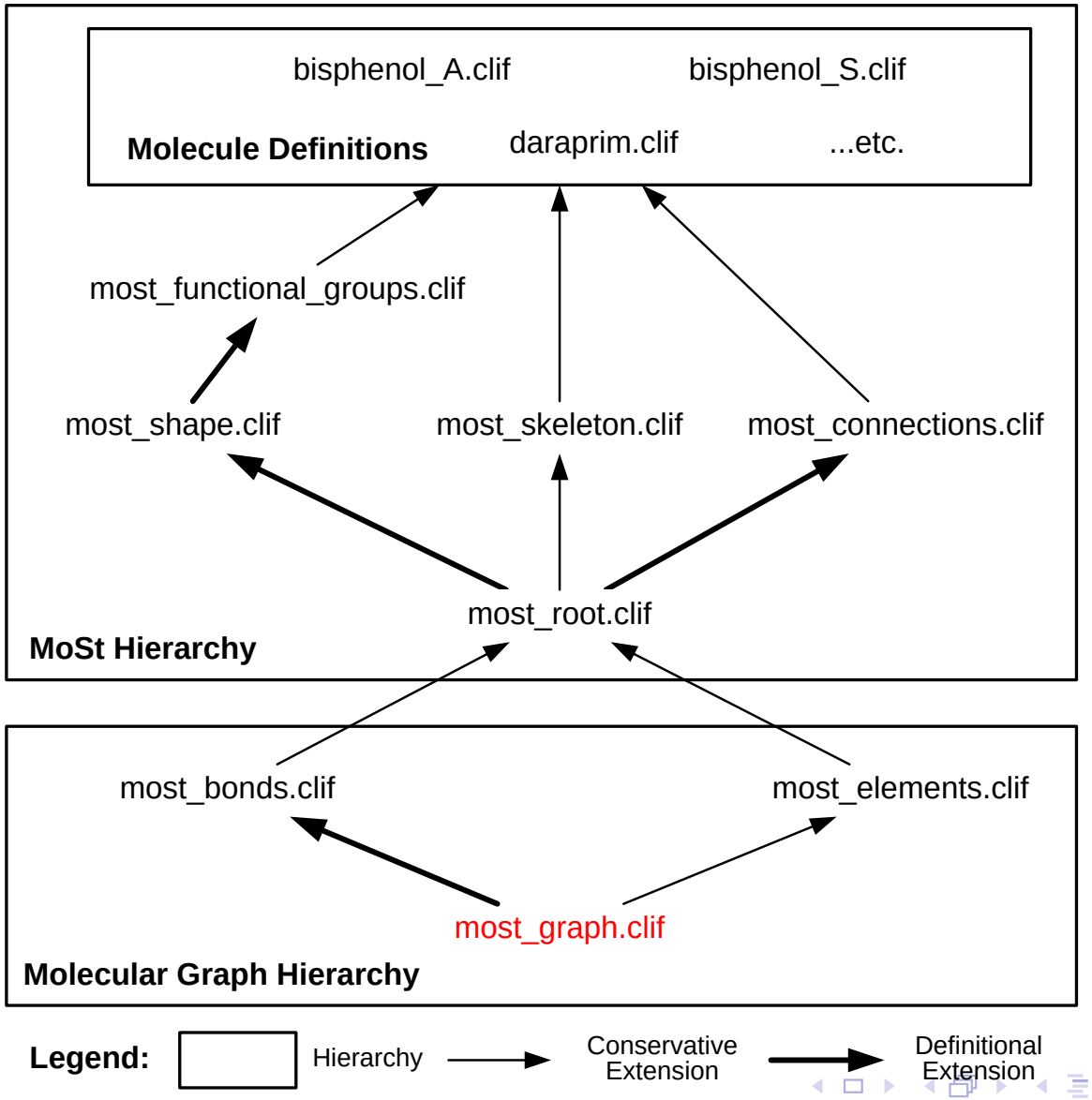
With this ontology, we can represent:

- Rings
- Chains
- Functional groups
- Classes of bonds
- Classes of elements
- Connections between functional groups

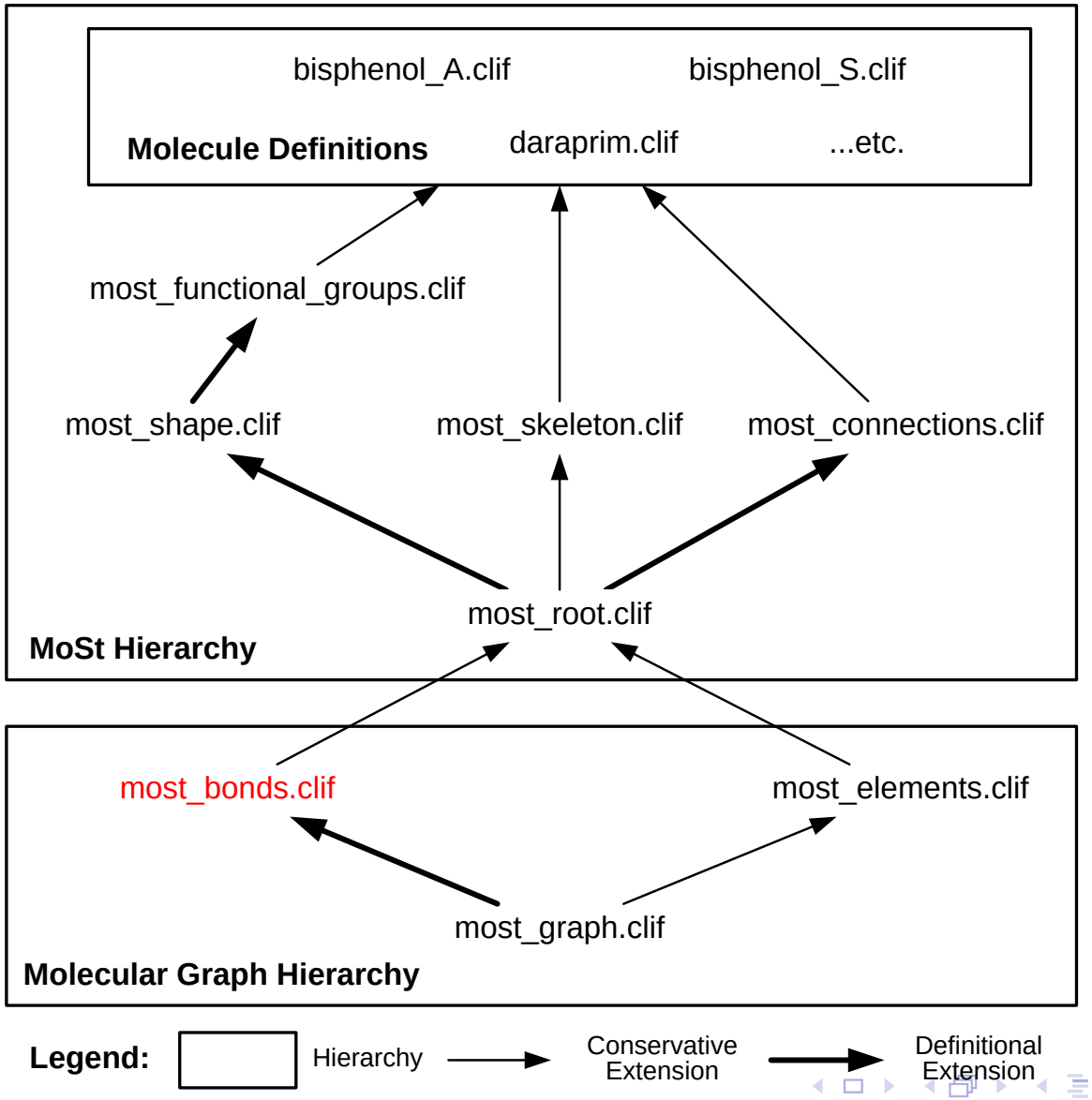
Organization of Theories



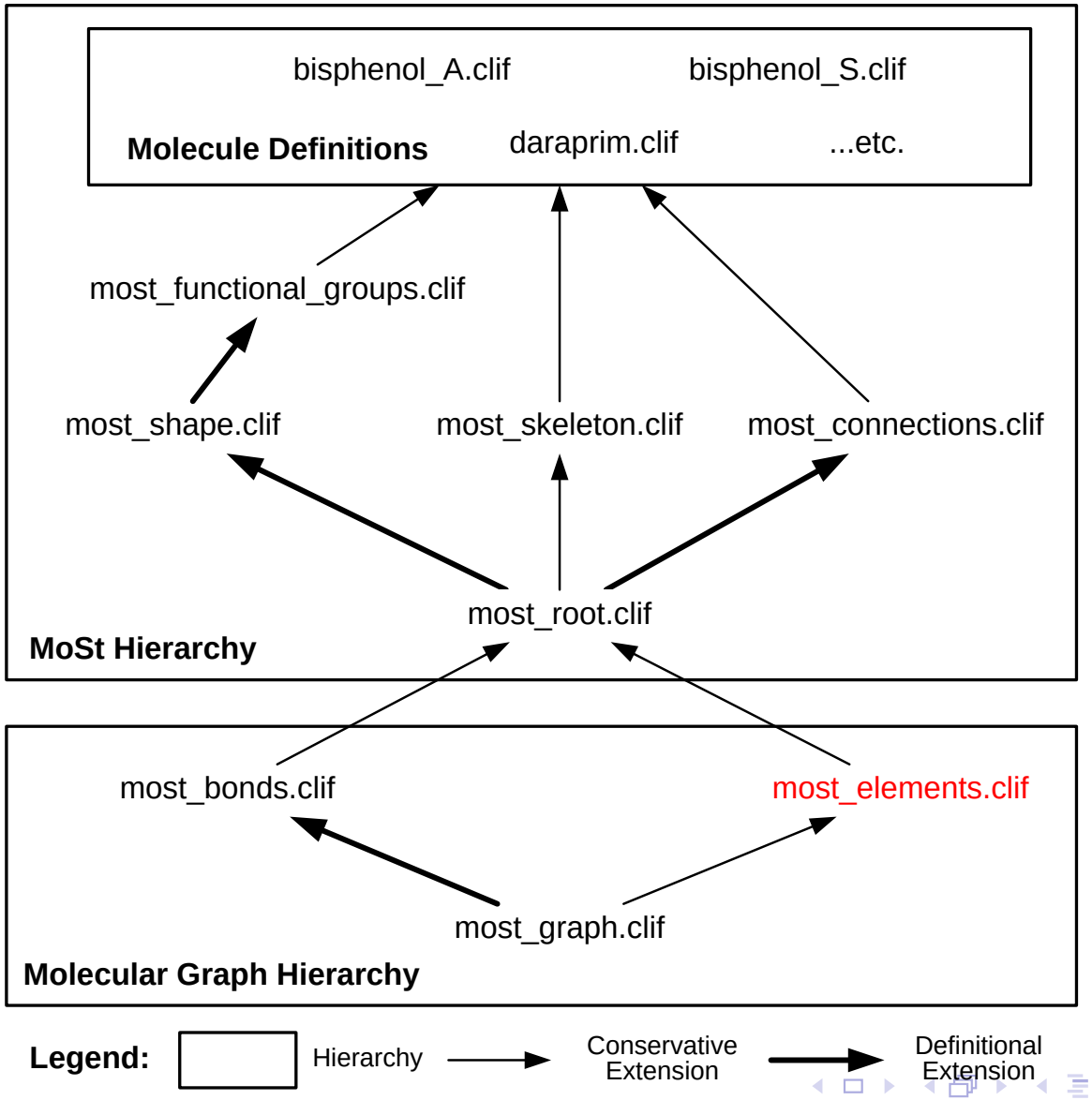
Organization of Theories



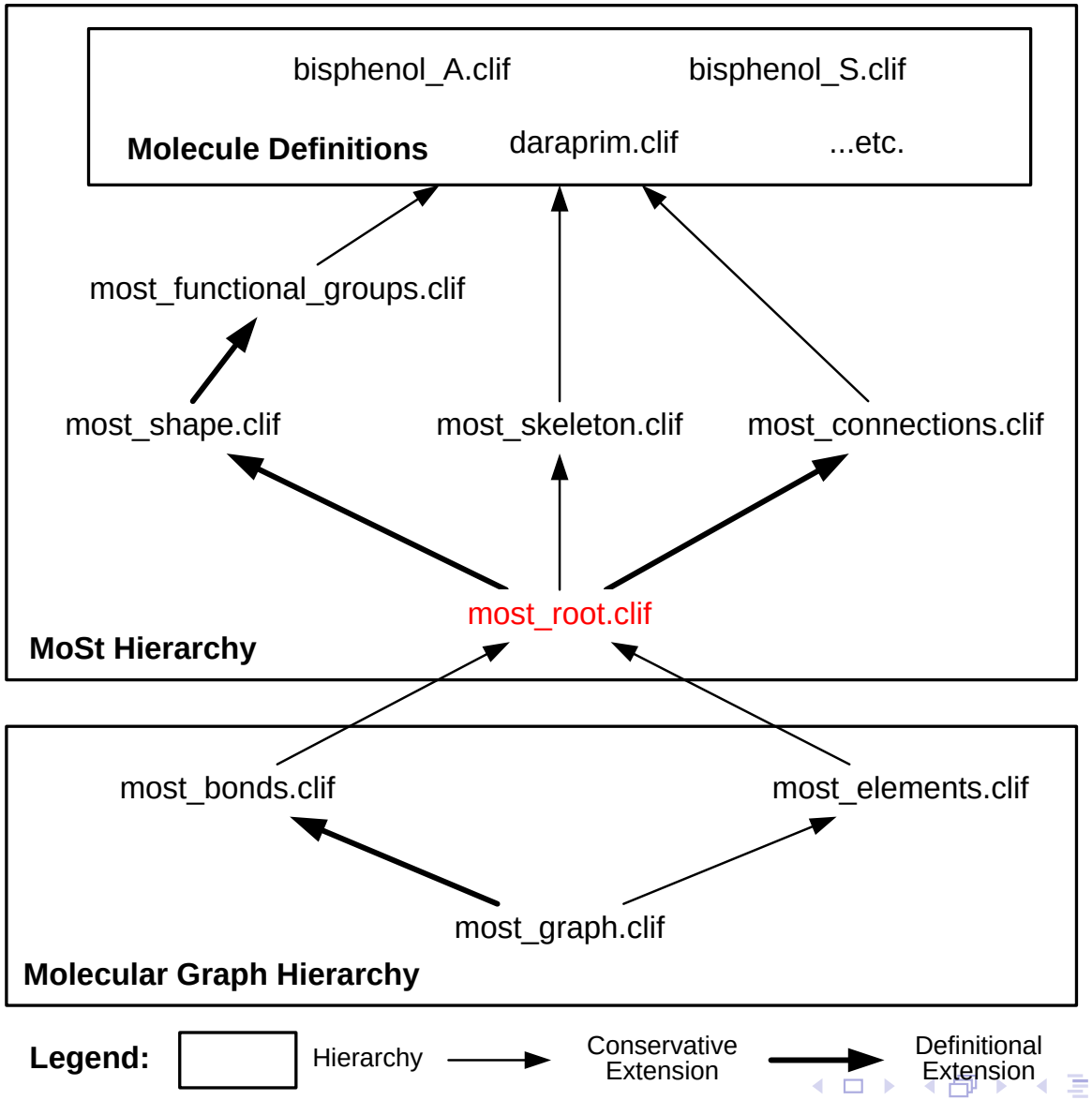
Organization of Theories



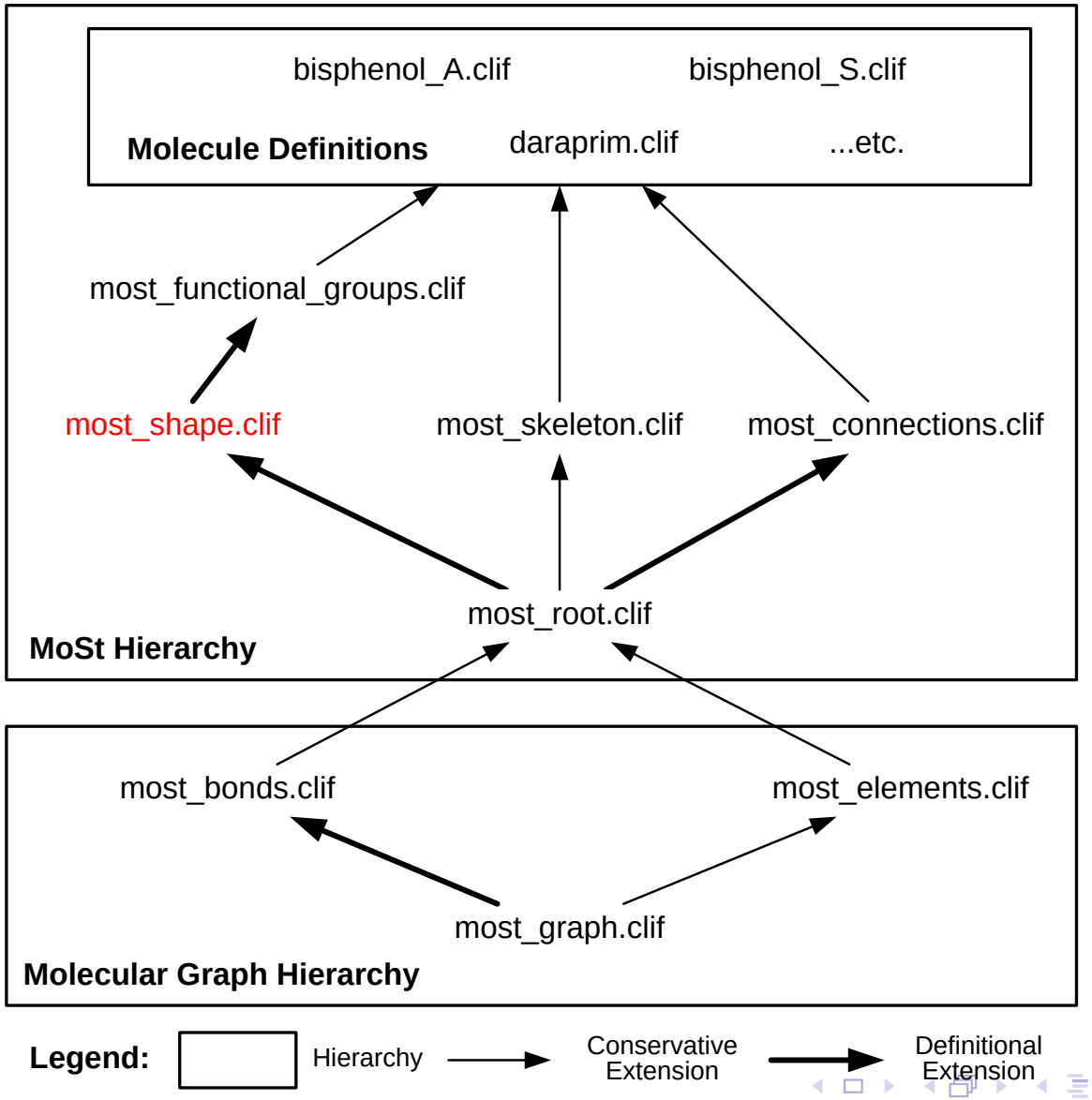
Organization of Theories



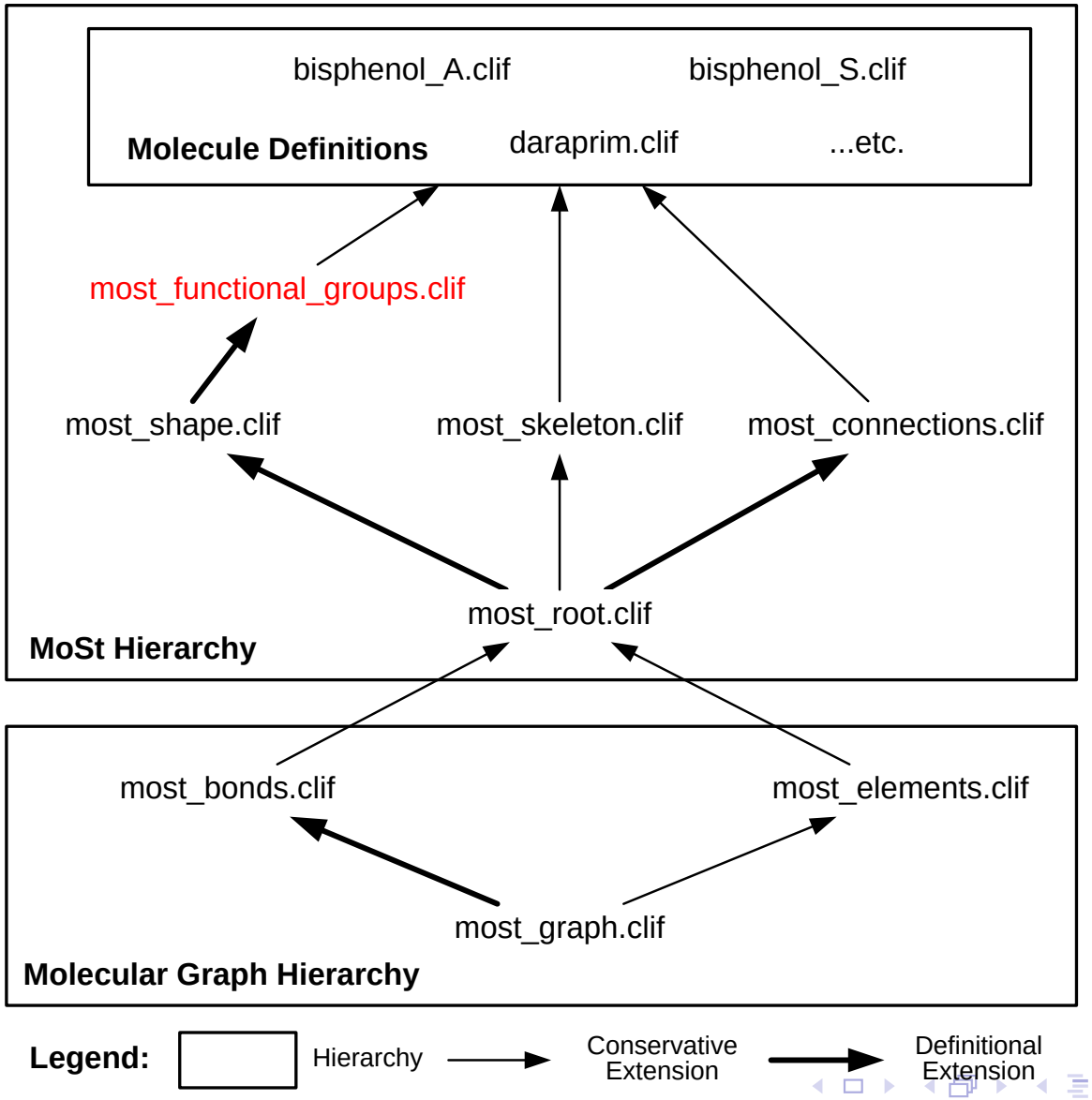
Organization of Theories



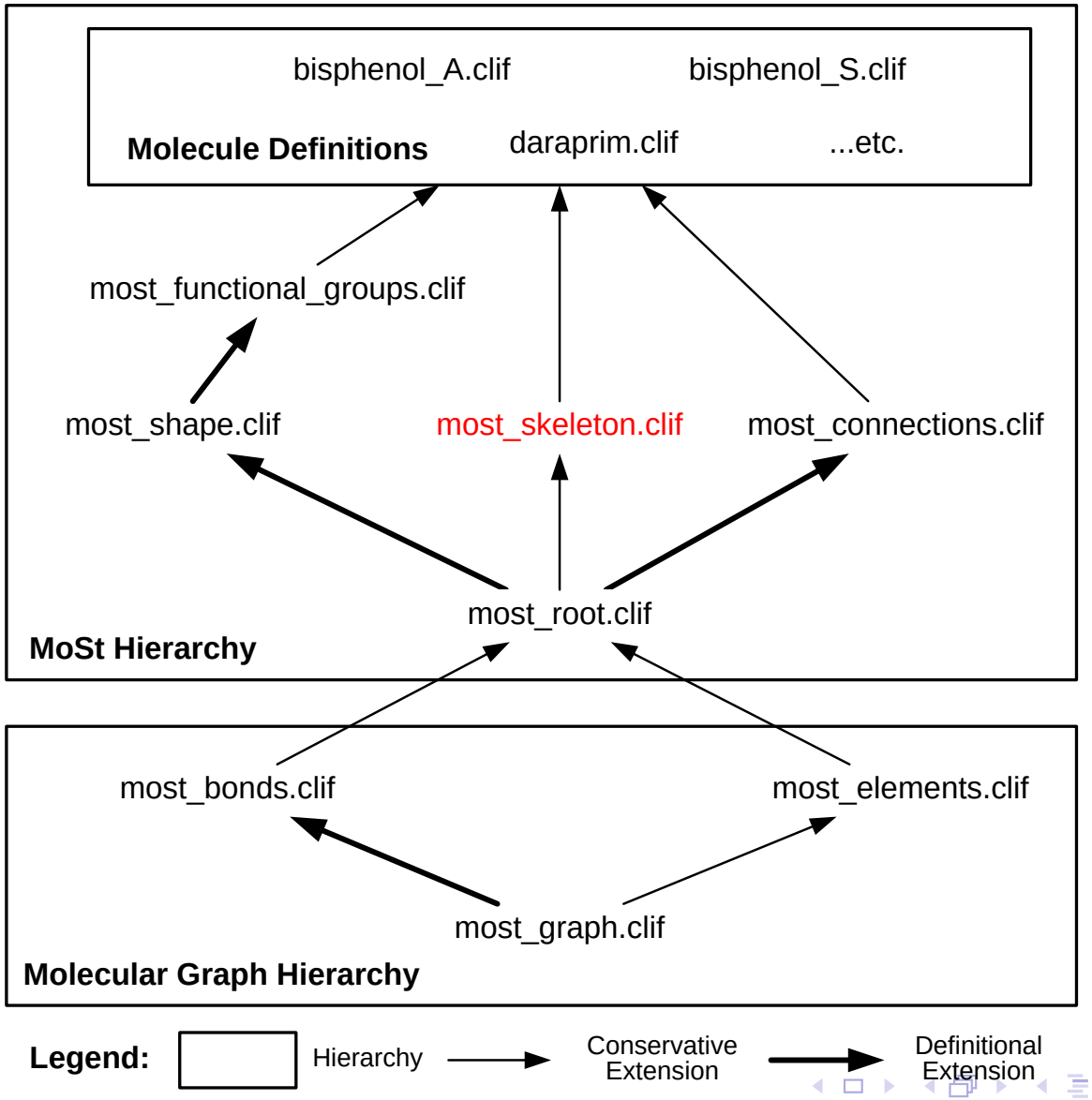
Organization of Theories



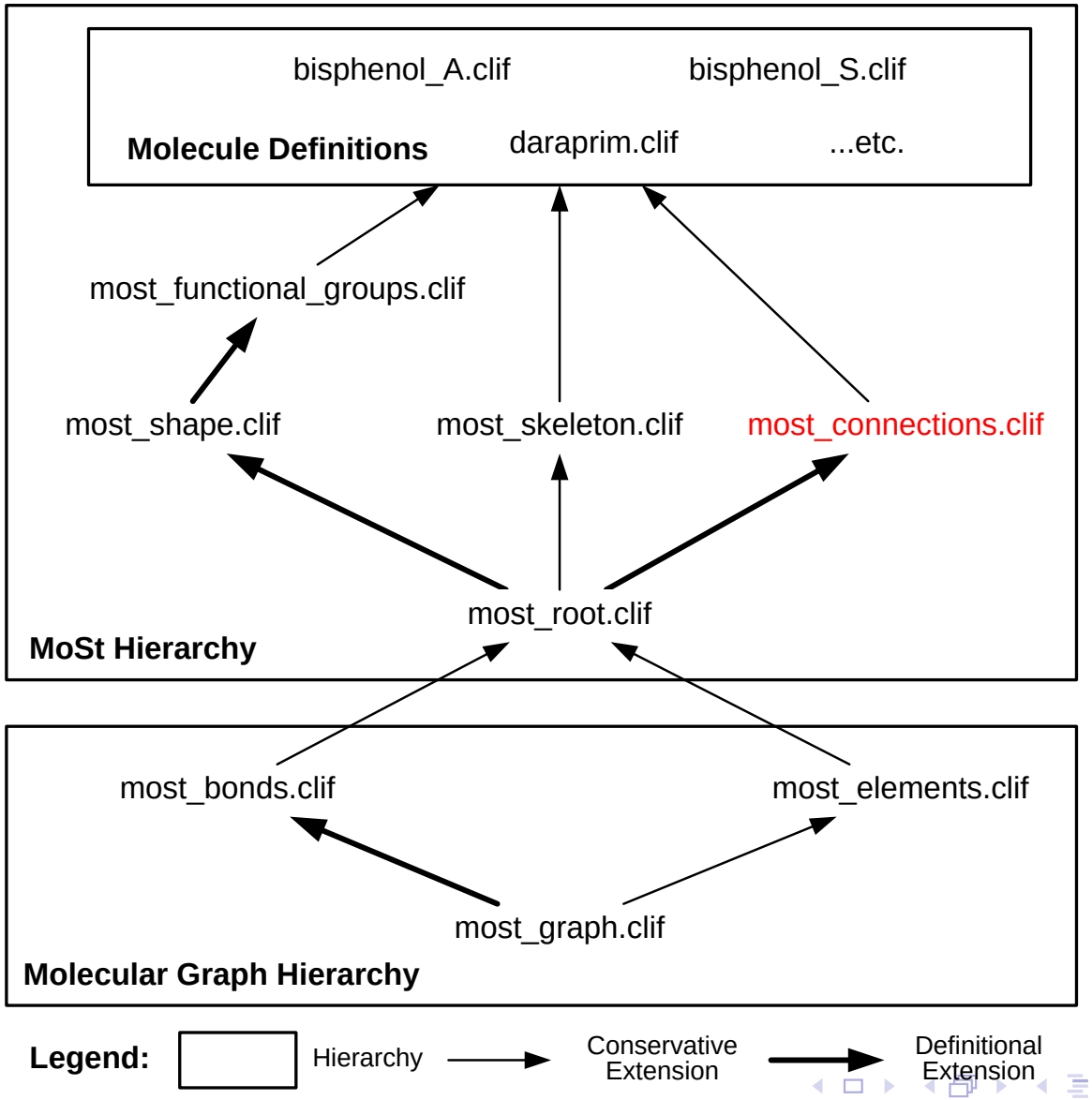
Organization of Theories



Organization of Theories

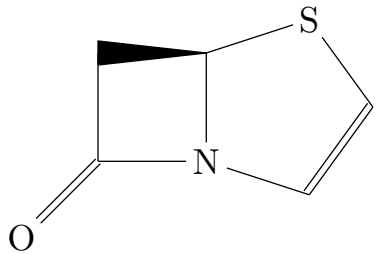


Organization of Theories



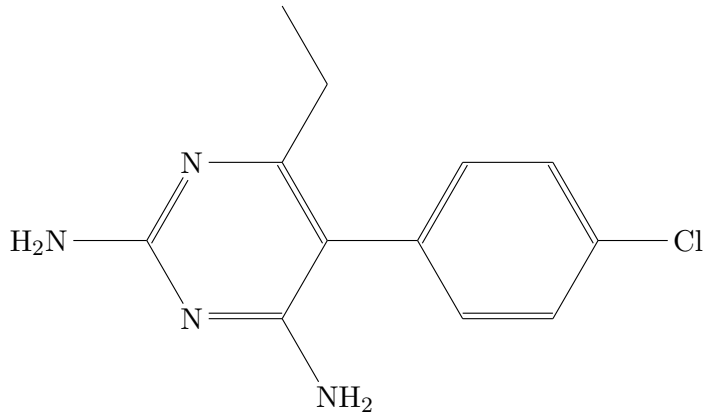
Example: Penem (C_5H_5NOS)

- IUPAC name: (5R)-4-thia-1-azabicyclo[3.2.0]hept-2-en-7-one
- Core structure found in penicillins.



$$\forall x\ penem(x) \equiv \exists g_1 \exists g_2\, skeleton(x) \wedge$$
$$thiazolidine(g_1) \wedge beta\_lactam(g_2) \wedge$$
$$inSkeleton(g_1, x) \wedge inSkeleton(g_2, x) \wedge$$
$$fused(g_1, g_2)$$

Example: Pyrimethamine/Daraprim ($C_{12}H_{13}ClN_4$)



$$\forall x\ pyrimethamine(x) \equiv molecule(x) \wedge$$
$$\exists w \exists y \exists a_1 \exists a_2 \exists u \exists z\ chlorophenyl(w) \wedge$$
$$ethyl\_pyrimidine(y) \wedge amine(a_1) \wedge amine(a_2) \wedge$$
$$containsSkeleton(x, w) \wedge containsSkeleton(x, y) \wedge$$
$$containsGroup(x, a_1) \wedge containsGroup(x, a_2) \wedge$$
$$inSkeleton(u, w) \wedge inSkeleton(z, y) \wedge$$
$$tether(u, z) \wedge tether(y, a_1) \wedge tether(y, a_2)$$

Drug Synthesis

- Drug synthesis corresponds to the process of constructing models of the ontology.
- Models are constructed by composing functional groups together via the three connection relationships (fusion, spiro, tether).

Summary

- We proposed a set of competency questions and requirements for an ontology of molecular structure to support drug synthesis.
- We have an axiomatization and verification for such an ontology.

Future Work

1. MOlecular Structure Ontology (MoSt)
   - Implementation of queries developed from competency questions.
2. Molecular Reaction Ontology (MoRe)
   - Axiomatization of a process ontology for chemical reactions that change the *shape* of molecules.
   - How does the structure change when a molecule undergoes a process?
     - For example: 1,4-dioxaspiro[4.5]decane[1] ($C_8H_{14}O_2$) is formed from cyclohexanone ($C_6H_{10}O$) and the ethanediol ($C_2H_6O_2$).