Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

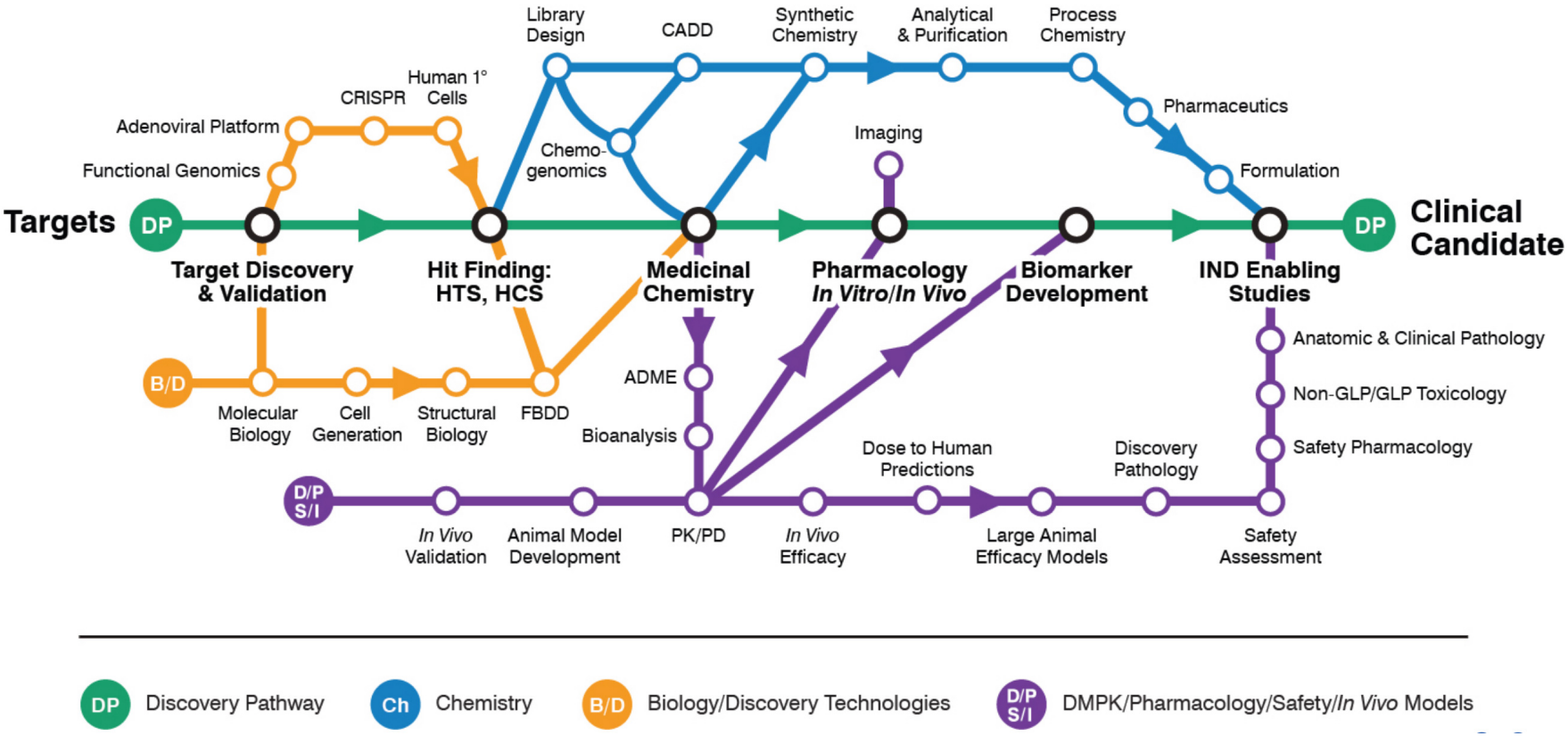
An Ontology for Medicinal Chemistry

Ph.D. Defence

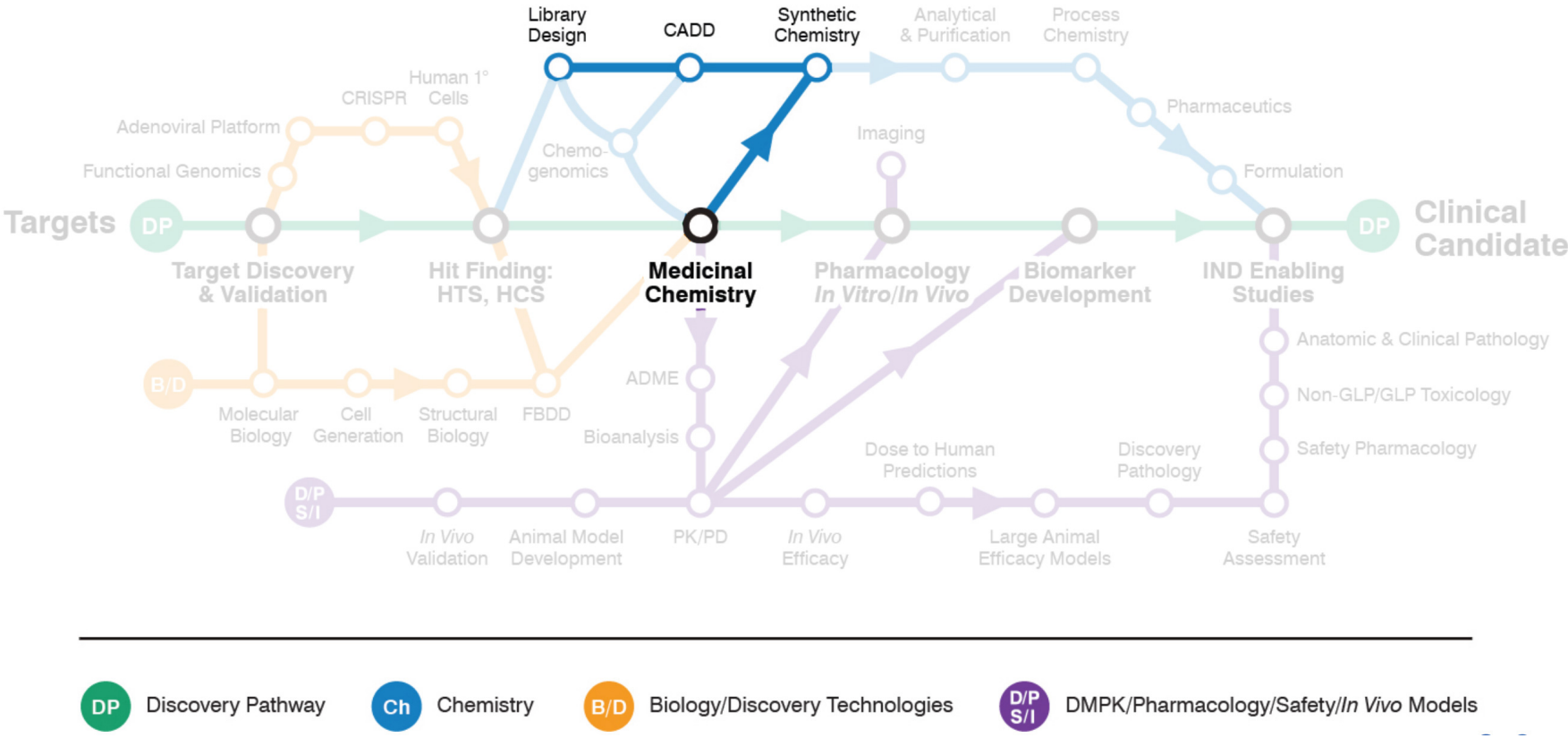
Carmen S. Chui

December 14, 2018

Drug Design & Discovery Roadmap from [Cha17]



Drug Design & Discovery Roadmap from [Cha17]



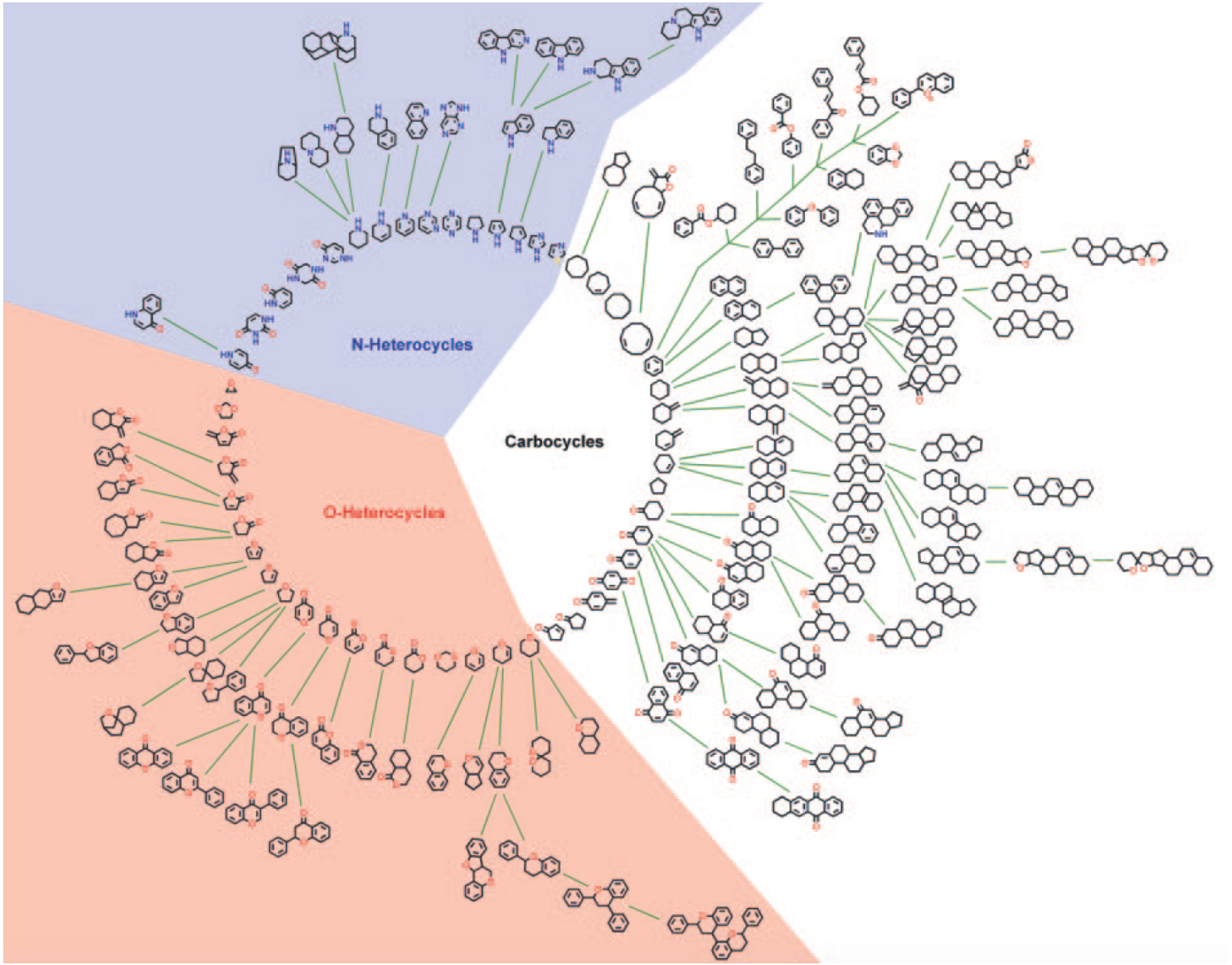
Research Objective

Objective

Existing work in cheminformatics discusses the notion of 'chemical space' to describe all possible organic molecules to be considered when searching for new drugs [RA12].

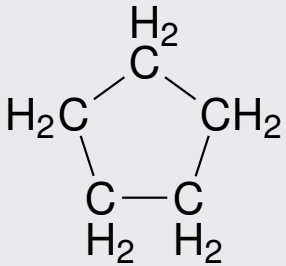
We want to provide **ontological foundations for chemical space**, where the central idea is that chemical space is characterized by the *shape* and *structure* of molecules.

Chemical Space: Scaffold Tree in [Koc+05]

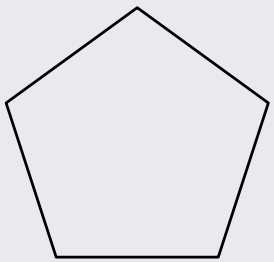


Motivations: What do we mean by shape?

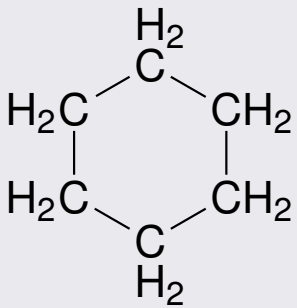
2D Shapes & Polygons



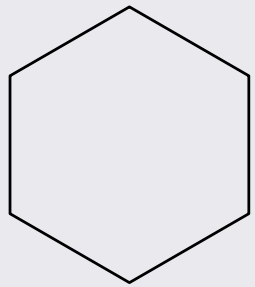
(a) Cyclopentane



(b) Pentagon



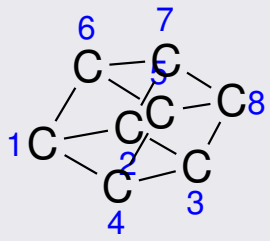
(c) Cyclohexane



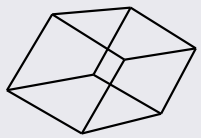
(d) Hexagon

Motivations: What do we mean by shape?

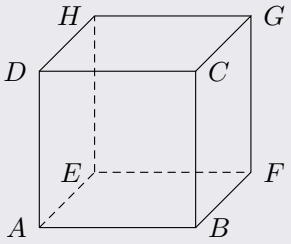
3D Shapes (Polyhedra) & Molecules



(a) Cubane

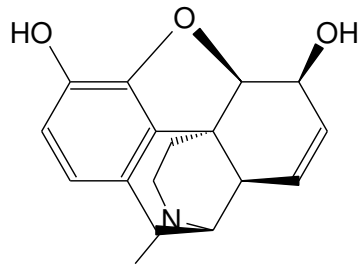


(b) Skeletal formula.



(c) A cube

Current Approaches to Represent Molecular Shape



Chemical name [Nat15]:

Morphine; Morphinum; Morphia; Morphium; Morphin

IUPAC name:

(4R,4aR,7S,7aR,12bS)-3-methyl-2,4,4a,7,7a,13-hexahydro-1H-4,12-methanobenzofuro[3,2-e]isoquinoline-7,9-diol

SMILES:

CN1CC[C@]23C4=C5C=CC(O)=C4O[C@H]2[C@@H](O)C=C[C@H]3[C@H]1C5

InChI identifier:

InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16-,17-/m0/s1

InChI key: BQJCRHHNABKAKU-KBQPJGBKSA-N

Overview of Contributions

To navigate and characterize chemical space, the following contributions have been made:

- C-1 Requirements for the Ontology & Its Models
- C-2 Axiomatization & Verification of MoSt
- C-3 Techniques for Decomposing & Re-Composing Molecules
- C-4 Drug Design as Model Construction
- C-5 Model-Theoretic Search Techniques

C-1 Requirements for the Ontology & Its Models

Competency questions guided the overall design of the MOlecular Structure ontology (MoSt)

Requirements & Semantic Conditions for Representing Shape

- Molecules must be represented as graphs
- Components of molecules must be elements of the domain
- Attachments between functional groups (spiro, tether, fusion) must also be represented

Requirements for the Models of the Ontology

- 1-to-1 correspondence of models of MoSt with molecules
- Intended models of the ontology are molecules
- Unintended models of the ontology are not molecules

C-1 Requirements for the Ontology & Its Models

Competency questions guided the overall design of the MOlecular Structure ontology (MoSt)

Requirements & Semantic Conditions for Representing Shape

- Molecules must be represented as graphs
- Components of molecules must be elements of the domain
- Attachments between functional groups (spiro, tether, fusion) must also be represented

Requirements for the Models of the Ontology

- 1-to-1 correspondence of models of MoSt with molecules
- Intended models of the ontology are molecules
- Unintended models of the ontology are not molecules

C-1 Requirements for the Ontology & Its Models

Competency questions guided the overall design of the MOlecular Structure ontology (MoSt)

Requirements & Semantic Conditions for Representing Shape

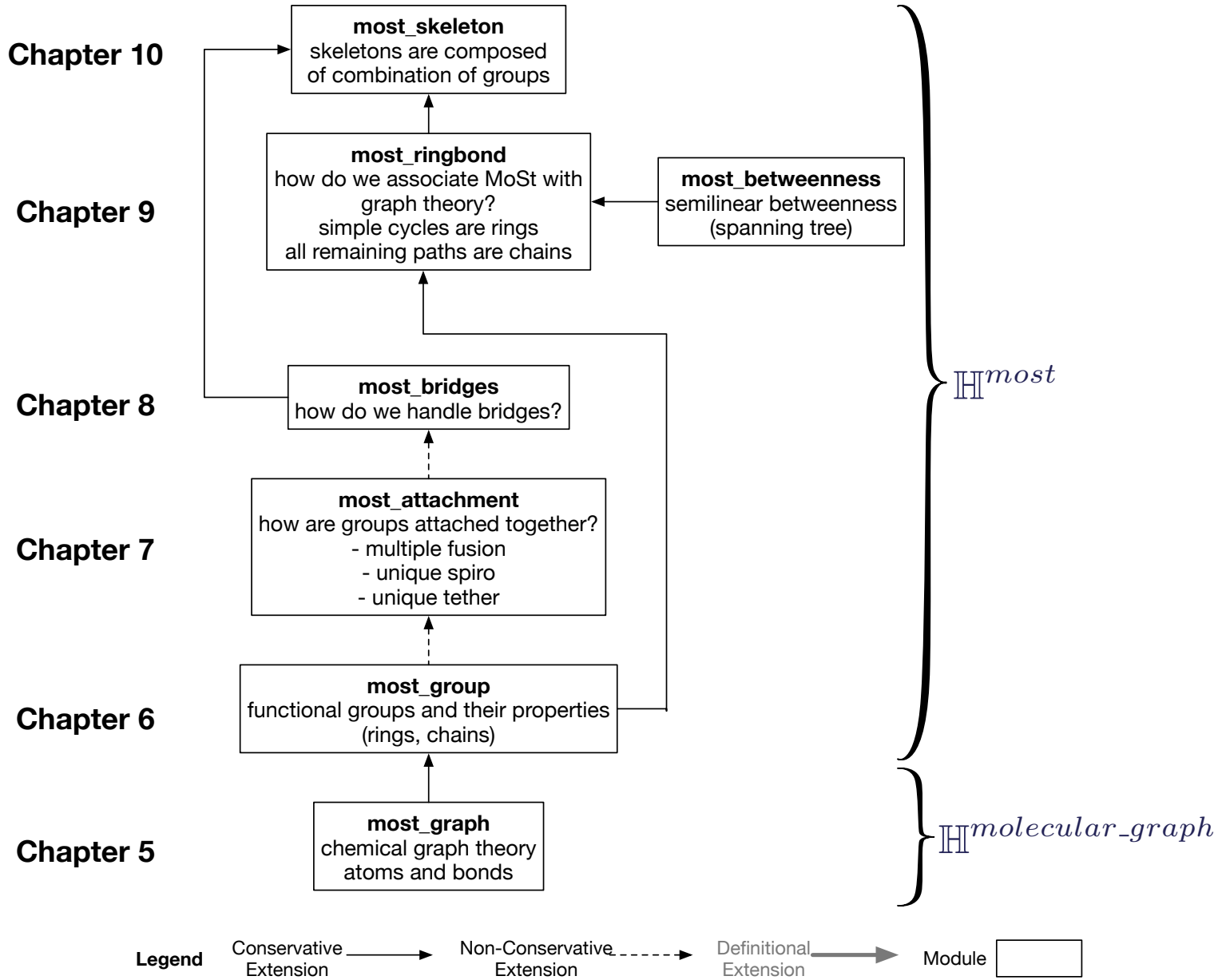
- Molecules must be represented as graphs
- Components of molecules must be elements of the domain
- Attachments between functional groups (spiro, tether, fusion) must also be represented

Requirements for the Models of the Ontology

- 1-to-1 correspondence of models of MoSt with molecules
- Intended models of the ontology are molecules
- Unintended models of the ontology are not molecules

C-2 Axiomatization & Verification of MoSt

We present a *first-order* axiomatization of MoSt, organized as such:



C-2 Axiomatization & Verification of MoSt (cont.)

- Verification results show that models of MoSt are **synonymous** with tripartite incidence structures found in COLORE
  - This gives us a complete classification of all the models of MoSt
- Because of this synonymy, we can take advantage of techniques for the construction and decomposition of models of the ontology
- We *inherit* techniques for building models from the mathematical incidence theories

C-3 Techniques for Decomposing & Re-Composing Molecules

We present:

- Techniques for decomposing molecules into their primitive functional groups with a **Decomposition Theorem**
- Techniques for (re)composing molecules from primitive functional groups with an **Attachment Theorem**
- Procedures for decomposing the underlying molecular graph into its building blocks

C-3 Techniques for Decomposing & Re-Composing Molecules

We present:

- Techniques for decomposing molecules into their primitive functional groups with a **Decomposition Theorem**
- Techniques for (re)composing molecules from primitive functional groups with an **Attachment Theorem**
- Procedures for decomposing the underlying molecular graph into its building blocks

C-3 Techniques for Decomposing & Re-Composing Molecules

We present:

- Techniques for decomposing molecules into their primitive functional groups with a **Decomposition Theorem**
- Techniques for (re)composing molecules from primitive functional groups with an **Attachment Theorem**
- Procedures for decomposing the underlying molecular graph into its building blocks

C-4 Drug Design as Model Construction

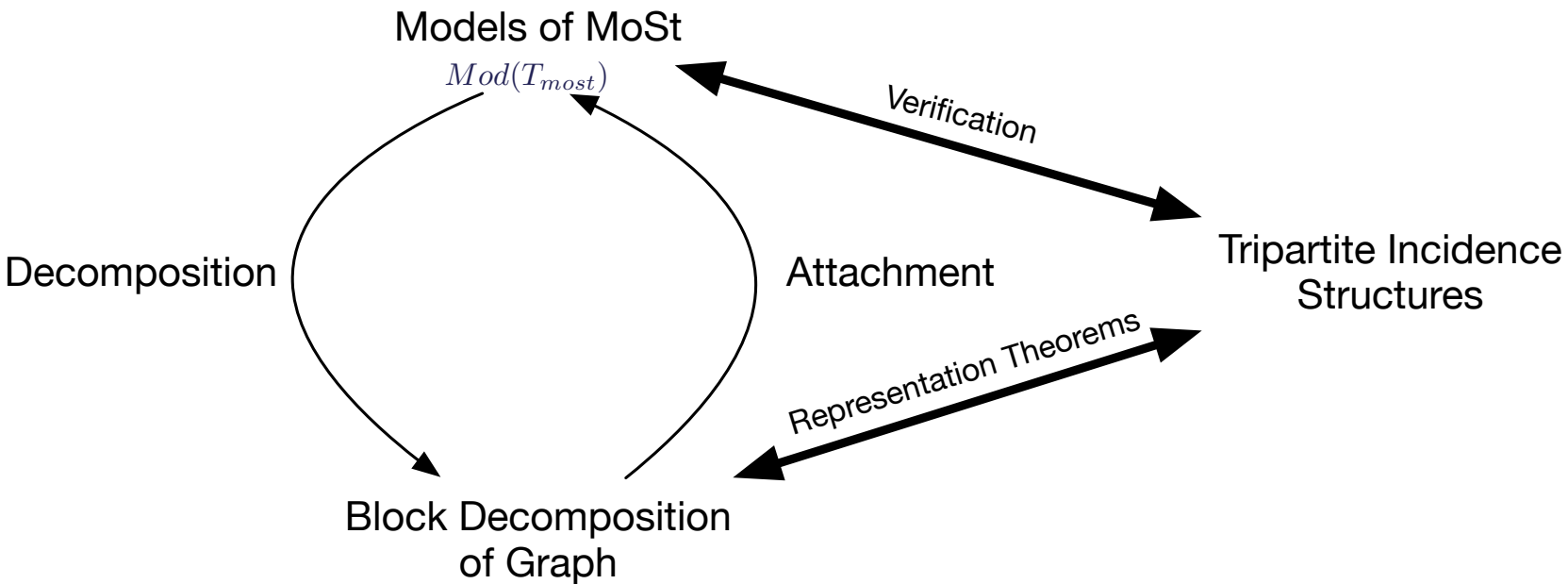
Think of drug design as a **process of building first-order models**:

- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence structures used in the verification

C-4 Drug Design as Model Construction

Think of drug design as a **process of building first-order models**:

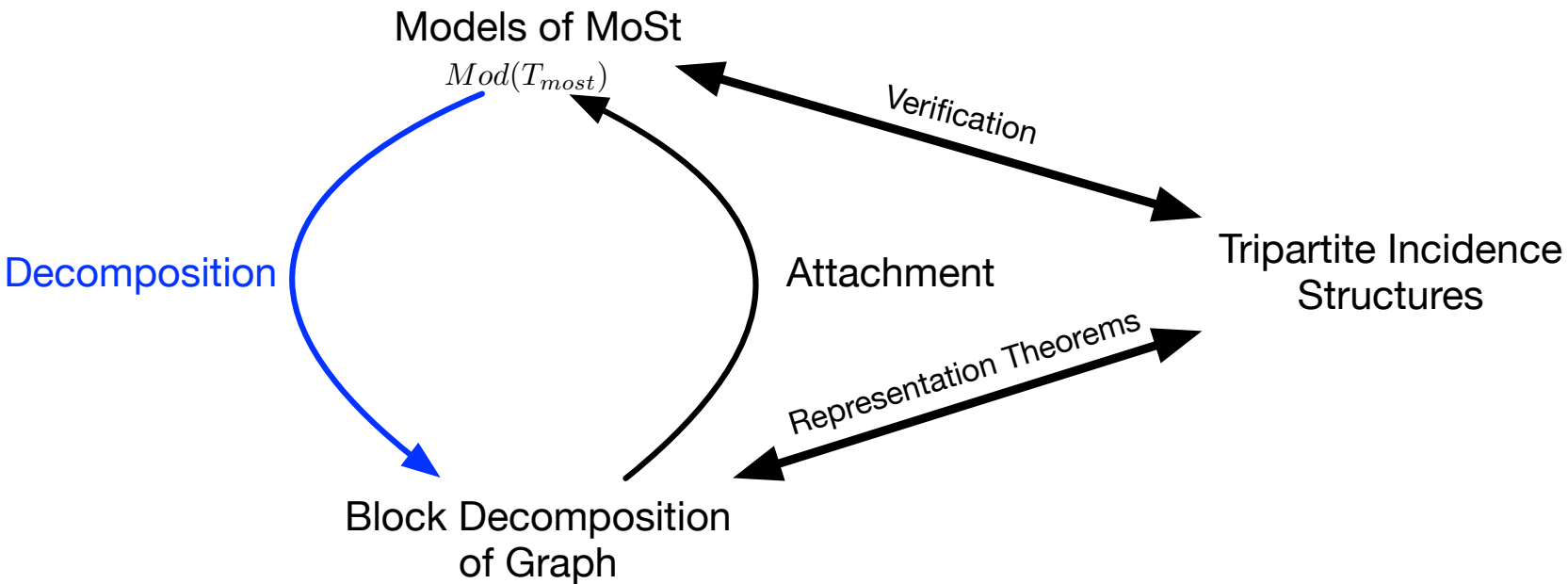
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence structures used in the verification



C-4 Drug Design as Model Construction

Think of drug design as a **process of building first-order models**:

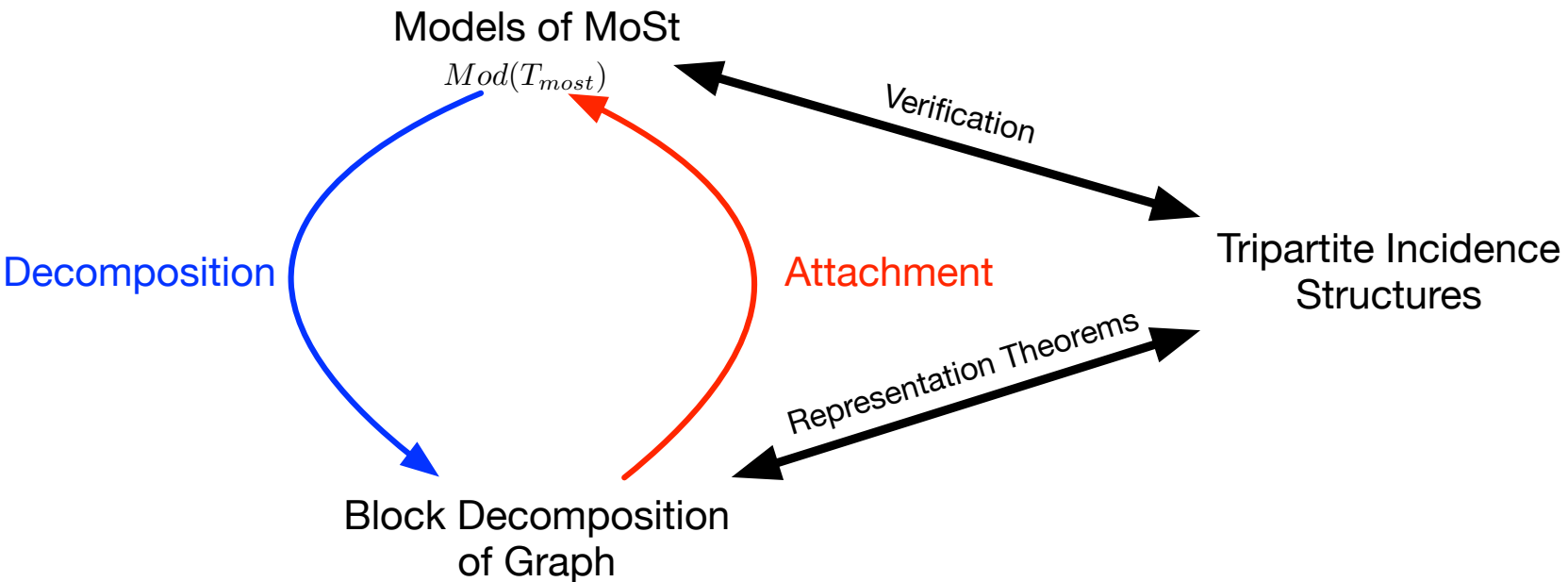
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence structures used in the verification



C-4 Drug Design as Model Construction

Think of drug design as a **process of building first-order models**:

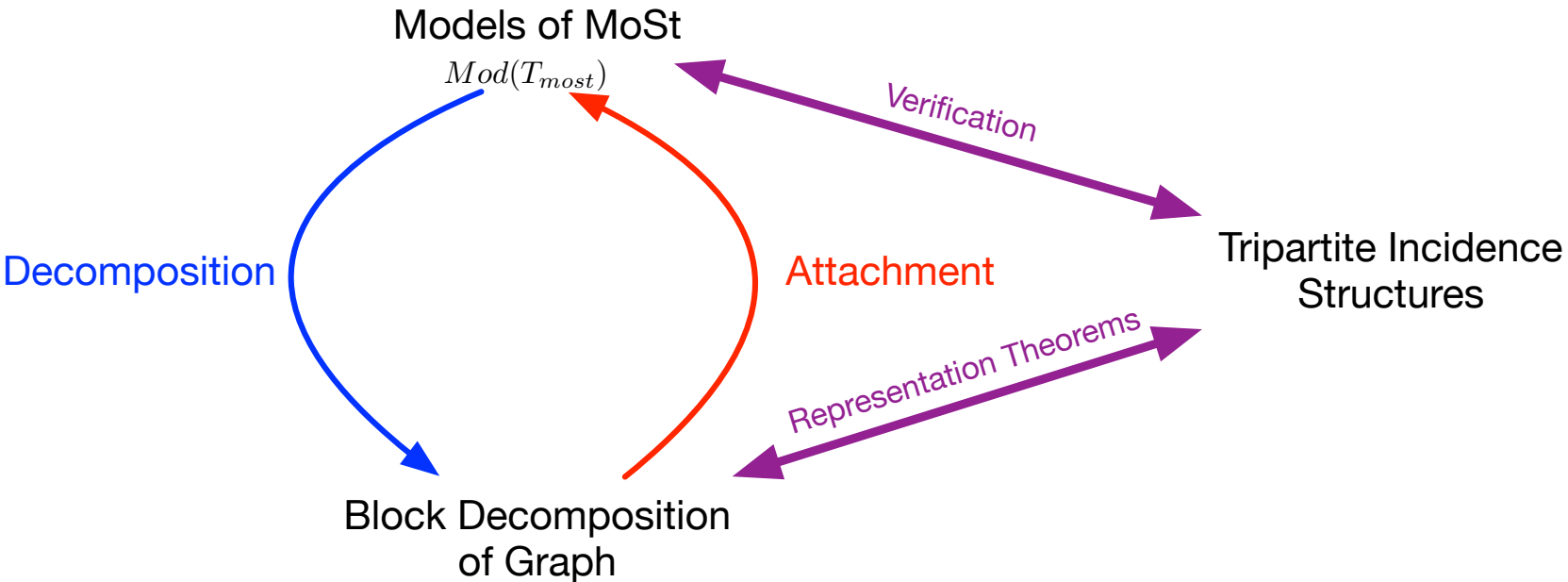
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence structures used in the verification



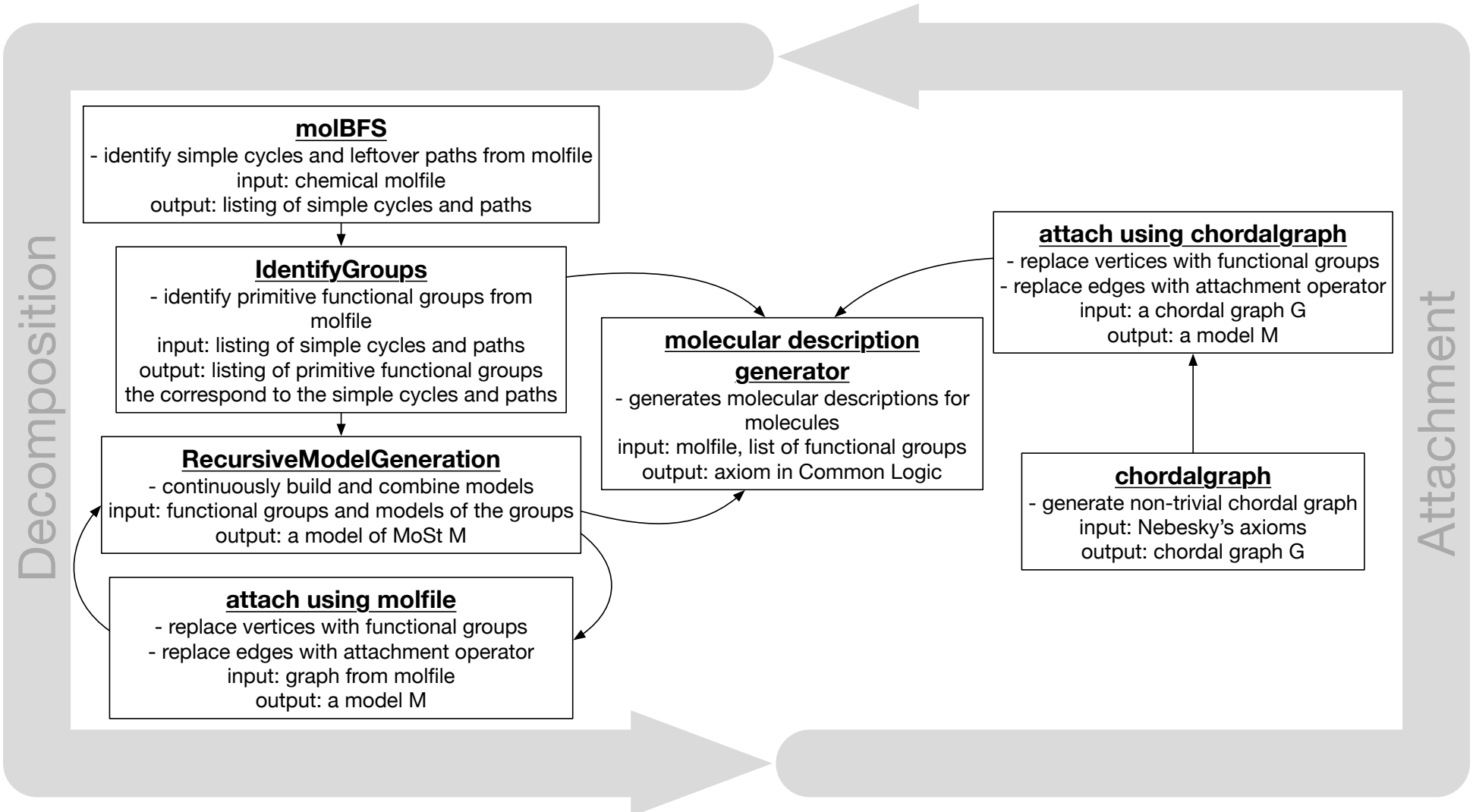
C-4 Drug Design as Model Construction

Think of drug design as a **process of building first-order models**:

- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence structures used in the verification



C-4 Drug Discovery as Model Construction (cont.)



Model-Theoretic Search Techniques

Navigating Chemical Space $\simeq$ Constraint Satisfaction Programming

Queries against the knowledge base using MoSt helps us navigate the search space:

$$\text{molecular description of } \mathcal{M} \text{ is } Th(\mathcal{M})$$

$$\text{a sentence } \in \Sigma(MoSt)$$

Example: Existential Queries

- Find a molecule that contains at least one ring.

$$KB \models \exists x \; ring(x)$$

- Find a molecule that contains at least one fork and at least one ring.

$$KB \models \exists x \exists y \; (x \neq y) \wedge fork(x) \wedge ring(y)$$

- Find a molecule that contains a ring that is fused to another ring.

$$KB \models \exists x \exists y \; (x \neq y) \wedge ring(x) \wedge ring(y) \wedge fused(x, y)$$

Example: Universal Queries

- All groups in the molecule are rings.

$$KB \models \forall x\, group(x) \supset ring(x)$$

- All groups in the molecule are fused.

$$KB \models \forall x \forall y\, group(x) \wedge group(y) \wedge (x \neq y) \wedge fused(x, y)$$

- All groups in the molecule are tethered.

$$KB \models \forall x \forall y\, group(x) \wedge group(y) \wedge (x \neq y) \wedge tethered(x, y)$$

Model-Theoretic Search Techniques

Example: Composite Queries

- All groups in the molecule are fused to some other ring.

$$KB \models \forall x\, group(x) \supset ring(x) \wedge \exists y\, ring(y) \wedge (x \neq y) \wedge fused(x, y)$$

- All groups are tethered to at least one other group.

$$KB \models \forall x\, group(x) \wedge \exists y\, group(y) \wedge tether(x, y) \wedge (x \neq y)$$

Summary of Contributions

Did we provide an ontological foundation for navigating chemical space?

- ✓ Requirements for a molecular structure ontology with C-1
- ✓ Design and Verification of MoSt with C-2
- ✓ New Techniques for Designing Molecules via Model Construction with C-3, C-4
- ✓ An Alternative Approach to Navigating Chemical Space with C-5

Summary of Contributions

Did we provide an ontological foundation for navigating chemical space?

- ✓ Requirements for a molecular structure ontology with C-1
- ✓ Design and Verification of MoSt with C-2
- ✓ New Techniques for Designing Molecules via Model Construction with C-3, C-4
- ✓ An Alternative Approach to Navigating Chemical Space with C-5

Summary of Contributions

Did we provide an ontological foundation for navigating chemical space?

- ✓ Requirements for a molecular structure ontology with C-1
- ✓ Design and Verification of MoSt with C-2
- ✓ New Techniques for Designing Molecules via Model Construction with C-3, C-4
- ✓ An Alternative Approach to Navigating Chemical Space with C-5

Summary of Contributions

Did we provide an ontological foundation for navigating chemical space?

- ✓ Requirements for a molecular structure ontology with C-1
- ✓ Design and Verification of MoSt with C-2
- ✓ New Techniques for Designing Molecules via Model Construction with C-3, C-4
- ✓ An Alternative Approach to Navigating Chemical Space with C-5

Summary of Contributions

Did we provide an ontological foundation for navigating chemical space?

- ✓ Requirements for a molecular structure ontology with C-1
- ✓ Design and Verification of MoSt with C-2
- ✓ New Techniques for Designing Molecules via Model Construction with C-3, C-4
- ✓ An Alternative Approach to Navigating Chemical Space with C-5

Open Questions & Future Work

- Decidability of MoSt
- Mereology on Skeletons
- Molecular Reactions Ontology (MoRe): A Process Ontology
- Reasoning About Molecules (RoMe): A Software Environment
- Integration with (Cheminformatics) Software Tools & Query Languages

Thank You!

Any Questions?

References & Additional Links #1

Charles River Laboratories, Inc. *The Benefits of Outsourcing Drug Discovery to an End-to-End CRO*. May 1, 2017. URL: https://www.criver.com/resources/benefits-outsourcing-drug-discovery-end-end-cro.

Marcus A. Koch et al. "Charting biologically relevant chemical space: A structural classification of natural products (SCONP)". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.48 (2005), pp. 17272–17277. DOI: 10.1073/pnas.0503647102. URL: http://www.pnas.org/content/102/48/17272.abstract.

References & Additional Links #2

National Center for Biotechnology Information. *PubChem Compound Database - Morphine (CID=5288826)*. 2015. URL: https://pubchem.ncbi.nlm.nih.gov/compound/5288826.

Jean-Louis Reymond and Mahendra Awale. "Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database". In: *ACS Chemical Neuroscience* 3.9 (2012). PMID: 23019491, pp. 649–657. URL: https://doi.org/10.1021/cn3000422%20https://doi.org/10.1021/cn3000422.

Vladsinger. *Isomerism*. Dec. 1, 2018. URL: https://commons.wikimedia.org/wiki/File:Isomerism.svg.

Why Topological Structure?

- Chemical spaces and the scaffold tree approach do not talk about the full geometry (such as stereochemistry pertaining to isomers, bond angles, etc.) since only the 'core' is examined
- Instead, we wanted the ontology to be geared toward supporting the scaffold tree approach

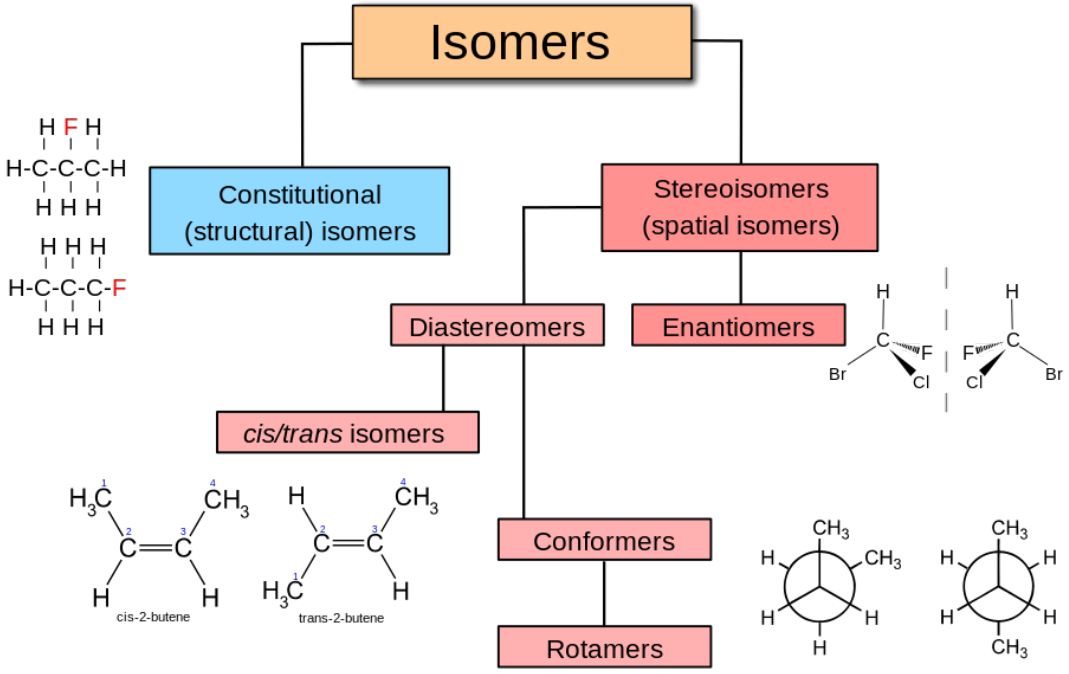
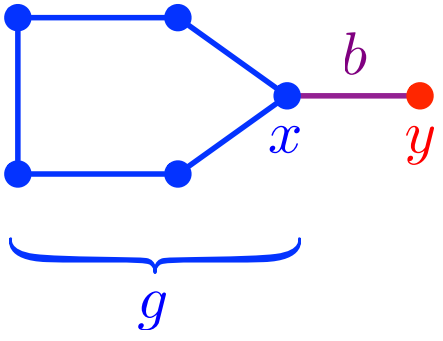


Figure 2: Image from [Vla18]

Why An Ontology?

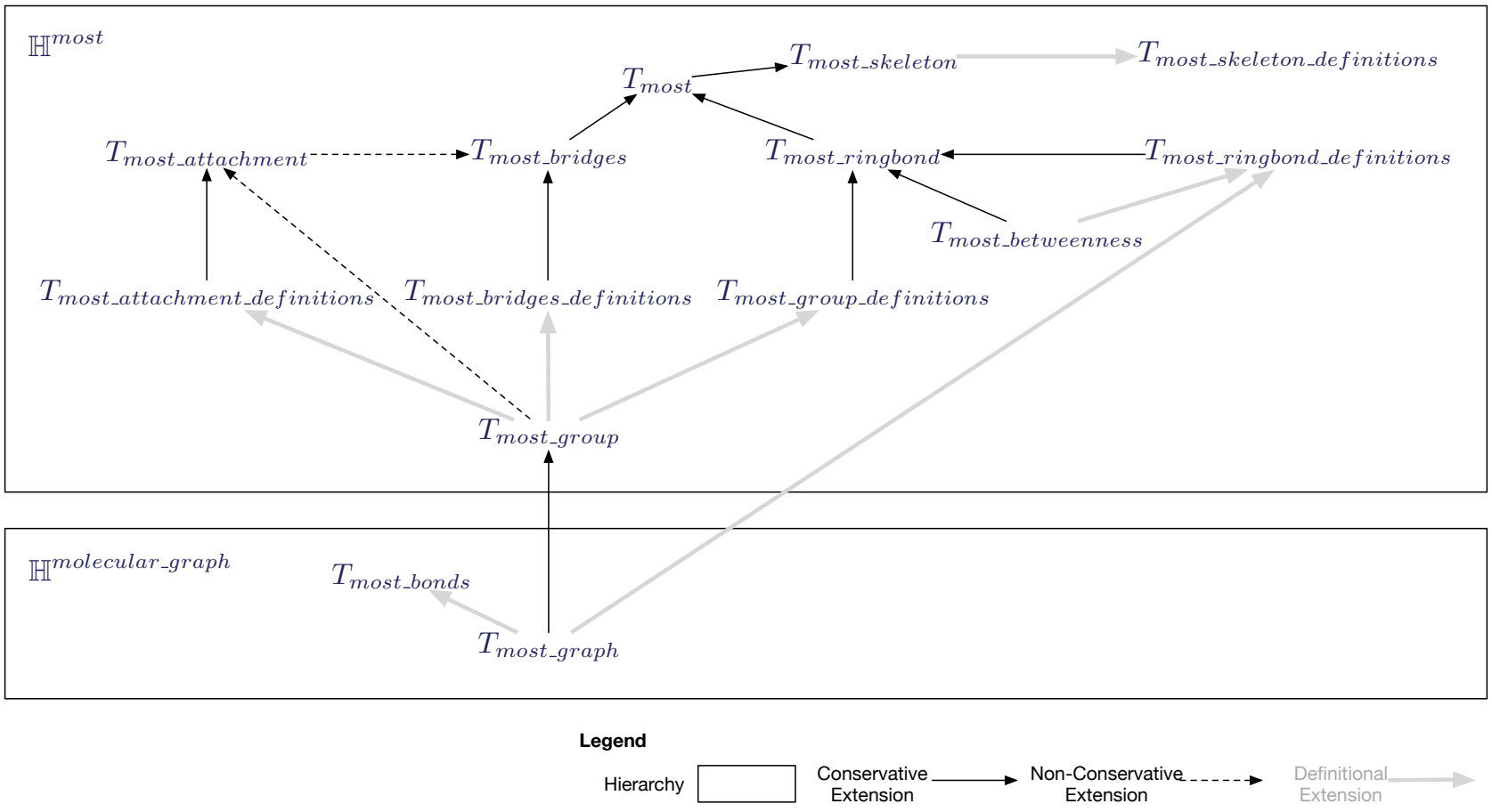
- Additional insights using axioms that describe commonsense intuitions about structure and relationships between the various attachments and elements of the domain
  - Example: *For a fork atom that is in a group, there exists an atom and a bond that is not part of that group.*

$$\forall x \forall g\ (fork(x) \wedge mol(x, g) \wedge group(g)) \supset\ \exists b \exists y\ atom(y) \wedge bond(b) \wedge mol(x, b) \wedge mol(y, b) \wedge \neg mol(y, g).$$



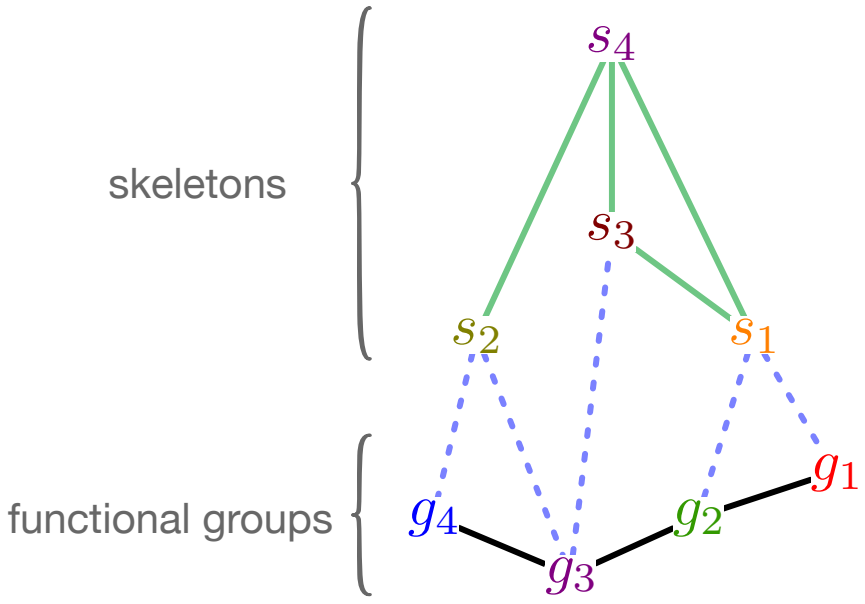
- Make use of available first-order reasoners - not much reasoning can be done with class/subclass relationships (in OWL)

Organization of MoSt

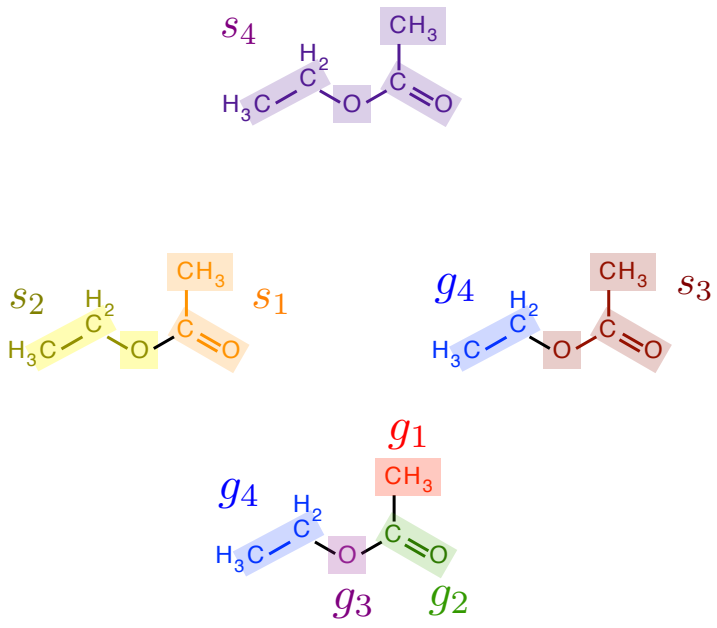


Skeletons

Breakdown of Skeletons & Groups



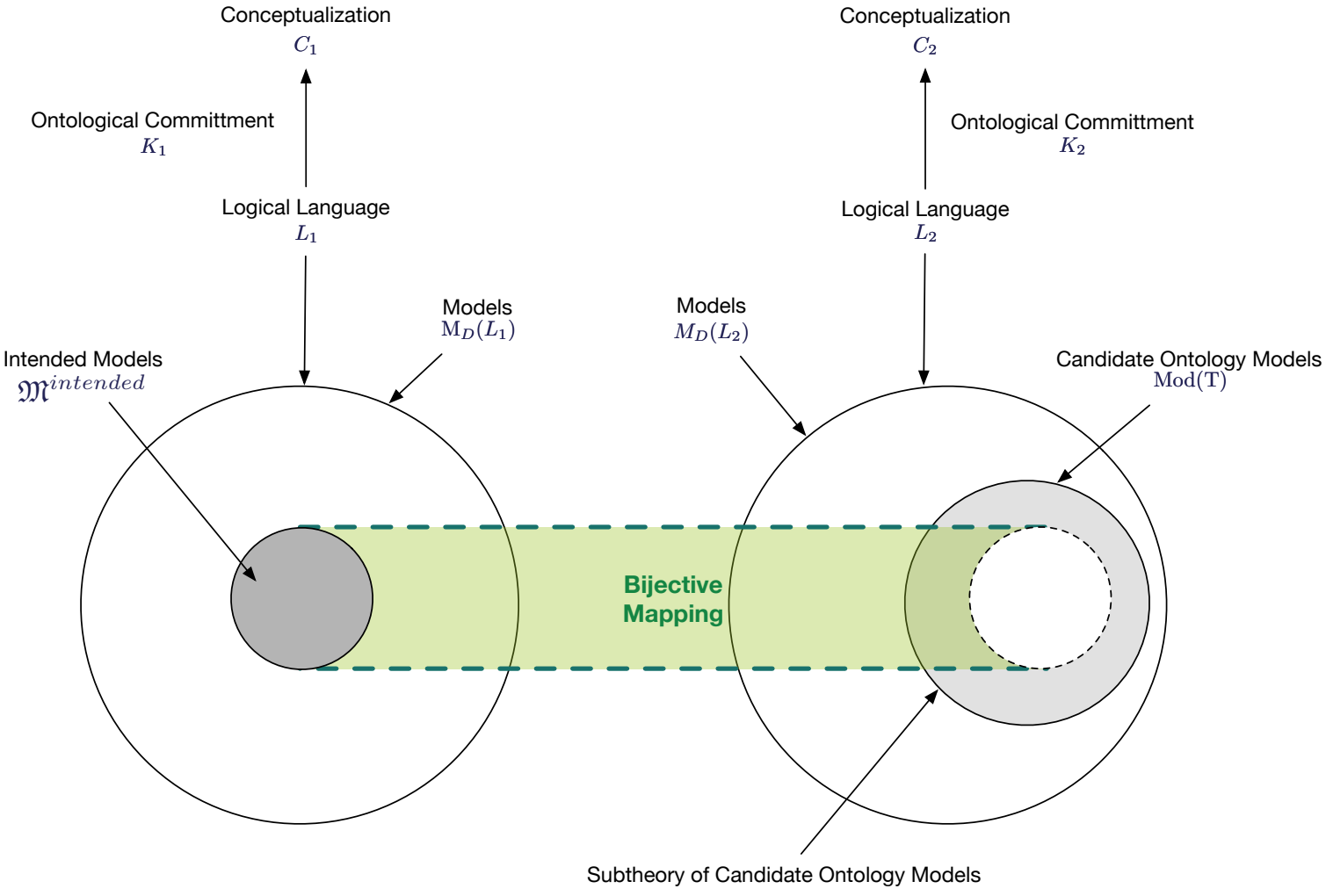
Corresponding Skeletal Diagrams



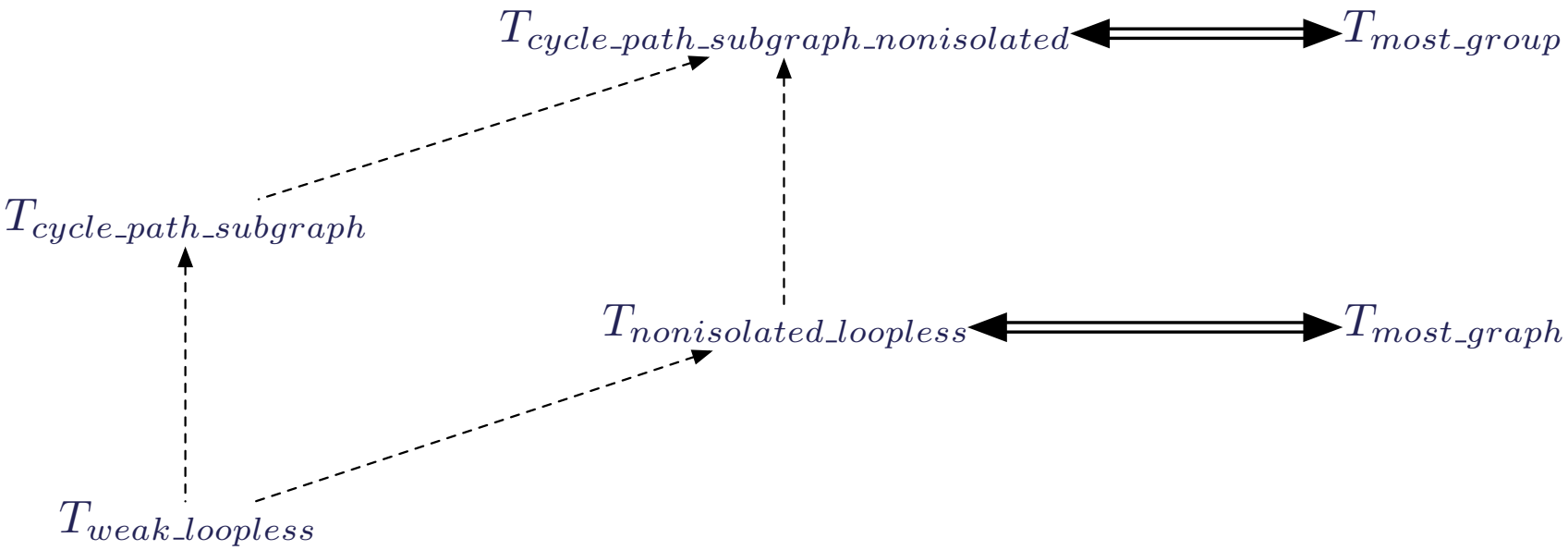
Legend

connectedness ——— incidence - - - - - mereology ———

Verification & Bijective Mappings



Verification Theories & Their Relationships



Legend

Conservative Extension — Non-Conservative Extension — Definitional Extension — Definably Equivalent