Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

An Ontology for Medicinal Chemistry

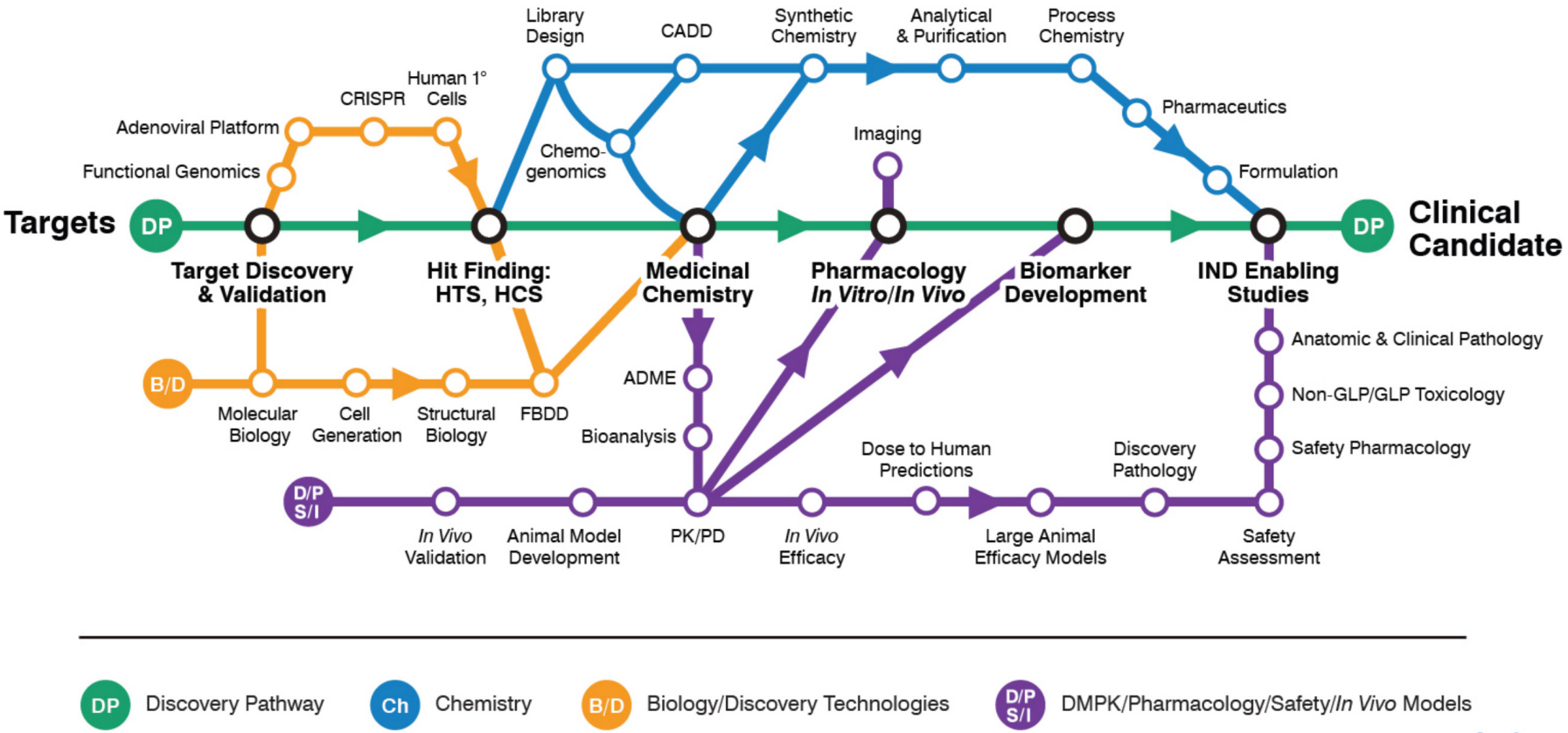
Ph.D. Defence

Carmen S. Chui

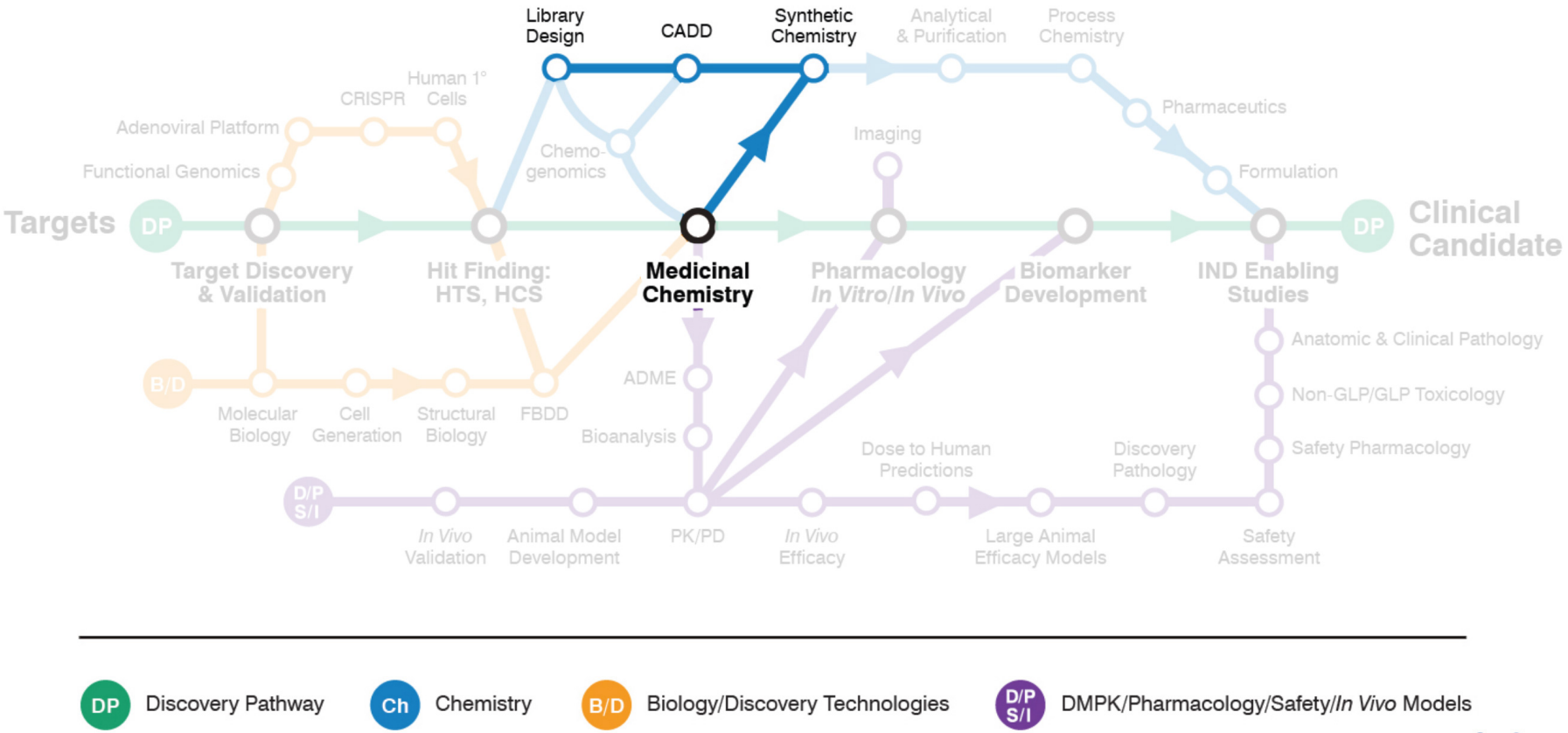
March 26, 2019

SGS Room 111

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]

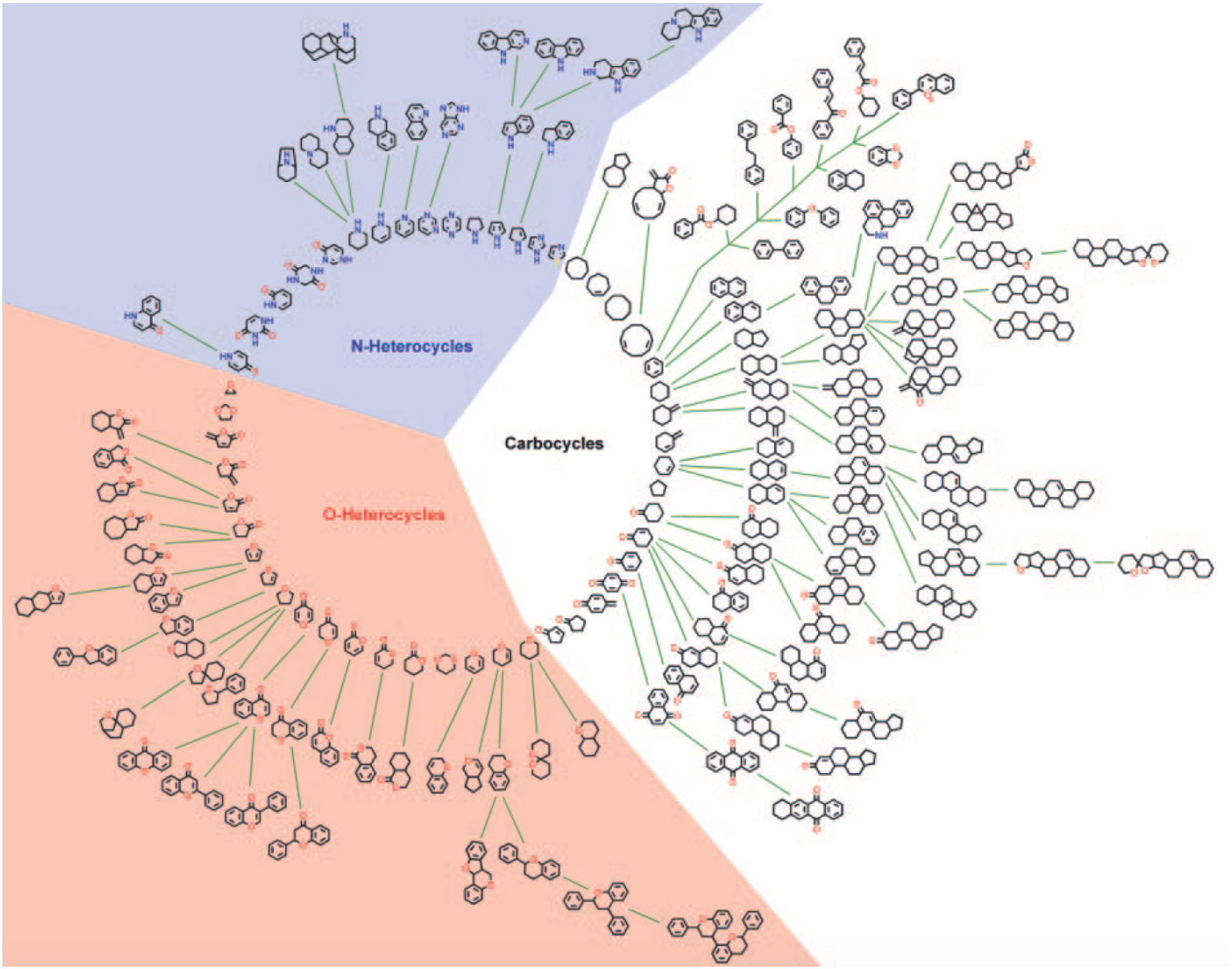
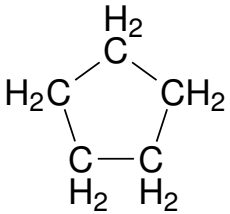
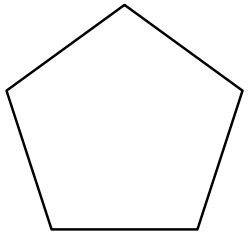


Figure 1: Scaffold Tree for Natural Products (Figure 1 in [Koc+05])

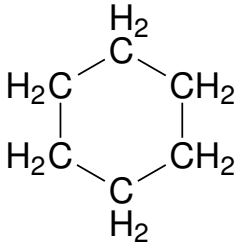
What do we mean by shape?



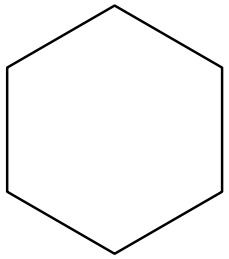
(a) Cyclopentane



(b) Pentagon

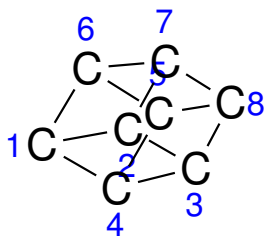


(c) Cyclohexane

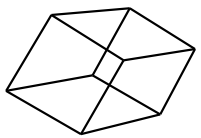


(d) Hexagon

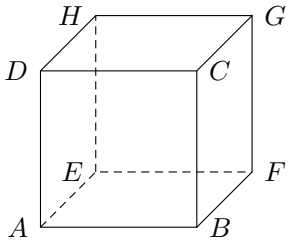
Figure 2: 2D Shapes (Polygons) & Molecules



(a) Cubane



(b) Skeletal formula.



(c) A cube

Figure 3: 3D Shapes (Polyhedra) & Molecules

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

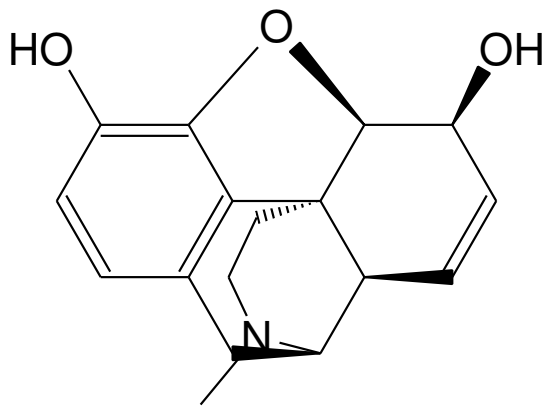
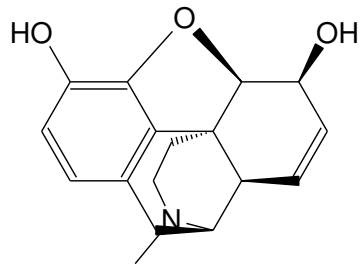


Figure 4: Morphine ($C_{17}H_{19}NO_3$)

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:

1. Design, Verification, and Validation of MoSt
2. New Techniques for Designing Molecules via Model Construction
3. An Alternative Approach to Navigating Chemical Space

1. Design, Verification, and Validation of MoSt

Requirements for the Ontology & Its Models

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

Axiomatization of MoSt

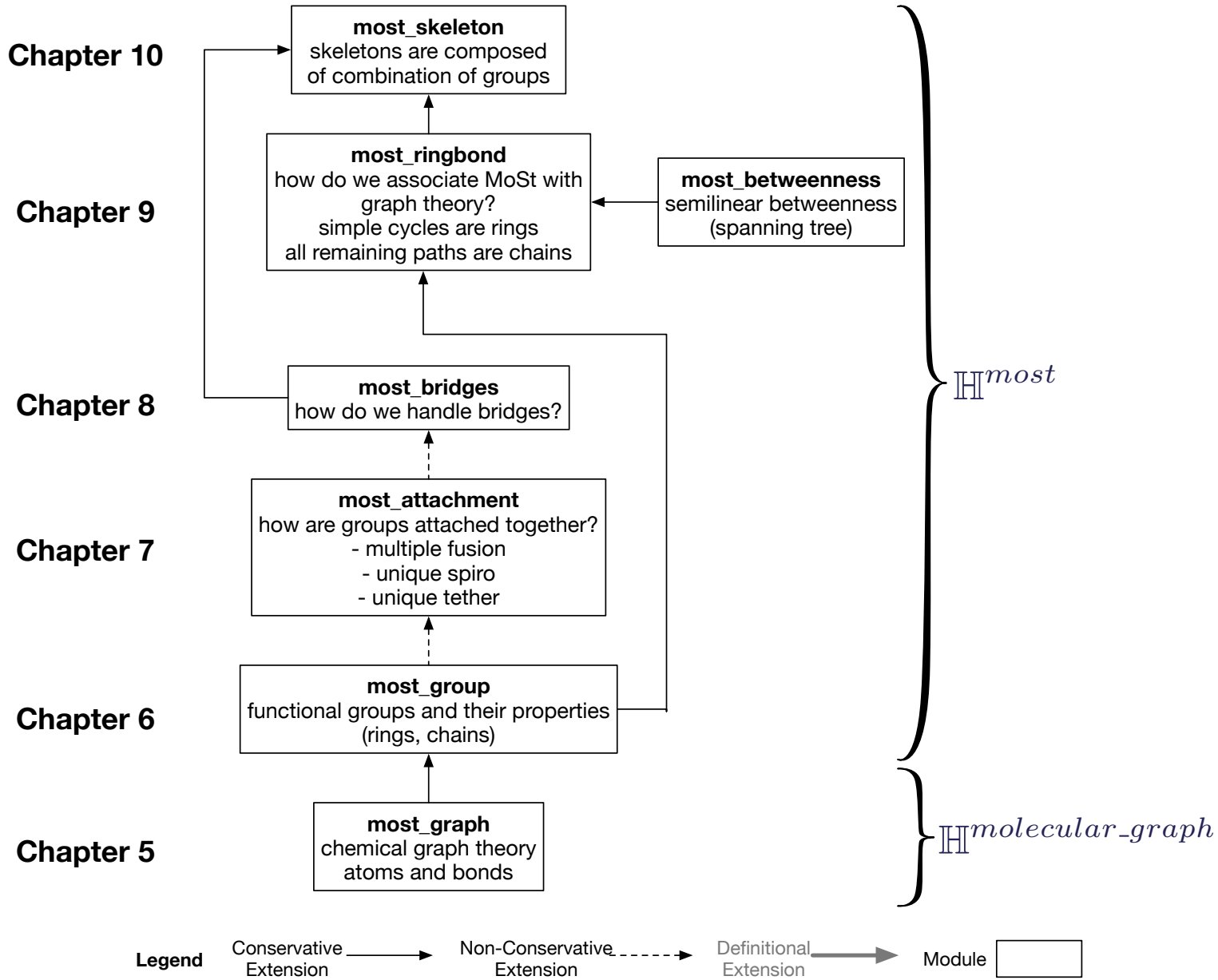


Figure 5: Axiomatization of MoSt (by chapter)

Verification of MoSt

- Verification results show that models of MoSt are **synonymous** with tripartite incidence structures found in COLORE
- We *inherit* techniques for building and decomposing models from the mathematical incidence theories

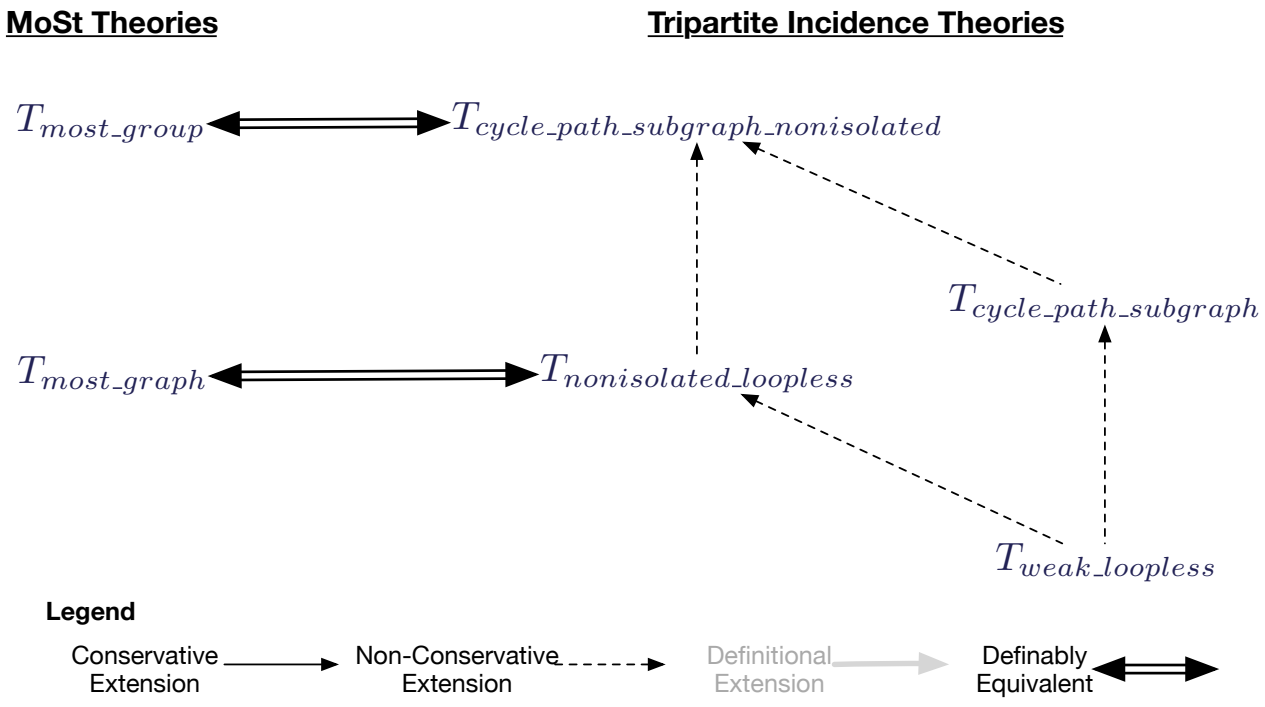
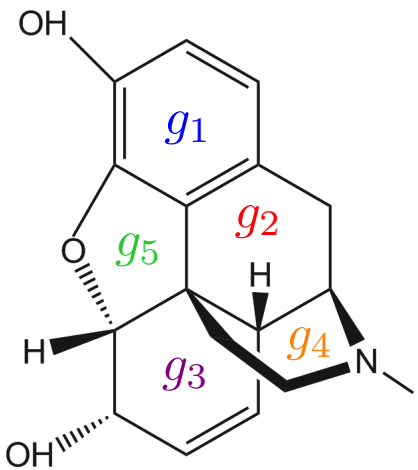


Figure 6: Complete classification of models of MoSt through verification.

Example of Axiomatization



(a) Structure

$$\forall x\ morphine(x) \equiv \exists g_1 \exists g_2 \exists g_3 \exists g_4 \exists g_5\ skeleton(x) \wedge phenol(g_1) \wedge$$
$$cyclohexane(g_2) \wedge cyclohexenol(g_3) \wedge$$
$$1methylpiperidine(g_4) \wedge furan(g_5) \wedge$$
$$multiply\_fused(g_1, g_2) \wedge multiply\_fused(g_1, g_5) \wedge$$
$$multiply\_fused(g_2, g_3) \wedge multiply\_fused(g_2, g_4) \wedge$$
$$multiply\_fused(g_2, g_5) \wedge multiply\_fused(g_3, g_2) \wedge$$
$$multiply\_fused(g_3, g_4) \wedge multiply\_fused(g_3, g_5) \wedge$$
$$multiply\_fused(g_4, g_2) \wedge multiply\_fused(g_4, g_3) \wedge$$
$$mol(g_1, x) \wedge mol(g_2, x) \wedge$$
$$mol(g_3, x) \wedge mol(g_4, x) \wedge mol(g_5, x)$$

(b) Axiomatization

Figure 7: Structure of morphine ($C_{17}H_{19}NO_3$)

Contribution Summary

- Identified requirements for a molecular structure ontology
- Provided an axiomatization of a molecular structure ontology (MoSt)
- Verified and validated the ontology with respect to its intended models

2. New Techniques for Designing Molecules via Model Construction

Leveraging the Ontology in Drug Design

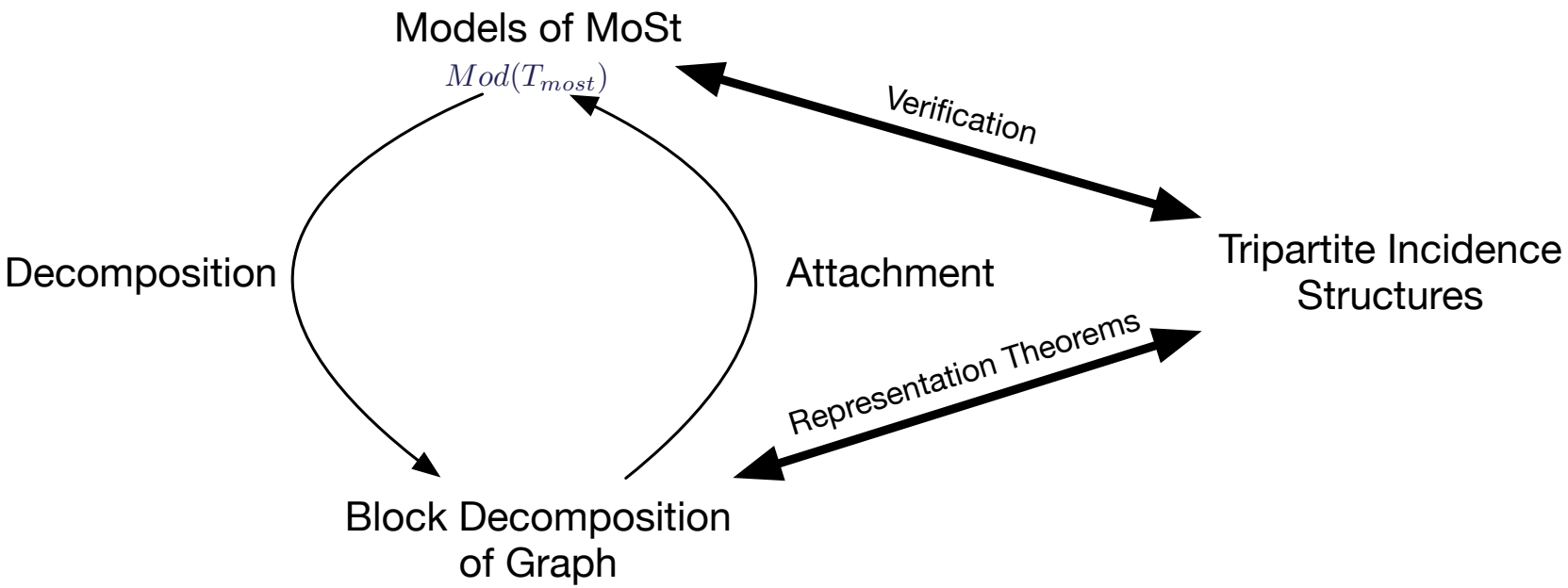
The ontology plays **dual** roles:

1. Verification shows that models correspond to graphs in geometry: we can construct models of MoSt for drug design. However, simply having the ontology itself does not tell us how we can use it for drug design.
2. We can exploit graph-theoretic properties of the models to come up with new model construction techniques.

Drug Design as Model Construction

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

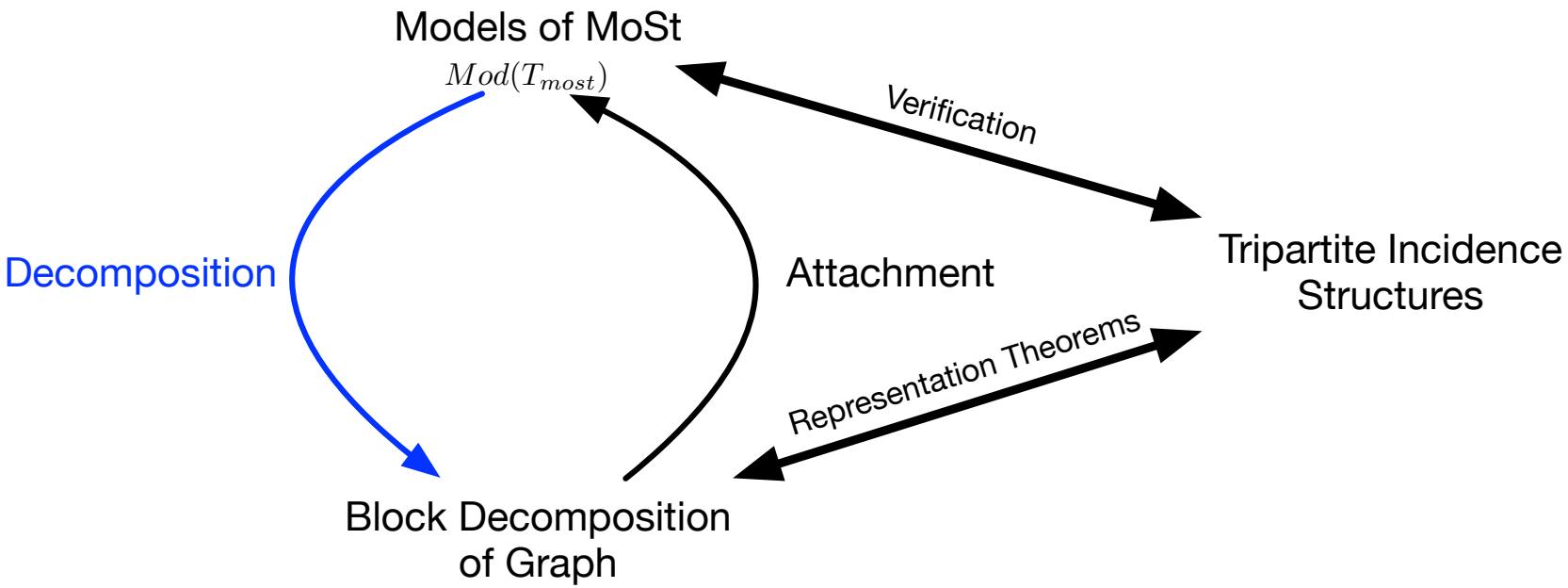
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence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



Drug Design as Model Construction

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

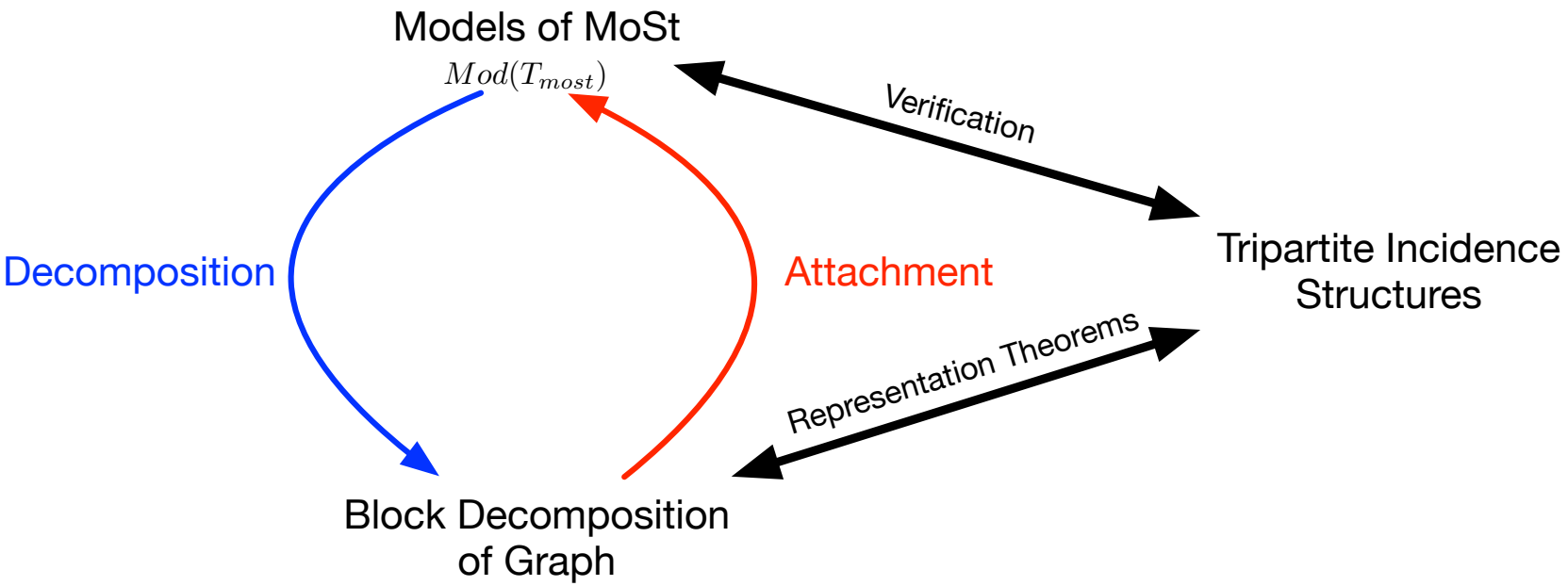
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence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



Drug Design as Model Construction

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

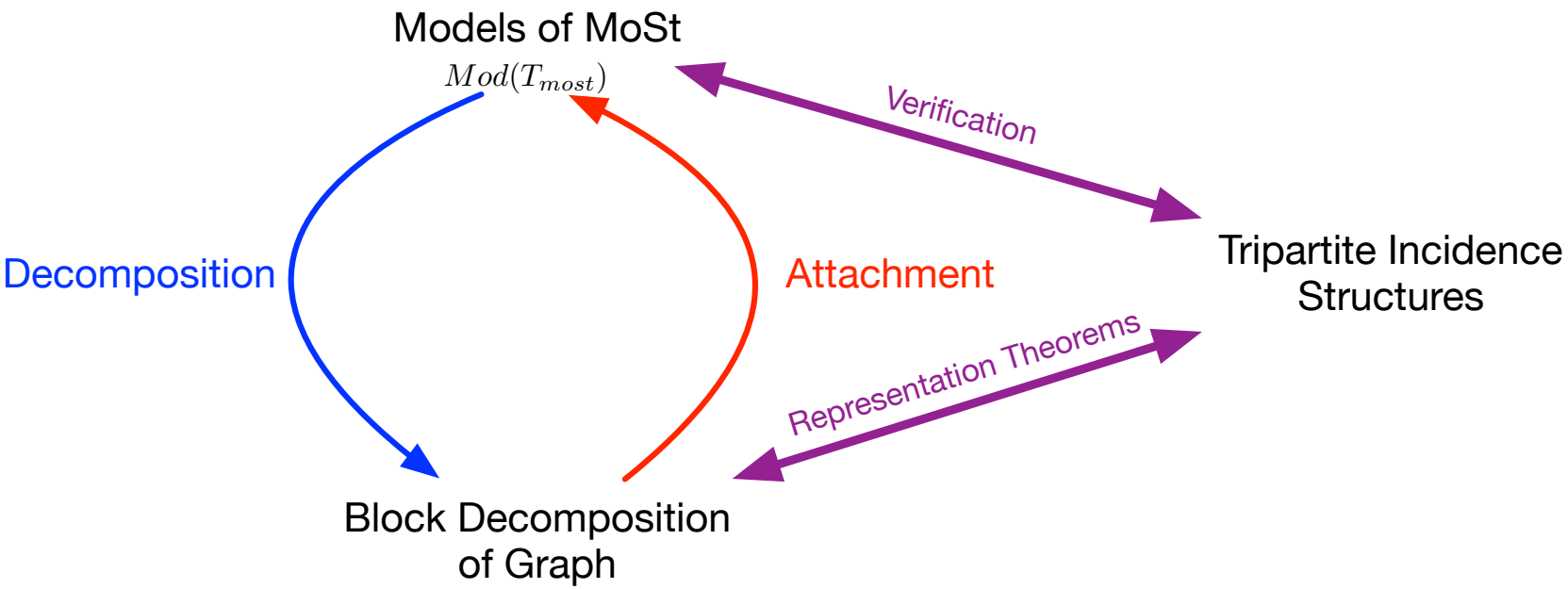
- A model of MoSt, $Mod(T_{most})$, can be decomposed into 2-connected graph components
- From these components, we can re-compose the graph via the attachment relationships
- Models of MoSt and the block decompositions of the underlying molecular graph are **synonymous** with the tripartite incidence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



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



Drug Discovery as Model Construction (cont.)

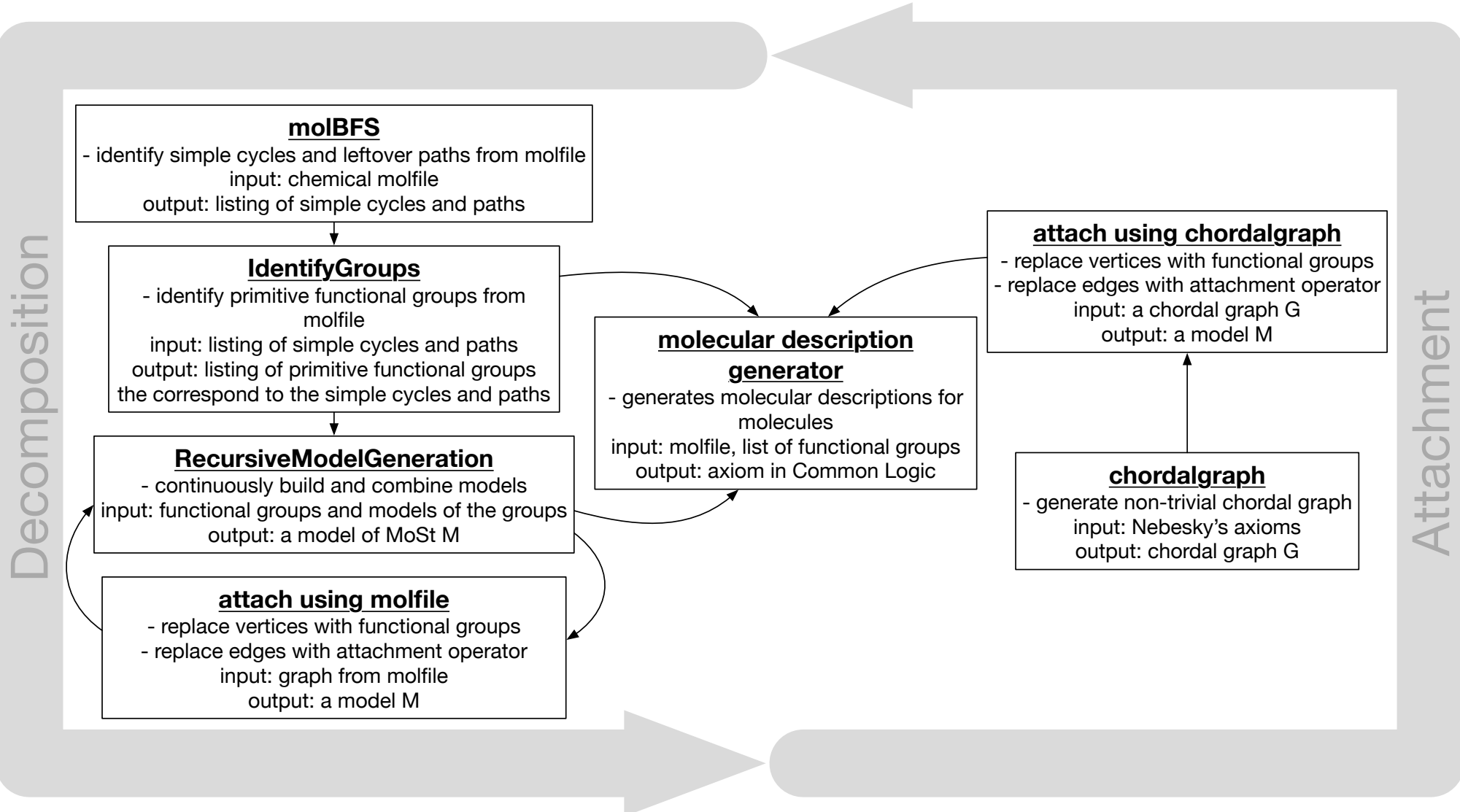
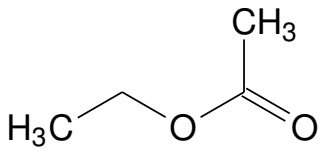


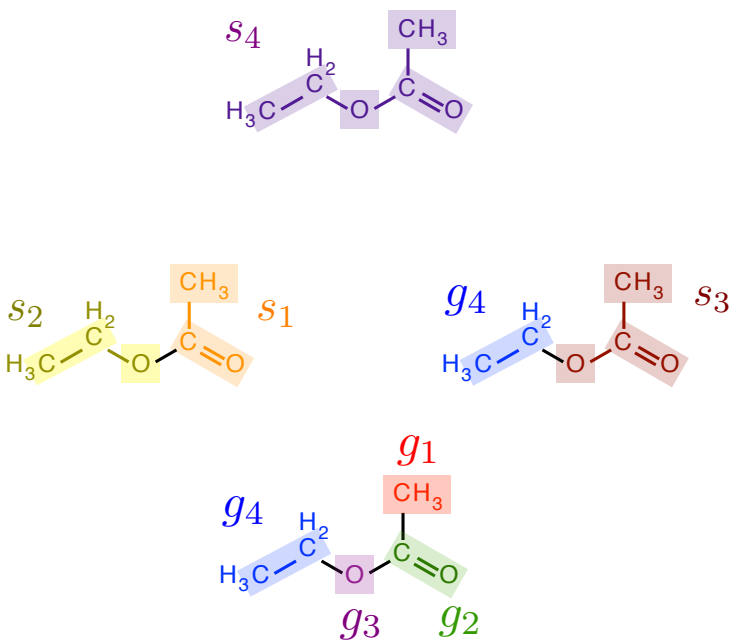
Figure 8: Generating models via attachment

Breaking Down Skeletons

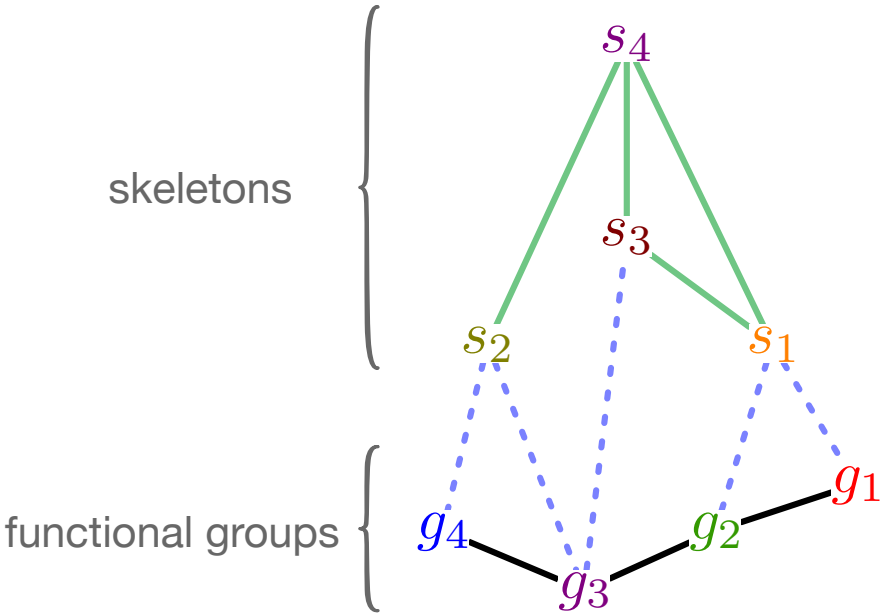
Consider Ethyl Acetate ($C_4H_8O_2$):



Corresponding Skeletal Diagrams

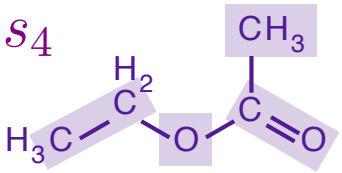


Breakdown of Skeletons & Groups

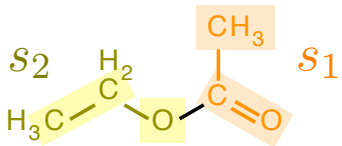


Legend connectedness ——— incidence - - - - - - mereology ———

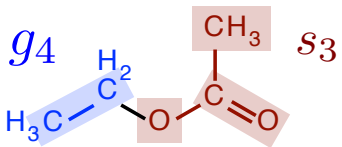
Breaking Down Skeletons (cont.)



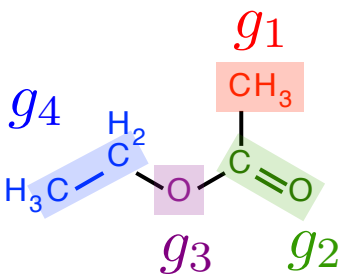
$$\forall x\ ethyl\_acetate(x) \supset\ skeleton(x)$$



$$\begin{aligned}\forall x\ ethyl\_acetate(x) \equiv \exists s_1 \exists s_2 \exists b_1\ & skeleton(x) \wedge \\ & acetic\_acid(s_1) \wedge ethanol(s_2) \wedge \\ & mol(s_1, x) \wedge mol(s_2, x) \wedge \\ & tether(s_1, s_2, b_1)\end{aligned}$$



$$\begin{aligned}\forall x\ ethyl\_acetate(x) \equiv \exists g_4 \exists s_3 \exists b_1\ & skeleton(x) \wedge \\ & acetyl\_oxy(s_3) \wedge ethane(g_4) \wedge \\ & mol(s_3, x) \wedge mol(g_4, x) \wedge \\ & tether(s_1, s_2, b_1)\end{aligned}$$



$$\begin{aligned}\forall x\ ethyl\_acetate(x) \equiv \exists g_1 \exists g_2 \exists g_3 \exists g_4 \exists b_1 \exists b_2 \exists b_3\ & skeleton(x) \wedge \\ & methyl(g_1) \wedge carbonyl(g_2) \wedge ether(g_3) \wedge \\ & ethane(g_4) \wedge mol(g_1, x) \wedge mol(g_2, x) \wedge \\ & mol(g_3, x) \wedge mol(g_4, x) \wedge tether(g_1, g_2, b_1) \wedge \\ & tether(g_2, g_3, b_2) \wedge tether(g_3, g_4, b_3)\end{aligned}$$

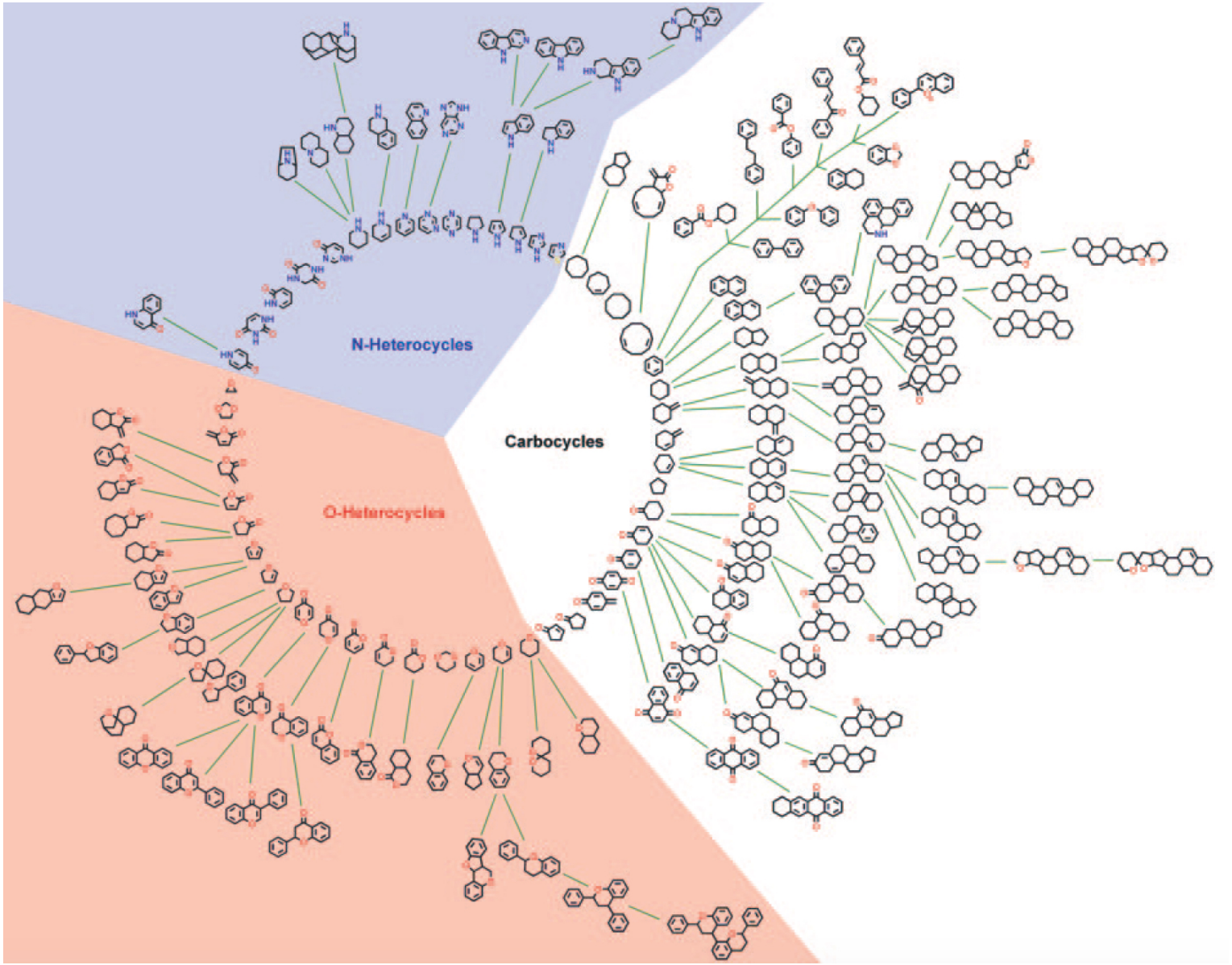
Contribution Summary

Based on the graph-theoretic properties of the axioms and the verification, we have presented:

- The process of drug design as the process of building first-order models of MoSt
- New techniques and procedures to decompose and re-compose molecules and primitive functional groups

3. An Alternative Approach to Navigating Chemical Space

Why Navigate Chemical Space?



Why Navigate Chemical Space?

- Chemical space is the property space spanned by all possible molecules and compounds, where its size is unknown
- Have ontologies, along with a model-theoretic approach, been considered to aid in the search tasks?
- We can use models of MoSt can be used to help narrow the search space

Navigating Chemical Space $\simeq$ Constraint Satisfaction Programming

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

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

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

Model-Theoretic Search Techniques

We can look at drug design in these contexts:

1. **Information retrieval**: querying a knowledge base (KB) of existing molecules via **entailment**

$$KB \models \text{(molecular description or molecular constraint)}$$

2. **Generating molecules**: solving satisfiability queries using MoSt by **building models** that satisfy molecular constraints

$$T_{most} \cup \text{(molecular constraint) is consistent}$$

$$T_{most} \not\models \neg\text{(molecular constraint)}$$

Using Queries to Find & Design New Molecules

Example: Information Retrieval (Entailment)

Find a molecule that contains at least one ring in the knowledge base that contains existing molecules.

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

Example: Satisfiability Problem (Model Construction)

Design a molecule that contains at least one ring.

$$T_{most} \not\models \neg(\exists x \; ring(x))$$

Using Queries to Find & Design New Molecules

Example: Information Retrieval (Entailment)

Find a molecule in the knowledge base where all groups are tethered to at least one other group.

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

Example: Satisfiability Problem (Model Construction)

Design a molecule where all groups are tethered to at least one other group.

$$T_{most} \not\models \neg(\forall x\, group(x) \wedge \exists y\, group(y) \wedge tether(x, y) \wedge (x \neq y))$$

Contribution Summary

Based on our earlier discussion of leveraging graph-theoretic properties of MoSt's models, we have:

- Shown how we can navigate the chemical search space using the axioms of MoSt
- Presented retrieval queries and satisfiability problems to narrow the search space based on molecular constraints

Open Questions & Future Work

Open Questions

- Decidability of MoSt
- Mereology on Skeletons

Future Work

- Molecular Reactions Ontology (MoRe): A Process Ontology
- Reasoning About Molecules (RoMe): A Software Environment
- Integration with (Cheminformatics) Software Tools & Query Languages

Summary of Contributions

We have provided an ontological foundation for navigating chemical space, with the following contributions:

1. Design, Verification, and Validation of MoSt, satisfying the requirements for a molecular structure ontology
2. New Techniques for Designing Molecules via Model Construction
3. An Alternative Approach to Navigating Chemical Space

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.