Designing potential extensions from G-SRS to ChEBI to identify natural product-drug interactions

Sanya Bathla Taneja
University of Pittsburgh
sbt12@pitt.edu
July 25, 2021

Natural Products Promoted for Complementary Health

Kratom leaf

Green Tea

Cannabidiol (found in cannabis)

Grapefruit juice

Goldenseal

Natural Products in Biomedical Ontologies

Food Ontology (FOODON)

NCIT

Drug Ontology (DRON)

Green Tea

Natural Products in Biomedical Ontologies

Food Ontology (FOODON)

Green Tea Leaf: FOODON_03304639

- Hierarchical structure
- Display and preferred names
- Synonym
- Annotations related to quality, formation, and derivation of green tea product.

NCIT

Green Tea: NCIT_C67048

- Hierarchical structure
- Display and preferred names
- Semantic type
- External references
  - UMLS
  - FDA UNII
  - NCI Drug

Plant or Plant Part Extract
NCIT_C106105

is-a

Green Tea
NCIT_C67048

Drug Ontology (DRON)

Green Tea Extract: DRON_00018919 (and Green Tea Extract Drug Capsule)

- Hierarchical structure
- Drug capsule information

Drugs/Chemicals in Biomedical Ontologies

ChEBI

Simvastatin: CHEBI_9150

- Hierarchical structure
- External references
  - PMID
  - KEGG
  - Wikipedia
  - Drug Central
  - Drug Bank
  - LINCS…
- IUPAC name
- Synonyms
- Chemical properties: charge, formula, inchi key, mass, smiles
- Chemical roles and metabolites
- Biological roles
- Applications

NCIT

Simvastatin: NCIT_C29454

- Hierarchical structure
- External references
  - ChEBI
  - FDA UNII
  - NCI Drug Dictionary
  - UMLS
- Semantic Type
- Synonyms
- Chemical formula
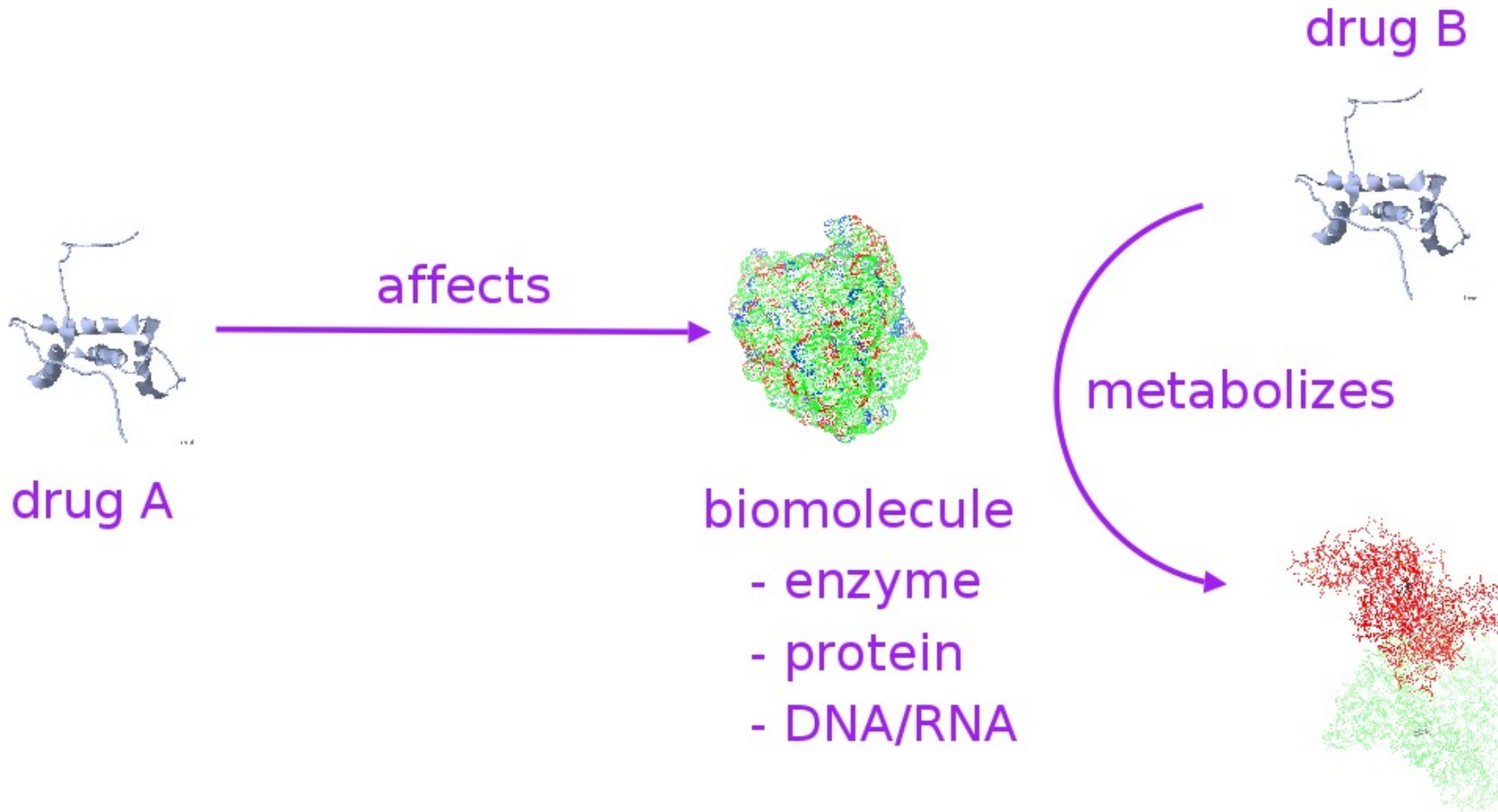- Pharmacokinetic information (inhibition, metabolism, targets)
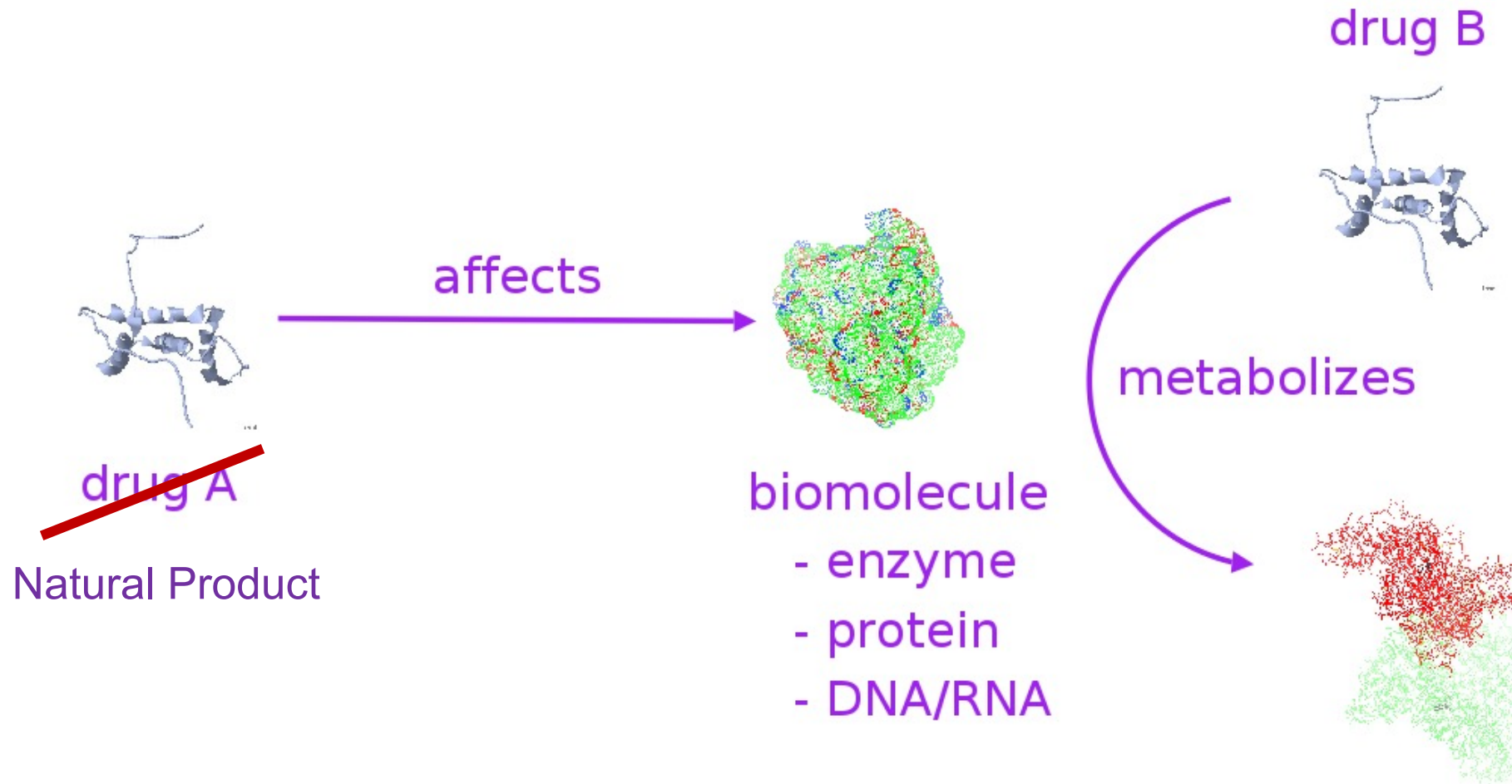
DRON

Simvastatin: DRON_00751919

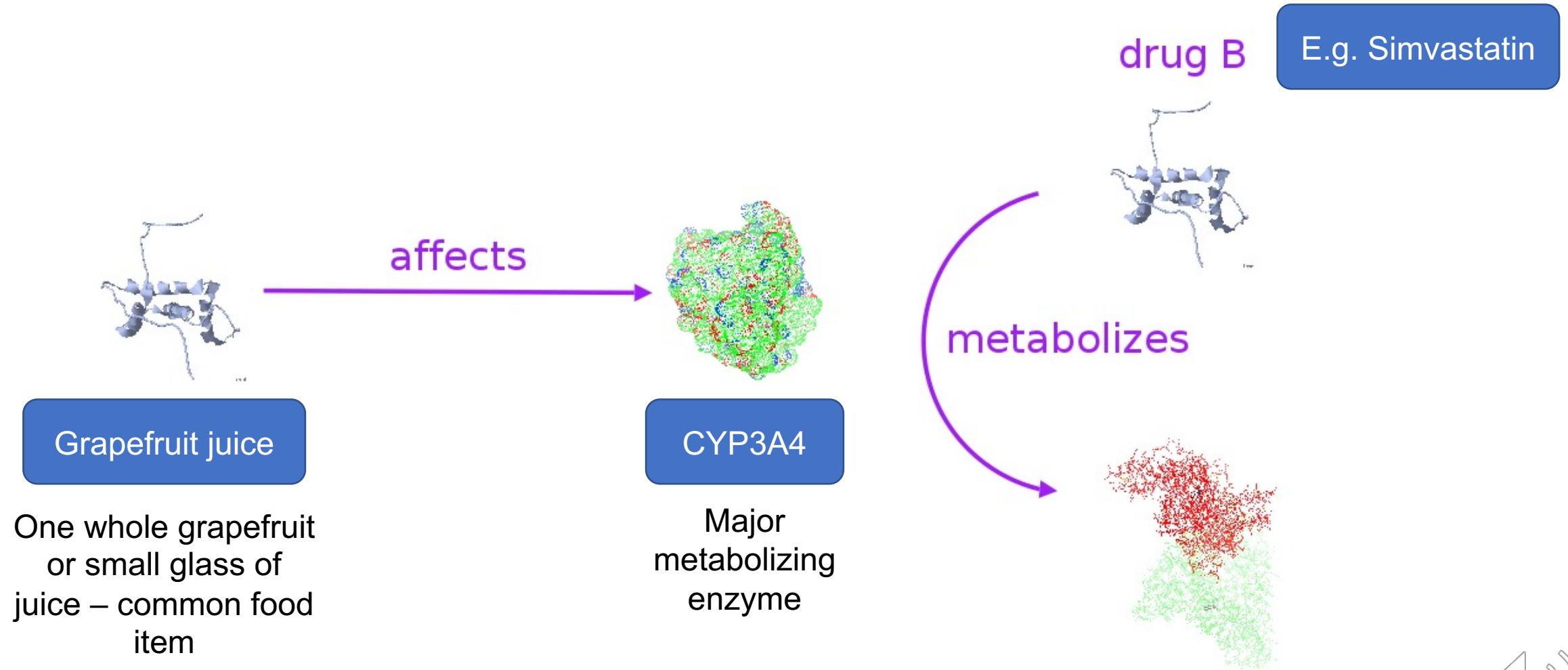- Hierarchical structure
- Drug capsule information
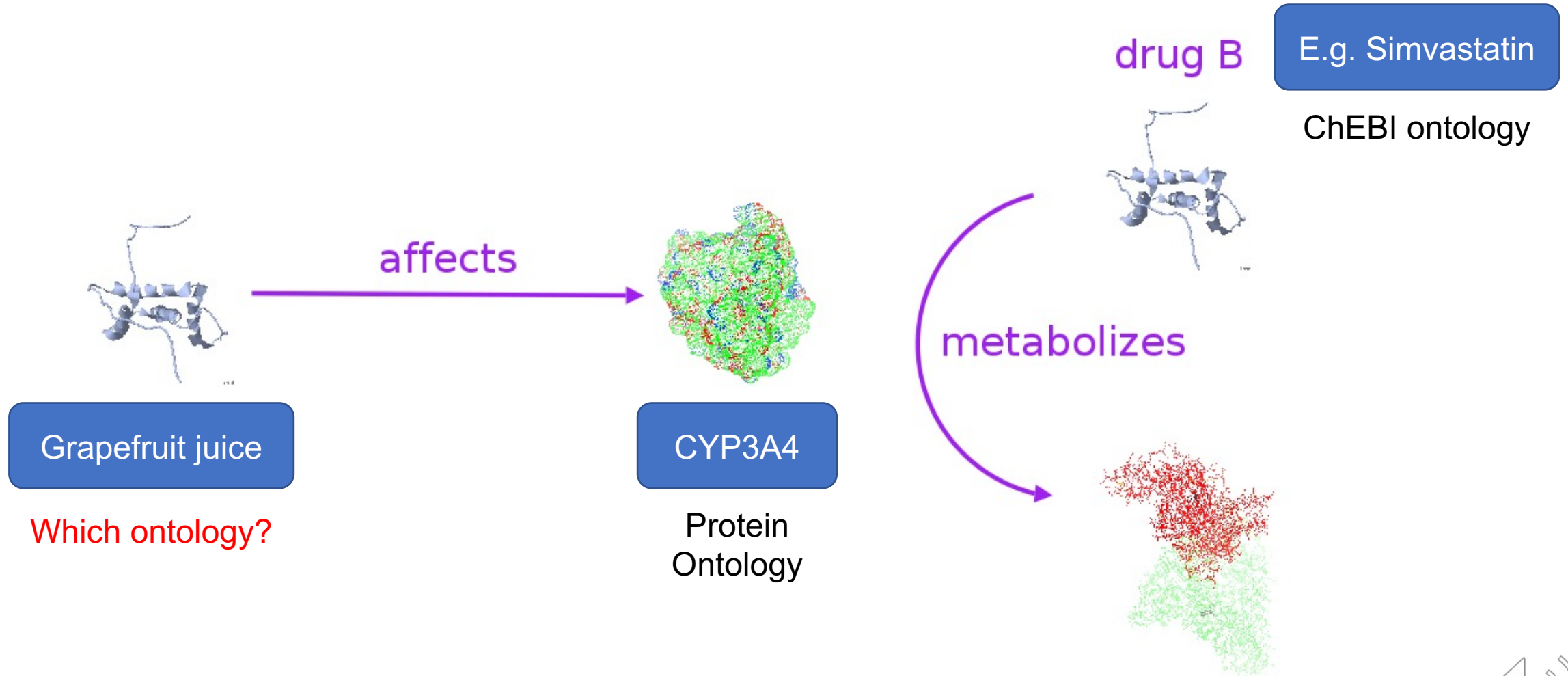- RxCUI available

Pharmacokinetic Drug Interactions

drug A

affects

biomolecule
- enzyme
- protein
- DNA/RNA

metabolizes

drug B

Pharmacokinetic NP-Drug Interactions

drug B

affects

metabolizes

drug A

Natural Product

biomolecule
- enzyme
- protein
- DNA/RNA

Pharmacokinetic NP-Drug Interactions

E.g. Simvastatin

drug B

affects

metabolizes

Grapefruit juice

One whole grapefruit or small glass of juice – common food item

CYP3A4

Major metabolizing enzyme

Pharmacokinetic NP-Drug Interactions

affects

metabolizes

drug B

E.g. Simvastatin

ChEBI ontology

Grapefruit juice

Which ontology?

CYP3A4

Protein Ontology

Goals

- Biomedical data integration of natural products (NPs) in the Open Biological and Biomedical Ontology (OBO) Foundry ontologies

- Design logical extensions to include NPs, NP constituents, and related pharmacokinetic information in biomedical ontologies

- Facilitate discovery of potential natural product-drug interactions (NPDIs)
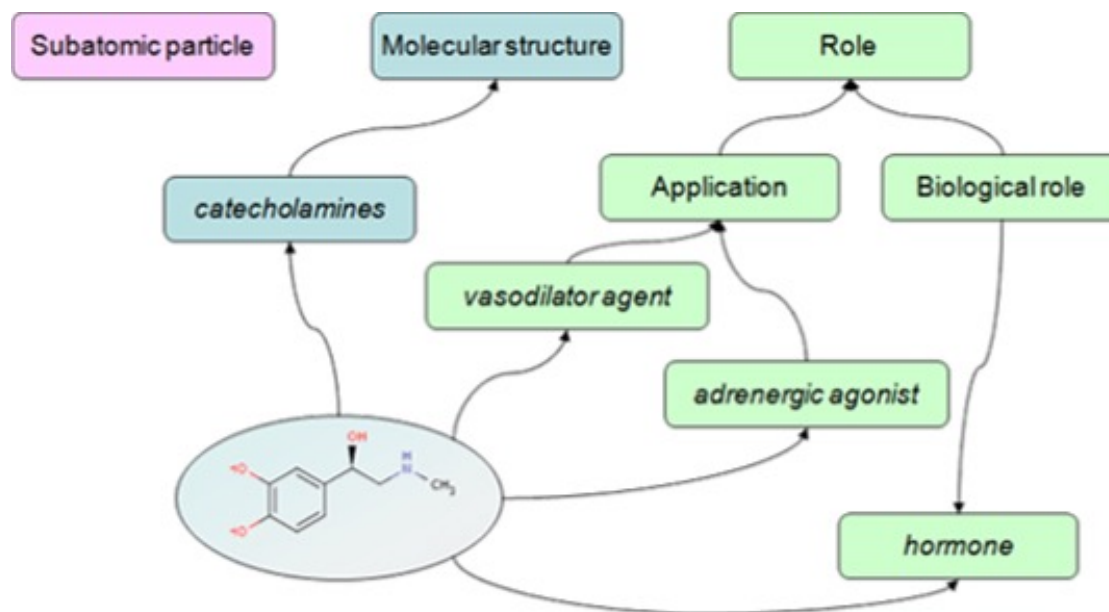
Workflow

Data Extraction → Semantic Representations → Logical Statements → Enhanced Ontology

Data Extraction

ChEBI: Chemical Entities of Biological Interest

- Dictionary of molecular entities
- Identifiers, names, synonyms, chemical characteristics
- "Relations" between entities – has_functional_parent, has_role, is_enantiomer_of
- Large number of drugs and chemicals
- Only some natural product constituents (mitragynine – kratom, catechin(s)– green tea)

References: https://www.ebi.ac.uk/training/online/courses/chebi-the-online-chemical-dictionary-for-small-molecules/chebi-ontology/

Information Requirements

- Standard identifiers
- Access to domain relevant information (genus, species, constituents, chemical characteristics)
- Comprehensive coverage of natural products
- Non-ambiguous names
- Availability of pharmacokinetic information

Potential data sources:

- **Global Substance Registration System (G-SRS)**
- Dietary Supplement Label Database (DSLD)
- Licensed Natural Health Products Database (Canada)
- Unified Medical Language System (UMLS)
- Others (NDF-RT, RxNorm, FOODON, Natural Medicines Database, MESH, SNOMED-CT)

Global Substance Registration System (G-SRS)

- Led by the Ginas project - National Institutes of Health/National Center for Advancing Translational Science (NCATS) and US Food and Drug Administration (FDA)

- 25,000 structurally diverse substances or natural products

- NP constituents and metabolites

- External database references (DSLD, NCIT)

- Available as a SQL database and web service
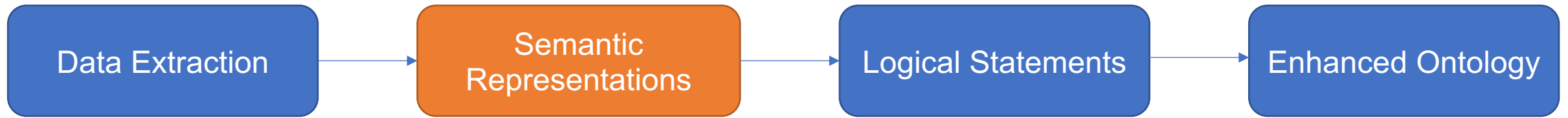

G-SRS
Ver. 2.7.1
Search

Overview	>
Names 18	>
Classification 4	>
Identifiers 10	>
Metabolites 28	>
Active Moiety 1	>
Constituents 8	>
Variant Concepts 2	>
Audit Info	>
References 42	>

GREEN TEA LEAF

Overview

Substance Class   Structurally Diverse

Record UNII   W2ZU1RY8B0

Record Protection Status   Public record 🔓

Record Status   Validated (UNII)

Source Materials Class   ORGANISM

Source Materials Type   PLANT

Source Materials Parent

CAMELLIA SINENSIS WHOLE

14

Data Extraction

Center of Excellence for Natural Product-Drug Interaction Research (NaPDI Center)

- Assess clinical relevance of pharmacokinetic natural product-drug interactions (NPDIs)

- Address existing gaps in scientific literature about NPs and NPDIs

- Publicly accessible data repository with scientific results, raw data, and recommended approaches related to pharmacokinetic NPDIs
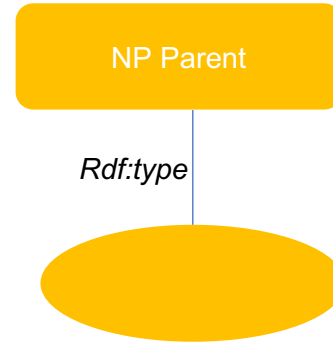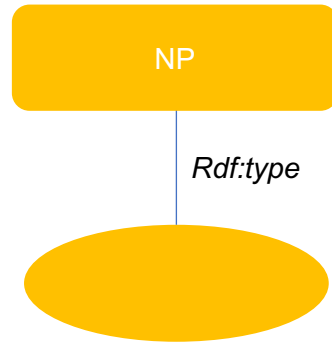
Workflow

Data Extraction → Semantic Representations → Logical Statements → Enhanced Ontology
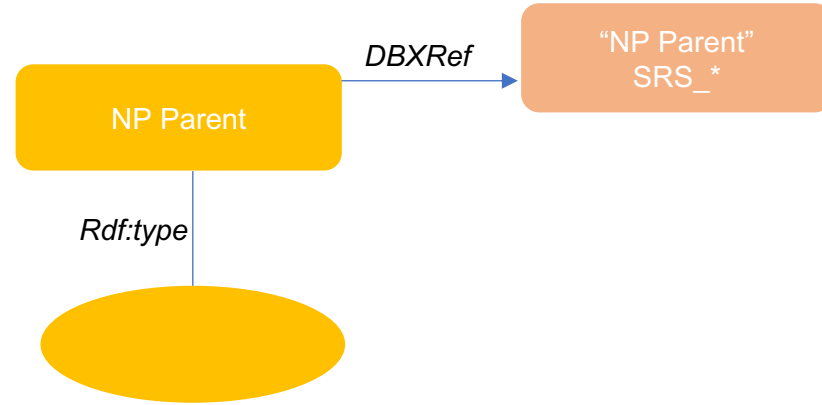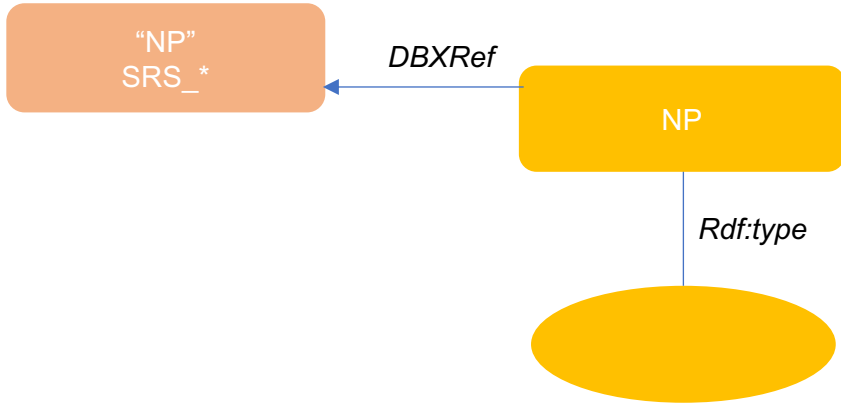
Semantic Representation

- Add novel classes for each natural product (NP) and its parent substance
    - subClassOf 'Plant Anatomical Entity' (Plant Ontology)

- Determine class-class and instance-instance relationships

- Relation Ontology (RO), ChEBI, Basic Formal Ontology (BFO) terms to establish relations between entities

- Generate semantic representation patterns based on available information -
    - NP constituent class exists in ChEBI ontology and G-SRS
    - NP constituent exists in G-SRS only
    - NP metabolite class exists in ChEBI ontology and G-SRS
    - NP metabolites exist in G-SRS only
    - NP has metabolites in G-SRS only
    - Known in vitro and/or clinical information related to enzymes and transporters

Semantic Representation

```
┌──────────────┐              ┌──────────────┐
│      NP      │              │  NP Parent   │
└──────────────┘              └──────────────┘
        │                             │
     Rdf:type                     Rdf:type
        │                             │
      ⬭                             ⬭
```
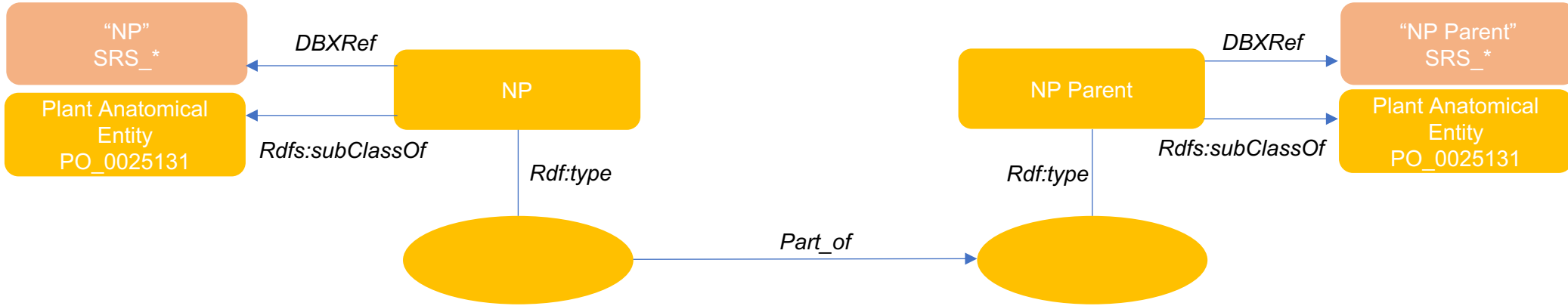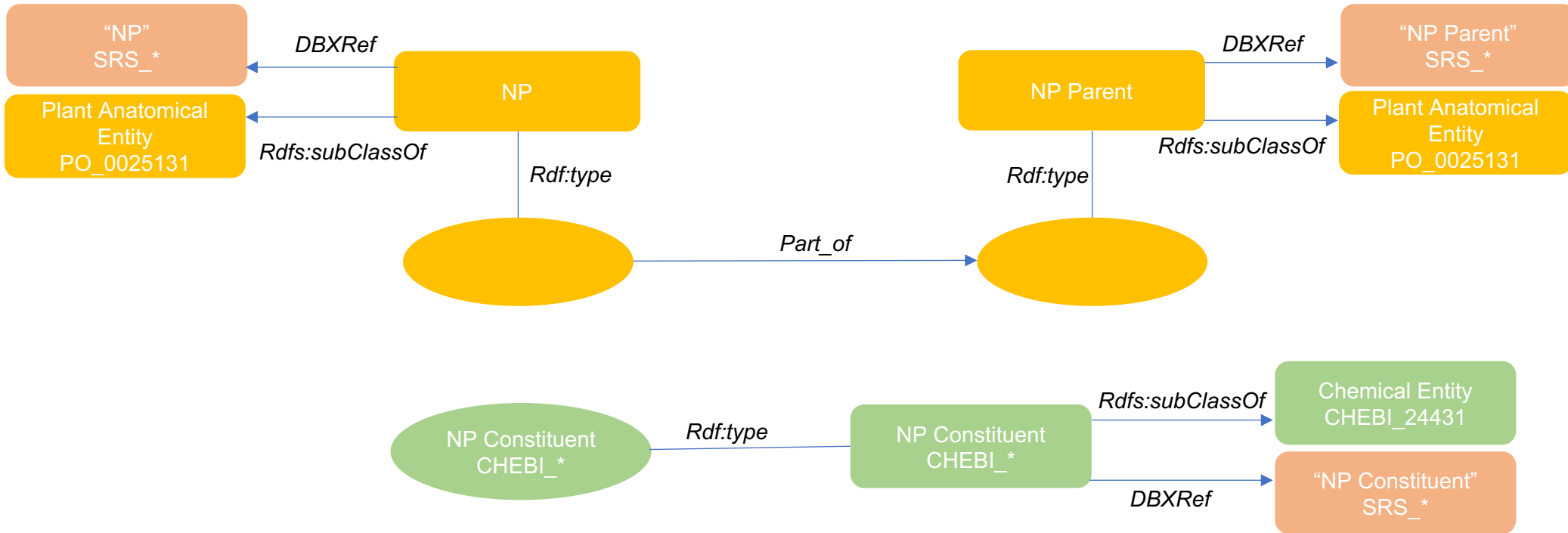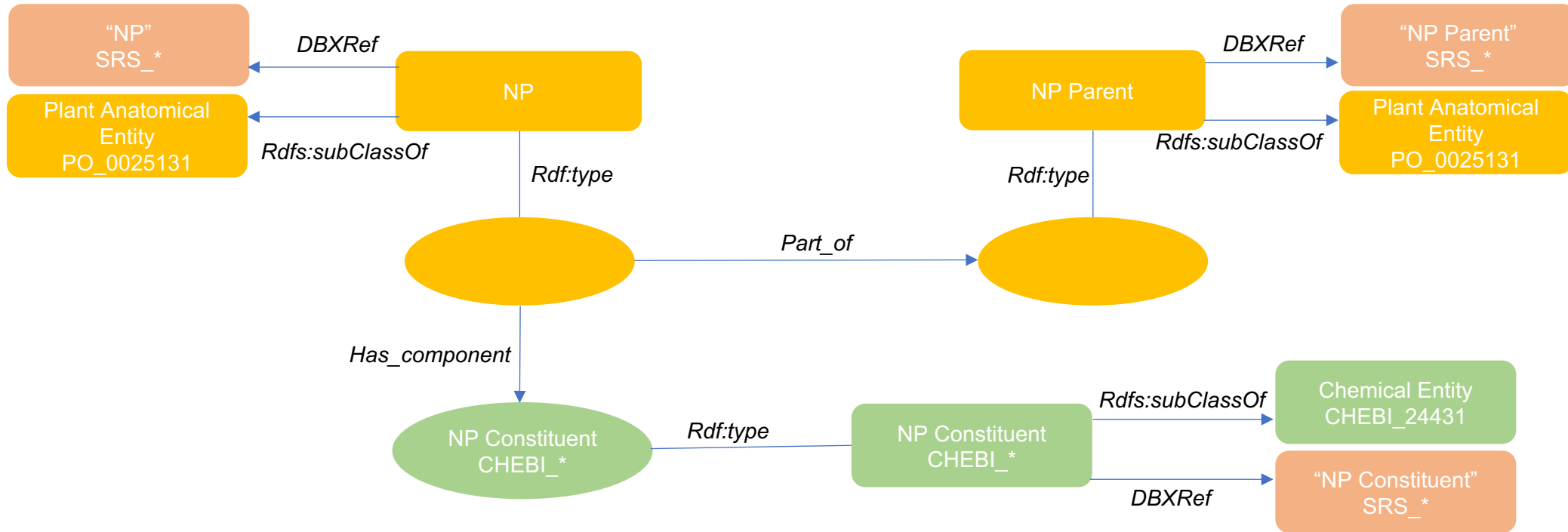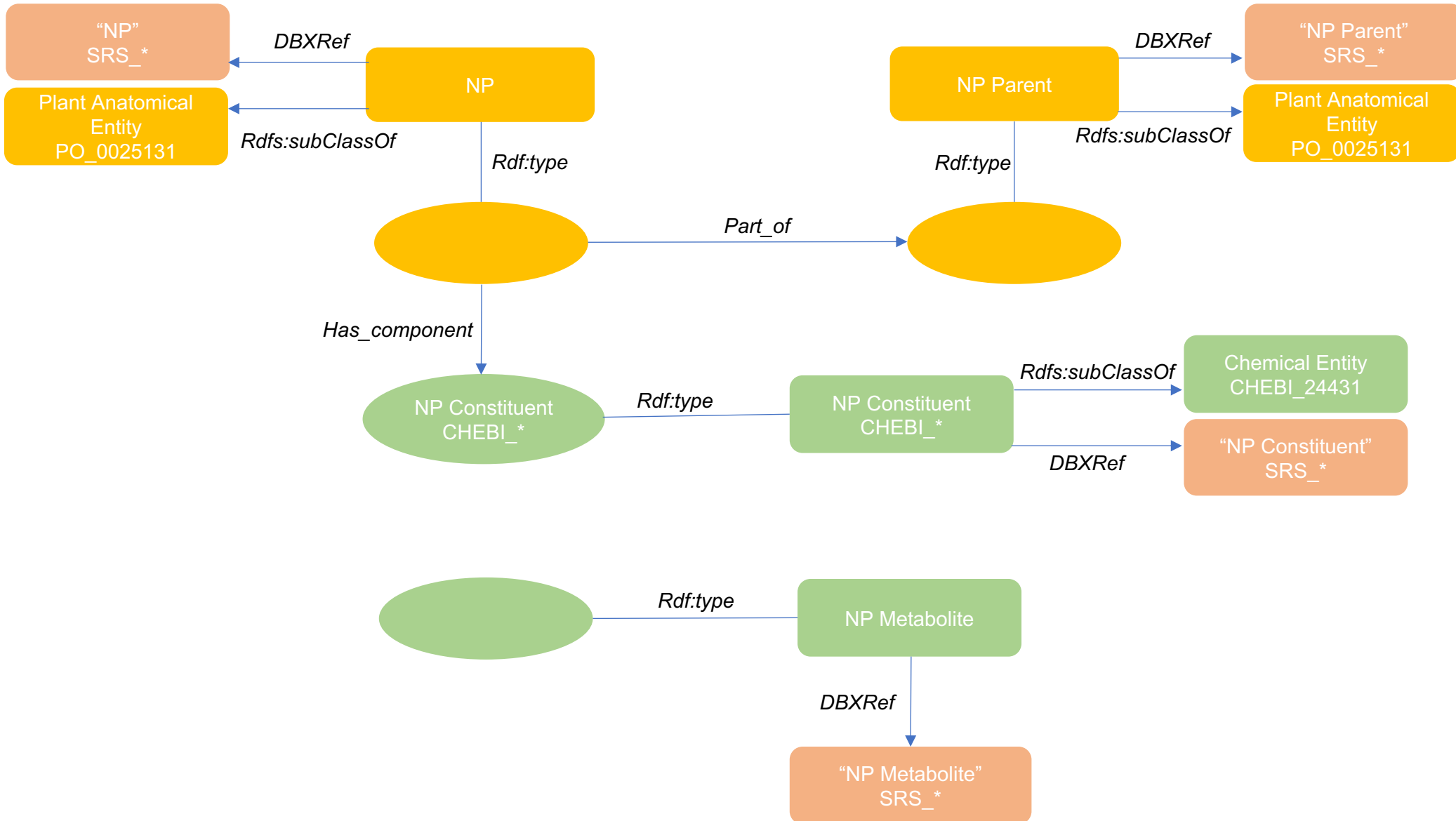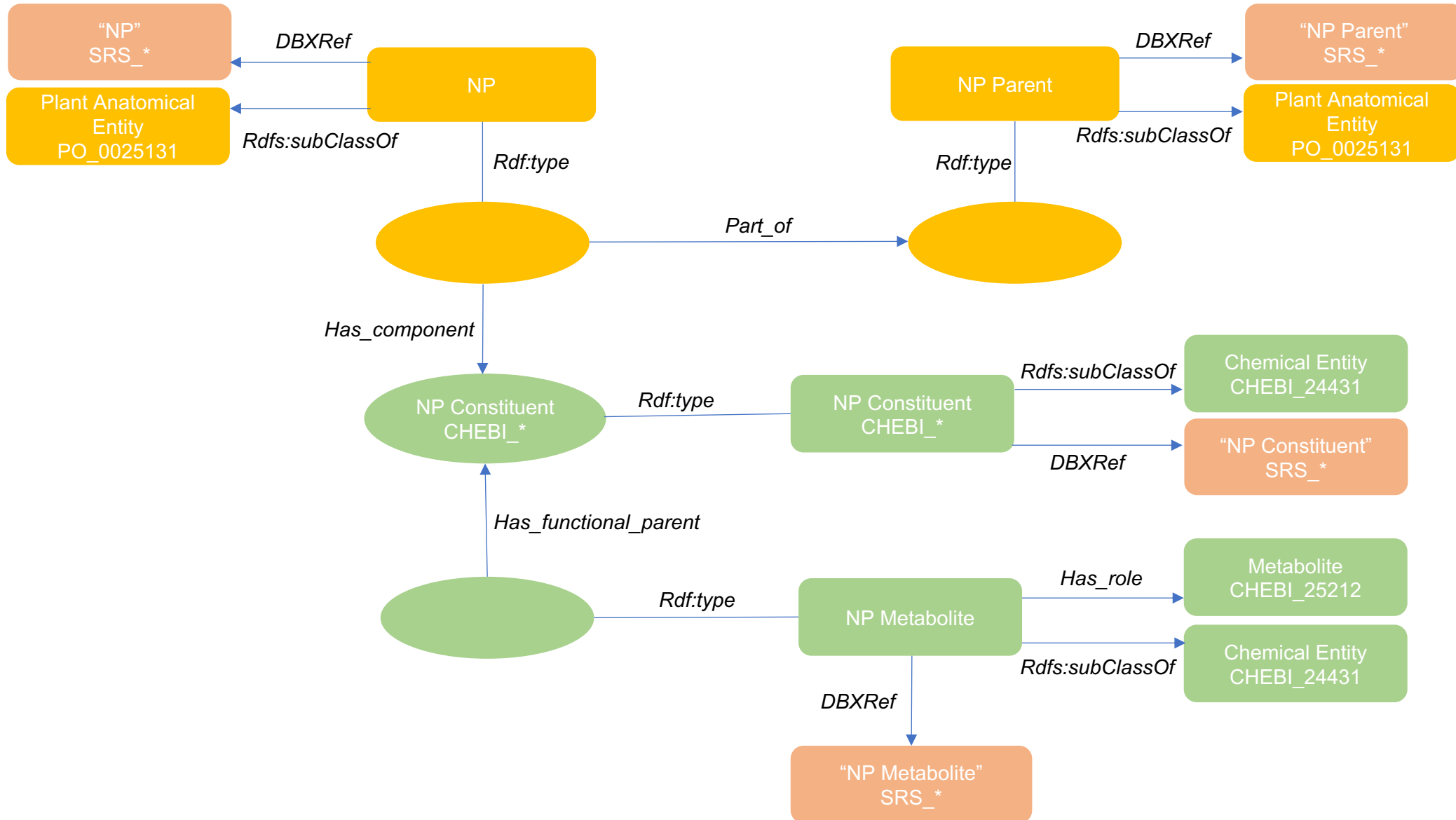
Semantic Representation

Semantic Representation

Semantic Representation

Semantic Representation
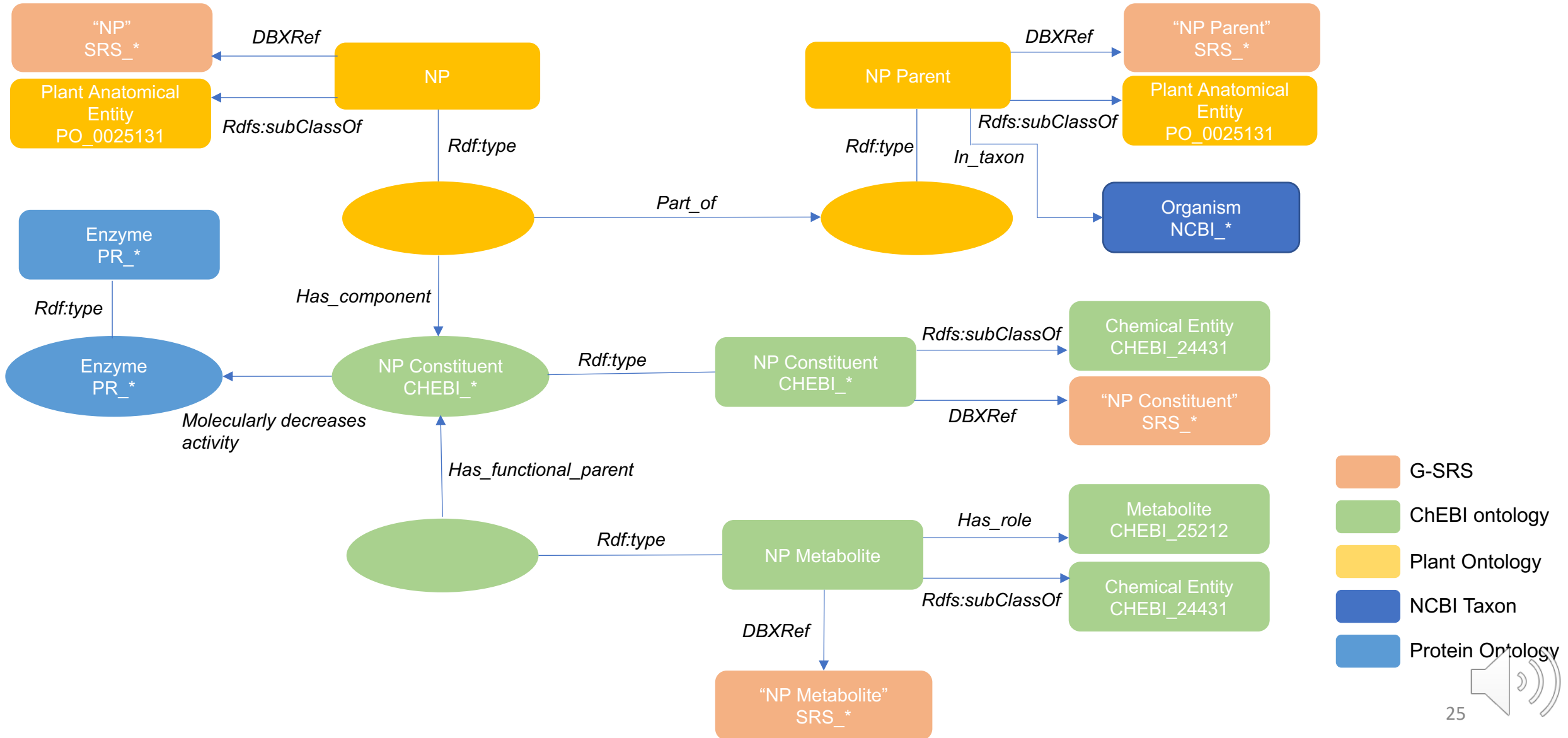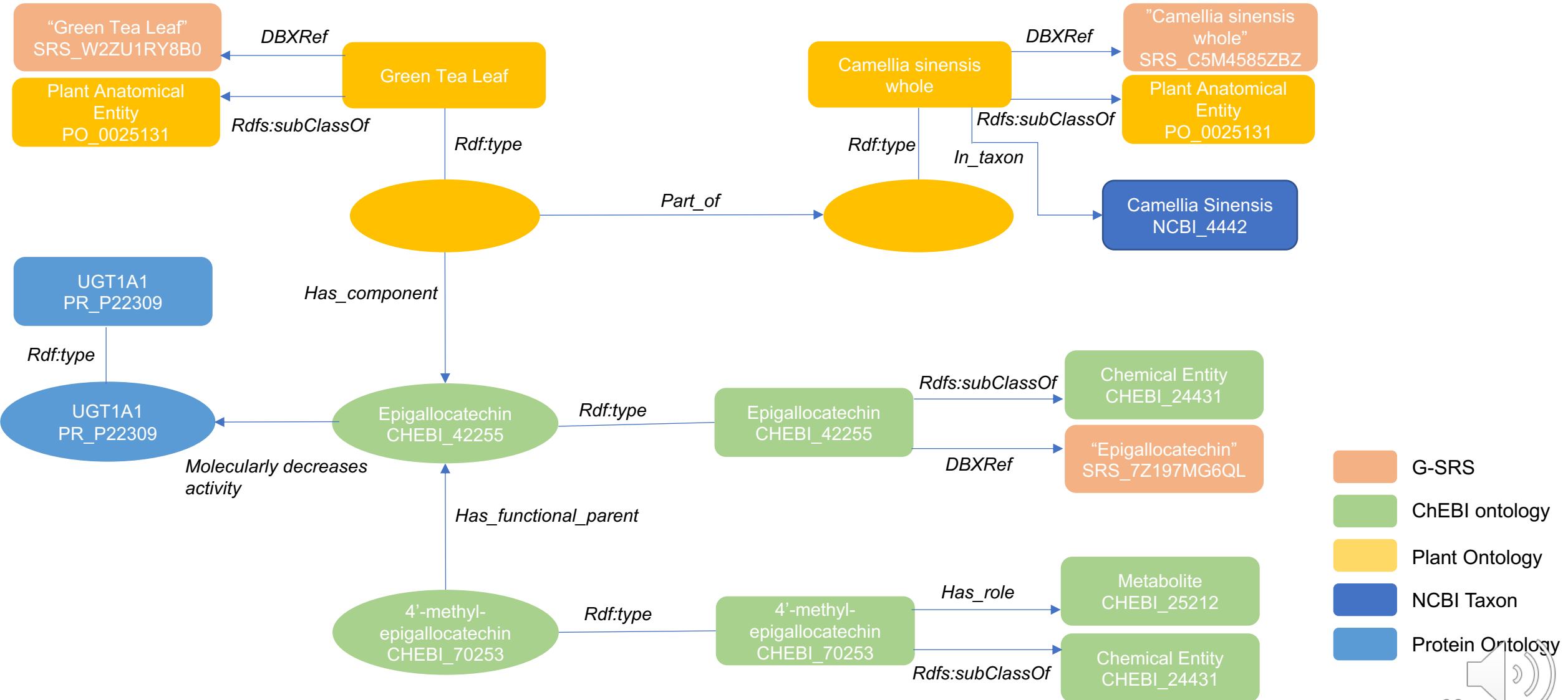
Semantic Representation

Semantic Representation

Semantic Representation

25

Semantic Representation – Green Tea

Workflow

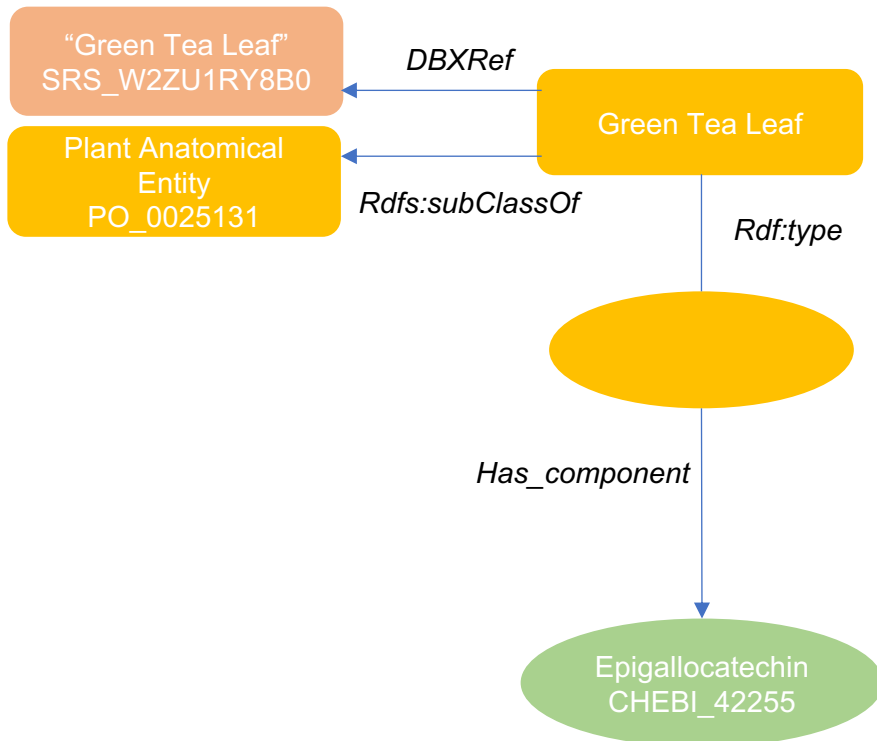Data Extraction → Semantic Representations → Logical Statements → Enhanced Ontology

Logical Statements

- Translated semantic representation patterns to logical statements in Web Ontology Language (OWL).

- Serialized statements using RDFLib library in Python (version 3) to RDF/XML triples.

- Blank nodes to represent existential variables.


```
BNode1 OBO:database_cross_reference SRS:Green Tea
Leaf

BNode1 RDFS:subClassOf PO:Plant Anatomical Entity

BNode1 rdf:type UUID2

BNode2 RO:has_component CHEBI:Epigallocatechin
```

Enhanced ChEBI Ontology

ChEBI (lite) ontology
- Classes = 156,098
- Individuals = 10
- Axioms = 1,200,981

Enhanced ChEBI ontology with kratom and green tea
- Classes = 156,113
- Individuals = 13
- Axioms = 1,201,077

- Merged triples with ChEBI (lite) ontology using OWLTools.

- Applied Hermit Reasoner (1.4.3.456) to verify logical consistency and infer novel axioms in the ontology.

Future work and work in progress

- Generate patterns and extract data in an automated manner for comprehensive natural product inclusion in the ontology.

- Differentiate between in vitro and clinical pharmacokinetic results.

- Create evidence strategies for knowledge using artifacts from the Drug-Drug Interaction and Evidence Ontology (DIDEO) and the Scientific Evidence and Provenance Information Ontology (SEPIO).

- Include the enhanced ontology in large-scale biomedical knowledge graph combined with machine reading for hypothesis generation for NPDIs.

- Code and additional representations at https://github.com/dbmi-pitt/NaPDI-pv/tree/master/ontology_map.

Thank you!

Richard Boyce, PhD, University of Pittsburgh
Tiffany Callahan, MPH, University of Colorado Anschutz Medical Campus
Mathias Brochhausen, PhD, University of Arkansas for Medical Sciences
Mary Paine, PhD, Washington State University
Sandra Kane-Gill, PharmD, University of Pittsburgh

Discussion