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

Sanya B. Taneja[*1], Tiffany J. Callahan[2], Mathias Brochhausen[3], Mary F. Paine[4,5], Sandra L. Kane-Gill[1,5], Richard D. Boyce[1,5]*

[1]University of Pittsburgh, Pittsburgh, Pennsylvania; [2]University of Colorado Anschutz Medical Campus, Aurora, Colorado; [3]University of Arkansas for Medical Sciences, Little Rock, Arkansas; [4]Washington State University, Spokane, Washington; [5]Center of Excellence for Natural Product Drug Interaction Research, Spokane, Washington

ABSTRACT

Botanical and other natural products (NPs) are not as widely represented in biomedical ontologies compared to conventional drugs. The growing use of NPs that have been implicated in clinically significant pharmacokinetic NP–drug interactions (NPDIs) renders addressing this knowledge gap imperative. In this study, we designed potential logical extensions to the Chemical Entities of Biological Interest (ChEBI) ontology that map information about NPs and NP constituents from the Global Substance Registration System (G-SRS). We extracted information from the G-SRS database using SQL; created semantically consistent logical representations for the case NPs - kratom, goldenseal, and green tea; and integrated them within the ChEBI ontology. The merged ontology contains NP information in computable form and is compatible with the principles of the Open Biomedical Ontologies Foundry. The potential logical extensions are the first step in advancing research related to NPDIs using biomedical ontologies and knowledge graphs.

1 INTRODUCTION

Complementary health approaches involving the use of botanical and other natural products (NPs) are increasing in popularity in the US, with records indicating regular consumption by up to 18% of adults (Clarke et al., 2015). Although there is myriad information about drug ingredients in US Food and Drug Administration (FDA)-approved drugs, NPs are generally absent from common drug databases (Drug Bank, Drug Central) and biomedical ontologies. These drug databases and ontologies, such as the Drug-Drug Interaction and Evidence Ontology (DIDEO) and the Chemical Entities of Biological Interest (ChEBI), are frequently used as sources of gold standard information for computational research on drug-drug interactions and drug repurposing. However, the lack of information about NPs in these sources precludes the ability to computationally investigate pharmacokinetic NP-drug interactions (NPDIs) with rigor. Pharmacokinetic NPDIs occur when a NP alters the absorption, distribution, metabolism, and/or excretion of a co-consumed object drug, potentially resulting in reduced treatment efficacy or adverse events (Paine et al., 2018). Among the 40 top-selling herbal medicinal products in the US, ten have been associated with clinically significant pharmacokinetic NPDIs (Spanakis et al., 2019). Computational investigation of NPDIs involving NPs first requires integration of logical representations with existing biomedical resources.

The FDA Global Substance Registration System (G-SRS) developed by the FDA and National Center for Advancing Translational Sciences, is used for registering and documenting information about substances present in medicines (Peryea et al., 2021). The G-SRS classifies NPs as structurally diverse substances and further defines the part of the organism from which the NP was prepared. The wealth of information available in the G-SRS about NPs, NP constituents, metabolites, and external references to common NP databases makes the G-SRS an ideal source for including information about NPs in the ChEBI ontology.

In this study, we used the NPs kratom, goldenseal, and green tea as cases to design potential logical extensions to map NPs to the ChEBI ontology with a goal to advance computational research on NPDIs.

2 METHODS

We used a three-step approach to design the logical extensions from the G-SRS to the ChEBI ontology. First, we identified information about NPs, NP constituents, and metabolites in the G-SRS to be included in the ChEBI ontology. Second, we created semantic representations for each NP and identified patterns in the representations. Third, we translated the representations to logical statements that were merged into the ChEBI ontology.

To identify NPs in the G-SRS, we matched the Latin binomial names (e.g., *Mitragyna speciosa*) to the substances in the G-SRS. To generate semantic representations of each NP, we then extracted information about the parent substance of the NP, constituents, metabolites, and salts of each NP.

ChEBI is an ontology of chemical compounds that is part of the OBO Foundry (Hastings et al., 2016). ChEBI contains

important information about the chemical characteristics of NP constituents such as mitragynine (in kratom), berberine and hydrastine (in goldenseal), and catechins (in green tea). ChEBI also contains information regarding the biological and chemical roles of these compounds. However, gaps exist in knowledge related to interactions of the various constituents, missing metabolites and salts, and parent NP substances. For instance, the chemical characteristics of mitragynine can be found in the ChEBI ontology. ChEBI does not include the metabolites of kratom (such as 7-hydroxy-mitragynine) or information about its pharmacokinetic interactions (inhibition of the cytochrome P450 enzymes in vitro). We used the NP constituents in ChEBI (when present) and roles to map substances from the G-SRS. We are also working to include information about pharmacokinetic NPDIs including chemical characterization of NPs, metabolomics analyses, and in vitro and clinical pharmacokinetic experimental results from the Center of Excellence for NPDI Research data repository (Birer-Williams et al., 2020). Additionally, we used entities from the Protein Ontology (PRO) and Gene Ontology (GO) to semantically represent pharmacokinetic information about NPs and their constituents and the National Center for Biotechnology Information Taxonomy (NCBITaxon) to identify the organism related to each NP.

2.1 Semantic Representations

Figure 1 shows the semantic representation patterns for potential extensions of NPs and NP constituents from G-SRS to the ChEBI ontology. Our basic approach was to add novel classes for the NP sources and constituents. We then used the relations between the classes to represent relationships that would apply to all instances of the classes. We used relations from the relation ontology (RO) *(has component, has role, in taxon)*, basic formal ontology (BFO) *(part of)*, and ChEBI *(has functional parent)* to link the new classes in the ontology. To incorporate other external references or database cross-references (DBXRef) to substances in the G-SRS, we used the relation DBXRef (http://www.geneontology.org/formats/oboInOwl#DbXref). The blank nodes (BNode) in Figure 1 represent existential variables.

2.2 Merging Ontologies

We queried data from the G-SRS using SQL. Then, we wrote a Python program to translate the semantic representations described above to logical statements in the Web Ontology Language (OWL). We merged the triples with the ChEBI (Lite) ontology using Protégé (5.5.0) (Musen, 2015), serialized the statements to triples in the RDF/XML format, and applied the Hermit Reasoner (1.4.3.456) (Glimm et al., 2014) to verify the logical consistency of the novel classes as well as to infer novel ontology axioms.

3 RESULTS

We mapped information about three NPs – kratom, goldenseal, and green tea – from the G-SRS to the ChEBI ontology. The ChEBI Lite ontology consisted of 290,466 logical axioms and 156,098 classes. The merged ontology contained 291,019 logical axioms, 156,335 classes, and 40 individuals. Examples of novel triples in the ontology include the following -

- BNode1 DBXRef (http://www.geneontology.org/formats/oboInOwl#DbXref) Mitragyna speciosa leaf (http://gsrs.ncats.nih.gov/ginas/app/substance/dac1ac7a-f1bb-42d7-ab9c-0bf06d0d9825)
- BNode1 has component (http://purl.obolibrary.org/obo/RO_0002180) Mitragynine (http://purl.obolibrary.org/obo/CHEBI_6956)
- BNode1 in taxon (http://purl.obolibrary.org/obo/RO_0002162) Mitragyna peciosa (http://purl.obolibrary.org/obo/NCBITaxon_170351)
- Mitragynine (http://purl.obolibrary.org/obo/CHEBI_6956) subClassOf (http://www.w3.org/2000/01/rdf-schema#subClassOf) Chemical Entity (http://purl.obolibrary.org/obo/CHEBI_24431)

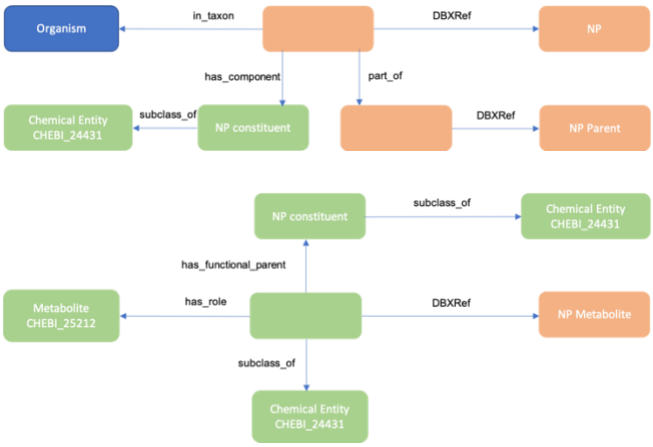

Figure 1. Semantic representation patterns for NPs.

4 DISCUSSION

This study is the first to design potential logical extensions to the ChEBI ontology to include information about NPs from the G-SRS. Prior work in this domain has focused on extending DIDEO to include information artifacts related to NPDIs (Judkins et al., 2018). The Pharmacognosy Ontology aims to revive the Natural Products Alert database and connect it to existing resources in the domain such as data on

bioactivity (Bisson et al., n.d.). Our work differs by introducing classes to include information about specific NPs, NP constituents, and metabolites within the OBO Foundry. Our overall goal is to use the enhanced ChEBI ontology in a larger, heterogeneous knowledge graph to generate hypotheses for pharmacokinetic NPDIs. We are progressing toward our goal of including all NPs with potential pharmacokinetic NPDIs available in the G-SRS in the ontology. In future work, we will integrate the ontology with classes and definitions in DIDEO to track evidence sources for potential NPDIs. Further, we will refine our enhanced ontology by including instances of each NP class in order to represent NP-specific facts. The code and additional examples of representations are available from https://github.com/dbmi-pitt/NaPDI-pv/tree/master/ontology_map.

5 ACKNOWLEDGMENTS

This work was funded by the National Institutes of Health National Center for Complementary and Integrative Health [Grant U54 AT008909].

REFERENCES

Birer-Williams, C., Gufford, B. T., Chou, E., Alilio, M., VanAlstine, S., Morley, R. E., McCune, J. S., Paine, M. F., & Boyce, R. D. (2020). A New Data Repository for Pharmacokinetic Natural Product-Drug Interactions: From Chemical Characterization to Clinical Studies. *Drug Metabolism and Disposition*, *48*(10), 1104–1112. https://doi.org/10.1124/dmd.120.000054

Bisson, J., McAlpine, J., Graham, J., & Pauli, G. F. (n.d.). *NAPRALERT, from an historical information silo to a linked resource able to address the new challenges in Natural Products Chemistry and Pharmacognosy*. 2.

Clarke, T. C., Black, L. I., Stussman, B. J., Barnes, P. M., & Nahin, R. L. (2015). Trends in the Use of Complementary Health Approaches Among Adults: United States, 2002–2012. *National Health Statistics Reports*, *79*, 1–16.

Glimm, B., Horrocks, I., Motik, B., Stoilos, G., & Wang, Z. (2014). HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning*, *53*(3), 245–269. https://doi.org/10.1007/s10817-014-9305-1

Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P., & Steinbeck, C. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Research*, *44*(D1), D1214-1219. https://doi.org/10.1093/nar/gkv1031

Judkins, J., Tay-Sontheimer, J., Boyce, R. D., & Brochhausen, M. (2018). Extending the DIDEO ontology to include entities from the natural product drug interaction domain of discourse. *Journal of Biomedical Semantics*, *9*(1), 15. https://doi.org/10.1186/s13326-018-0183-z

Musen, M. A. (2015). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, *1*(4), 4–12. https://doi.org/10.1145/2757001.2757003

Paine, M. F., Shen, D. D., & McCune, J. S. (2018). Recommended Approaches for Pharmacokinetic Natural Product-Drug Interaction Research: A NaPDI Center Commentary. *Drug Metabolism and Disposition*, *46*(7), 1041–1045. https://doi.org/10.1124/dmd.117.079962

Peryea, T., Southall, N., Miller, M., Katzel, D., Anderson, N., Neyra, J., Stemann, S., Nguyễn, Đ.-T., Amugoda, D., Newatia, A., Ghazzaoui, R., Johanson, E., Diederik, H., Callahan, L., & Switzer, F. (2021). Global Substance Registration System: Consistent scientific descriptions for substances related to health. *Nucleic Acids Research*, *49*(D1), D1179–D1185. https://doi.org/10.1093/nar/gkaa962

Spanakis, M., Sfakianakis, S., Sakkalis, V., & Spanakis, E. G. (2019). PharmActa: Empowering Patients to Avoid Clinical Significant Drug–Herb Interactions. *Medicines*, *6*(1). https://doi.org/10.3390/medicines6010026