# Extension of Roles in the ChEBI Ontology

*This manuscript (permalink) was automatically generated from chemical-roles/manuscript@63a34cd on June 30, 2020.*

## Authors


- **Charles Tapley Hoyt**
  0000-0003-4423-4370 · cthoyt · cthoyt
  Enveda Therapeutics

- **Christopher J Mungall**
  0000-0002-6601-2165 · cmungall
  Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

- **Nicole Vasilevsky**
  0000-0001-5208-3432 · nicolevasilevsky

- **Daniel Domingo-Fernández**
  0000-0002-2046-6145 · ddomingof · daniel_sunday
  Enveda Therapeutics

- **Matthew Healy**
  0000-0001-6439-5038
  Enveda Therapeutics

- **Viswa Colluru**
  0000-0002-6208-9288
  Enveda Therapeutics

## Motivation

The Chemical Entities of Biological Interest [1,2] ontology models chemicals, their classes, their roles, and their interrelations (Figure 1, left). While many roles correspond to how their substituent chemicals affect proteins and other biological entities (see Figure 3), this information is not formalized nor structured in the ChEBI ontology. Yan *et al.* (2011) previously described how these correspondences could be theoretically formalized [3]. This article proposes a concrete schema and axioms through which these roles can be linked to their target entities (Figure 1, right), a suite of open source, reusable curation tools, and ultimately a manually curated database of relationships between chemical roles and their targets.

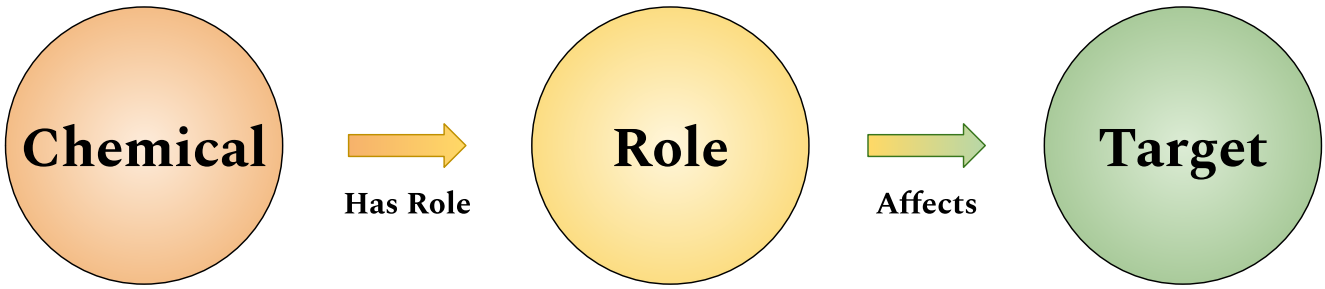


**Figure 1:** Schema for inference of chemicals' relations to targets via roles. Targets may be other chemicals, proteins, protein families, protein complexes, pathways, pathologies, or organisms.

Throughout this article, the term *role* (in the context of the ChEBI ontology) will be used in the colloquial sense described by Batchelor et al. (2010) [4] rather than the formal sense prescribed by the Basic Formal Ontology [5,6].

This article has been typeset with Manubot 7. Along with each named entity, we are piloting Manubot's new feature to link to entities via Identifiers.org compact URIs (CURIEs).

## Extension of the ChEBI Schema

Before generalizing the rules for the relationships between targets, it is best to examine a specific example. Chemicals that have the role *p53 activator* [8] obviously affect a certain target. The type of effect is directly activates [9]. The type of the target is protein. The target itself is the protein encoded by the *TP53* [10] gene. Because the chemical CBL0137 [11] has the role *p53 activator*, it can be inferred that *CBL0137 directly activates TP53*.

While Yan *et al.* exemplified the formalization of the relationship between roles from the ChEBI ontology and their targets using the Manchester syntax, this article will only informally describe axioms. It is intended that these descriptions are easily understandable such that the focus can remain on the practical utility of the relations curated and inferred during the work described here. The following informal amalgamation of a shape expression (common to graph query languages like SPARQL) and propositional logic outlines how to infer if a chemical C is an activator of protein P given it has the role R and role R is related to the activation of protein P.

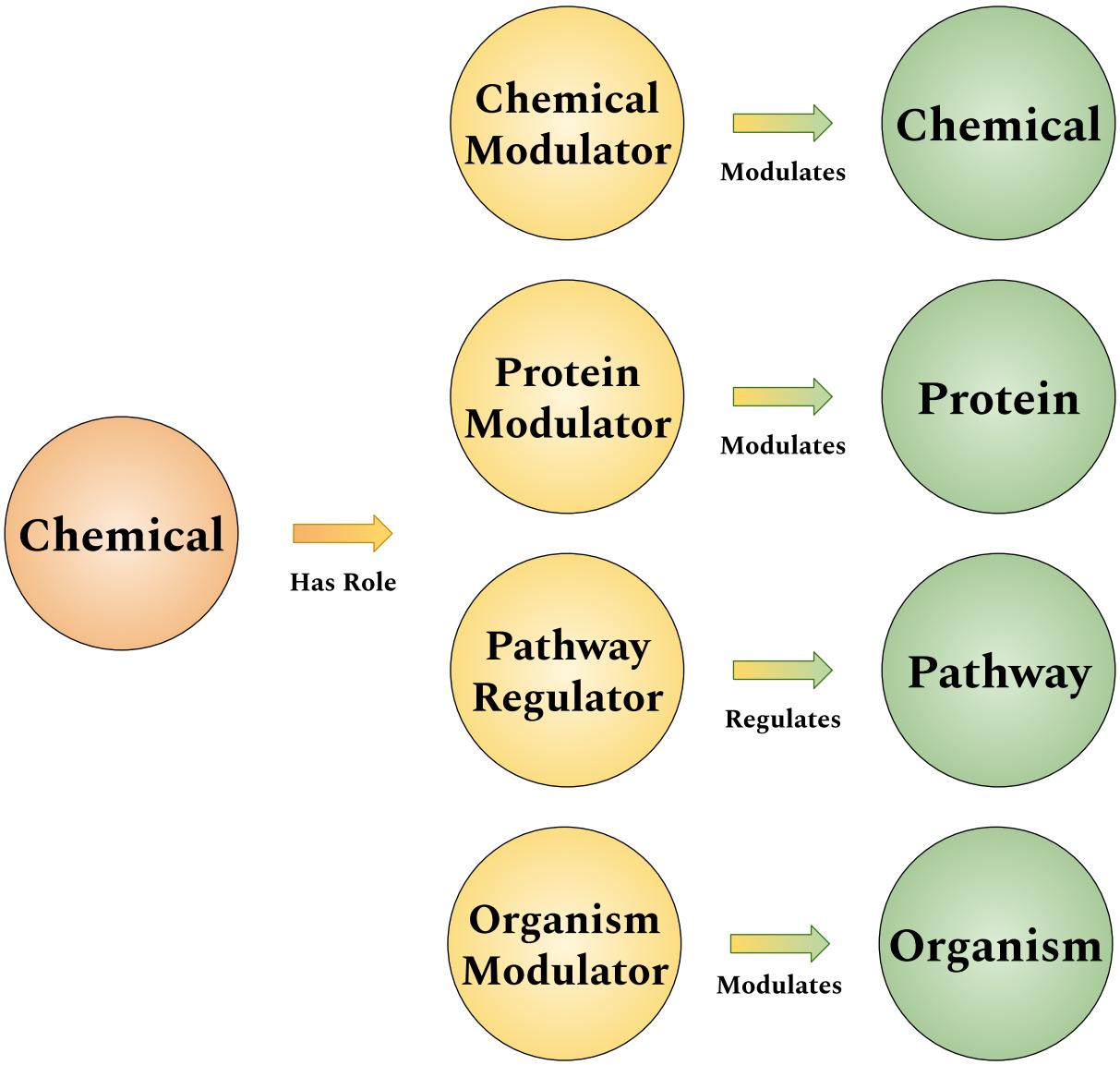


**Figure 2:** Generalization of the ChEBI schema to include relationships between several categories of roles and their target entities.

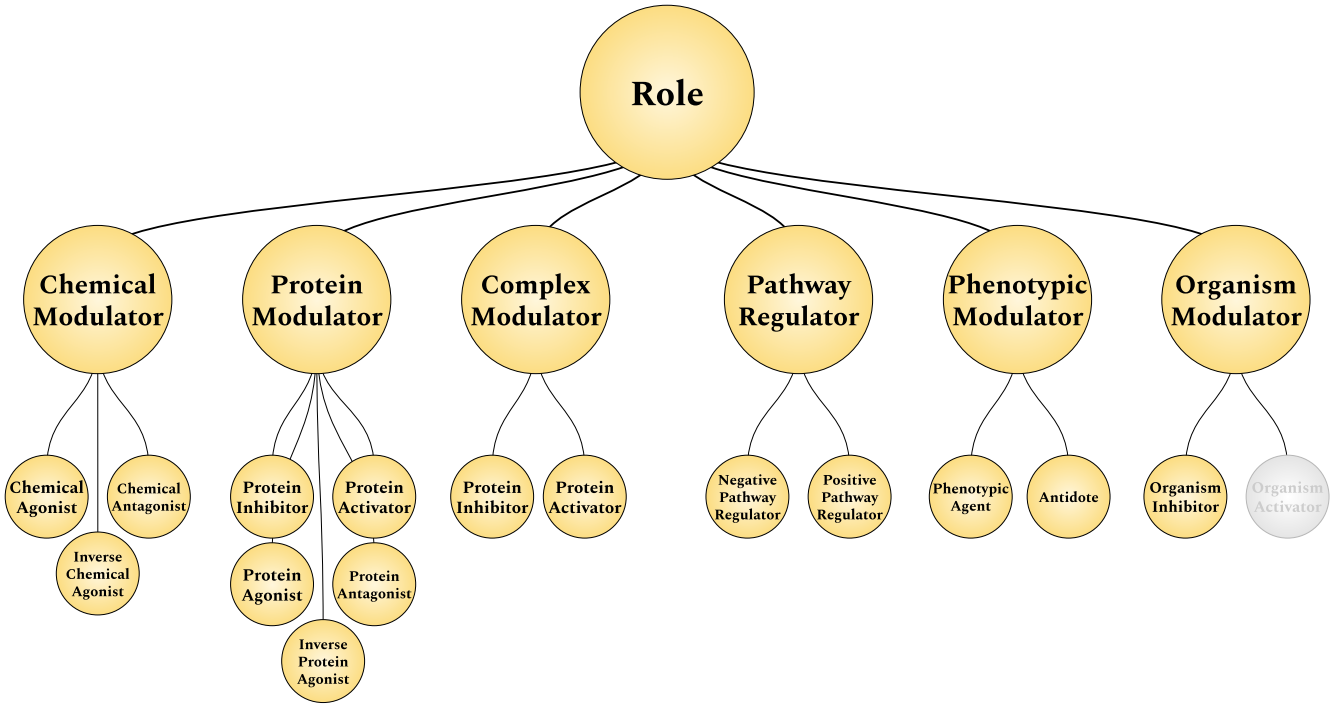


**Figure 3:** A hierarchy of chemical roles. Each role corresponds to a chemical having a given relationship. Within the ChEBI hierarchy, the role (CHEBI:50906) has three children: chemical role (CHEBI:51086), biological role (CHEBI:24432), and application (CHEBI:33232). Our interest lies with the latter two.

# Curation Methods

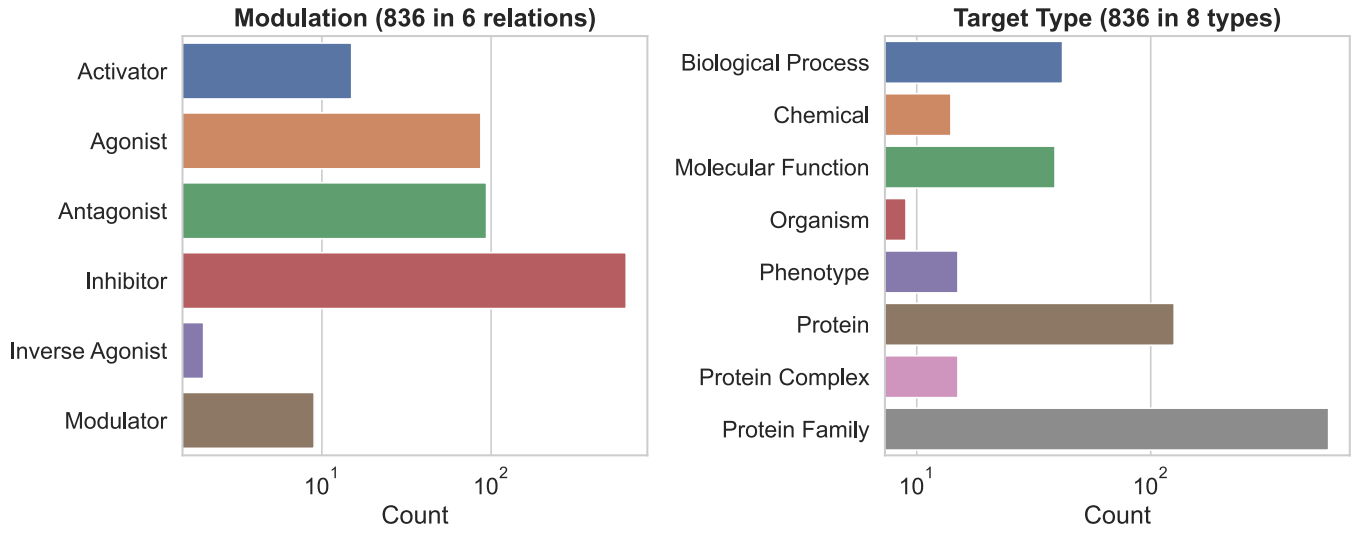


**Figure 4:** Summaries over the types of modulations and target entity types covered by the database so far.

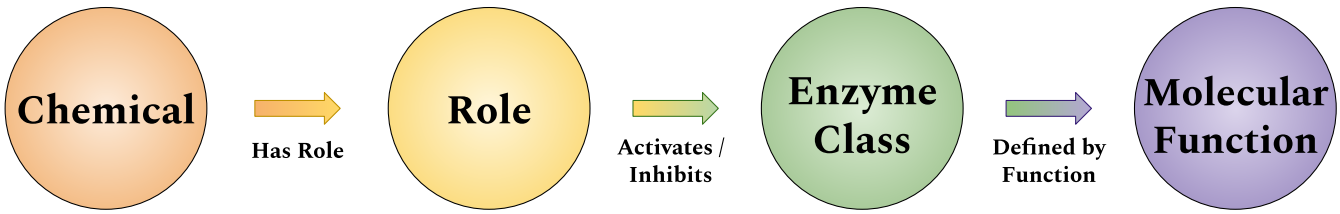


**Figure 5:** Schema for inference of chemicals' relations to molecular functions via roles and enzyme classes

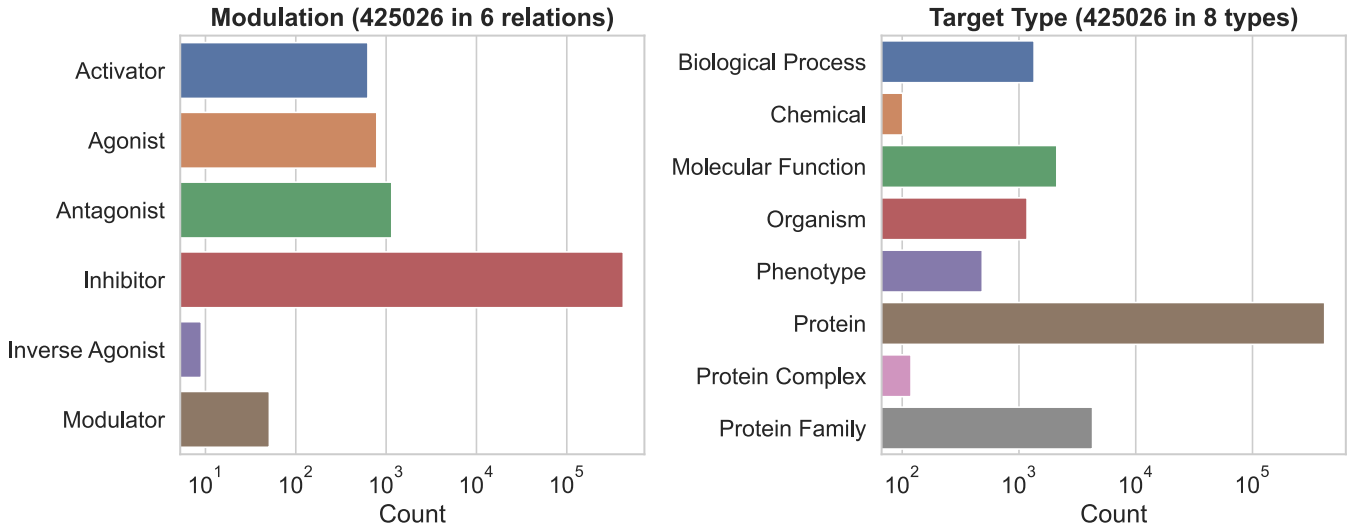


**Figure 6:** Summaries over the types of modulations and target entity types in inferred relationships. Inference can be optionally extended over role hierarchies, which produces millions of relations.

# Interoperability

# Concluding Remarks

# References


1. **The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013.**
   Janna Hastings, Paula de Matos, Adriano Dekker, Marcus Ennis, Bhavana Harsha, Namrata Kale, Venkatesh Muthukrishnan, Gareth Owen, Steve Turner, Mark Williams, Christoph Steinbeck
   *Nucleic acids research* (2012-11-24) https://www.ncbi.nlm.nih.gov/pubmed/23180789
   DOI: 10.1093/nar/gks1146 · PMID: 23180789 · PMCID: PMC3531142

2. **ChEBI in 2016: Improved services and an expanding collection of metabolites.**
   Janna Hastings, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, Christoph Steinbeck
   *Nucleic acids research* (2015-10-13) https://www.ncbi.nlm.nih.gov/pubmed/26467479
   DOI: 10.1093/nar/gkv1031 · PMID: 26467479 · PMCID: PMC4702775

3. **Use of multiple ontologies to characterize the bioactivity of small molecules**
   Yan Ying, Janna Hastings, Jee-Hyub Kim, Stefan Schulz, Christoph Steinbeck, Dietrich Rebholz-Schuhmann
   http://ceur-ws.org/Vol-833/paper68.pdf

4. **Formal ontology in information systems: proceedings of the Sixth International Conference (Fois 2010)**
   Antony Galton, Riichiro Mizoguchi (editors)
   *IOS Press* (2010)
   ISBN: 9781607505341

5. **Functions in Basic Formal Ontology**
   Andrew D. Spear, Werner Ceusters, Barry Smith
   *Applied Ontology* (2016-06-22) https://doi.org/f8z4js
   DOI: 10.3233/ao-160164

6. **Structured Descriptions of Roles, Activities,and Procedures in the Roman Constitution**
   Yoonmi Chu, Robert B. Allen
   *arXiv* (2015-02-16) https://arxiv.org/abs/1502.04108

7. **Open collaborative writing with Manubot**
   Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter
   *PLOS Computational Biology* (2019-06-24) https://doi.org/c7np
   DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653

8. **p53 activator (CHEBI:77731)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:77731

9. http://purl.obolibrary.org/obo/RO_0002406

10. https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/11998

11. **CBL0137 (CHEBI:138650)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:138650