Extension of Roles in the ChEBI Ontology

Authors

- **Charles Tapley Hoyt**
  ⓘD [0000-0003-4423-4370](#) · ◯ [cthoyt](#) · 🐦 [cthoyt](#)
  Enveda Therapeutics

- **Christopher J Mungall**
  ⓘD [0000-0002-6601-2165](#) · ◯ [cmungall](#) · 🐦 [chrismungall](#)
  Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

- **Nicole Vasilevsky**
  ⓘD [0000-0001-5208-3432](#) · ◯ [nicolevasilevsky](#) · 🐦 [N_Vasilevsky](#)

- **Daniel Domingo-Fernández**
  ⓘD [0000-0002-2046-6145](#) · ◯ [ddomingof](#) · 🐦 [daniel_sunday](#)
  Enveda Therapeutics

- **Matthew Healy**
  ⓘD [0000-0001-6439-5038](#)
  Enveda Therapeutics

- **Viswa Colluru**
  ⓘD [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. Ying *et al.* [3] previously described how these correspondences could be theoretically formalized. 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.
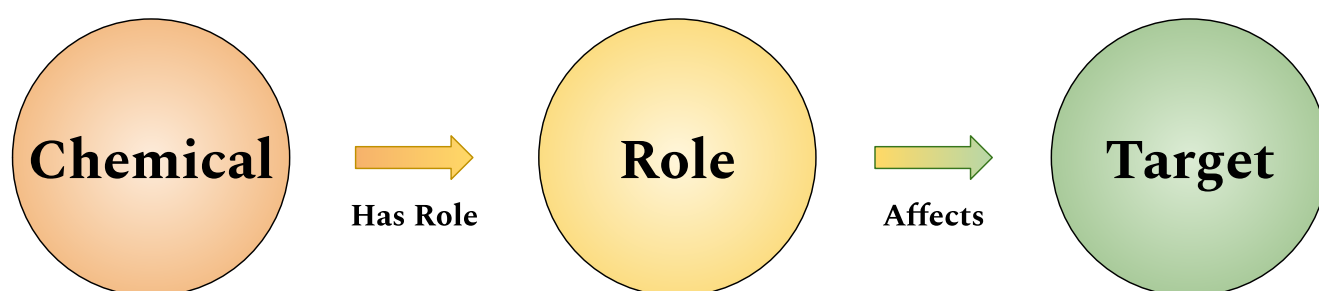

Figure 1: Schema for inference of chemicals' relations to targets via roles. Targets may be other chemicals, proteins, protein families, protein complexes, pathways, pathologies, or organisms.

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.* [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 [10]. The target itself is the protein encoded by the TP53 [11] gene. Because the chemical CBL0137 [12] has the role p53 activator [8], it can be inferred that CBL0137 [12] directly activates TP53 [11].

While Ying *et al.* [3] exemplified the formalization of the relationship between roles from the ChEBI ontology and their targets using the Manchester syntax to convey OWL-DL, 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 predicate 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**.

Table 1: Shape expression and predicate logic rules for inferring chemicals that activate proteins linked to a given role (could be stated more concisely in OWL as *has_role o role_for_directly_activates -> directly_activates*).

Antecedent	Consequent

	Antecedent	Consequent
ChEBI	**C** a Chemical **R** a Role **C** has role **R**	
Curated	**P** a Protein **R** role for directly activates **P**	**C** directly activates **P**

The ChEBI ontology can be used to determine if chemical **C** has the role **R**, but not whether **R** is related to protein **P**, or any other protein for that matter. This motivates the generation of a database of relations between ChEBI roles that activate proteins to support this type of inference. After generalizing to other relationships and entity types, we generated such a database and included it with this article.

However, there are ChEBI roles that imply relations to proteins besides directly activates, so this definition needs to be generalized. First, it should be noted that the relationship between role **R** and protein **P**, "role for directly activates" was not taken from an ontology. It was named based on the relation between chemical **C** and protein **P** that role **R** implies. In practice, this means that "role for" is added as a prefix. In this example, role **R** that corresponds to the relation "directly activates" between chemical C and protein **P** is called "role for directly activates". A database of relations between ChEBI roles and proteins should include the directly activates relation, but also others such as the directly inhibits [13], agonist of, antagonist of, and inverse agonist of relations[1]. The shape expression and predicate logic can be updated to reflect the relation m between role **R** and protein **P** that corresponds to the externally defined relation **M** between chemical **C** and protein **P**.

Shape expression and predicate logic rules generalized for other relationships from Table . {#tbl:table-2}

	Antecedent	Consequent
ChEBI	**C** a Chemical **R** a Role **C** has role **R**	
Curated	**P** a Protein **R m P** **m** correspondsTo **M**	**C M P**

Again, since **m** is dependent on external relationship **M**, it is acceptable to use relationships **M** during curation to mitigate the need for additional ontologizing. Therefore, an entry in the database could contain the role, relationship, and protein as shown in Table 2.

Table 2: An example database entry.

Role	Relationship	Protein
CHEBI:77731	RO:0002406	HGNC:11998

Further, there are ChEBI roles that imply relations to entities other than proteins, so this definition must be generalized. Rather than discussing relationships between role **R** and protein **P**, it would suffice to discuss relationships between role **R** and target **T**. The relationships themselves that are appropriate should already have well-defined semantics that dictate what types of entities are valid for the range. This article will use externally defined relationships from sources like the Relation Ontology (RO; [14]) when possible and avoid discussion of these semantics. In many cases, roles **R** can be categorized based on the types of targets **T** that they affect (Figure {2).
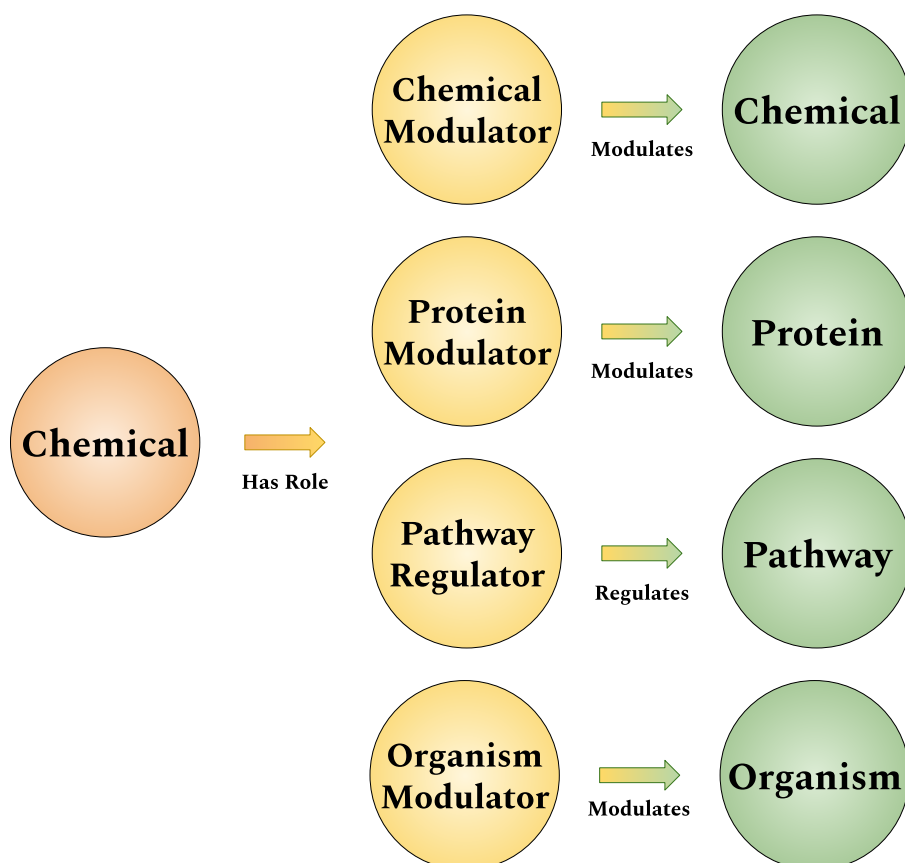
Figure 2: Generalization of the ChEBI schema to include relationships between several categories of roles and their target entities.

In many cases, it is unclear what the type of entity a given ChEBI role implies. Therefore, the database also includes explicit annotation of entity types. This is also helpful when using databases that contain multiple entity types, such as the Gene Ontology (GO) [15], and should serve to dissuade users from inferring entity types based on databases. Now, an entry in the database should contain a role, relation, type, and target as shown in.

Table 3: An example database entry with entry type.

Role	Relationship	Type	Target
CHEBI:77731	RO:0002406	Protein	HGNC:11998

While not strictly necessary for the reasoning described above, it could be advantageous to enrich the existing role hierarchy with further roles that directly correspond to a given relationship and entity type (Figure 3), such that maintenance of children of these terms would only necessitate the curation of the target itself.
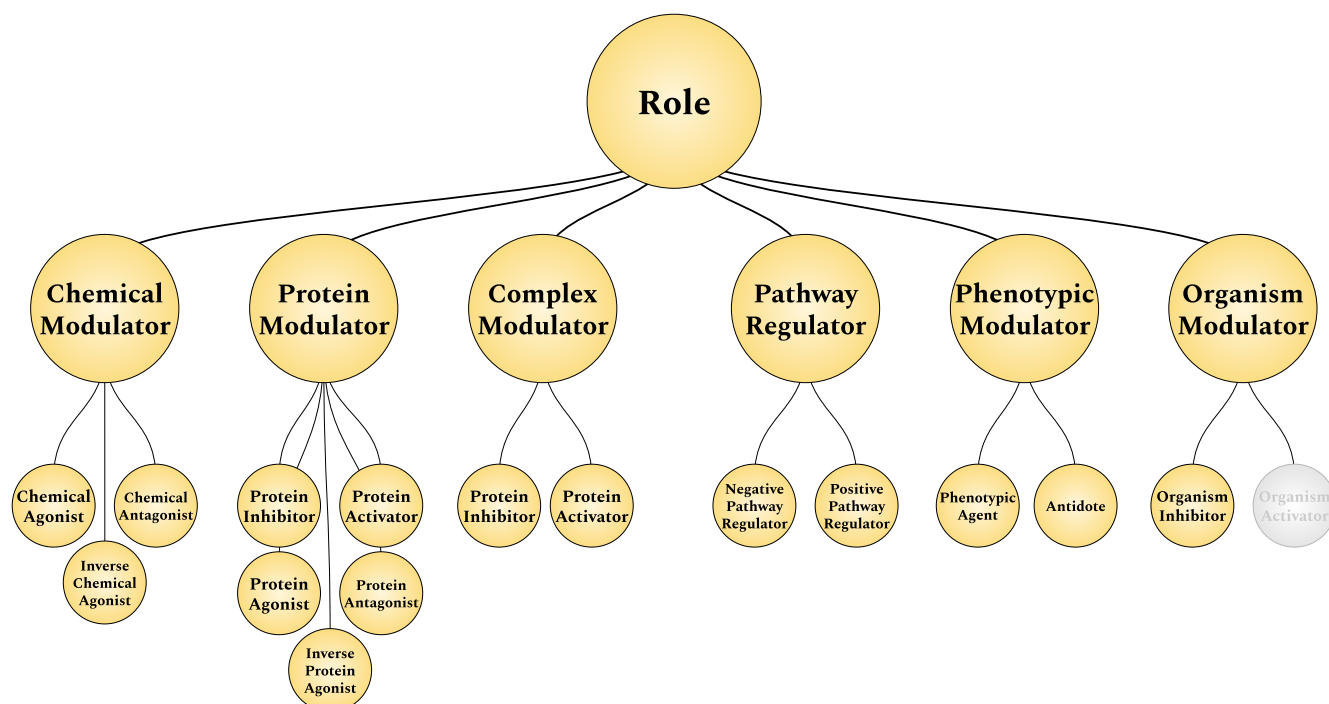
Figure 3: A hierarchy of chemical roles. Each role corresponds to a chemical having a given relationship. Within the ChEBI hierarchy, the role [16] has three children: chemical role [17], biological role [18], and application [19]. Our interest lies with the latter two.

For example, the role enzyme inhibitor [20] clearly corresponds to the entity type being protein family and the relation being directly inhibits [13]. Therefore, descendents such as EC 3.4.24.71 (endothelin-converting enzyme 1) inhibitor [21] can directly be inferred to directly inhibit a protein family. All that would be necessary is to annotate the identity of that family, endothelin-converting enzyme 1 [22]. While this is already well-defined for enzyme inhibitors, it is not for other roles. Figure 3 proposes a possible hierarchy of additional roles corresponding to relations and target types. Note that proteins and protein families are considered equivalent in this context. This article will not fully explore this concept, though during curation, several oddities were identified, such as the fact that pathway inhibitor [23] is a subclass of enzyme inhibitor [20].

Curation Methods

Following a reasonable formalization of the schema for the extension to ChEBI, a simple set of curation guidelines were generated.

1. Identify a ChEBI role such that each has a well-defined relation and target.
2. Ground the relation to a controlled vocabulary like the RO or another carefully annotated one.
3. Identify the target's type. This is often non-trivial, especially in the cases of agonists, antagonists, and inverse agonists. For example, GABA antagonist [24] refers to the family of Gamma-aminobutyric acid (GABA) receptors [25] while it may be mistaken to mean the chemical γ-aminobutyric acid [26]. Alternatively, some roles really do seem to refer to the chemicals themselves such as vitamin B1 antagonist [27].
4. Ground the entity to a controlled vocabulary. Prefer those with easily accessible data such as those available through the OBO Foundry [28]. In the many cases when protein complexes and protein families cannot be grounded to a pre-existing resource, submit additional curation to the FamPlex resource [29] following their curation guidelines.

Below is described how roles were identified in automated or semi-automated fashion and their names were pre-processed for search. In many cases, the Gilda grounding tool [30] was used to prioritize candidate groundings from select databases.

Affix-Based Role Identification

There are a number of naming conventions within ChEBI that were leveraged to quickly identify entities to be curated. Many roles that imply common semantics share common affixes (i.e., prefixes or suffixes). In the case of enzyme classes, the nomenclature is standardized so well that the EC number could be parsed out and relations could be generated automatically. In a few cases, this was not possible, and either additional rules were added to the parser or in several cases issues were filed to the ChEBI tracker [31] to make improvements. Table 5 enumerates several of these affixes at the end of the manuscript.

Adaptation of ChIRO

Previously, the ChEBI Integrated Role Ontology (ChIRO) [32] was generated to axiomize the relationships between select ChEBI roles and other OBO ontologies; primarily the Molecular Function branch of the GO. Many of its ~200 entries connected EC inhibitors and activators to the corresponding molecular function in GO. Additionally, a small number of exemplar axioms were added for other branches of GO and other ontologies, for example, anti-apoptosis utilizes the GO Biological Process apoptotic process [33]. The ChEBI role anti-arrythmia drug [34] connects to a phenotype ontology class for Arrythmia [35]. However, curation of ChIRO was manual and due to limited resources the initial set of axioms were never expanded. ChIRO did not take advantage of lexical regularities, which could have been used to automate the axiomatization of large numbers of terms, in particular EC inhibitors and activators to the corresponding enzyme class, as shown below. Throughout the process of incorporating ChIRO into this work, several improvements were suggested through its issue tracker [36].

Hierarchy-Driven Role Identification

While affix-based search had relatively high precision (with the exception of -agent, which listed many applications that were not relevant in this context), most entries do not have a naming convention. Luckily, ChEBI is an ontology with pre-defined hierarchies defining roles including enzyme inhibitor [20], central nervous system drug [37], antidote [38], and many others. For example, the 8 sub-roles of antidote [38] were all identified by affix, but none of the subclasses of central nervous system depressant [39] were.

ChEBI's roles hierarchy is divided into three parts: chemical role [17], biological role [18], and application [19]. Within the latter two are the most interesting roles for curation, but not all are relevant. Therefore, a blacklist was developed including irrelevant applications like protecting group [40], dye (CHEBI:37958), and MALDI matrix material [41] as well as irrelevant or misclassified biological roles like provitamin [42], insect attractant [43], and α-D-GalN-(1→4)-[α-DD-Hep-(1→2)]-α-D-GalA-yl group [44]. The remaining elements from these two role hierarchies were curated.

As curation over the remaining parts of the ChEBI hierarchy resulted in the largest number of suggestions, the curation scripts were careful to remove results from previously-curated terms to reduce curator fatigue. Additionally, a list of false positives was developed and maintained throughout curation to reduce the number of propositions and ultimately reduce curator fatigue.

Summary Statistics

There are several ways of summarizing the 828 roles that have been curated so far. Below, the Figure [4](#) summarizes the types of modulations and target entity types.
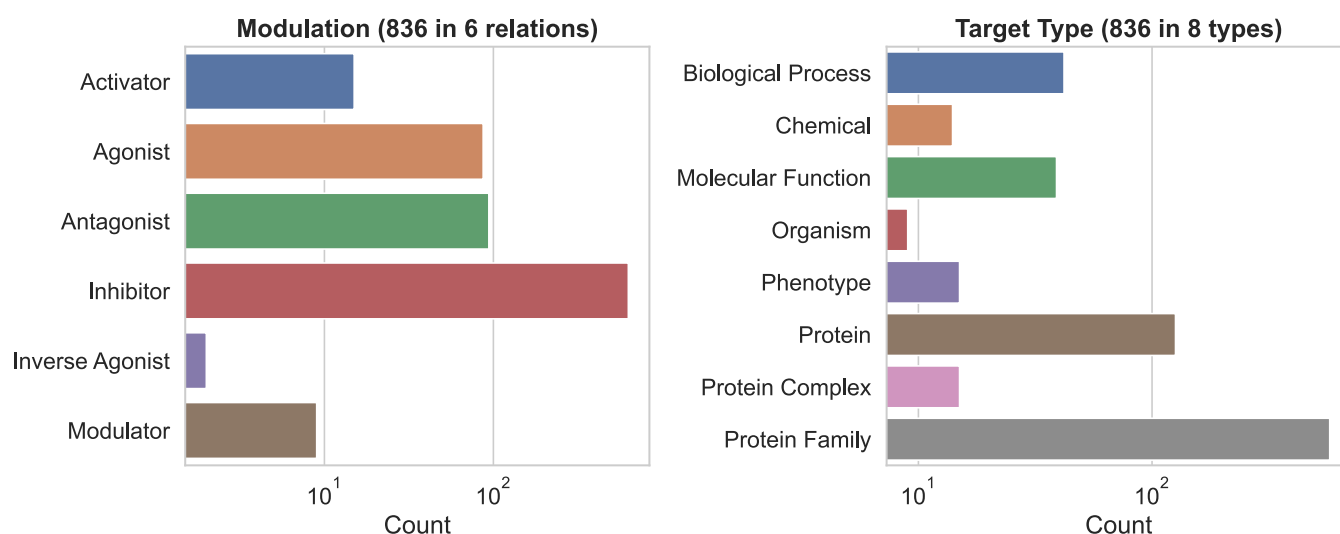

Figure 4: Summaries over the types of modulations and target entity types covered by the database so far.

Inference

Before outlining a generalized strategy for inference, the liver X receptor inverse agonist [45] will be presented as an example role corresponding to the inverse agonism of the LXR family of proteins. Its definition implies that chemicals that have this role should be inferred to affect the two members of this family LXRα [46] and LXRβ [47]. Therefore, the shape expression from Table ?? can be extended to handle protein families over which inference should be extended to their respective members in Table [1](#).

Table 1: Shape expression and predicate logic rules for reasoning over protein family memberships.

	Antecedent	Consequent
ChEBI	**C** a Chemical **R** a Role **C** has role **R**	
Curated	**G** a Protein Family **R m G** **m** correspondsTo **M**	
External	**P** a Protein **P** a **G**	**C M P**

However, we have a number of hierarchical resources besides proteins, and sometimes these resources are several levels deep. Additionally, chemicals and roles both have their own hierarchies by ChEBI. In Table [2](#), the `a*` notation is to show variable length traversals through "is a" relationships. This means if `A a B` and `B a C`, then A a* C is true. This holds for chains of arbitrary length. In the context of the `a*` operator, it is also true that `A a A`. Finally, Table [2](#) presents a fully generalized inference framework.

Table 2: A fully generalized shape expression and predicate logic rules for inference over hierarchies.

	Antecedent	Consequent

	Antecedent	Consequent
ChEBI	**C** a Chemical **R** a Role **c** a* **C** **r** a* **R** **c** has role **r**	
Curated	**T** a Target **r m T** **m** correspondsTo **M**	
External	**t** a* **T**	**c M T**

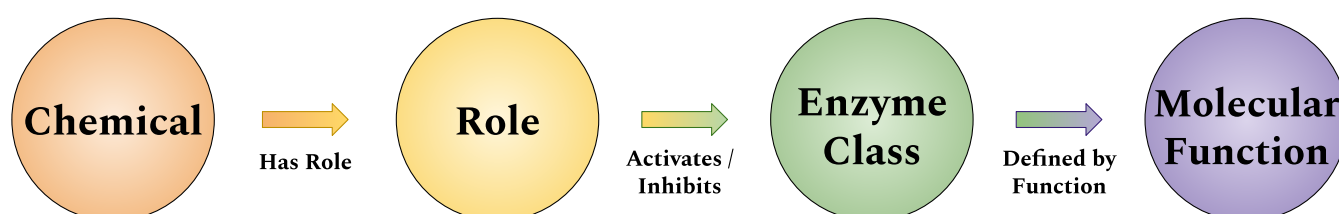## Linking Roles to Gene Ontology Molecular Functions through Enzyme Classes

Figure 1: Schema for inference of chemicals' relations to molecular functions via roles and enzyme classes

Following the exhaustive mapping from chemical roles classifying activators or inhibitors of given enzyme classes, the EC2GO mapping [48] provided by GO can be applied to infer on to which molecular function each chemical role exerts its effect. Like in the first example where relation m between role **R** and target **T** was defined based on relation **M** between chemical **C** and target **T**, we define relation M_F to correspond to the relationship between a chemical and a molecular function corresponding to a given enzyme as shown in Table 3.

Table 3: Shape expression and predicate logic rules inferring chemical relations to molecular functions.

	Antecedent	Consequent
ChEBI	**C** a Chemical **R** a Role **C** has role **R**	
Curated	**E** an Enzyme **R m E** **m** correspondsTo **M** **m** correspondsTo$_F$ M_F	
GO	**F** a Molecular Function **E** hasFunction **F**	**C M E** **C** M_F **F**

Implementation

The acquisition of external resources and inference over hierarchies is accomplished by a simple script whose installation and execution is mediated by the Tox [49] configuration available at https://github.com/chemical-roles/chemical-roles. It relies heavily on the PyOBO Python package [50] to acquire hierarchical data sources from OBO Foundry and generate OBO-like structures for several other resources, notably HGNC, GO, InterPro, etc.

Summary Statistics

Inference over entity hierarchies resulted in over 424K new relationships. Figure 2 summarizes them by modulation and target entity type. Notice that each has increased by several orders of magnitude essentially for free by leveraging a vast amount of previously curated, high quality content.
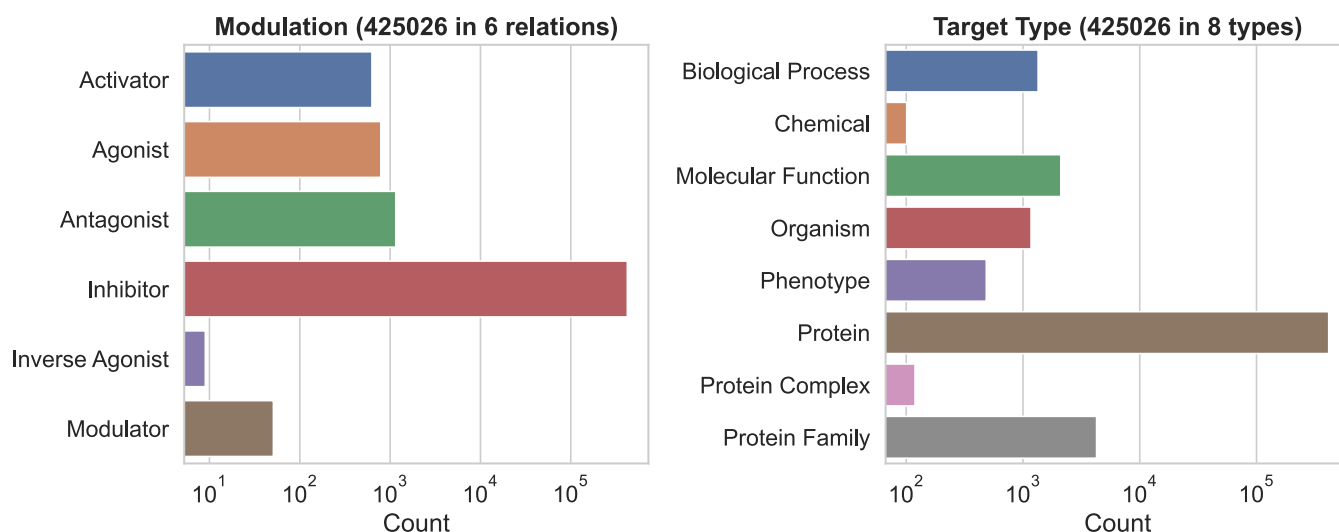

Figure 2: Summaries over the types of modulations and target entity types in inferred relationships. Inference can be optionally extended over role hierarchies, which produces millions of relations.

Interoperability

While the manually curated and inferred relations arising from this work are distributed in a simple format appropriate for wide reuse, we highlight two use cases below.

Uptake by ChIRO and OBO

One of the goals of OBO is to facilitate modular ontology development, and provide a set of richly interlinked ontologies. For example, many ontologies such as GO and HPO are axiomatized using ChEBI [**???**] for example, gliotoxin catabolism [51] has-input some gliotoxin [52]. ChEBI is not itself axiomatized in this way. The ChIRO ontology was originally created as an overlay on top of ChEBI to provide this additional level of linkage. This provides multiple advantages - users interested in concepts such as Arrythmia [35] in HPO can navigate to the drugs that exacerbate or inhibit it. In particular, the use of axioms that connect from ChEBI back to GO allow for pathway queries when used in combination with GO-CAMs [53], for example reasoning from a gene that is upstream of a process that regulates biosynthesis of a particular chemical that itself acts as an inhibitor of another GO activity. A workflow has been developed using Dead Simple OWL Design Patterns (DOSDP) [**???**] and the Ontology Development Kit (ODK) [54]. DOSDPs are a generic framework that allows for generation of OWL terms and axioms using specific patterns. The patterns are represented in a YAML file and contain slots for variables such as chemical classes from ChEBI. The specific fillers for each variable in the pattern are represented as rows in a tab-separated values (TSV) file, which are then run through a pipeline via the ODK to generate new terms that include logical axioms that were specified by the pattern, that can be subsequently imported into the ontology. Examples from ChIRO can be found at https://github.com/obophenotype/chiro/tree/master/src/patterns.

Conversion to the Biological Expression Language

The Biological Expression Language (BEL) [**???**] is a domain specific language (DSL) for storing rich biomedical relations, supporting data integration [doi:10.1101/631812]. The relationships curated and inferred in this work have a one-to-one correspondence with BEL statements as described in Table [4](#). The internal BEL DSL in the PyBEL software package [**???**] was used to implement a conversion that is included with each export.

Table 4: Mapping of relationships to BEL statements.

Relation	BEL Example
Activation, Agonist	`a(X) directlyIncreases act(p(Y))`
Inhibition, Inverse Agonist, Antagonist	`a(X) directlyDecreases act(p(Y))`
Modulator	`a(X) regulates act(p(Y))`

Concluding Remarks

This article identified an area of improvement in the ChEBI ontology, presented a schema for curating that data, scripts for assisting in curation, and ultimately a curated database of role-modulation-target triples. It also presented an informal framework for inference and an implementation for doing that inference. The results of curation and inference are both available directly through the GitHub repository at https://github.com/chemical-roles/chemical-roles.

Like with most curation, this work is not complete. Curation will continue using the procedures described until exhaustion - about another 800 roles remain. It would also be advantageous to generalize this curation technique from the roles in ChEBI to the roles in MeSH, a resource notorious for its lack of cross-references with other public biological data sources. Several examples have already been included in the database, an ontology of chemicals from PubChem and MeSH pharmacological roles similar to ChEBI's is available through the PubChem FTP server [**???**] that enable similar inference to that presented here.

Finally, this resource should be made as interoperable with others as possible. It was not designed to compete with ChEBI or ChIRO, but rather to complement and enrich them. A future goal is to automate the upload of the results to WikiData through a bot to bolster their chemical data integration efforts. Most importantly is to enable the ChEBI curators and the greater OBO community to best intake these relations into their ontologies and inference frameworks to benefit their work.

Resource Availability

All code used in the curation workflow is available at https://github.com/chemical-roles/chemical-roles under the MIT License. All manually curated data comprising the database are also available at https://github.com/chemical-roles/chemical-roles under the CC0 License. All code and data are archived on Zenodo at https://doi.org/10.5281/zenodo.3883800.

Acknowledgements

This work would not have been possible without the work done by the excellent team of curators and maintainers of ChEBI, especially Janna Hastings and Adnan Malik.

Author Contributions

CTH designed the study. CTH, NV, MH, and DDF curated content. CTH, CM, NV and VC wrote the manuscript. All authors contributed to editing and revision. All authors have approved the final manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

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. **Protein Ontology - protein - Classes | NCBO BioPortal**http://bioportal.bioontology.org/ontologies/PR?p=classes&conceptid=PR%3A000000001

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

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

13. http://purl.obolibrary.org/obo/RO_0002408

14. **oborel/obo-relations**
oborel
(2020-06-11) https://github.com/oborel/obo-relations

15. **The Gene Ontology Resource: 20 years and still GOing strong**
The Gene Ontology Consortium
Nucleic Acids Research (2019-01-08) https://doi.org/gf63mb
DOI: 10.1093/nar/gky1055 · PMID: 30395331 · PMCID: PMC6323945

16. **role (CHEBI:50906)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:50906

17. **chemical role (CHEBI:51086)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:51086

18. **biological role (CHEBI:24432)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:24432

19. **application (CHEBI:33232)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:33232

20. **enzyme inhibitor (CHEBI:23924)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:23924

21. **EC 3.4.24.71 (endothelin-converting enzyme 1) inhibitor (CHEBI:82974)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:82974

22. **IntEnz - EC 3.4.24.71**https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=3.4.24.71

23. **pathway inhibitor (CHEBI:76932)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:76932

24. **GABA antagonist (CHEBI:65259)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:65259

25. **Gene group | HUGO Gene Nomenclature Committee**https://www.genenames.org/data/genegroup/#!/group/562

26. **gamma-aminobutyric acid (CHEBI:16865)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:16865

27. **vitamin B(1) antagonist (CHEBI:78248)**https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:78248

28. **The OBO Foundry**http://www.obofoundry.org/

29. **FamPlex: a resource for entity recognition and relationship resolution of human protein families and complexes in biomedical text mining**
John A. Bachman, Benjamin M. Gyori, Peter K. Sorger
BMC Bioinformatics (2018-06-28) https://doi.org/gg3knp
DOI: 10.1186/s12859-018-2211-5 · PMID: 29954318 · PMCID: PMC6022344

30. **indralab/gilda**
INDRA Labs
(2020-06-25) https://github.com/indralab/gilda

31. **ebi-chebi/ChEBI**
GitHub

https://github.com/ebi-chebi/ChEBI

32. **obophenotype/chiro**
   obophenotype
   (2020-04-09) https://github.com/obophenotype/chiro

33. **AmiGO 2: Term Details for "apoptotic process"**
   (GO:0006915) http://amigo.geneontology.org/amigo/term/GO:0006915

34. **anti-arrhythmia drug (CHEBI:38070)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:38070

35. **Human Phenotype Ontology** https://hpo.jax.org/app/browse/term/HP:0011675

36. **obophenotype/chiro**
   GitHub
   https://github.com/obophenotype/chiro

37. **central nervous system drug (CHEBI:35470)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:35470

38. **antidote (CHEBI:50247)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:50247

39. **central nervous system depressant (CHEBI:35488)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:35488

40. **protecting group (CHEBI:51087)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:51087

41. **MALDI matrix material (CHEBI:64345)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:64345

42. **provitamin (CHEBI:50188)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:50188

43. **insect attractant (CHEBI:24850)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:24850

44. **alpha-D-GalN-(1->4)-[alpha-DD-Hep-(1->2)]-alpha-D-GalA-yl group**
   (CHEBI:136367) https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:136367

45. **liver X receptor inverse agonist (CHEBI:90846)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:90846

46. https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/7966

47. https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/7965

48. https://geneontology.org/ontology/external2go/ec2go

49. **Welcome to the tox automation project — tox 3.16.2.dev1**
   documentation https://tox.readthedocs.io/en/latest/

50. **pyobo/pyobo**
   PyOBO
   (2020-06-25) https://github.com/pyobo/pyobo

51. **AmiGO 2: Term Details for "gliotoxin catabolic process" (GO:2001309)** http://amigo.geneontology.org/amigo/term/GO:2001309

52. **gliotoxin (CHEBI:5385)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:5385

53. **Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems**
   Paul D. Thomas, David P. Hill, Huaiyu Mi, David Osumi-Sutherland, Kimberly Van Auken, Seth Carbon, James P. Balhoff, Laurent-Philippe Albou, Benjamin Good, Pascale Gaudet, … Christopher J. Mungall
   Nature Genetics (2019-09-23) https://doi.org/ggcfst
   DOI: 10.1038/s41588-019-0500-1 · PMID: 31548717 · PMCID: PMC7012280

54. **INCATools/ontology-development-kit**
   INCATools
   (2020-06-15) https://github.com/INCATools/ontology-development-kit

55. **anti-inflammatory drug (CHEBI:35472)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:35472

56. **isobavachalcone (CHEBI:28106)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:5984

57. **anti-HIV-1 agent (CHEBI:64947)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:64947

58. **antidote to opioid poisoning (CHEBI:90755)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:90755

59. **soluble guanylate cyclase activator (CHEBI:76022)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:76022

60. **UDP-glucuronosyltransferase activator (CHEBI:77324)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:77324

61. **EC 2.7.11.31 {[hydroxymethylglutaryl-CoA reductase (NADPH)] kinase} activator (CHEBI:85053)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:85053

62. **hypoxia-inducible factor pathway activator (CHEBI:134083)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:134083

63. **Hsp70 inducer (CHEBI:78606)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:78606

64. **apoptosis inducer (CHEBI:68495)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:68495

65. **interferon inducer (CHEBI:36710)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:36710

66. **MTP inhibitor (CHEBI:72298)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:72298

67. **eukaryotic initiation factor 4F inhibitor (CHEBI:137079)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:137079

68. **glycine transporter inhibitor (CHEBI:85425)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:85425

69. **EC 1.1.1.188 (prostaglandin-F synthase) inhibitor (CHEBI:77425)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:77425

70. **autophagy inhibitor (CHEBI:88230)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:88230

71. **aryl hydrocarbon receptor agonist (CHEBI:72768)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:72768

72. **cannabinoid receptor agonist (CHEBI:67072)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:67072

73. **aryl hydrocarbon receptor antagonist (CHEBI:76998)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:76998

74. **cannabinoid receptor antagonist (CHEBI:73413)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:73413

75. **5-hydroxytryptamine 2A receptor inverse agonist (CHEBI:133016)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:133016

76. **inverse agonist (CHEBI:90847)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:90847

77. **retinoic acid receptor modulator (CHEBI:90708)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:90708

78. **gamma-secretase modulator (CHEBI:48668)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:48668

79. **PPAR modulator (CHEBI:70781)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:70781

80. **tubulin modulator (CHEBI:60832)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:60832

81. **antiglaucoma drug (CHEBI:39456)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:39456

82. **trypanocidal drug (CHEBI:36335)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:36335

83. **tranquilizing drug (CHEBI:35473)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:35473

84. **antihypotensive agent (CHEBI:137431)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:137431

85. **VR nerve agent (CHEBI:140422)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:140422

86. **sensitiser (CHEBI:139492)** https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:139492

Appendix

Table 5: Affixes used for initial targeted curation. In the first column, P corresponds to prefix, S to suffix, and A to both in the case of "EC-...-inhibitor/activator".

	Affix	**Description**

	Affix	Description
P	anti-	Various agents against biological processes (e.g., anti-inflammatory drug [55]), pathologies (e.g., antihyperplasia drug [56]), and organisms (e.g., anti-HIV-1 agent [57]).
P	antidote to-	Antidotes for poisons (e.g., antidote to opioid poisoning [58]).
A	EC-...-inhibitor or EC-...- activator	Inhibitors of enzymes. The ChEBI hierarchy of roles implicitly models the enzyme classification with the root term being enzyme inhibitor [20]. However, some terms have the suffix of -activator, which do not have corresponding roles nor a corresponding root term enzyme activator.
S	-activator	Activators of proteins (e.g., p53 activator [8]), protein complexes (e.g., soluble guanylate cyclase activator [59]), protein families (e.g., UDP-glucuronosyltransferase activator [60]), enzyme classes (e.g., EC 2.7.11.31 {[hydroxymethylglutaryl-CoA reductase (NADPH)] kinase} activator [61]), pathways (e.g., hypoxia-inducible factor pathway activator [62])
S	-inducer	Inducers of protein families (e.g., Hsp70 inducer [63]), pathways (e.g. apoptosis inducer [64]). The difference in semantics between activator and inducer is unclear, and it seems that activator was preferred by ChEBI curators. Only three entries, interferon inducer [65] being the third, have this suffix.

	Affix	Description
S	-inhibitor	Inhibitors of proteins (e.g., MTP inhibitor [66]), protein complexes (e.g., eukaryotic initiation factor 4F inhibitor [67]), protein families (e.g. glycine transporter inhibitor [68]), enzyme classes (e.g., EC 1.1.1.188 (prostaglandin-F synthase) inhibitor [69]), pathways (e.g., autophagy inhibitor [70])
S	-agonist	Agonists of proteins (e.g., aryl hydrocarbon receptor agonist [71]), protein families (e.g., cannabinoid receptor agonist [72])
S	-antagonist	Antagonists of proteins (e.g., aryl hydrocarbon receptor antagonist [73]), protein families (e.g., cannabinoid receptor antagonist [74])
S	-inverse agonist	Inverse agonists of proteins (e.g., 5-hydroxytryptamine 2A receptor inverse agonist [75]), protein families (e.g., liver X receptor inverse agonist [45]). Unfortunately, these are the only two such entries. They both inherit from inverse agonist [76].
S	-modulator	Modulators (lacking the polarity described by activators or inhibitors) of proteins (e.g., retinoic acid receptor modulator [77]), protein complexes (e.g., γ-secretase modulator [78]), protein families (e.g., PPAR modulator [79]), cellular components (e.g., tubulin modulator [80])
S	-suppressor / -suppressant	Like -inducer, there are only a handful of examples. The preferred suffix is -inhibitor
S	-deactivator	Surprisingly, there were no terms in ChEBI using this suffix

	Affix	Description
S	-drug	Drugs treating pathways (e.g., anti-inflammatory drug [55]), pathologies (e.g., antiglaucoma drug [81]), infections (e.g., trypanocidal drug [82]), and psychoses (e.g., tranquilizing drug [83]). However, most drugs treating psychoses are children of central nervous system drug [37] and do not end in the suffix -drug.
S	-agent	Agents cover a wide variety of biology from drugs (e.g., antihypotensive agent [84]) to toxins/poisons (e.g., VR nerve agent [85]) and other applications.
S	-stabilizer / -stabiliser	This search was included because of the notes from ChIRO, but no terms in ChEBI corresponded to it.
S	-destabilizer / -destabiliser	This search was included because of the notes from ChIRO, but no terms in ChEBI corresponded to it.
S	-sensitizer / -sensitiser	This search was included because of the notes from ChIRO, but no terms in ChEBI had this suffix besides the role sensitiser [86]

1. It should be noted that the pharmacological relationships agonist, antagonist, and inverse agonist are well-defined in ChEBI, but not so much so in standardized ontologies of relationships between biochemical entities such as the Relation Ontology. See GitHub discussions at oborel/obo-relations#369 and oborel/obo-relations#371.↩