

Extension of Roles in the ChEBI Ontology

This manuscript ([permalink](#)) was automatically generated from [chemical-roles/manuscript@75dca77](#) on June 30, 2020.

Authors

- **Charles Tapley Hoyt**

 [0000-0003-4423-4370](#) ·  [cthozt](#) ·  [cthozt](#)

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. 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.

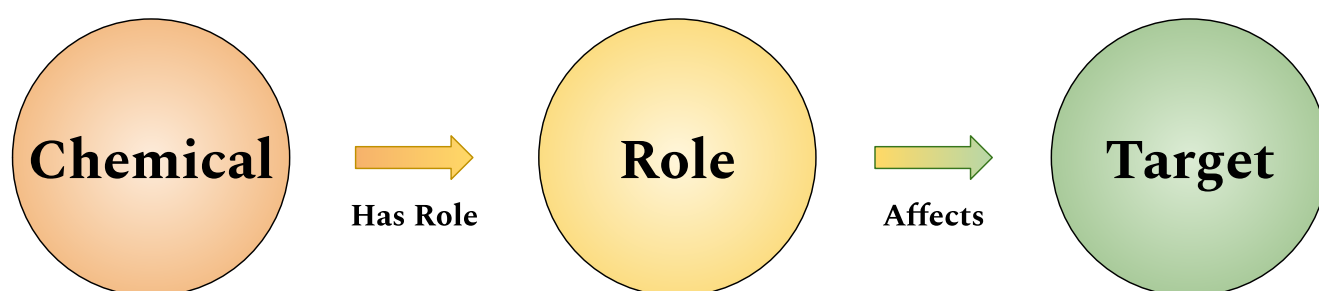


Figure 1: Schema for inference of chemicals' relations to targets via roles. Targets may be other chemicals, proteins, protein families, protein complexes, pathways, pathologies, or organisms.

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. (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 [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. This 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. 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 . {#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}).

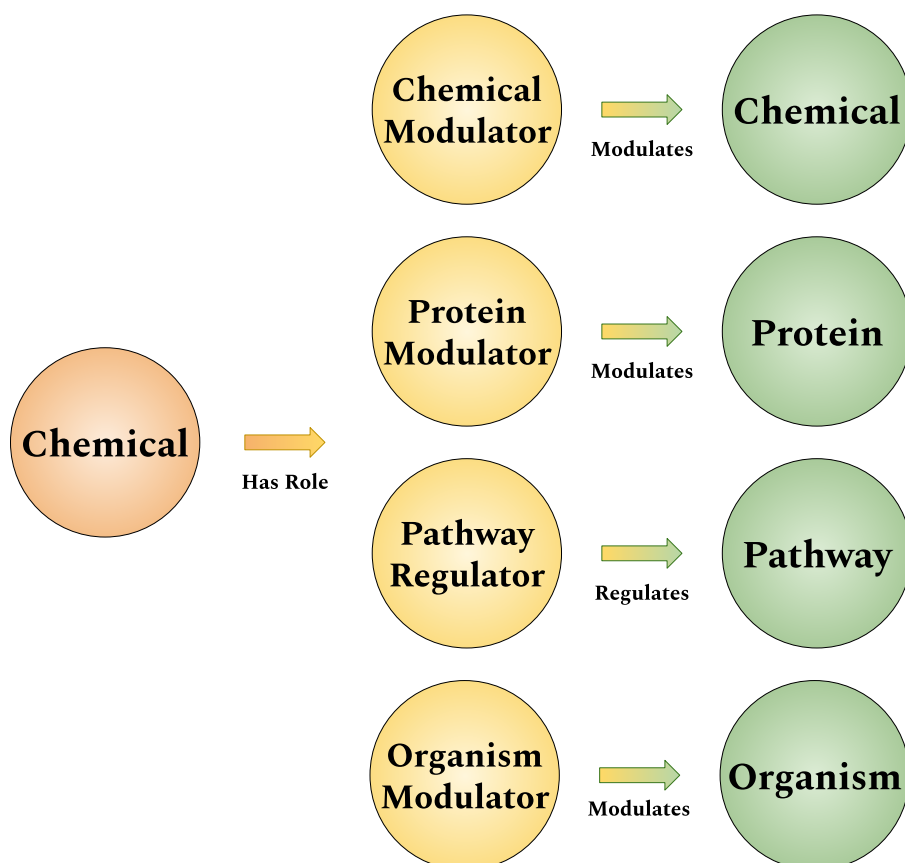


Figure 2: Generalization of the ChEBI schema to include relationships between several categories of roles and their target entities.

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.

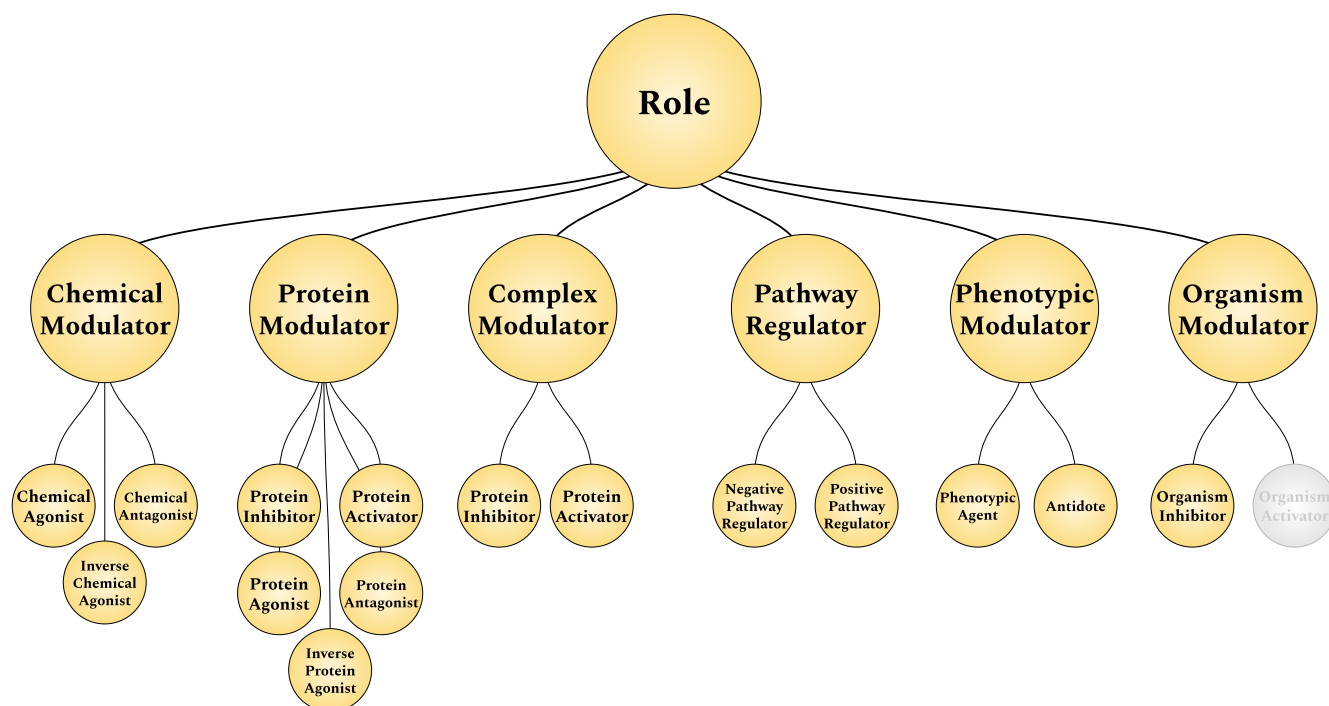


Figure 3: A hierarchy of chemical roles. Each role corresponds to a chemical having a given relationship. Within the ChEBI hierarchy, the role [16] has three children: chemical role [17], biological role [18], and application [19]. Our interest lies with the latter two.

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

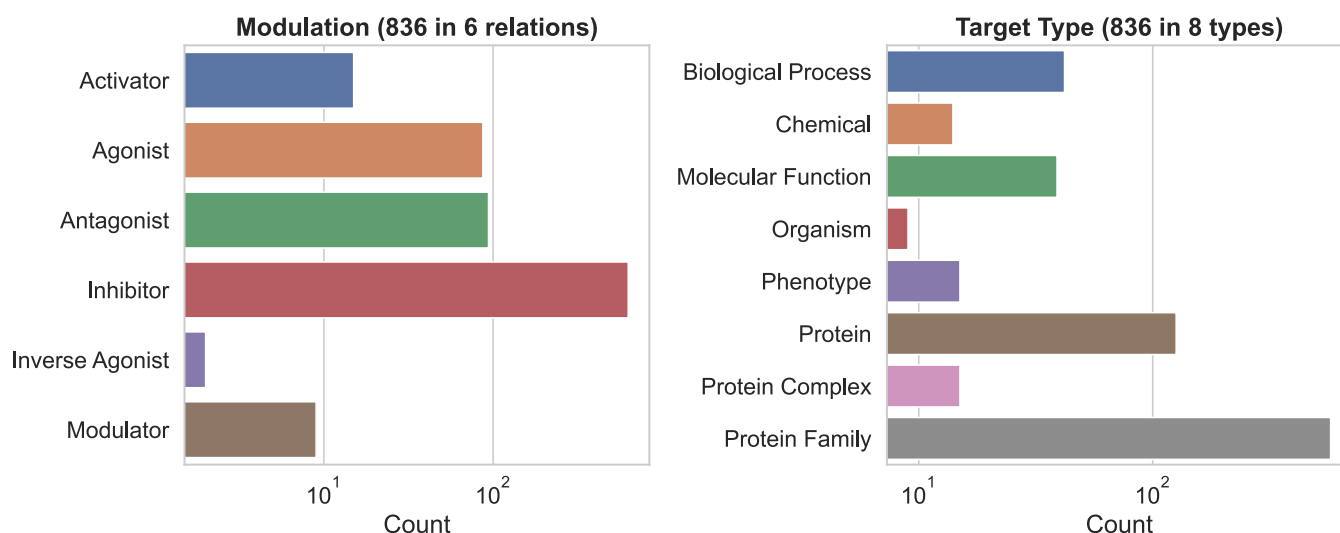


Figure 4: Summaries over the types of modulations and target entity types covered by the database so far.

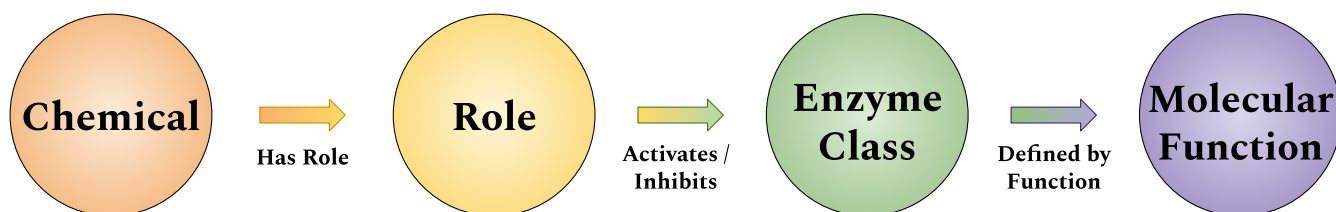


Figure 5: Schema for inference of chemicals' relations to molecular functions via roles and enzyme classes

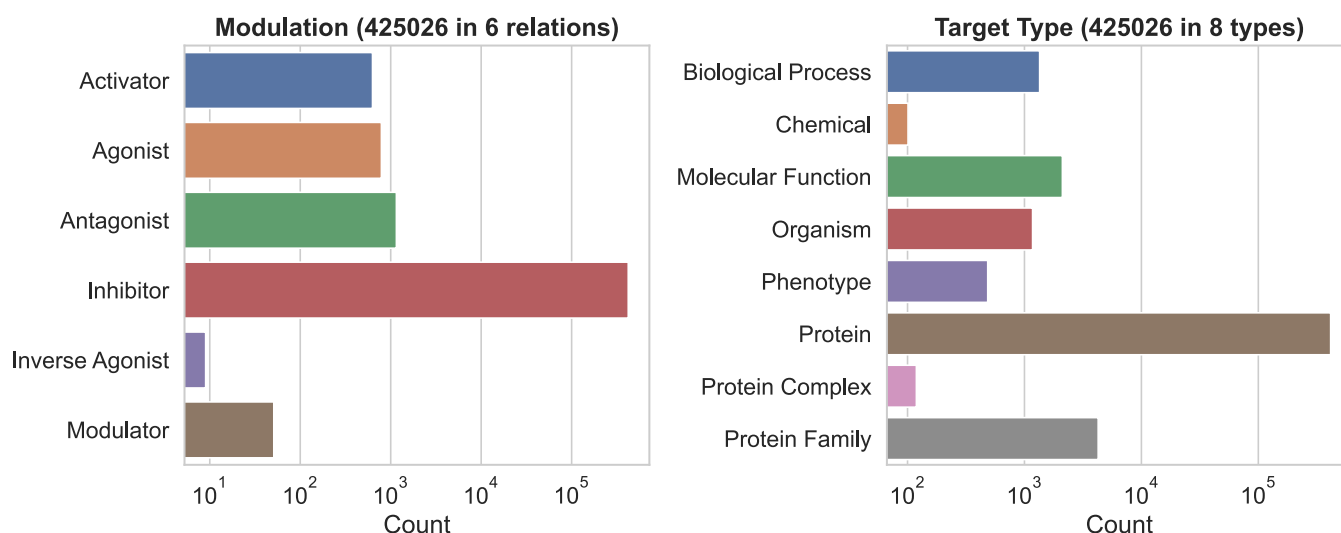


Figure 6: Summaries over the types of modulations and target entity types in inferred relationships. Inference can be optionally extended over role hierarchies, which produces millions of relations.

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](https://doi.org/10.1093/nar/gks1146) · PMID: [23180789](https://pubmed.ncbi.nlm.nih.gov/23180789/) · PMCID: [PMC3531142](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1093/nar/gkv1031) · PMID: [26467479](https://pubmed.ncbi.nlm.nih.gov/26467479/) · PMCID: [PMC4702775](https://pubmed.ncbi.nlm.nih.gov/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](https://www.isbn-international.org/product/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](https://doi.org/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](https://doi.org/10.1371/journal.pcbi.1007128) · PMID: [31233491](https://pubmed.ncbi.nlm.nih.gov/31233491/) · PMCID: [PMC6611653](https://pubmed.ncbi.nlm.nih.gov/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](https://doi.org/10.1093/nar/gky1055) · PMID: [30395331](https://pubmed.ncbi.nlm.nih.gov/30395331/) · PMCID: [PMC6323945](https://pubmed.ncbi.nlm.nih.gov/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>