Drugs in ChEBI

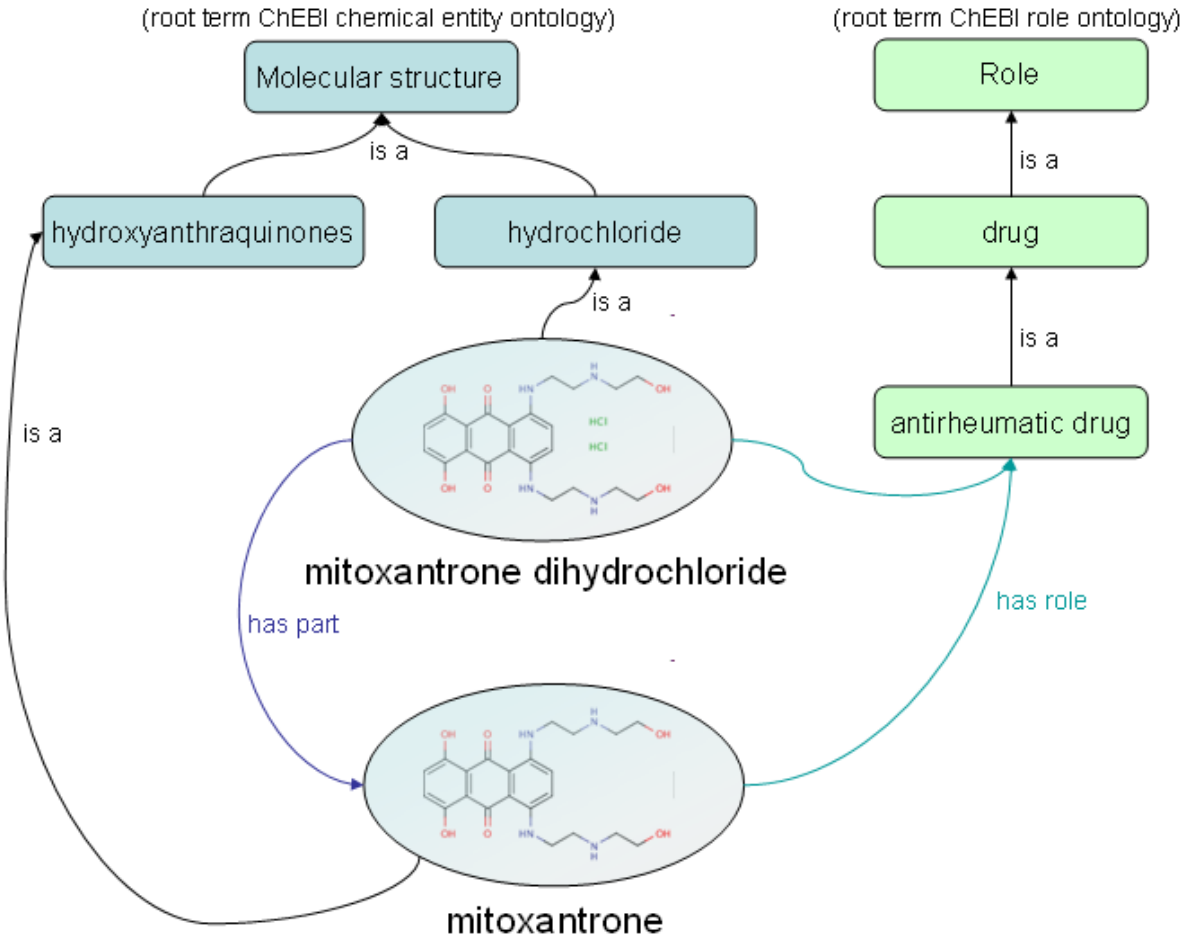
ChEBI is a database and ontology for chemical entities in a biological context. Part of what sets ChEBI aside from other publicly available chemical data resources is the provision of the chemical ontology in which knowledge and information about the chemical entities is explicitly encoded through formal relations between entities in the ontology. The ChEBI ontology currently incorporates ten relationship types, but for the purposes of this document the relevant relationship types which will be referred to within the examples are 'is a' (the primary classification / subsumption relationship), 'has part' (the primary compositional relationship), 'is enantiomer of' (which links the two enantiomers of a chiral molecule), and 'has role' (which links a chemical entity to a role or bioactivity in the role ontology).

This document details the current curation procedures and standards for capturing drug data in ChEBI together with examples of annotated drug entries.

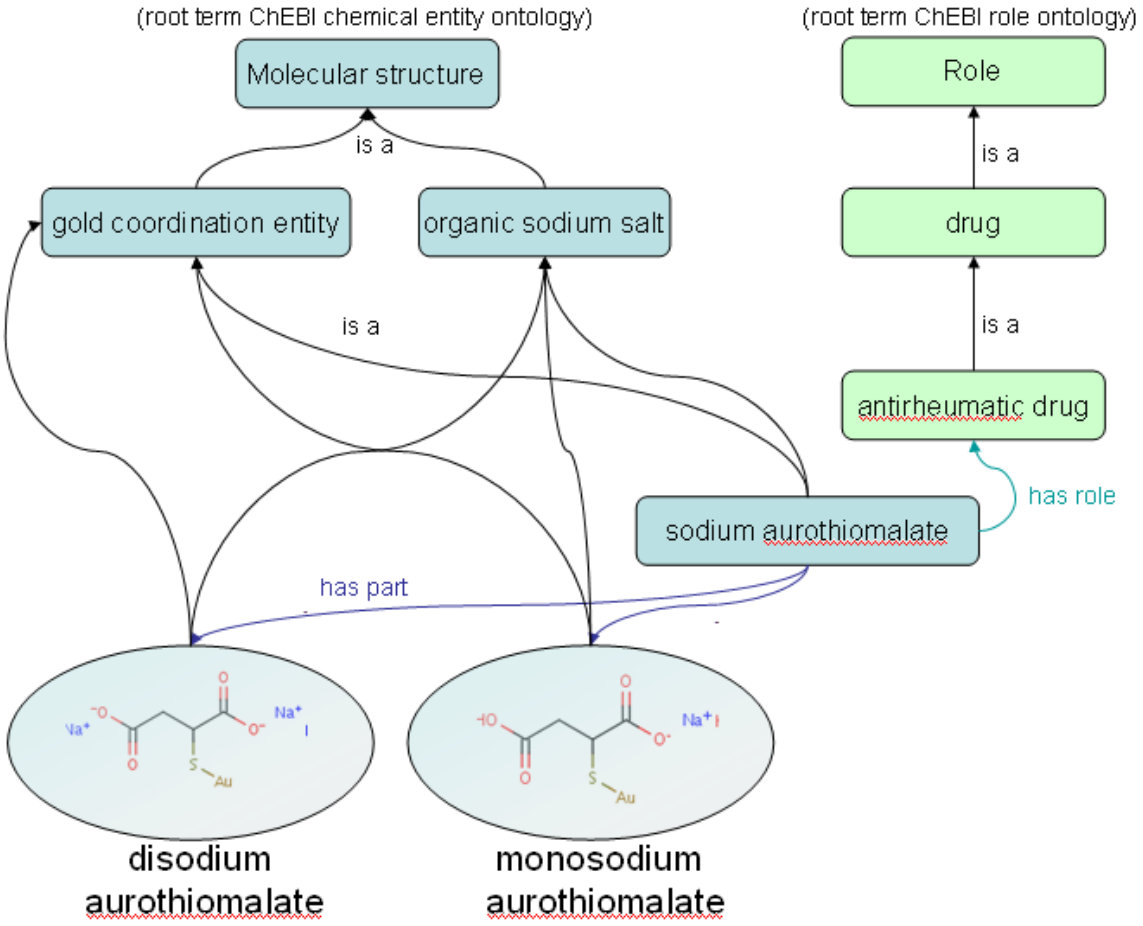
Structures and identity

The main data resource for annotating ChEBI Drugs is DrugBank. It should be noted that individual DrugBank entries may be linked to multiple ChEBI entities, as different salts, hydrates and isomers from a given compound share a common ID in DrugBank but are annotated as different entities in ChEBI.

For example, DrugBank entry DB01204 (Mitoxantrone, an antineoplastic drug) links to two ChEBI entries, one for the active molecule itself, and one for the primary salt form that it is delivered in. These are illustrated together with their interrelationships in the diagram below. Many intervening relationships in the 'is a' hierarchy to the roots are omitted for clarity; the relationship from 'hydroxyanthraquinones' to the root term 'molecular structure' actually traverses several 'is a' relationships. This is not a problem though because 'is a' is transitive. The 'has part' relation is used to link the salt to the active molecule, and the 'has role' relationship links both the salt and the active molecule to the 'antineoplastic drug' role.



In addition to salt forms, mixtures are explicitly modeled as separate ChEBI entitites. For example, sodium aurothiomalate (CHEBI:5516) is an antirheumatic drug which is delivered as a mixture of two salt forms. The ontology for this mixture looks as illustrated in the next figure (again, this shows only a subset of the relationships of the relevant terms and paths to root are truncated):

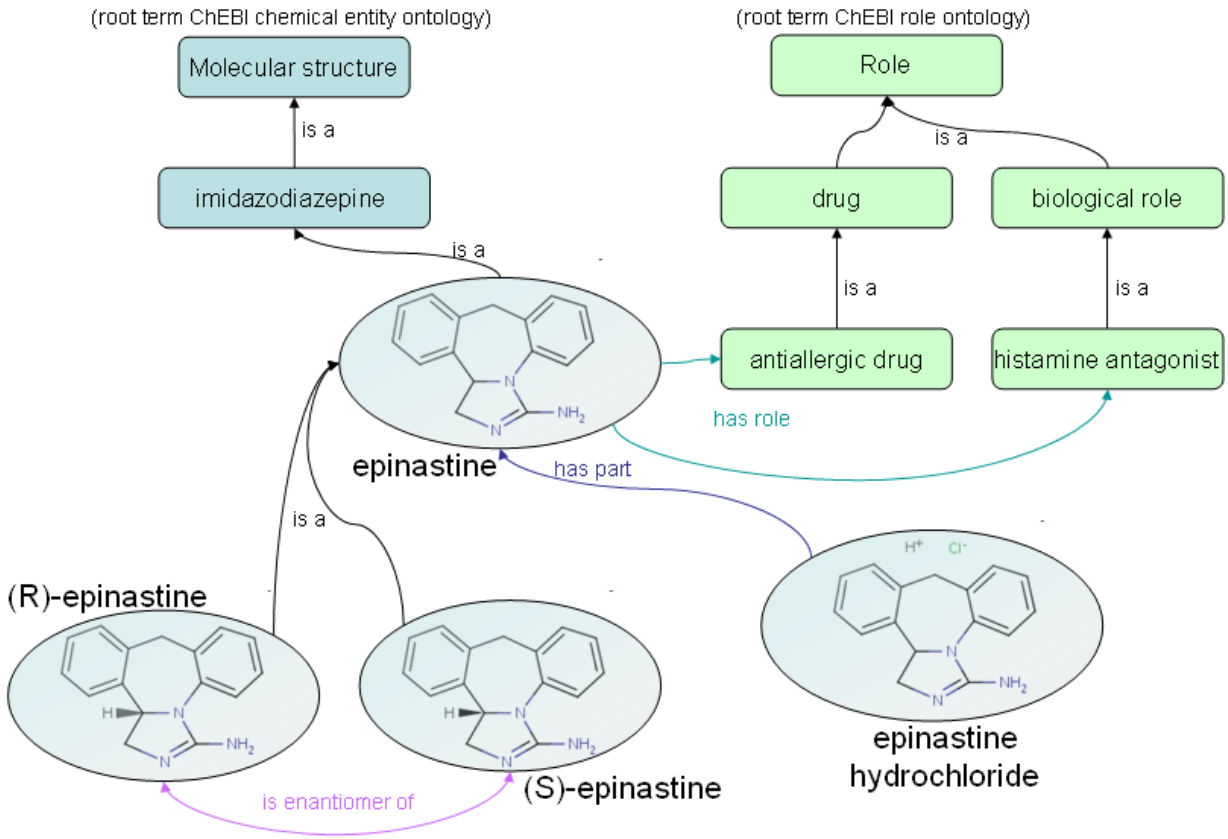


Additional annotations such as synonyms, patent links and literature citations, as well as cross-references to other databases of biological interest, are also captured as per the standard ChEBI annotation procedures.

Racemates

A racemic mixture, or racemate, is one that has equal amounts of left- and right-handed enantiomers of a chiral molecule. The trend in the pharmaceutical development is increasingly moving towards the development of single isomers rather than racemates; however, there are around 400 racemic drugs approved at present. Any drug which is a racemic compound is annotated as three ChEBI entities: (1) a non-stereospecified entity, (2) the left-handed enantiomer, (3) the right-handed enantiomer. Each of the two enantiomers would be linked to the non-stereospecified entity via an 'is a' relationship and to each other via an 'is enantiomer of' relationship. A unique DrugBank ID is crossreferenced to the three entities except when a specific DrugBank ID exists for one or both enantiomers. The same applies to the INN of the racemic compound: this will be the same for all three ChEBI entities except when a specific INN exists for one or both isomers. The illustration below shows a drug with two enantiomeric forms and the interrelationships

both to the stereounspecified parent, the role ontology, the structure ontology and a salt form.



Brand names

As our primary source, we initially input all DrugBank brand names in ChEBI but soon realised that most of them are deprecated. We have made an effort trying to distinguish between brand names currently in use (those authorised and/or commercialised in at least one country) and those which are deprecated (indicated in ChEBI by a yellow triangle surrounding an exclamation mark). However, this is a very manual and time-consuming process, so currently drug brand names are **not** annotated in ChEBI. We have found RxNorm to be a potential resource from where to fetch automatically non-deprecated brand names. To date RxNorm includes FDA-approved brand names but only a few approved by Governmental Drug Administrations in other countries. To our knowledge, brand names in RxNorm are assigned correctly to salts, hydrates and isomers of a given compound, so in future we might use this resource to annotate brand names automatically to ChEBI Drugs.

INNs

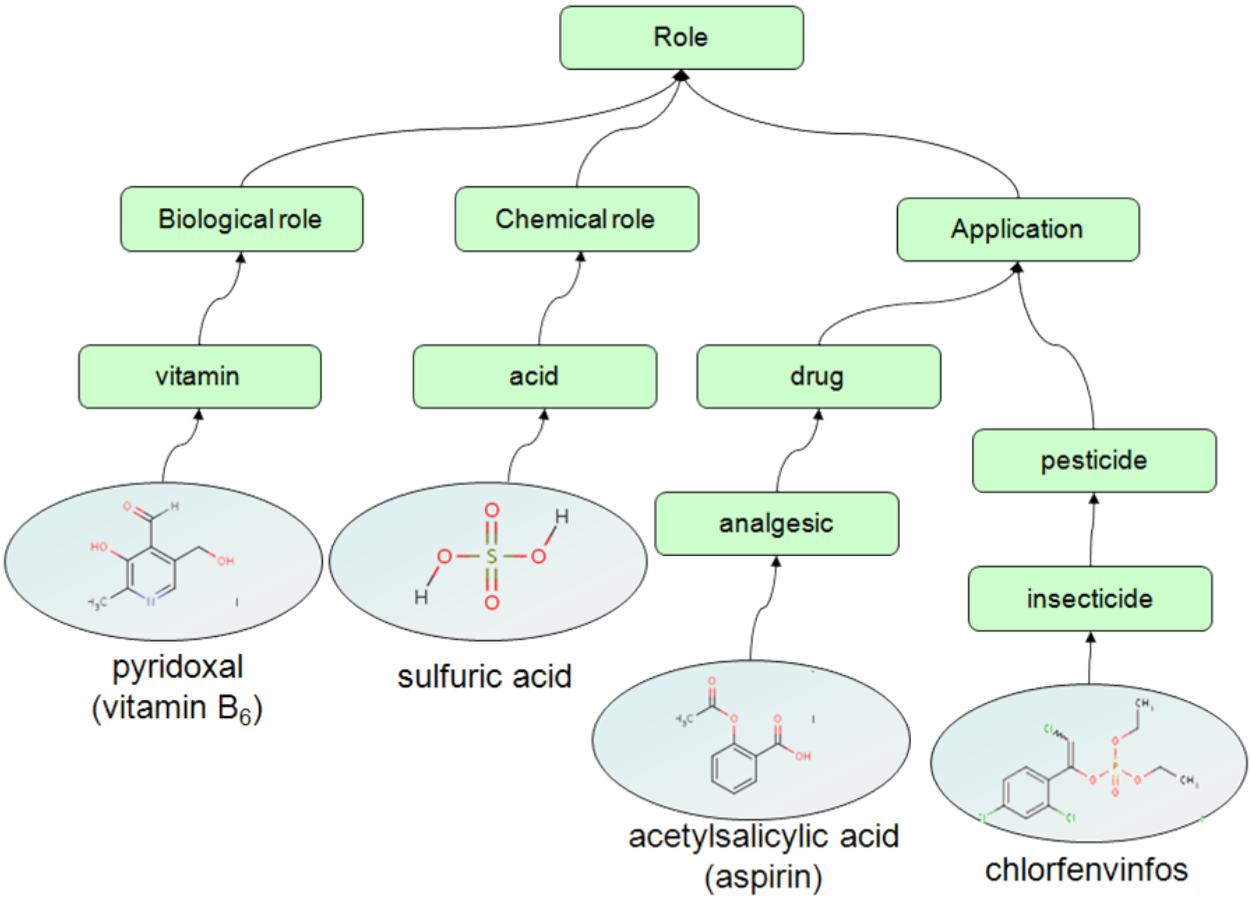
International Nonproprietary Names (INNs) facilitate the identification of pharmaceutical substances or active pharmaceutical ingredients. Each INN is a unique name that is globally recognized and is public property. The World Health Organization collaborates closely with INN experts and national nomenclature committees to select a single name of worldwide

acceptability for each active substance that is to be marketed as a pharmaceutical. Nonproprietary names, also known as generic names, are annotated for all ChEBI Drugs from [WHO MedNet] services.

Role (activity and drug) classification

ChEBI captures information about the biological activity of small molecules in our 'role' ontology. The root term 'role' is a catch-all for dispositional properties of a molecule - properties that inhere in a molecule by virtue of the capacity of the object to affect or to be affected by other things, when participating in a process. Within this ontology, ChEBI differentiates three primary different categories of dispositional properties, namely 'biological role' which refers to the functions which molecules play in a biological context, such as 'agonist' and 'inhibitor'; 'chemical role' which refers to brute chemical functions such as 'acid' and 'base'; and applications to which molecules can be put by humans, such as 'pesticide'. Molecules are linked to roles via the 'has role' ontology relationship.

The following diagram gives an overview of the primary divisions in the ChEBI role ontology together with examples of each and molecules which are linked to that role via the 'has role' ontology relationship.



For each drug entity in ChEBI or more roles are associated with the entity via the has_role relationship to the ChEBI role ontology to capture the biological action or the drug type. It

should be mentioned that drug types are such by virtue of the formal system and process by which drugs are approved for certain treatments, while biological action simply is such by virtue of what occurs within a biological system when that molecule is imbibed. These two axes of classification can be thought of as corresponding to the "formal classification" of the drug and the "mode of action" of the drug. It is clear that for something to be listed as a drug, a drug classification will be available, however, the mode of action is not always known. It is also clearly the case that a particular molecule may be approved as multiple different kinds of drug, and furthermore may have multiple modes of action.

A sample of some of the drug types in ChEBI are: antiinfective drug, diuretic, antiglaucoma drug, antiemetic, fibrin modulating drug, reproductive control drug, nutraceutical, anti-ulcer drug. A sample of some of the biological roles in ChEBI are: xenobiotic, molecular messenger, nodulation factor, osmolyte, toxin, provitamin, fixative, immunomodulator, epitope. Full lists may be retrieved by browing the ChEBI ontology or searching via the web interface or web services. Furthermore, the ChEBI public web interface provides a mechanism for searching for all compounds associated with a given bioactivity or drug class in the ontology. This is achieved by selecting the 'Advanced Search' option 'Filter by ontology term' with the 'has role' relationship specified and the target drug or bioactivity class selected. This functionality is also available via the ChEBI web service method getAllOntologyChildrenInPath.

Funding and outlook

ChEBI received 6 months' curation funding dedicated to curate DrugBank data in 2009. However, this funding was not extended, and we now address DrugBank curation on an ad-hoc basis along with our other curation projects, and gradually we are increasing our drug coverage. ChEBI currently has four full-time curators and another one will be joining the team soon, but of course the backlog of project commitments and user requests remains exceptionally high.