Recent Developments in the ChEBI Ontology

Janna Hastings, Paula de Matos, Adriano Dekker, Marcus Ennis, Kenneth Haug, Zara Josephs, Gareth Owen, Steve Turner, Christoph Steinbeck

Chemoinformatics and Metabolism, European Bioinformatics Institute, Cambridge, UK

Introduction

ChEBI — Chemical Entities of Biological Interest — is an ontology of chemical entities such as molecules and ions, and their roles in biological contexts [1]. As of April 2011, it contains in total around 25,000 classes. Here, we report on recent developments and changes in the ontology, and give a brief view on ongoing work that will lead to changes in the future.

1 Recent changes

1.1 Mapping to upper-level ontology BFO

In order to comply with our goal of increasing interoperability with other ontologies in the biomedical domain, ChEBI has undertaken to provide a mapping to the upper level ontology BFO (Basic Formal Ontology) [2]. Mapping multiple ontologies beneath a common upper level allows easier linking between ontologies, since it reduces ambiguities in interpretations through a clear ontological commitment.

The ChEBI mapping to BFO is illustrated in Figure 1, and provided as an OWL file, which is downloadable from:

ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/

1.2 Renaming of 'Molecular Structure' Root

ChEBI renamed the root term of the sub-ontology in which chemical entities such as molecules and ions are defined, from 'molecular structure' to 'chemical entity'. While the historical name accurately reflected the organising principle of the sub-ontology (the classification of entities therein is on the basis of structural features), it was not adequate for purposes of automated reasoning, since it led to incorrect inferences through the transitivity of the *is a* relationship,

such as:

caffeine is a molecular structure.

After the modification, we have the correct inference:

caffeine is a chemical entity.

1.3 Expanded Substance Hierarchy

In order to adequately deal with user-requested mixtures and polymers within the ontology, ChEBI has expanded its 'chemical substance' hierarchy. This reflects a slight change in scope relative to earlier versions of the ontology, which tried to explicitly exclude aggregate chemical substances. The changes has been introduced to allow adequate classification of some of the entities which were strongly requested by our users, and for which no other suitable ontology yet existed.

We have created a new upper-level term beneath 'chemical entity': chemical substance. We further differentiate between pure and mixed substances. An example of a pure substance is a macroscopic homogeneous collection of molecular entities (such as, say, while a mixture contains non-homogeneous collection - composed of at least two different sorts of entity. In particular, this allows us to correctly model racemic mixtures, which are crucial in the adequate representation of drugs, since many active substances found in drugs are formulated as racemic mixtures. Most chemical databases skirt the issue of representing racemic mixtures, or do so inconsistently. (For a discussion of this point, see: http://chem-blaics.blogspot.com/2011/02/chemical-data-curatio n-yes-it-is-that.html)

Our preferred ontology representation for racemic mixtures is illustrated in Figure 2.

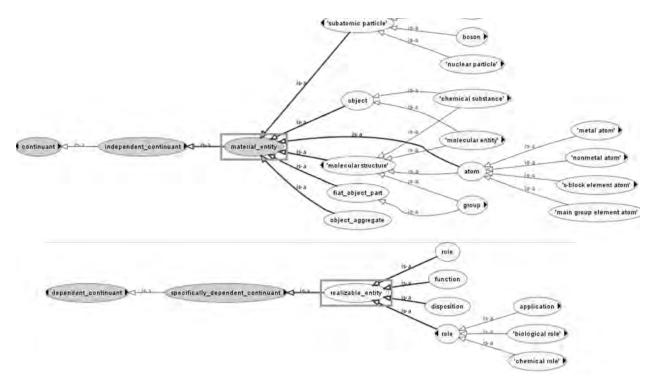


Figure 1. BFO Mapping.

The figure illustrates the mapping between upper-level ChEBI classes and the BFO terms to which they map, separated between ChEBI chemical entities and subatomic particles as independent continuants and ChEBI roles as realizable entities. Note that ChEBI 'role' does not map to BFO:role but to BFO:realizable entity.

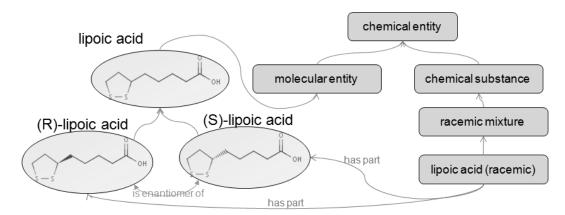


Figure 2. Racemic mixtures: Mixtures in ChEBI are explicitly modelled and their parts are linked via 'has part' relations.

1.4 Refactoring Natural Products

Natural products are of substantial interest in drug discovery and metabolism research, since they represent molecules that in many cases have been shaped by natural selection to be bioactive in highly specific ways. However, the core definition of what specifically constitutes a natural product is seldom rendered explicit, and differs from community to community. Some

candidate definitions among the many possibilities, in sequence from more inclusive to more exclusive, are:

- 1. All chemicals that can be isolated from a living organism;
- 2. All metabolites (primary and secondary);
- 3. Secondary metabolites only;

4. Secondary metabolites in plants only.

ChEBI currently includes classes related to natural products in two different places in the ontology. Firstly, common natural product families are explicitly classified in the chemical entity ontology, and secondly, 'metabolite' is specified in the role ontology. An example of a natural product class in the chemical entity ontology is:

cinchonine *is a* heterocyclic natural product *is a* natural product.

The text definition¹ is as follows: "Cinchonan in which a hydrogen at position 9 is substituted by hydroxy (S configuration). It occurs in the bark of most varieties of Cinchona shrubs, and is frequently used for directing chirality in asymmetric synthesis."

Due to the inherent ambiguity, current curation efforts involve the deprecation of classes explicitly containing 'natural product' in their name, and instead classifying molecules such as cinchonine explicitly as secondary metabolites. Future work will involve adding the species as explicit context to the definition of metabolites.

2 Ongoing Work and Future Changes

2.1 Focus on immunology

A large-scale ongoing curation effort in collaboration with the La Jolla Institute for Allergy and Immunology (LIAI, http://www.liai.org/) is focused on annotating compounds relevant for immunology, such as those which act as antigens and immunogens. ChEBI has so far annotated more than 1,000 such compounds.

2.2 Relationship Definition and Re-Evaluation

ChEBI is undergoing a major re-evaluation of the relationships which it makes use of, in order to bring them in line with the RO [3] where possible, and to provide formal definitions for chemistry-specific relationships. As part of this ongoing process, ChEBI will introduce RO relations such as *disjoint from*, and may

deprecate some of the chemistry-specific relations such as *has parent hydride* if they prove resistant to full logical definition.

2.3 Disentangling Role and Chemical Entity

Prior to 2009, the is a relationship in ChEBI was overloaded, linking molecular entities with chemical classes and specifying the 'roles' that chemical entities can enact in various contexts. To address this, the relationship has role was introduced and used to link molecular entities to roles, for example, the molecular entity acetylsalicylic acid (CHEBI:15365) has role non-narcotic analgesic (CHEBI:35481). The disentanglement was performed programmatically, and subsequent manual curation was required to clean up some cases where errors occurred, such as when a chemical entity lacked a structure and was only classified with a role parent. Current curation efforts are underway to fully define classes which are specified with both structural and role-based features. such as the entity tricyclic antidepressant (CHEBI:36809), which defined as is a organic tricyclic compound and has role antidepressant.

References

- de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S., Steinbeck, C.: Chemical Entities of Biological Interest: an update. Nucl. Acids Res. 38, D249–D254 (2010)
- 2. Grenon, P., Smith, B.: SNAP and SPAN: Towards dynamic spatial ontology. Spatial Cognition & Computation: An Interdisciplinary Journal 4(1), 69–104 (2004).
- 3. Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A., Rosse, C.: Relations in biomedical ontologies. Genome Biology 6(R46) (2005)

¹ Sourced from ChEBI version 78, http://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:27509