ICBO: International Conference on Biomedical Ontology
July 24-26, 2009 ▪ Buffalo, New York, USA

An OWL-DL Ontology for Classification of Lipids

Hong-Sang Low[1], Christopher J.O. Baker[2], Alexander Garcia[3], Markus R. Wenk[1]
[1]National University of Singapore, Singapore
[2]University of New Brunswick, Saint John, Canada
[3]University of Bremen, Bremen, Germany

Abstract

Lipids can be systematically classified according to functional properties, structural features, biochemical origin or biological system. However Lipid nomenclature has yet to become a robust research tool since no rigorous definitions exist for membership of specific lipid classes. Lipids need to be defined in a manner that is systematic yet at the same time semantically explicit. We report on the reuse of existing lipid nomenclature, ontology describing chemical structure and the extension of the OWL-DL Lipid Ontology to support the classification of lipid molecules. We applied definitions, DL-axioms, to describe lipids classes and illustrate suitability of the ontology for the classification of Fatty Acyl lipids and Mycolic acids.

Introduction

IUPAC-IUBMB proposed a systematic nomenclature for lipids which received limited adoption by the lipid community. The proposed classification was complicated and prone to erroneous application by scientists. Moreover the naming scheme was not extended and does not adequately represent many novel lipid classes discovered in the recent decades. As a result lipids still lack systematic classification and a nomenclature that is universally adopted by the biomedical research community. The LIPIDMAPS consortium[1] aims to resolve this by introducing a scientifically robust, comprehensive and extensible classification system evolved from the IUPAC nomenclature. This classification scheme organizes lipids from different phyla and synthetic domains yet uptake by the lipid community has been slow and the literature is steeped with instances of lipid synonyms that fail to reflect the new nomenclature.

Hierarchical Classification of Lipid Nomenclature

Lipids are organic compounds and can be systematically classified according to various features e.g. atomic connectivity, physicochemical properties, presence of functional groups, or types of bioactivities. Albeit an important contribution, the LIPIDMAPS central repository of lipids has primarily used is-a relationships[2] to categorize lipids and many definitions describing LIPIDMAPS lipid classes remain implicit. Moreover they are often dependent on a chemical diagram in the form a molecular graphic file that can only be accurately classified by a trained lipid expert. No rigorous definition, independent of a graphical diagram, exists and the graphical definitions are not flexible, nor are they extensible. Changes in such definitions require the redrawing of the chemical diagram/definition. Subsequently, communicating, storing and transferring of such structural definitions in the current format is inefficient and there is much reliance on trained experts. There is therefore a need for lipids to be defined in a manner that is systematic and explicit. A rigorous definition would involve a minimal necessary and sufficient declaration for class membership that could adequately describe a lipid without requiring a molecular structure diagram.

Description logics (DL) describe a domain using class descriptions according to a logic based semantic. In previous work DL has been used to represent chemical knowledge[3,4]. Using DL, it is possible to define a lipid with necessary conditions such that an alpha mycolic acid is defined as a lipid that minimally has alpha-hydroxyl acid and cyclopropane groups. Moreover, we can define necessary and sufficient conditions limiting the definition of an alpha mycolic acid to a lipid that has only alpha-hydroxyl acid and cyclopropane functional groups. Consequently molecules that have functional groups other than alpha-hydroxyl acid group and cyclopropane groups cannot be considered as an alpha mycolic acid.

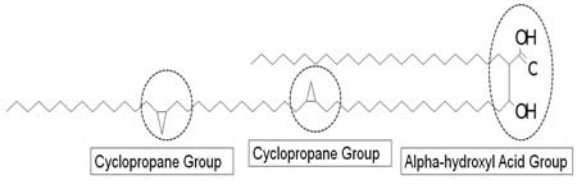

Figure 1. An example of alpha mycolic acid

The Lipid Ontology

The Lipid Ontology was exclusively developed to conceptualize and capture knowledge in the domain of lipids through the use of concepts, relations, instances and constraints on concepts[5]. It was designed to provide a common terminology for the

lipid domain, a basis for interoperability between information systems and to support navigation of text mining results from lipid literature. The ontology has been extended to describe the LIPIDMAPS nomenclature classification explicitly using description logics (OWL-DL) and to support reasoning and inference tasks. Prior to extending the ontology for classification tasks we reviewed existing chemo-ontologies[6,7] for reusable components. We reviewed the Chemical Ontology[6] for reuse of functional group specifications in the *Organic_Group* hierarchy. We enriched the Lipid Ontology with 32 functional groups from Chemical Ontology and 63 new concepts were added under the *Organic_Group* super-concept. The *Organic_Group* hierarchy was reorganized and asserted with new is-a relationships. From Chemical Ontology, we also used hasPart to relate concepts of lipids to concepts under *Organic_Group*. In reviewing the ChEBI[7] Ontology we identified that it is currently undergoing major revisions to correct inconsistent use of 'IsA' and 'IsPartOf' properties. We opted not to re-use its organization and relationship definitions, moreover to represent a systematic lipid nomenclature using formal logical definition of classes, we do not yet need all the relationship definitions found in ChEBI To further facilitate the reuse of the formal definitions in the lipid ontology we provide a high level alignment to ChEBI using SAMBO[12]. The alignment is available online at: http://www.lipidprofiles.com/LipidOntology/Others/SAMBO_0.rdf

Functional Groups Used in Lipid Classification

Lipids can have a wide range of distinct functional organic groups that should be accommodated in their conceptualization and classification. Distinct combinations of these organic groups underpin the definitions of lipid classes and membership of lipid classes can be restricted by formal descriptions which refer to functional groups. While the Chemical Ontology[6] describes basic functional groups, a wider range of functional groups are needed to describe lipids. To equip the Lipid Ontology for use as a classification tool we added 400 DL definitions to all lipid classes, with the exception of polyketides (Table 1). Primarily we re-used, from Chemical Ontology, the axiom "*Organic_Compound* hasPart *Organic_Group*" to relate *Lipid* concepts to *Organic_Group* concepts. We then defined concepts to describe lipid functional groups, namely *Organic_Group* and *Ring_System*. *Organic_Group* has three sub-groups (i) *Simple_Organic_Group*, (ii) *Complex_Organic _Group*, (iii) *Chain_Group*. *Simple_Organic_Group* subsumes concepts that describe basic functional groups whereas *Complex_Organic_Group* encapsulates glycans and amino acids. Glycans, in particular, are used to classify lipids such as sacharrolipids, and other sugar-linked lipids such as sphingolipids. *Chain_Group* consists of the *Carbon_Chain_ Group* and the *Sphingoid_Base_ Chain_Group*. The *Sphingoid_ Base_Chain_Group* is used exclusively for sphingolipids whereas *Carbon_Chain_ Group* is applied to other lipid classes accordingly. The *Ring_System* consists of (i) *Isoprenoid_ring_ derivative*, (ii) *Monocyclic_Ring_ Group* and (iii) *Polycyclic_Ring_System*. These concepts are used to define lipids that have one or more rings, primarily sterol, prenol and other ring lipids. In Lipid Ontology these concepts are extensively used to provide the necessary structural descriptions to define the identity of lipid-based compounds.

Total No. of Classes	715
No. of Lipid Classes	428
Primitive Lipid Classes	162
Defined Lipid Classes	266
Total No. of Restrictions	901
Total No. of Properties	41
DL Expressivity	ALCHIQ(D)

Table 1. Summary of the current Lipid Ontology

Hierarchical Classification of Lipids

Lipid concepts are organized hierarchically with the super-classes restricted by generic necessary conditions. More specific necessary conditions are used to define membership requirements for sub classes of lipid. At the end of a hierarchy, lipid classes are restricted by necessary and sufficient conditions and closure axioms. Super-classes are not closed by closure axiom to avoid inconsistency among disjointed sibling classes. More specific lipid classes are defined in two ways. In the first approach we specify the subclass of the present class to restrict the definition of a lipid. Necessary conditions such as "hasPart some *Carboxylic_Acid_derivative_ Group*" can be further specified by the subclass of *Carboxylic_Acid_Derivative_Group*, e.g. aldehyde. The second approach uses a Cardinality Axiom that restricts the number of a particular concept to be allowed in a restriction. Lipid classes can be defined by the number of certain functional group concept or *Chain_Group* concept. For example, a triacylglycerol is an acylglycerol with 3 acyl chains. Its superclass is restricted with an existential axiom "has some *Acyl_ Chain*". This is further specified with the following cardinality axiom "hasAcyl_Chain exactly 3". We are currently exploring Qualified Cardinality Axioms in

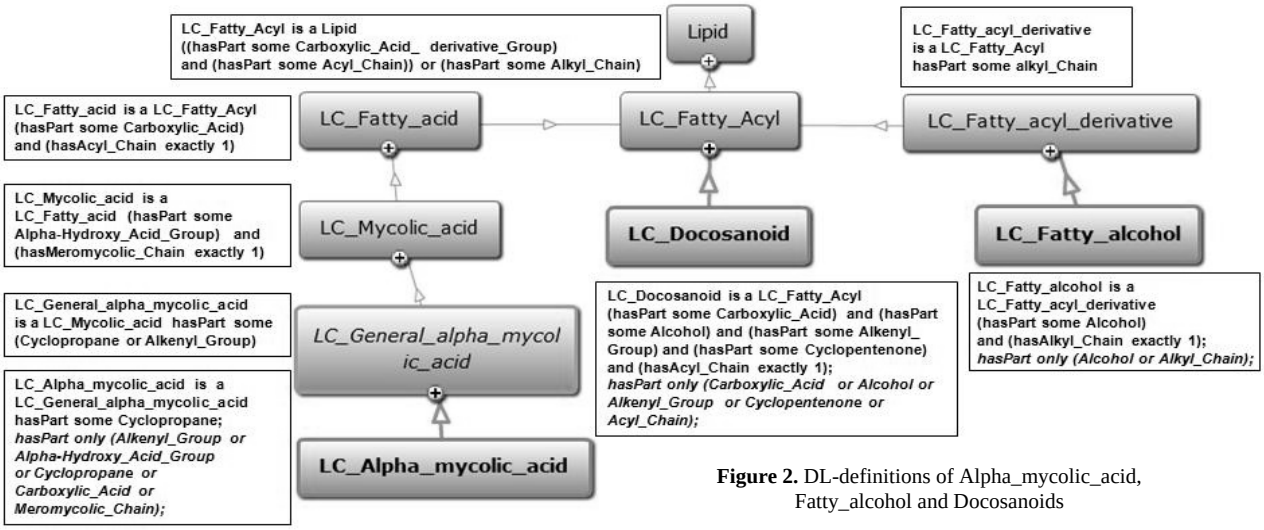

Figure 2. DL-definitions of Alpha_mycolic_acid, Fatty_alcohol and Docosanoids

OWL 2.0 as a means of defining lipid classes. For example we can define the class *Triacylglycerol* with "hasPart exactly 3 *Acyl_Chain*" without additional properties such as hasAcyl_Chain. However, an *Acyl_Chain* would have a carboxylic acid functional group encapsulate within it. When defining lipids with "hasPart exactly 3 *Acyl_Chain*", care must be taken not to add functional groups that entail others.

DL Axioms for the definition of Fatty_Acyl

Fatty acyls are a diverse lipid group synthesized by chain-elongation of an acetyl-CoA primer with malonyl-CoA/methylmalonyl-CoA groups[1]. We define a fatty acyl as a lipid that has at least one *Carboxylic_Acid_derivative_Group* and at least one *Acyl_Chain*. *Docosanoid* is a subclass of fatty acyls that inherits from *Fatty_Acyl* class the *Carboxylic_Acid_derivative_Group* as well as *Acyl_Chain*. The *Carboxylic_Acid_derivative_ Group* in *Docosanoid* is further specified to be a *Carboxylic_Acid*, whereas the *Acyl_Chain* is specified with a cardinality axiom and the property hasAcyl_Chain. Consequently, *Docosanoid* is defined to have only 1 *Acyl_Chain*. In addition, *Docosanoid* can have multiple and distinct functional groups such as *Carboxylic_Acid*, *Alkenyl_Group*, *Alcohol* and *Cyclopentenone*. These functional groups are associated with the class *Docosanoid* via the property "hasPart" in conjuction with the existential axiom "some". A closure axiom is needed to restrict the type of relationship constraints allowed for a lipid class. The closure axiom is applied to the class *Docosanoid* so that lipids of this class can only have the following functional groups, namely, *Carboxylic_Acid*, *Alkenyl_Group*, *Alcohol*, *Cyclopentenone* and *Acyl_Chain*. (Figure 2). As LIPIDMAPS nomenclature classifies lipids based on chemical structure or biosynthetic origin, lipids such as fatty alcohols are classified as fatty acyls in spite of having no *Acyl_Group*. When considered structurally, this classification of lipids is not systematic. We address this shortcoming in LIPIDMAPS nomenclature by expanding the definition of *Fatty_Acyl* to include *Alkyl_Chain*, a characteristic structure of un-usual *Fatty_Acyl* classes. In doing so a *Fatty_alcohol* inherits an *Alkyl_Chain* from *Fatty_Acyl* and is further defined to have a single *Alkyl_Chain* in the necessary and sufficient condition. This definition includes a "hasPart" property that connects *Fatty_alcohol* to an *Alcohol* class allowing inclusion of a lipid without an acyl group as a member of *Fatty_Acyl* (Figure 2). In addition we create a new lipid class, namely *Fatty_Acyl_derivative*, a subclass of *Fatty_Acyl*. Using the flexibility of OWL-DL, we can begin to address inconsistencies in lipid classification grounded in lipid definitions that are non systematic.

Extension of the Mycolic Acid Class

Mycolic acids are a key component of the cell wall of *Mycobacterium tubeculosis sps*. and are implicated mycobacterial disease. By 1998 there existed 500+ known chemical structures of related mycolates[8] and yet LIPIDMAPS currently contains only 3 mycolic acid records. Consequently many mycolic acids with known structures have yet to be systematically classified. Classification of these lipids is critical for system-level analysis of mycobacterial pathogenesis. We illustrate extension of Lipid Ontology to include new *Mycolic_Acid* classes and demonstrate classification of a real instance of an alpha mycolate (Figure 1) to the appropriate class. Based on LIPIDMAPS nomenclature, we assign *Mycolic_acid* as a member of *Fatty_Acid* and extend *Mycolic_acid*

classification to 9 defined subclasses (Table 2 http://www.lipidprofiles.com/LipidOntology/Others/Table2.jpg), distributed among three primitive superclasses. Alpha mycolic acid is a mycolic acid that has cyclopropane and alpha-hydroxyl acid (a special class of carboxylic acid) groups. The carboxylic acid group is a member of the acyl group, an ester group. Therefore, according to the classification scheme below, alpha mycolic acid must be a member of *Fatty_Acyl*. Among members of *Fatty_Acyl*, only *Octadecanoid, Docosanoid, Eicosanoid* and *Fatty_Acid* have carboxylic acid. Alpha mycolic acid does not have a cycloketone group and therefore, it cannot be *Docosanoid, Eicosanoid* or *Octadecanoid* and must be a member of *Fatty_Acid*. Among members of *Fatty_Acid*, only *Mycolic_acid* has an *Alpha-Hydroxy_Acid_Group* and a *Meromycolic_ Chain*. Therefore, alpha mycolic acid is classified under this class of *Fatty_Acid*. Since *Alpha_mycolic_acid* is the only class that accepts mycolic acid with Cyclopropane, the lipid in Figure 1 is classified as a member of *Alpha_mycolic_acid*. (Figure 2).

Conclusion

Lipid research is increasingly integrated within systems level biology such as lipidomics[9] where lipid definition and classification are required before annotation of chemical functions can be applied. In this paper we have sought to address the ongoing challenge of classifying lipids through the adoption of W3C standard knowledge representation and the application of DL axioms. In other domains of metabolomics, e.g. glycomics, the adoption of ontologies such as, GlycO – a focused ontology representing complex carbohydrates, have enabled correlation of structural features of glycans to the biosynthesis and metabolism[10]. We initiated the process of defining lipids according to appropriate functional groups with the intent of using the ontology for classification of lipids. Ontology driven classification has been applied to proteins[11] and small molecules[6] through the coordination of protein domain or pharmacophore analysis, OWL-DL ontology, and DL reasoning. By adding precisely defined DL-axioms to the lipid ontology we can apply a similar approach for the automated classification of lipids. Our approach is extensible to accommodate novel lipids and we extended the use of DL-axioms to classify all lipid classes (except for polyketides). In support of mycobacterial disease research, we extended lipid nomenclature and classification of mycolic acids. We have made available systematic and formalized OWL-DL definitions of lipids for testing the appropriateness of existing nomenclature to lipid structures. This will serve as a reusable standard for lipid researchers and the lipid bioinformatics community. The Lipid Ontology is available online at NCBO's Bioportal and at: http://www.lipidprofiles.com/LipidOntology/LiPrO-02042009.owl

Acknowledgements

NUS Office of Life Science (R-183-000-607-712), ARF (R-183-000-160-112), BMRC A*STAR (R-183-000-134-305), Singapore NRF under CRP award No. 2007-04, NBIF, New Brunswick, Canada.

References

1. Fahy E, *et al*. A comprehensive classification system for lipids. *J. Lipid Res.* 2005, 46: 839–62.
2. Sud M, *et al*. LMSD: LIPID MAPS structure database. *Nucl. Acids Res.* 2007, 35: D527–32.
3. Baader FI, *et al*. Description logics as ontology languages for the semantic web. In *Festschrift Jörg Siekmann, LNAI* 2003.
4. Villanueva-Rosales N, Dumontier M. Describing chemical functional groups in OWL-DL for the classification of chemical compounds. OWLED 2007, co-loc. ESWC2007 Innsbruck, Austria.
5. Baker CJO, *et al*. Towards ontology-driven navigation of the lipid *bibliosphere. BMC Bioinformatics*. 2008, 9 (Suppl 1):S5.
6. Feldman HJ, *et al*. CO: A Chemical Ontology for Identification of Functional Groups and Semantic Comparison of Small Molecules. *FEBS Letters.* 2005, 579: 4685–4691.
7. Degtyarenko K, *et al*. ChEBI: A database and ontology for chemical entities of biological interest. *NAR.* 2008, 36: D344–D350.
8. Barry CE, *et al*. Mycolic acids: structure, biosynthesis and physiological functions. *Prog. Lip. Res.* 1998, 37: 143–179.
9. Wenk MR. The emerging field of Lipidomics. *Nat. Rev. Drug Discov.* 2005, 4: 594–610.
10. Thomas CJ, *et al*. Modular ontology design using canonical building blocks in biochemistry domain. FOIS pp. 115–127, 2006.
11. Wolstencroft K, *et al*. Protein classification using ontology classification. *Bioinf. 2006*, 22: e530–8.
12. http://www.ida.liu.se/~iislab/projects/SAMBO/