IMPROVING THE STRUCTURE OF THE GENE ONTOLOGY

A thesis submitted to the University of Manchester
for the degree of Master of Science
in the Faculty of Science and Engineering

2004

Mikel Egaña Aranguren
School of Biological Sciences

Contents

List of Tables

List of Figures

Abstract

The Gene Ontology has given the scientific community a means of resolving the semantic heterogeneity of biological database annotations by providing an ontology that describes the functions of gene products of an idealised general cell. The Gene Ontology is a controlled vocabulary of terms, where the terms are related by the logical relationships *IsA* and *PartOf*, building a tree-like structure that can be queried. As a result of its good design and community participation, the Gene Ontology has grown in the last years to the point that hand-crafted curation is becoming more difficult, specially maintaining all the neccesary logical relationships for an accurate and sound representation of biological knowledge about functions of gene products. This project is the continuation of the GONG project (Gene Ontology Next Generation), which aims to demonstrate that migrating the Gene Ontology to a more expressive environment, like Description Logics, will lead to a more automated curation, detecting possible missing logical relationships. The GONG methodology has been applied to certain areas of the Gene Ontology (*binding*, *transporter activity* and *metabolism*), migrating them to a Description Logics environment (Ontology Web Language -OWL-) and pointing missing links by automated reasoning, demonstrating the utility of the approach for improvement of the Gene Ontology. Other possible applications of Description Logics for changing the structure of the Gene Ontology are given.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Copyright

Dedication

This work is dedicated to my grandfather, Juanito Aranguren, who died this year after a tough life fighting against fascism in the Spanish civil war and during the following dictatorship. He showed me that there are people in this world ready to give everything in the name of justice and freedom.

Acknowledgements

I would like to thank first Robert Stevens and Chris Wroe from Manchester University. Robert Stevens agreed to be my supervisor, encouraged me, suggested interesting points and checked the manuscript. Chris Wroe wrote the original scripts (including the installation of the UMLS lexical resources), helped me setting up the workflow, answered all my questions and checked the manuscript. Thanks must be given as well from Manchester University to Christopher Garwood for correcting the manuscript, Mathew Horridge for developing a special Protégé plugin that made importing and reasoning possible, and to G.D Gough for developing the Manchester University Computer Science Thesis LaTeX 2_ε class, which made the writing of this manuscript much easier after the neccesary changes to the class were made.

I would like to acknowledge the help or support of the following people outside Manchester University: Jane Lomax from EBI reviewed as promptly as she could my new relationships, Asha Teo was there in the tough moments, and my family gave me plenty of support during this year.

List of abbreviations

OWL: Ontology Web Language.
GO: Gene Ontology.
DL: Description Logics.
GONG: Gene Ontology Next Generation.
OBO: Open Biological Ontologies.
UMLS: Unified Medical Language System.
MeSH: Medical Subject Headings.
MGI: Mouse Genome Informatics.
SGD: *Saccharomyces* Genome Database.
DAG: Directed Acyclic Graph.
XML: eXtensible Markup language.
GOA: Gene Ontology Annotation project.
GOAT: Gene Ontology Annotation Tool.
KR: Knowledge Representation.
W3C: World Wide Web Consortium.
DAML: Darpa Agent Markup Language.
OIL: Ontology Inference Layer.
DARPA: Defense Advanced Research Projects Agency.
RDFS: Resource Description Framework Schema.
UML: Unified Modeling Language.
TaO: TAMBIS Ontology.
TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources.
RDF: Resource Description Framework.
CUI: Concept Unique Identifier.

Chapter 1

Introduction

The number of bioinformatics resources have grown immensely since the discipline first formed, presenting new problems and pushing computer science to new limits. At the same time, bioinformatics has adapted to the newest advances in computer science more rapidly than other disciplines [53] because of the peculiar characteristics of the discipline itself. The most important of these characteristics can be roughly resumed in addressing that bioinformatics deals with *large amounts* of *complex* information, unmanageable for a scientist without sophisticated knowledge management and information processing tools [78]. This is not surprising given that bioinformatics is evolving to address issues related to such a complex phenomenon as life itself at a purely informational level. Thus bioinformatics is a *knowledge driven* discipline [96, 52] where the core is the knowledge surrounding the sequences and data that are obtained in the sequencing projects, high-throughput experiments, and other techniques. Those data are growing at an unprecedented rate [100, 86] but the *knowledge* (reliable, accessible and manageable information) surrounding them is not growing at the same pace. There are different reasons for this lack of productive knowledge but they can be summarised by realizing that biological phenomena can be described in many different ways [42, 87] and this complexity has not been tackled in a semantical level. That means that usually the biologists are left with a giant domain of information that they cannot access, analyse, or integrate in a sensible way [31].

A way of minimising this problem comes from what it can be defined as *semantic computation*: making computers *understand*[1] the information they manage

[1]The word *understand* is not used here in its complete cognitive sense; it is referring to computational tools that manage the meanings of the concepts in logical ways, not artificial

reduces tremendously human work. Perhaps more importantly, apart of reducing human intervention semantic computation creates new and more powerful ways of managing information. The biggest exponent of this trend is the controversial semantic web[2] [10]. The semantic web aims to transform the actual web, focused on human processing, into a computer-processing web where computers *understand* the web content, making new services and processes possible. It has been defined as *the conceptual structuring of the Web in an explicit machine-readable way* [9]. Another exponent, closely related to the semantic web, are ontologies.

Ontologies can be defined as vocabularies with well defined terms that capture the knowledge about a domain; the biggest difference with a dictionary is that the terms in an ontology are related via logical relationships. Ontologies appeared in computer science in the artificial intelligence research community and since then their use has grown greatly.

In the case of bioinformatics, ontologies have been used as an integration tool to tackle the *complexity* and *immensity* of biological and medical information by way of converting it in useful knowledge. Ontologies have been widely applied in bioinformatics and medicine but the most famous example is the Gene Ontology. The Gene Ontology (GO) describes the molecular functions, biological processes, and cellular locations of gene products in an ontology consisting of:

- Terms: the *names* of the functions or related components of gene products, for example *protein serine/threonine kinase activity.*
- Term definitions: a brief statement describing the term. In the case of the term *protein serine/threonine kinase activity*, the definition is:

  > *Catalysis of the reaction: ATP + a protein serine/threonine = ADP + protein serine/threonine phosphate*

- Logical relationships relating terms: they can be of two types, *IsA* or *PartOf.* Logical relationships build a structure of terms (figure 1.1).

The Gene Ontology has had a considerable success, with many databases linking to it and several *satellite* projects extending its functionality beyond its original aims. As a consequence of this success, the Gene Ontology has become

intelligence agents that simulate complex and more cognitive tasks as pattern recognition, problem solving, etc. Undoubtedly both research domains are related but the distinction is made here to clarify the context of this project.

[2] `http://www.w3.org/2001/sw/`

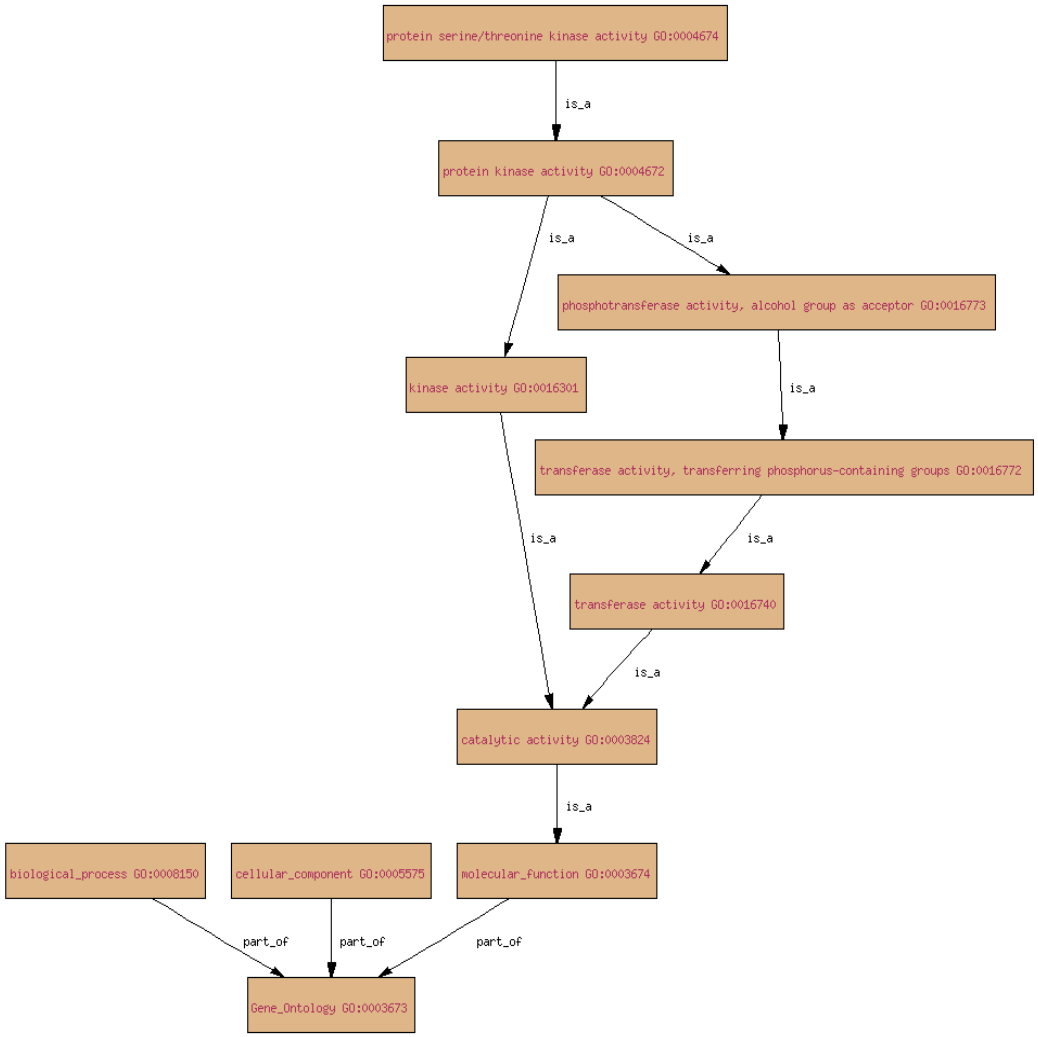

Figure 1.1:
Gene Ontology structure of the term *protein serine/threonine kinase activity*. The structure is constructed using logical relationships (*IsA* and *PartOf*). Taken from the AmiGO browser (See Chapter 2, section 2.2).

very large, with more than 17,000 terms by July of 2004. Maintaining such a big ontology by hand is extremely difficult, especially making sure that all of the neccesary relationships are present, i.e. detecting and fixing missing *IsA* or *PartOf* relationships between terms.

The principal aim of this project is to demonstrate that migrating the Gene Ontology to a more expressive ontological language based in Description Logics (DL) can help maintain semi-automatically an exhaustive structure of the ontology where all the neccesary relationships between terms are present. Description

Logics is a way of implementing ontologies, with a sound mathematical basis, more expressive than a simple *IsA* and *PartOf* structure, that allows the use of *reasoning*. In reasoning, a program called a *reasoner* can infer the consistency, the missing links and other aspects of a given ontology or ontologies. As a result the Gene Ontology will become a more efficient tool for knowledge management.

This document is organised into six chapters. Following this introduction the chapter two gives detailed descriptions of ontologies, the Gene Ontology and Description Logics. The chapter three gives an overview of the GONG (Gene Ontology Next Generation) methodology, adapted for this project. The chapter four gives the results obtained applying the methodology to certain areas of the Gene Ontology. The chapter five offers a discussion of those results and an analysis, with suggestions for further developments regarding the Gene Ontology and Description Logics. The conclusion of the project is given in the chapter six.

Chapter 2

Background

The aim of this chapter is to provide an overview of the technologies and resources involved in this project: ontologies, the Gene Ontology and Description Logics (OWL). The last section of the chapter links these elements with the methodology of the project (GONG), which is described in the chapter three.

2.1 Ontologies

2.1.1 What are Ontologies?

The term *Ontology* is controversial, and different people consider different things to be ontologies. There is a certain consensus in what an ontology is not. It is not a taxonomy (is not just a class-subclass hierarchy), a dictionary (includes relationships between terms), nor a knowledge base that includes individual objects. In the other hand, the range of what is considered to be an ontology is wide [16]. In this case ontologies will be considered as structured vocabularies with logical relationships between terms, where the terms are precisely identified and defined resulting in a structure of some kind.

The conceptual origin of ontologies can be traced back to early philosophers [92] because an ontology, in abstract terms, is no more (and no less) than a representation of reality. Therefore the term *ontology* is used in philosophy to name the discipline that tries to describe reality[1] but in more technical fields is used to name the description itself, not the discipline that describes reality[2]. The

[1]The term ontology comes from the Greek *ontos* (being) and *logos* (word) [16].

[2]See the paper by Smith *et al.* [91] for an interesting review of philosophical vs. purely technical ontology creation strategies. Smith *et al.* propose some ways of improving the Gene

first formal and explicit approach to ontologies in the more technical sense dates back to 1900, given by Husserl [72].

Ontologies entered the computer science field by way of artificial intelligence in the 1980's as a way of describing systems with knowledge about a domain [80]. The ontologies in that moment were defined as *the specification of a conceptualisation* [38], a definition that is widely used. Ontologies have the following aspects [3]:

Complete: ontologies are designed to capture the maximum amount of relevant concepts of the domain they represent.

Formality: ontologies are built using mathematical formalisms, making them suitable for computing tools.

Understandable by humans: ontologies are built on a more or less intuitive manner using natural language terms, making them accessible for scientists.

General: ontologies aim to represent conceptual domains independently of any specific use or implementation [14].

The design and creation of ontologies is a difficult process because many factors must be taken in account ranging from formal principles of ontology making to intuitiveness, including coverage of the domain, fidelity, computational tractability, etc. An important obstacle in ontology design and creation is the difficulty of biologists or domain experts to understand the ontology structure and design principles behind it. For the creation of an ontology to happen, a consensus of its content must be reached which is not an easy task [15]. They are plenty of tools for ontology development. See the paper by Lambrix *et al.* [70] for an analysis of the performance of the different tools, including some tools used for this project.

2.1.2 Use of ontologies

Ontologies, as they become a more important tool for information management, are being used in plenty of different fields, such as the semantic web [94], web

Ontology, inherited from traditional philosophy.

agents [43] and web services [35], GRID technology [99], e-commerce [68], data mining and text mining [74, 62], computer security [75], and many more.

Despite the critics [17] (See the opinion paper by Hunter [51] for a response) ontologies have a growing use in biological and medical sciences. The ontologies used in this field are of three types, although the division is not always clear:

Task oriented: they are designed for concrete tasks such as data mining [56, 23] or resources integration [6]. Ontologies are important in resource integration because they can be used to handle the semantic heterogeneity of the nearly 550 molecular biology databases existing today [34]. The semantic heterogeneity refers to differences in the databases where an entry can refer to different things or different entries refer to the same thing [93]. Other ontology based systems like SEMEDA [65] or TAMBIS [97] deal with the semantic heterogeneity of databases on other levels, such as the names of the attributes on each database[3].

Domain oriented: they capture the knowledge of a concrete field. The ontology, apart of being queried, can be used as the centre for other technologies.

Generic: they are high level ontologies with general terms that are used to integrate different ontologies.

In the case of domain conceptualisations, some of the most used ontologies can be found in the Open Biological Ontologies site[4] (OBO). OBO is an umbrella project for biological ontologies that includes different ontologies that satisfy the following criteria:

- The ontologies must be open (no restriction in use).
- The ontologies must be implemented in standard ways (languages like OWL).
- The ontologies must be orthogonal to each other (independent).
- The ontologies must have a unique identifier prefix.
- Definitions of the terms of the ontologies must be given.

[3] They are other problems with databases integration, ranging from purely technical (storage methods, query languages) to legal ones.

[4] `http://obo.sourceforge.net/`

Other ontologies cover medical terminology in projects like GALEN [85], Microarray data[5], metabolic pathways in projects like EcoCyc [57] or PATIKA [31], proteins with inorganic groups [29], specific diseases [40] and cancer diagnosis [50], structural genomics [71] or taxonomic descriptions [83], amongst many more. The Unified Medical Language System[6] (UMLS) [13] integrates different medical ontologies in a single system and provides different tools to manage them. From the medical ontologies included in the UMLS the Medical Subject Headings[7] (MeSH) is the one used for this project.

The number and diversity of ontologies will grow in the following years as they have been demonstrated to be a useful tool not only for resource integration but also as tools of knowledge generation or prediction [19]. For example, it has been demonstrated that functional annotation of new sequences based on sequence similarity is not optimal [32] and semantic methods of functional annotation [12] based in ontologies represent an improvement.

2.2 The Gene Ontology

2.2.1 Introduction to the Gene Ontology

The Gene Ontology[8] is the most successful use of ontologies in bioinformatics [26]. It has been demonstrated to be a sensible way of dealing with the semantic heterogeneity of gene product annotations in other databases, by providing an ontology of gene products' function terms that refer to an idealised[9] cell [28]. The Gene Ontology is the most detailed attempt at this moment to describe the functions of gene products [79]. Thus, the Gene Ontology is an invaluable tool for resource integration [64], apart of claryfing communication between scientists. The Gene Ontology is responsibility of the GO consortium, a joint project formed by different organism databases[10] that was started by FlyBase [25], Mouse Genome Informatics (MGI) [11] and the *Saccharomyces* Genome Database (SGD) [66].

[5] `http://mged.sourceforge.net/ontologies/index.php`
[6] `http://www.nlm.nih.gov/research/umls`
[7] `http://www.nlm.nih.gov/mesh/`
[8] `http://www.geneontology.org`
[9] The scope of GO is limited: it doesn't express individual genes.
[10] See the following web for a complete list of partners:
`http://www.geneontology.org/GO.consortiumlist.html`

2.2.2 Structure and technical properties of the Gene Ontology

The Gene Ontology consists of three orthogonal ontologies [27]:

Molecular Function: processes at the molecular level.

Biological Process: assemblies of various molecular functions.

Cellular Component: cellular locations and macromolecular complexes.

The whole ontology is implemented using Directed Acyclic Graphs (DAGs): multiple parent-child relationships are allowed in the structure, but cycles (a term being a child of itself) are prohibited. The *top* of the hierarchy is populated by general terms (*binding* or *cell*) and as we move *deeper* (more terms in the path) the terms become more specialised. The terms on the edge of the path are called *leaves* [90]. The terms in the path are usually called *nodes*. Each term has an unique identifier or GOid apart from the term name, like GO:0005488 for the term *binding*. There are two types of relationships in GO:

IsA: it describes more concrete properties and it is also known as *subsumption* relationship; one term *subsumes* the other[11]. It can be described as a term being a subclass of a bigger class: *mitotic chromosome* is a subclass of *chromosome*, therefore *mitotic chromosome* ***IsA*** *chromosome*. It should be noted that the *IsA* relationship doesn't mean a concrete real example or instance [102]; as the GO users guide says, *clogs are a subclass or is_a of shoes, while the shoes I have on my feet now are an instance of shoes*. The *IsA* relationship is *transitive*, which means that a child is a subclass of its grandfather because its father is a subclass of its grandfather: if *Dog* is a *Mammal* and *Mammal* is a *Vertebrate*, *Dog* is a *Vertebrate*. For a more in depth explanation of the philosophical assumptions behind the *IsA* relationship see the conference paper by Nicola Guarino and Christopher Welty [39].

[11] `http://www.geneontology.org/GO.usage.html#isa`

PartOf: this means that a child is a component (in the cellular component ontology) or a sub-process (in the biological process ontology) of its parent[12] [27]. This relationship is transitive as well. See the conference paper by Smith *et al.* [93] for a critical analysis of the *PartOf* relationship in the Gene Ontology.

An important assumption behind GO is the *true path rule*[13]: starting from a leaf all the relationships that go up in the tree along its path must be true. For example *chitin biosynthesis* cannot be a child of *cell wall biosynthesis* (in fungi) because chitin is also used in building the cuticle of arthropods. Querying for genes of cell wall biosynthesis gave results of genes related to cuticle building that were annotated to chitin biosynthesis [45] and the structure was rearranged to avoid that. Another important aspect of GO organisation is the use of the word *sensu*: it is used when a term can have different meanings [76]. For example, the term *cell wall* (GO:0005618) can be used to refer to bacteria, fungi, and plants. In biological sciences the same word is used to refer to the three types of cell wall but the cell walls have different properties. Therefore the word *sensu* is added to the term, meaning *in the sense of*: the term *cell wall* (GO:0005618) has three children: *cell wall (sensu Bacteria)* (GO:0009274), *cell wall (sensu Fungi)* (GO:0009277) and *cell wall (sensu Magnoliophyta)* (GO:0009505).

GO can be explored using various tools, the most common one being the AmiGO web interface[14]. DAG-EDIT is another popular tool for editing and exploring ontologies in DAGs[15], which is a standalone program written in JAVA. GO ontologies can be obtained in different ways[16], including OBO format, flat files, XML (eXtensible Markup language), MySQL tables, etc. Apart from ontologies, other resources are available. Slims are high level slimmed down ontologies for analysing gene group annotations [28]. Annotations of other databases to GO are available in a list[17]. The databases that include GO annotations are: SGD (*Saccharomyces cerevisiae*), FlyBase (*Drosophila melanogaster*), TAIR (*Arabidopsis thaliana*), WormBase (*Caenorhabditis elegans*), RGD (*Rattus norvegicus*),

[12] `http://www.geneontology.org/GO.usage.html#partof`

[13] `http://www.geneontology.org/GO.usage.html#truePathRule`

[14] `http://www.godatabase.org`

[15] See the following web for a list of all the GO related tools, some of which are mentioned further on in the document: `http://www.geneontology.org/GO.tools.html`. DAG-EDIT can be downloaded in `http://sourceforge.net/projects/geneontology`

[16] `http://www.geneontology.org/index.shtml#downloads`

[17] `http://www.geneontology.org/GO.current.annotations.shtml`

Gramene (*Oryza sativa*), ZFIN (*Danio rerio*), DictyBase (*Dictyostelium discoideum*), TIGR, Sanger GeneDB, GenBank and UniProt. Every GO annotation needs an evidence code[18] that states where the evidence for the annotation came from (e.g. *inferred from direct assay*, *inferred from electronic annotation*, etc.). Mappings of GO to other external systems (e.g. Enzyme Commission numbers, SWISS-PROT keywords) are also available[19]; recently GO has been mapped to the UMLS [76].

2.2.3 Evolution and ampliation of the Gene Ontology

The growth and success of GO has been spectacular in recent years because of its openness, community involvement, intuitive structure and other reasons pointed to in a paper by Bada *et al.* [5]. It is a very dynamic project and full-time curators include the large amount of change requests from the community, supervised by each organisms' database staff. Plenty of new resources include GO annotations[20]. As a consequence, its functionality has been augmented to include, amongst others:

- The Gene Ontology Annotation project (GOA): assigns GO codes to other database annotations [21, 20].
- The Gene Ontology Annotation Tool (GOAT): closely related to GONG (see section 2.4 and chapter 3), this project aims to create a tool that helps creating consistent annotation when using GO terms [4]. It works with Description Logics (see section 2.3).
- Automated [44, 84, 105, 37, 60] or integrated [55] gene annotation.
- Use of semantic similarity for sequence searching [108, 77].
- Categorisation of gene groups [54, 109, 107, 8, 2]; given a large set of genes a node or nodes on the Gene Ontology are used to somehow summarise their function [53].
- Categorisation of gene expression [33, 61, 101, 73, 89, 88] and statistical genomics [22].

[18] `http://www.geneontology.org/GO.evidence.html`
[19] `http://www.geneontology.org/GO.indices.html`
[20] `http://www.geneontology.org/GO.annotation.html`

- Prediction of protein function by coupling machine learning with the Gene Ontology [69, 63, 30] and prediction of subcellular location of a given protein [24].

2.3 Description Logics and OWL

2.3.1 Frames based and Description Logics based languages

As ontologies represent a conceptualisation, they are implemented using *knowledge representation* (KR) languages. There is a wide range of these languages, differing mainly in the philosophy behind them, computational tractability and expressivity [36]. Two paradigms can be addressed: the frames based languages and the Description Logics based languages.

The frames based philosophy is conceptually analogous to object oriented programming: a frame represents a concept. A concept is a set of objects or instances. Therefore, in JAVA programming a frame would be a class. Frames contain slots representing its attributes and facets describe the scope and other properties of the slots, such as the values they can have. Relationships in this systems are implemented as special slots that can have another frame as its value. In this way frames inherit properties, like classes that inherit properties in object oriented programming languages. The notion of frames as a KR tool was first proposed by Marvin Minsky [103] and since then it has been growing in the field of knowledge representation because of its ease of use and intuitive structure [72]. The best known frames based systems in bioinformatics are the EcoCyC [59] and MetaCyC [67] metabolic pathway databases [58], both members of the BioCyC project[21].

The Description Logics approach, as the name itself implies, relies on axioms to describe concepts and their relationships in ontologies. Therefore a Description Logics ontology is built by combining elements using logical statements rather than building the whole hierarchy from scratch [96]. The most attractive feature of Description Logics approaches, apart of the expressivity, is the possibility of using reasoning to infer the hierarchy (the subsumption relationships) and the consistency of the ontology. The reasoning is done by a program called a *reasoner*.

[21] `http://biocyc.org`

2.3.2 OWL: Ontology Web Language

Amongst KR languages, OWL[22] has properties of both approaches, and it is being adopted as a standard for web ontologies, being now a W3C (World Wide Web Consortium)[23] official recommendation [48]. The origin of OWL can be dated back to two different languages: DAML (Darpa Agent Markup Language) and OIL (Ontology Inference Layer). DAML was a project in the US funded by the DARPA (Defense Advanced Research Projects Agency) that included the markup language and some tools. OIL was primarily based in Europe, funded by the European Union's Information Society Technologies Program. The efforts converged in DAML+OIL, incorporating the best of both, which would become OWL [49]. OWL is written using RDFS (Resource Description Framework Schema) and it is totally oriented for use on the web. Therefore OWL was designed to fulfil the following aims [110]:

- OWL ontologies should be suitable for sharing; they should be public and different systems through the web should be able to refer to the same ontology.
- OWL ontologies should be able to evolve and a given resource should be able to point to the version of the ontology which is being used.
- OWL should be able to allow ontologies to interoperate between each other when the same concepts are represented in different ontologies, allowing a *web of ontologies*.
- It should be possible to detect inconsistencies between different ontologies that are contradictory.
- OWL aims to meet a balance between expressivity and computational tractability, which leads to reasoning. As the more expressive a language is, the less computationally tractable it becomes.
- OWL should be easy to use and intuitive.
- OWL should be compatible with other standards like XML or UML (Unified Modeling Language).

[22] `http://www.w3c.org/2004/OWL/`
[23] `http://www.w3c.org`

- OWL should be compatible with internationalisation (use in different languages).

OWL plays a central role in the semantic web, as Tim Berners-Lee, James Hendler and Ora Lassila stated [10]:

> *For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial-intelligence researchers have studied such systems since long before the Web was developed. Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to realize its full potential it must be linked into a single global system.*

Or, in more general terms, as James Hendler points [43]:

> *The Semantic Web, as I envision it evolving, will not be primarily comprised of nice neat ontologies that have been carefully constructed by expert artificial intelligence researchers. Rather, I envision a complex web of semantics ruled by the same sort of anarchy that currently rules the rest of the web. Rather than a few large, complex, consistent ontologies, shared by great numbers of users, I envision a great number of small ontological components largely created of pointers to each other and developed by web users in much the same way that web content is currently created.*

There are plenty of implementations and ontologies that already use OWL[24] or are being migrated from DAML+OIL to OWL. The cancer ontology from the US National Cancer Institute's Center for Bioinformatics is a browseable example of an actual ontology implemented in OWL[25]. Another example is the recently implemented ontology of mouse embryo anatomy [18], available at the Mouse Atlas project[26]. In a bigger scale, OWL has been proposed to create a high level ontology that integrates anatomy ontologies from OBO [1].

[24] See the following web for an updated list of implementations: `http://www.w3.org/2001/sw/WebOnt/impls`

[25] `http://www.mindswap.org/2003/CancerOntology/`

[26] `http://genex.hgu.mrc.ac.uk`

In OWL ontologies there are three main elements [46]:

Individuals: the actual objects of the domain. They are equivalent to instances in frames based systems or object oriented programming.

Properties: relations binding individuals. Properties can be of different types:

- Object properties: link individuals.
- Data type properties: link individuals to types of data (integers, for example.)
- Annotation properties: they are used to add extra information.

OWL offers some elements to describe relationships more accurately such as *property characteristics* (properties can be *functional*, *inverse functional*, *transitive* or *symmetric*) or property *domains* and *ranges* (properties link individuals from a certain *domain* to individuals of a certain *range*).

Classes: sets of individuals. The conditions for class membership of the individuals are stated precisely using *restrictions*: properties are used to create restrictions that define the conditions for membership of a class. Individuals can belong to more than one class and it can be explicitly stated that two classes are disjoint (an individual can't belong to both classes).

There are three types of OWL, depending on their expressivity:

OWL-Lite: it is the simplest type, only simple class hierarchies and simple restrictions are allowed.

OWL-DL: it is based in *First Order Logic* and therefore it is computationally tractable, thus automated reasoning can be applied. It is more expressive than OWL-Lite and this is the type of OWL used for this project.

OWL-Full: being the most expressive OWL type, the computational tractability is not guaranteed and reasoning is not possible.

2.3.3 An example of an ontology built using Description Logics

A good example of the Description Logics approach can be found in the paper by Stevens *et al.* [95] that describes the building process of the TaO ontology [6] of the TAMBIS project (Transparent Access to Multiple Bioinformatics Information Sources) [97]. The TAMBIS project aims to give the scientist a single interface to access different bioinformatics and biomedical resources. The TaO ontology describes information about molecular biology and the bioinformatics tasks applicable to those data. The TaO ontology was built using OIL in a iterative manner; some basic concepts were defined, some refinements added, reasoning was applied to the ontology, and the new relationships and changes given by the reasoner were added to the ontology. This cycle was repeated improving the ontology in each step. A concrete example can be given with the portion of TaO that represents the classification of enzymes and proteins; the first classification consisted of the class `protein` having as subclasses `holoprotein` and `enzyme` and the class `enzyme` having as a subclass the class `holoenzyme` (figure 2.1).

In the refinement process it was stated that, regarding classes:

- `holoprotein binds` a `prosthetic-group`.
- `enzyme catalyses` a `reaction`.
- `holoenzyme binds` a `prosthetic-group`.

In further refinements the following slot-constrains were implemented: `holoenzyme catalyses` a `reaction` in addition to `binds` a `prosthetic-group`. This allowed the reasoner to infer the proper classification: `holoenzyme` is a subclass of `enzyme`, but also a subclass of `holoprotein`. The reasoner inferred the correct classification and proposed new relationships.

In the same ontology, the reasoner was used to point out inconsistencies. In the beginning the biologists described `cofactor` as being a subclass of two classes, `metal-ion` and `small-molecule`. Later in the refinements it was asserted that the classes `metal ion` and `small molecule` were disjoint. As a result, the reasoner pointed that the classification of `cofactor` was inconsistent and it was corrected.

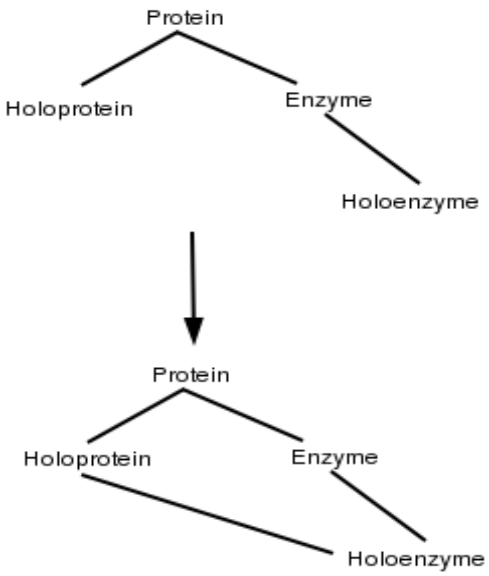

Figure 2.1:
Portion of the TaO ontology. The portion of the ontology shown refers to the classification of proteins and enzymes. Top: in the beginning, before reasoning and refinements were applied. Bottom: new relationships were added after the reasoning and refinements.

2.4 GO and GONG

The Description Logics approach described in section 2.3 can be applied to the Gene Ontology in the same manner as described in the TaO example: adding semantic content from outside the ontology and reasoning over it results in pointing new relationships and consistency lacks. That is the basis of the GONG project which is described in detail in the next chapter. In the case of GONG, the new content was provided by an external chemical ontology.

Chapter 3

Gene Ontology Next Generation

In this chapter a detailed description of the methodology used for the project is given. The chapter is organised according to the temporal order of processes, except the introduction (section 3.1) and the section referring to the programming environment (section 3.5).

3.1 Introduction to GONG

The general methodology applied in this project was the one described by Stevens *et al.* [104] in the GONG project[1].

The GONG procedure dissects a portion of the Gene Ontology into two classifications: the chemical and the functional. Both classifications can differ in complexity. In the case of the *metabolism* subtree, the functional classification has just three classes: *metabolism*, *catabolism* and *biosynthesis* (figure 3.1). The chemical classification is more complex, depending on the term being referring to. In the case of *acetylcholine* (from the GO term *acetylcholine biosynthesis*), it includes *amine*, *biogenic amine*, *ethanolamine*, *neurotransmitter*, *amino acid derivative* and *acetylcholine*. Both classifications are rewritten in OWL and related to each other using OWL restrictions. As an example, in the term *acetylcholine biosynthesis*, *biosynthesis* acts on an *acetylcholine*, therefore a property restriction called `acts_on` is created (`biosynthesis acts_on acetylcholine`).

In this manner both classifications remain independent but can be related to compose definitions of more complex GO concepts such as `acetylcholine biosynthesis` as shown in figure 3.1. When both classifications are built, the

[1] `http://gong.man.ac.uk/`

semantic content of each part of the term *acetylcholine biosynthesis* is made computationally accessible. In the original form of the term this information is implicit. Neither *acetylcholine* was classified as a type of chemical nor *biosynthesis* related to other metabolic actions in a purely functional classification. Now that this information is explicit two types of reasoning are applied: reasoning solely over the newly created ontology, which has both classifications, and reasoning over it including external content given by a chemical ontology, which gives a more complete chemical classification. The reasoner can infer new relationships in both cases in a similar process to the one pointed to in chapter 2, section 3.3, in the TaO ontology.

In more specific terms, the GONG methodology transforms the chosen part of the Gene Ontology (called subontology in further sections) into an OWL ontology (called subset ontology in further sections), adds any external ontology needed (called functional ontology and chemical ontology in further sections), and the reasoner operates over the resulting ontologies altogether, inferring new relationships. The methodology consists of various steps that are explained in detail in the following sections of this chapter. See figure 3.2 for an outline of the whole workflow. This methodology was applied to the following subontologies of GO: *binding* (GO:0005488), *transporter activity* (GO:0005215) and *metabolism* (GO:0008152). The same workflow will be applied to areas of GO that refer to development, using OBO ontologies for the complementary ontologies, in the near future.

3.2 STEP 1: Acquiring the subontologies in the correct format

The GO June (2004) version was used and downloaded in RDF (Resource Description Framework) format[2]. The whole ontology was converted to DAML+OIL [98, 47] using Oiled [7] by opening it and saving it in the appropriate format. The correspondence of GO RDF to DAML+OIL elements can be seen in table 3.1.

[2] `http://www.godatabase.org/dev/database/archive/2004-06-01/`

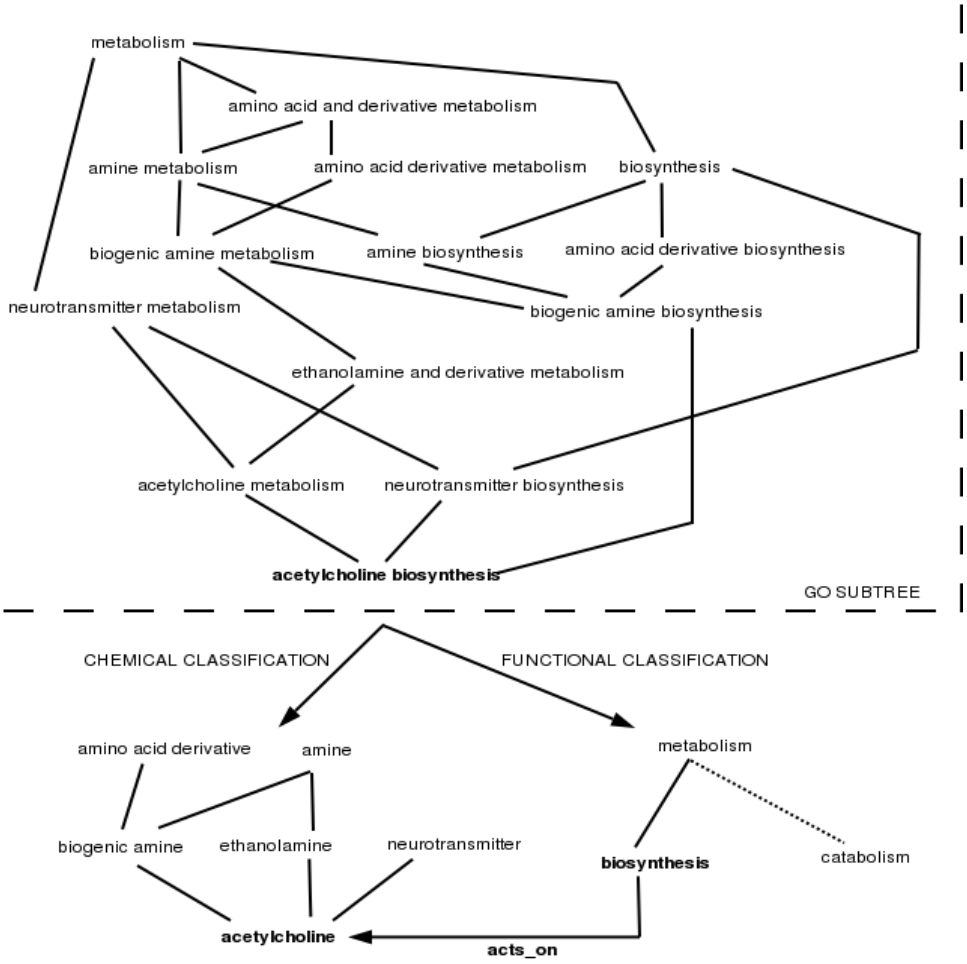

Figure 3.1:
Functional and chemical classification in *metabolism* for the term *acetylcholine biosynthesis*. The functional classification in the case of *metabolism* includes three elements: *catabolism*, *metabolism* and *biosynthesis*. *Catabolism* is included in the diagram for clarity, although is not present in the GO subtree of the example. The chemical classification (simplified in the diagram) is more complex, depending on the term.

GO RDF	DAML+OIL
<go:term>	<daml:Class>
<go:isa>	<daml:subClassOf><daml:Class>
<go:part-of>	<daml:subClassOf><daml:Restriction> <daml:onProperty><daml:ObjectProperty rdf:resource="go:part-of"/> <daml:hasClass><daml:Class>

Table 3.1:
Correspondence between GO XML and DAML+OIL. The term, *IsA* and *PartOf* relationships are included.

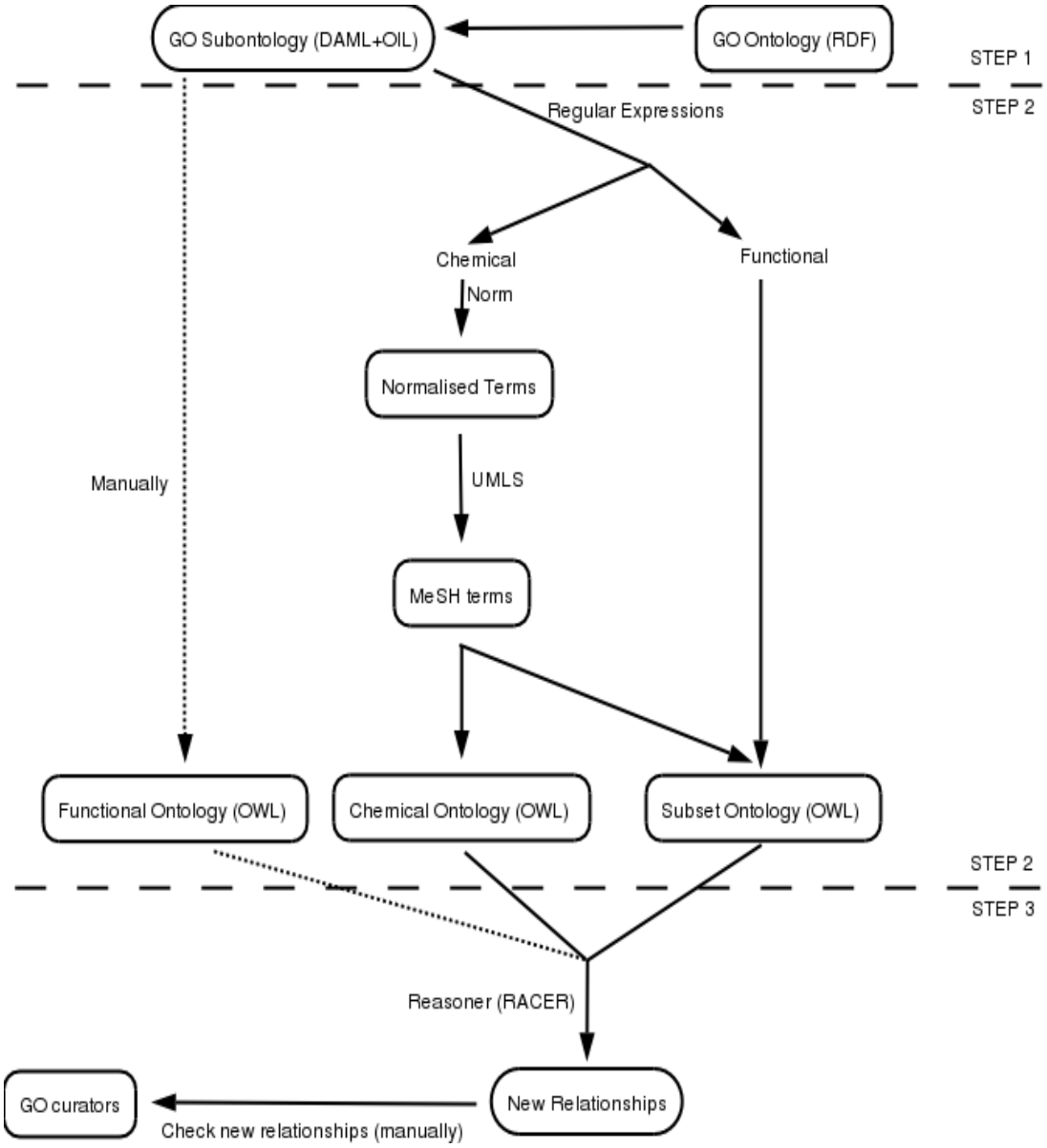

Figure 3.2:
General workflow of the GONG project. The workflow consists of three main steps, divided by horizontal non-continuous lines. The process starts in the top-right with the GO ontology. The dashed line represents a process that takes place only in the case of the *transporter activity* subontology.

3.3 STEP 2: Dissection of the subontology: creation of the subset, chemical and functional ontologies

The aim of the dissection step was to divide each of the GO terms semantically and build the neccesary OWL ontologies: the subset ontology and the chemical ontology. The whole process can be appreciated following two examples,

the dissection of the terms *glutamate binding* (GO:0016595) and *alanine:sodium symporter activity* (GO:0015655), from the *binding* and *transporter activity* subontologies, respectively (figures 3.3 and 3.4). The process was fully automated.

The splitting was done by matching each class name in the subontology against regular expressions (table 3.2). Two semantic axes were considered which appeared in the three subontologies: chemicals and functions (analogous to the chemical and functional classifications in the *metabolism* example in section 3.1). Each GO term in those three subontologies is formed by a kind of chemical (organic, inorganic, macromolecule, carbohydrate, neurotransmitter, ...) and a function acting on/from/by it (biosynthesis, receptor activity, ...). Both axes in the two term examples can be seen in table 3.3. Figures 3.3 and 3.4 show how new restrictions were added to each OWL class based on the result of each regular expression match.

For every term captured by any of the regular expressions, two parallel (and related) processes begun, the creation of the chemical ontology from and external source of information and the addition of relationships pointing to the chemical classes in the subset ontology:

Chemical ontology: this was done in two stages using the chemical terms captured by the regular expressions (following the example, *glutamate* in one case and *alanine* and *sodium* in the other):

- Lexical normalisation: the term had to be in the correct form (e.g. transformed from *glutamate* to *Glutamates*) in order to match entries in the MeSH database (see next step). This was done using Norm from the UMLS tools.

- Query the MeSH: by querying the MeSH database with the normalised chemical term, the hierarchy of that chemical and its correct ID (called CUI -Concept Unique Identifier-) were returned. The chemical ontology was built with that information, creating a class for each chemical term.

Subset ontology: the subset ontology was built using the original subontology classes, modified to OWL format. The chemical and functional axis were included, but for the chemical subterm[3], the MeSH IDs (CUIs) and normalised subterms

[3]The meaning of *subterm* is a string within a term, *glutamate* is a subterm of *glutamate binding*.

were used instead of the original ones. The reasoner was thus able to combine the subset and the chemical ontologies and reason over them. The restrictions in the functional axis pointed to the classes of the chemical ontology. The modification to OWL can be described in plain language (see table 3.4 for plain language example and table 3.5 for the entire mapping of DAML+OIL to OWL including the OWL statements ACTS_ON, HAS_INTERMEDIATE and HAS_PRODUCT).

Functional ontology: in the case of the *transporter activity* subontology the functional axis wasn't properly translated to the subset OWL ontology, so a small functional ontology was manually created (figure 3.5).

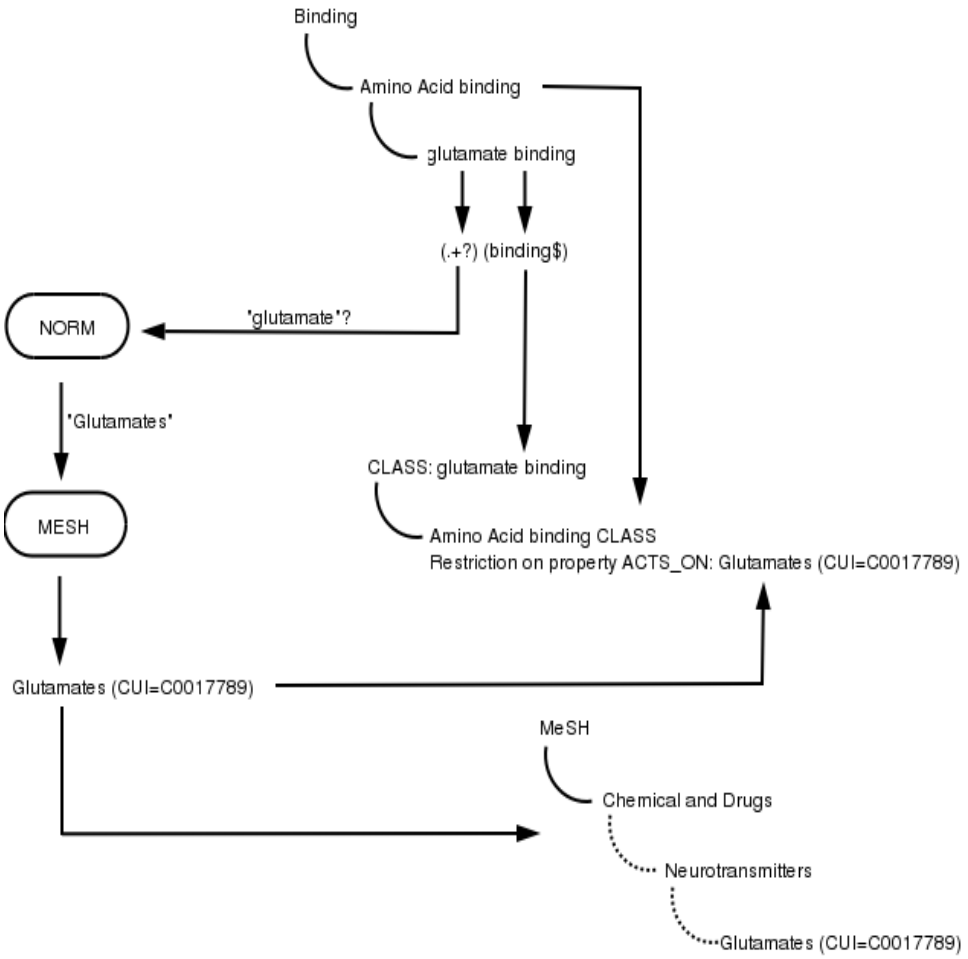

Figure 3.3:
Whole process of dissecting the term *glutamate binding*. The arcs represent relationships within ontologies. Dashed arcs represent paths in an ontology that include more elements than the ones shown (some terms were deleted for simplicity). The whole process starts in the top ontology: the term *glutamate binding* is captured by the regular expression and the dissection takes place. *glutamate* is used to query the MeSH database and therefore build a chemical ontology using *Glutamates* with its ID (CUI). The *glutamate binding* class is transformed into OWL, using at the same time *Glutamates* to build the restriction in the ACTS_ON property.

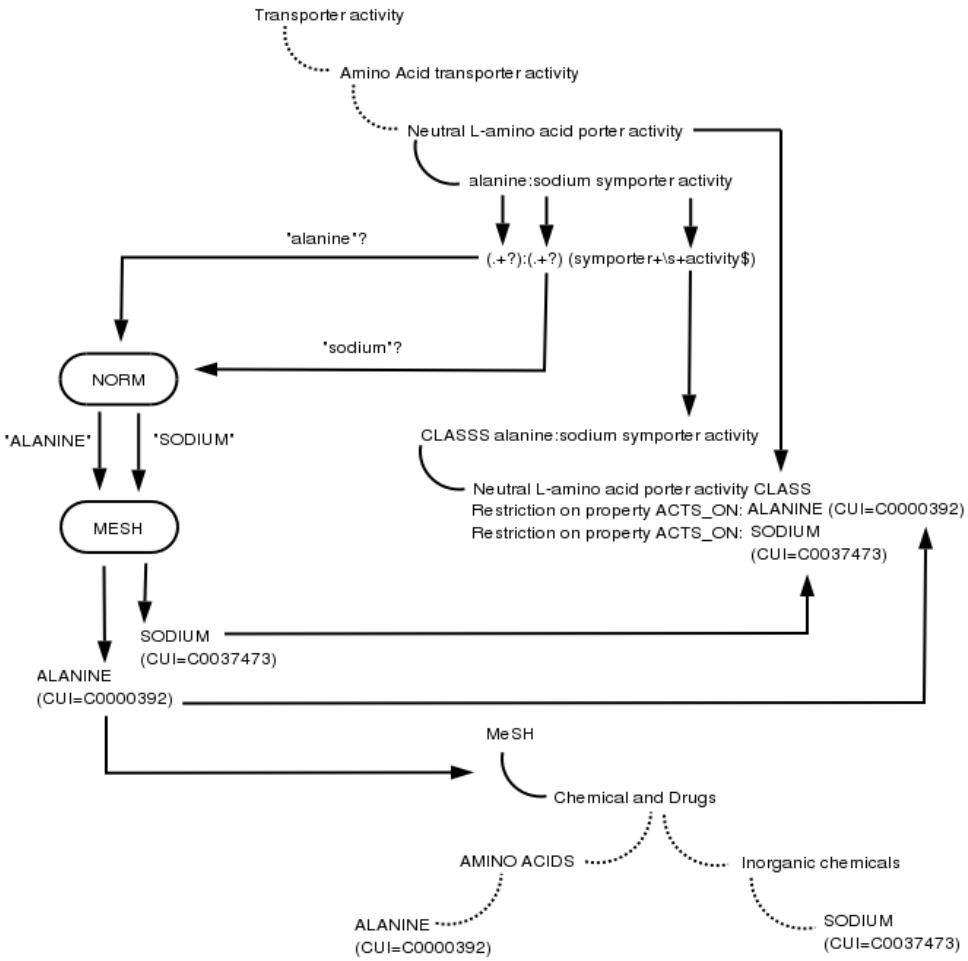

Figure 3.4:
Whole process of dissecting the term *alanine:sodium transporter activity.* The arcs represent relationships within ontologies; the dashed arcs include more elements than the ones shown. Note that in this case, the MeSH database is queried twice, once for each chemical, if compared with the dissection of *glutamate binding* (Figure 3.3). Both queries are included in the chemical ontology and in the subset ontology.

Regular Expression	**Captured example**
***binding* subontology**	
(.+?) (binding$)	Glutamate binding
(.+?) ([a-z]+er+\s+activity$)	ISG15 carrier activity
(.+?) ([a-z]+ion+\s+activity$)	protein homodimerization activity
(.+?) ([a-z]+or+\s+activity$)	oxigen sensor activity
(.+?) ([a-z]+ist+\s+activity$)	receptor antagonist activity
(.+?) ([a-z]+ion+\s+guide+\s+activity$)	RNA 2'-O-ribose methylation guide activity
(.+?) (.+?) ([a-z]+mone+\s+activity$)	mating pheromone activity
***transporter activity* subontology**	
(.+?) (channel+\s+activity$)	calcium channel activity
(.+?)-(transporting) (ATPase+\s+activity$)	lipopolysaccharide-transporting ATPase activity
(.+?):(.+?) (antiporter+\s+activity$)	acetylcholine:hydrogen antiporter activity
(.+?):(.+?) (symporter+\s+activity$)	glutamate:sodium symporter activity
(.+?) transporting (porin+\s+activity$)	oligosaccharide transporting porin activity
(.+?) (.+?) (permease+\s+activity$)	arabinose efflux permease activity
(.+?) (\S+ase+\s+activity$)	protein-N(PI)-phosphohistidine-sugar phosphotransferase activity
(.+?) ([a-z]+er+\s+activity$)	amide transporter activity
(.+?) ([a-z]+or+\s+activity$)	electron acceptor activity
***metabolism* subontology**	
(.+?) ([a-z]+ism$\|[a-z]+ing$|[a-z]+tion$\|[a-z]+sis$|[a-z]+age$)	neurotransmitter biosynthesis
(.+?) ([a-z]+ism|[a-z]+tion) to (.+)	glutamate catabolism to 2-oxoglutarate
(.+?) ([a-z]+ism|[a-z]+tion|[a-z]+sis|linkage|[a-z\-]+ing)[\,]+?via (.+)	L-alanine biosynthesis via ornithine
(.+?) ([a-z]+ism|[a-z]+sis)[\,]+?from (.+)	glycine biosynthesis from serine

Table 3.2:
Regular expressions used for capturing terms from *binding*, *transporter activity* and *metabolism* subontologies. The left column includes the regular expressions whereas the right column includes examples of the captured terms.

Functional axis	**Chemical axis**
***binding* subontology** (*glutamate binding*)	
Gene Ontology molecular function **binding** amino acid **binding** glutamate **binding**	Gene Ontology molecular function binding **amino acid** binding **glutamate** binding
***transporter activity* subontology** (*alanine:sodium symporter activity*)	
Gene Ontology molecular function **transporter activity** amine/polyamine **transporter activity** amino acid **transporter activity** neutral amino acid **transporter activity** neutral L-amino acid **porter activity** alanine:sodium **symporter activity**	Gene Ontology molecular function transporter activity **amine/polyamine** transporter activity **amino acid** transporter activity **neutral amino acid** transporter activity **neutral L-amino acid** porter activity **alanine**:sodium symporter activity
	Gene Ontology molecular function transporter activity **ion** transporter activity **cation** transporter activity **cation**:amino acid symporter activity alanine:**sodium** symporter activity

Table 3.3:
Chemical and functional axis for *binding* and *transporter activity* subontologies. In the case of *alanine:sodium symporter activity* only part of the GO hierarchy is shown for the sake of simplicity.

GO	**OWL**
glutamate binding	The class *glutamate binding* is a subclass of *amino acid binding* which acts solely on Glutamates
alanine:sodium symporter activity	The class *alanine:sodium symporter activity* is a subclass of *symporter activity*, *neutral L-amino acid porter activity* and *cation:amino acid symporter activity* which acts solely on ALANINE and SODIUM

Table 3.4:
Human readable OWL. Examples of the OWL statements in the subset ontology for the terms *glutamate binding* and *alanine:sodium symporter activity* in a simplified plain language.

DAML + OIL	OWL
<daml:Class>	<owl:Class>
<daml:subClassOf><daml:Class>	<rdfs:subClassOf> <owl:Class /> </rdfs:subClassOf> Other conditions can be added in the captured terms to express their relationships to other classes using equivalent classes: <owl:equivalentClass> <owl:Class> <owl:intersectionOf rdf:parseType="Collection"> <owl:Class/> </owl:intersectionOf> </owl:Class> </owl:equivalentClass>
<daml:subClassOf><daml:Restriction> <daml:onProperty><daml:ObjectProperty rdf:resource="go:part-of"/> <daml:hasClass><daml:Class>	<rdfs:subClassOf><owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about="go:part-of"/> </owl:onProperty> </owl:Restriction></rdfs:subClassOf>
	ACTS_ON: <owl:Restriction> <owl:someValuesFrom rdf:resource="mesh:CUI"/> <owl:onProperty rdf:resource="go:ACTS_ON"/> </owl:Restriction> HAS_PRODUCT: <owl:Restriction> <owl:someValuesFrom rdf:resource="mesh:CUI"/> <owl:onProperty rdf:resource="go:HAS_PRODUCT"/> </owl:Restriction> HAS_INTERMEDIATE: <owl:Restriction> <owl:someValuesFrom rdf:resource="mesh:CUI"/> <owl:onProperty rdf:resource="go:HAS_INTERMEDIATE"/> </owl:Restriction> HAS_SUBSTRATE: <owl:Restriction> <owl:someValuesFrom rdf:resource="mesh:CUI"/> <owl:onProperty rdf:resource="go:HAS_SUBSTRATE"/> </owl:Restriction>

Table 3.5:
Mapping of DAML+OIL to OWL. The term (class) and *IsA* and *PartOf* relationships are included with the additional OWL specific statements for the subset ontology: ACTS_ON, HAS_PRODUCT, HAS_INTERMEDIATE and HAS_SUBSTRATE. *CUI* refers to the MeSH IDs.

3.4 STEP 3: Merging and reasoning over the ontologies

Once all the ontologies had been created, they were merged (imported) in Protégé[4] [81] (figure 3.5). The reasoning process was done using RACER[5] [41] via Protégé. An outline of the simplified reasoning process can be seen in figure 3.6. The new relationships inferred by the reasoner were reviewed by the author for biological inaccuracies. The legitimate new relationships were send to GO curators for reviewing and possible inclusion in GO. This was done using the SourceForge GO Curator Requests Tracker[6].

3.5 Programming environment

The workflow was built using a collection of Jython scripts[7]. The HP labs Jena toolkit[8] and OiLed libraries were used to manipulate the ontologies. The UMLS database was implemented using MySQL[9].

[4] http://protege.stanford.edu/
[5] http://www.sts.tu-harburg.de/~r.f.moeller/racer/
[6] https://sourceforge.net/tracker/?atid=440764&group_id=36855&func=browse
[7] http://www.jython.org/
[8] http://www.hpl.hp.com/semweb/jena.htm
[9] http://www.mysql.com/

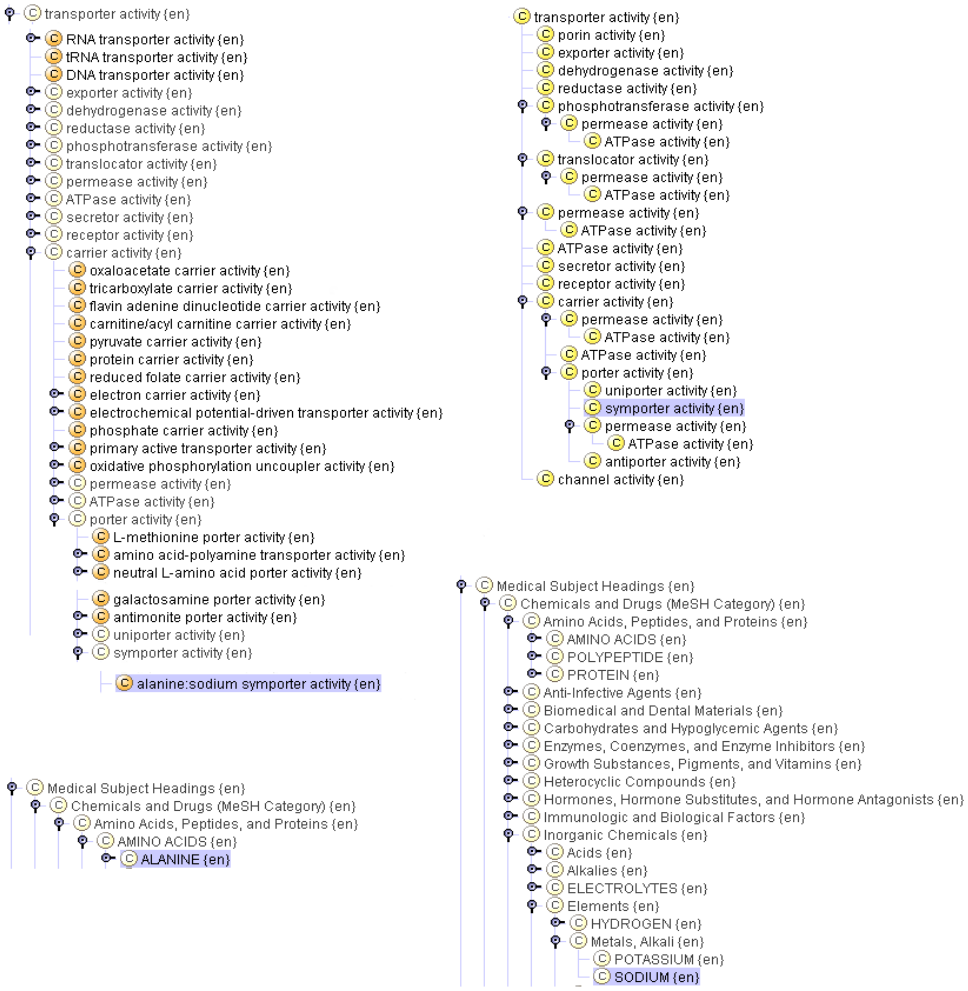

Figure 3.5:
The subset, chemical and functional ontologies resulting from the dissection of the *transporter activity* subontology, in the case of the term *alanine:sodium symporter activity*. Classes are represented by the circles with a *C* inside together with the class name. Left top: the subset ontology. Right top: the manually created functional ontology. Left bottom: the classification of ALANINE in the chemical ontology. Right bottom: the classification of SODIUM in the chemical ontology. The four appeared as parts of the same ontology as a result of importing them: in the subset ontology (left top) some class items appeared in clearer colours which means that they cannot be edited in Protégé because they have been imported. This can be noticed comparing the functional ontology with the subset ontology: the classes of the functional ontology (*carrier activity*, *ATPase activity*, ...) appear in clearer colours in the subset ontology. Taken from Protégé.

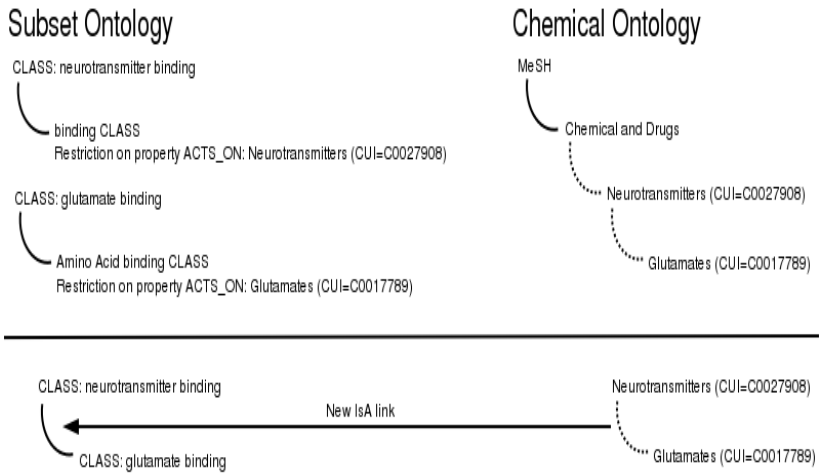

Figure 3.6:
Simplified reasoning process. The arcs represent relationships within ontologies. Dashed arcs represent paths in an ontology that include more elements than the ones shown. Note that the two ontologies share the same IDs (CUIs) for the chemicals. The reasoner thus infers the new *IsA* link based on the chemical classification.

Chapter 4

Results

In this chapter a brief overview of the results (the new relationships inferred by the reasoner) is given. All the new relationships can be seen in Appendices A, B and C.

4.1 Result percentages

The results in percentages can be seen in table 4.1. They are divided following the GONG workflow steps. Starting from the GO subontology, the following percentages are given:

Captured terms: the terms captured by the regular expressions.

Changed terms: the terms that were changed by the reasoner. These changes included one or more new parentage relationships (New *IsA* relationships) or new positions in the hierarchy.

Changes accepted by the author: the changes that made biological sense for the author.

Changes accepted by the GO curators: the changes that were accepted for being included in the GO ontology.

In the case of *binding*, 771 terms from 789 were captured by the regular expressions; from those ones 134 were changed and 67 of those changes were

	Captured	Changed	Accepted by author	Accepted by GO curators
binding	98%	17%	8%	5%
transporter activity	94%	21%	11%	-
metabolism	90%	20%	8%	-

Table 4.1:
Result percentages for *binding*, *transporter activity* and *metabolism* subontologies. The percentages express the relative amount of terms (classes in OWL).

accepted by the author and send to GO for review: 37 were accepted for GO inclusion. In *transporter activity* and *metabolism*, 810 terms from 859 and 3248 terms from 3625 were captured respectively (see table 4.1 for the rest of the percentages). The results from *transporter activity* and *metabolism* will be sent to GO for review in the near future (see Appendix A, B and C).

4.2 New links accepted for GO inclusion

The new links from the *binding* subontology accepted by the GO staff can be seen in table 4.2.

New IsA link: Glycine binding (GO:0016594) IsA neurotransmitter binding (GO:0042165)
New position: Glycosaminoglican binding (GO:0005539) should be under polysaccharide binding (GO:0001871) instead of binding (GO:0005488)
New position: hemoglobin binding (GO:0030492) should be under protein binding (GO:0005515) instead of binding (GO:0005488)
New position: ISG15 carrier activity (GO:0019793) should be under protein carrier activity (GO:0008320) instead of protein binding (GO:0005515)
New IsA link: melanocyte stimulating hormone receptor activity (GO:0004980) IsA hormone binding (GO:0042562)
New IsA link: epidermal growth factor binding (GO:0048408) IsA hormone binding (GO:0042562)
New IsA link: FAD binding (GO:0050660) IsA adenyl nucleotide binding (GO:0030554)
New IsA link: FMN binding (GO:0010181) IsA nucleotide binding (GO:0000166)
New position: Galanin receptor activity (GO:0004966) should be under neuropeptide receptor activity (GO:0008188) instead of peptide receptor activity, G-protein coupled (GO:0008528)
New IsA link: neuropeptide binding (GO:0042923) IsA neurotransmitter binding (GO:0042165)
New IsA link: neuropeptide receptor activity (GO:0008188) IsA neurotransmitter receptor activity (GO:0030594)
New IsA link: peptide YY receptor activity (GO:0001601) IsA hormone binding (GO:0042562)
New IsA link: peroxisome targeting signal receptor activity (GO:0005051) IsA peptide receptor activity (GO:0001653)
New position: pyridoxal phosphate binding (GO:0030170) should be under vitamin binding (GO:0019842) instead of binding (GO:0005488)
New IsA link: Retinoic acid receptor binding (GO:0042974) IsA transcription factor binding (GO:0008134)
New position: Ribonucleoprotein binding (GO:0043021) should be under protein binding (GO:0005515) instead of binding (GO:0005488)
New IsA link: snRNA modification guide activity (GO:0030566) IsA snRNA binding (GO:0017069)
New IsA link: rRNA modification guide activity (GO:0030556) IsA rRNA binding (GO:0019843)
New IsA link: tRNA modification guide activity (GO:0030557) IsA tRNA binding (GO:0000049)
New IsA link: somatostatin receptor activity (GO:0004994) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: somatostatin receptor activity (GO:0004994) IsA hormone binding (GO:0042562)

Table 4.2:
New relationships of the *binding* subontology accepted by the GO staff. *New position* means that one *IsA* link should be deleted and another one added. *New IsA link* means that another *IsA* link should be added in the following manner: *child* ***IsA*** *parent*.

New IsA link: Adrenocorticotropin receptor activity (GO:0004978) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: Adrenocorticotropin receptor activity (GO:0004978) IsA hormone binding (GO:0042562)
New position: apolipoprotein E receptor binding (GO:0050749) should be under low-density lipoprotein receptor binding (GO:0050750) instead of receptor binding (GO:0005102)
New IsA link: beta-endorphin receptor activity (GO:0004979) IsA neuropeptide receptor activity (GO:0008188)
New position: cadherin binding (GO:0045296) should be under cell adhesion molecule binding (GO:0050839) instead of protein binding (GO:0005515)
New IsA link: Cholecystokinin receptor activity (GO:0004951) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: ecdysone binding (GO:0035100) IsA steroid binding (GO:0005496)
New IsA link: Glutamate binding (GO:0016595) IsA neurotransmitter binding (GO:0042165)
New position: tau protein binding (GO:0048156) should be under cytoskeletal protein binding (GO:0008092) instead of protein binding (GO:0005515)
New IsA link: thyroid hormone receptor binding (GO:0046966) IsA transcription factor binding (GO:0008134)
New IsA link: Transforming Growth Factor beta binding (GO:0050431) IsA growth factor binding (GO:0019838)
New IsA link: Transforming Growth Factor beta binding (GO:0050431) IsA peptide binding (GO:0042277)
New IsA link: tRNA modification guide activity (GO:0030557) IsA tRNA binding (GO:0000049)
New IsA link: ubiquinone binding (GO:0048039) IsA coenzyme binding (GO:0050662)
New IsA link: vasopressin receptor activity (GO:0005000) IsA hormone binding (GO:0042562)
New IsA link: vitamin D receptor binding (GO:0042809) transcription factor binding (GO:0008134)

Table 4.3:
New relationships of the *binding* subontology accepted by the GO staff (Cont.)

Chapter 5

Discussion

The GONG workflow has been successful because new relationships for GO were suggested and accepted by the GO team. However the results were affected in several stages of the workflow. The following sections give an analysis of how the results were affected during the workflow and why were they either accepted or rejected. Some Description Logics applications apart from (but related to) GONG are given in the last section of the chapter.

5.1 Problems capturing the terms

When capturing the terms with the regular expressions, the semantic content that would be accessible for the reasoner was greatly determined by the design of the regular expressions. In general terms the approach worked because of the highly stereotyped syntactical structure of the GO terms, with a high occurrence of strings (subterms) and substrings in a regular manner [82]. However there were two problems:

- Some terms were not captured at all (2% in the case of the *binding* subset) which means that GO terms are not that syntactically stereotyped. Examples include: *MHC class II protein binding, via lateral surface* and *translation release factor activity, codon nonspecific*, amongst others.

- Even when captured, in some cases the semantic subtetlies of the term were not translated to the final ontology. For example, the term *RNA polymerase II transcription factor activity, enhancer binding* was split in the subterms *RNA polymerase II transcription factor activity, enhancer*

(chemical axis) and *binding* (functional axis). The first subterm did not yield any result in the MeSH querying because of its complexity. Therefore it wasn't included in the subset OWL ontology. On the other hand, there were subterms that could be used taking advantage of OWL's expressivity. Subterms like *enhancer* and *transcription factor* could be codified using other relationships apart of ACTS_ON yielding new relationships.

5.2 Errors in GO

As a consequence of reviewing the new relationships, some minor errors were found in GO and were sent with the new relationships for reviewing. These were accepted and fixed:

- There were two similar terms, *acyl-CoA or acyl binding* (GO:0005541) and *acyl binding* (GO:0000035), with the same definitions creating redundancy.
- Two terms have incorrect parents:
  - The term *organophosphate:inorganic phosphate antiporter activity* (GO:0015315) had the term *solute:hydrogen antiporter activity* (GO:0015299) as parent: neither organophosphate or inorganic phosphate are hydrogen.
  - The term *potassium:amino acid transporter activity* (GO:0017032) had an incorrect definition:

    > *Catalysis of the reaction: amino acid(out) + +(out) = amino acid(in) + Na+(in)*

    The same term had an incorrect parent (*solute:sodium symporter activity* (GO:0015370)) because the parent referred to sodium instead of potassium.
- The term *organic anion transporter activity* (GO:0008514) had an error in the definition; cations instead of anions were mentioned in it:

  > *Enables the directed movement of organic anions into, out of, within or between cells. Organic* ***cations*** *are atoms or small molecules with a negative charge which contain carbon in covalent linkage.*

5.3 New relationships rejected before sending them to GO

Many of the new relationships inferred by the reasoner were rejected by the author for the following reasons:

Different criteria in GO and MeSH: the chemical ontology built in the workflow included all the possible links in a chemical classification from MeSH. As a consequence many of the inferred new relationships were based on pure chemical criteria and were not suitable for GO which is a functionally oriented ontology. A clear example is the following:

```
New IsA link:
nucleotide binding (GO:0000166) IsA
carbohydrate binding (GO:0030246)
```

In a chemical sense, the binding of a nucleotide is a type of carbohydrate binding because nucleotides are made in part of carbohydrates. Functionally, however, they are different concepts and therefore the relationships of this type were rejected.

There was a difference too between purely chemical considerations in GO and MeSH:

```
New IsA link:
oligosaccharide transporter activity (GO:0015157) IsA
polysaccharide transporter activity (GO:0015159)
```

In the GO *transporter activity* subontology oligosaccharides are not a subclass of polysaccharides (what happens in MeSH), therefore this new relationship was rejected.

Breakage of the true path rule: some of the new relationships were not true in all the cases:

```
New IsA link:
cell adhesion molecule binding (GO:0050839)
IsA antigen binding (GO:0003823)
```

This new relationship was rejected because not all the cell adhesion molecules can be considered antigens.

Insufficient semantic content: as a consequence of not properly capturing the semantic content with the regular expressions, some of the new relationships were not correct:

```
New IsA link: cytochrome c oxidase biogenesis (GO:0008535)
IsA cytochrome biogenesis (GO:0017004)
```

The MeSH querying interpreted *cytochrome c oxidase biogenesis* as a type of *cytochrome*, but cytochrome c oxidase is an oxidase that acts in cytochrome type c oxidazing it, not a cytochrome type. Therefore *cytochrome c oxidase biogenesis* is not a type of *cytochrome biogenesis* and this relationship was rejected.

Insufficient semantic content in the well captured terms: even in the case of well captured terms, some inferred relationships were wrong because of the semantic content of the terms themselves. Some terms have certain subtetlies that cannot be tackled with the regular expression approach because these subtetlies are either mentioned on the term definitions or not mentioned at all:

```
New IsA link: CAM photosynthesis (GO:0009761)
IsA glycoprotein biosynthesis (GO:0009101)
```

CAM in this case refers to *Crassulacean Acid Metabolism*, not to *Cell Adhesion Molecule* which is a type of protein in MeSH. This relationship was rejected because CAM is a type of plant metabolism, not a type of protein.

Terms already in the GO hierarchy: as the GO IsA relationship is transitive some of the new relationships inferred by the reasoner were redundant because the term was already in the hierarchy:

```
New IsA link:
L-arabinose/beta-D-thiogalactopyranoside:hydrogen
antiporter activity (GO:0015524) IsA
arabinose transporter activity (GO:0042900)
```

arabinose transporter activity (GO:0042900) is on the hierarchy that leads to *L-arabinose/beta-D-thiogalactopyranoside:hydrogen antiporter activity* (GO:0015524):

Gene Ontology
molecular function
transporter activity
carbohydrate transporter activity
monosaccharide transporter activity
pentose transporter activity
arabinose transporter activity
L-arabinose transporter activity
L-arabinose/beta-D-thiogalactopyranoside:hydrogen antiporter activity

Obsolete terms: some of new relationships referred to obsolete terms:

```
New IsA link: CDP reduction (GO:0006250)
IsA CDP catabolism (GO:0046706)

New IsA link: coenzyme A catabolism (GO:0006765)
IsA glycoside catabolism (GO:0016139)
```

CDP reduction (GO:0006250) and *coenzyme A catabolism* (GO:0006765) are obsolete, therefore adding new relationships to them is pointless.

More concise hierarchy positions already in GO: Some of the new relationships had a more appropriate counterpart already in GO:

```
New IsA link: phosphoserine binding (GO:0050815)
IsA amino acid binding (GO:0016597)
```

The term *protein amino acid binding (GO:0045308)* was already in the hierarchy leading to *phosphoserine binding (GO:0050815)* and it is a better term than *amino acid binding (GO:0016597).* Adding the new link is unneccesary.

5.4 Reasoning results without the chemical ontology in the *binding* subontology

Reasoning over the *binding* subset without importing the chemical ontology yielded some results. 19 terms were changed (in comparison to 134 importing the chemical ontology) and from them only two were finally accepted for GO inclusion:

```
New IsA link: Acyl carrier activity (GO:0000036) IsA
acyl binding (GO:0000035)
```

```
New position: ISG15 carrier activity (GO:0019793)
should be under protein carrier activity (GO:0008320)
instead of protein binding (GO:0005515)
```

This example shows that only by dissecting the terms in the functional and chemical axis and rewriting them in OWL GO inconsistencies were revealed.

5.5 New relationships accepted by the author for GO submission

The new relationships accepted by the author and therefore send to GO had the following properties:

Different positions deeper in the hierarchy: most of the *New position* relationships represented more specialised locations for the terms:

```
New position: Actinin binding (GO:0042805) should be under
cytoskeletal protein binding (GO:0008092) instead of
protein binding (GO:0005515)
```

Functional vs. chemical new relationships: some new relationships were based in the chemical classification, for example:

```
New IsA link: Activin binding (GO:0048185) IsA
hormone binding (GO:0042562)
```

The new relationships was inferred because Activin is a type of hormone in MeSH. Other results pointed to more functional new relationships. Thus, even if the reasoning process was based on a chemical classification new functions were implicitly revealed:

```
New IsA link: myosin binding (GO:0017022) IsA
enzyme binding (GO:0019899)
```

5.6 New relationships accepted by the GO staff

Surprisingly, some of the new relationships accepted by the GO staff included the neccesary new relationship in the term definition. For example:

```
New IsA link: neuropeptide binding (GO:0042923) IsA
neurotransmitter binding (GO:0042165)
```

The subterm *neuropeptide* is defined as being a neurotransmitter in the term definition:

> *Interacting selectively and stoichiometrically with neuropeptides, peptides with direct synaptic effects (peptide* ***neurotransmitters****) or indirect modulatory effects on the nervous system (peptide neuromodulators).*

This is an indicator of how difficult it is to maintain GO manually. New relationships that are obvious are not included by the curators.

5.7 New relationships rejected by the GO staff

Most of the new relationships not accepted by the GO staff were rejected due to the aforementioned difference in the chemical criteria of MeSH and GO. For example:

```
New IsA link: bradykinin receptor activity (GO:0004947)
IsA neuropeptide receptor activity (GO:0008188)
```

This relationship was rejected for not considering bradykinin to be a neuropeptide.

Other relationships were rejected due to breakage of the true path rule:

```
New IsA link: interleukin binding (GO:0019965)
IsA growth factor binding (GO:0019838)
```

This relationship was rejected because not all the interleukins are growth factors.

5.8 GO and Description Logics

One of the relationships rejected by the GO curators was the previously mentioned example of *myosin binding*:

```
New IsA link: myosin binding (GO:0017022)
IsA enzyme binding (GO:0019899)
```

It was rejected the argument being that even if the myosin possesses a catalytic activity, it could not be considered an enzyme. However this leaves the question of how to characterise the enzymatic activity of myosin unanswered. This is not just an isolated problem but reflects an structural limitation of the whole GO. It is an aspect that can be improved using a Description Logics approach for the whole GO. Description Logics allow a more fine-grained definition of relationships and a relation could be added in that case to codify the (secondary) catalytic function of a molecule that is not an enzyme. Another example of the same problem can be found in the ATPase activity when it is coupled to a transporter activity. In the actual version of GO this relationship is codified by adding a special subterm, *coupled to transmembrane movement of substances*:

ATPase activity, coupled to transmembrane movement of substances
  ATP-binding cassette (ABC) transporter activity
    ABC-type efflux permease activity
      lipopolysaccharide-transporting ATPase activity

A more efficient DL solution would be having a class *ATPase activity* and codyfing the coupling to the transmembrane movement with a special restriction (relationship).

Going further in possible applications of DL, Maziére *et al.* [79] have proposed a new scheme for describing biological processes. The model is based on combining *bricks* or very simple concepts, called BEAS (*Basic Elements of Action*): *a BEA is a description of an elementary action at the chemical level, which can be associated with a given molecule and involved in a biological process.* By combining a few BEAS, several complex biological processes and pathways can be described. A model of this type appears as the natural substrate for applying DL. Maziére *et al.* suggest that this scheme allows a better representation of the knowledge related to biological processes and give an example: the GO term *transmembrane receptor protein kinase activity* (GO:0019199) has two parents, *protein kinase activity* (GO:0004672) and *transmembrane receptor activity* (GO:0004888) however nothing is addressed regarding the type of relation that binds the transmembrane receptor activity and the protein kinase activity when in fact knowledge is widely available describing how this relationship takes place.

A related but different problem has been addressed as well in more general terms by Williams *et al.* [102] proposing different types of *PartOf* relationships for GO. A Description Logics approach could help in that path.

Another interesting change for GO that has been already proposed [106, 93] is the merging of the three subontologies to make links that answer questions such as in which cellular components a biological process takes place or if a molecular function is always a part of a biological process. A Description Logics approach offers the neccesary fine-grained expressivity for this purpose.

Chapter 6

Conclusion

GONG has been shown to be a good engineering environment for migrating GO to Description Logics, although new improvements can be suggested:

- More profound and sophisticated text processing should be included instead of relying just on regular expressions.
- The content of the term definitions should be exploited somehow. This appears as difficult task, but plenty of new semantic content could be unleashed.

Even if there were some errors in the workflow, the GONG approach has worked and has demonstrated that Description Logics is a more desirable environment for GO. New relationships were accepted by the GO team and possible application areas have been suggested. Just by applying a simple dissecting procedure without full coverage and a minimum programming and reviewing work, GO improved substantially, which shows the optimality of GONG and Description Logics for GO. However other factors must be taken in consideration: GONG is an automated procedure but not all the semantics are properly captured, so human intervention is needed. Human intervention introduces errors and is time consuming and expensive.

As a result of the GONG approach, GO will become a more comprehensive database where all the neccesary relationships are present. This can be appreciated with the following example: the term *ecdysone binding* (GO:0035100) does not have a relationship with the term *steroid binding* (GO:0005496), when in fact, ecdysone is an steroid. If the user queries the database for all the children of *steroid binding*, *ecdysone binding* will not appear but it should. The GONG

methodology detected that missing link and fixed it. Apart of making the Gene Ontology more reliable for querying, a DL approach such as GONG makes the Gene Ontology a better tool for knowledge representation. More complex relationships are available, giving a more adjusted picture of biological procceses and entities. If the Gene Ontology is migrated to a Description Logics environment, a consequence is that the tools and technologies that rely on the structure of the ontology itself will improve in terms of performance. There is a trade-off between the technical advantages of Description Logics (expressivity and reasoning) and its ease of use for biologists. Description Logics definitions are complex and need effective tools for authoring and maintaining. However as the Gene Ontology grows and more automated technologies are added to it, the Description logics approach will be more useful even considering the difficulties of biologists adopting it. As more ontologies are properly formalised the GONG approach will be more effective, taking the burden of curation away from the curator and relying more on reasoning procceses that combine different ontologies. In a broader and long term perspective, migrating GO to a Description Logics environment such as OWL will make it able to interact with other technologies such as the semantic web, including it on the next generation of the World Wide Web.

Bibliography

[1] JS Aitken, BL Webber, and JBL Bard. Part-of relationships in anatomy ontologies: a proposal for RDFS and OWL formalisations. In *9th Pacific Symposium on biocomputing (PSB)*, pages 624–636, 2004.

[2] Fátima Al-Shahrour, Ramón Díaz-Uriarte, and Joaquín Dopazo. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 20(4):578–580, 2004.

[3] Bill Andersen. What is an ontology? *Ontology Works* `http://www.ontologyworks.com`, 2001.

[4] M. Bada, R. McEntire, C. Wroe, and R. Stevens. GOAT: The Gene Ontology Annotation Tool. In *Proceedings of the 2003 UK e-Science All Hands Meeting*, pages 514–519, Nottingham, UK, 2003.

[5] Michael Bada, Robert Stevens, Carole Goble, Yolanda Gil, Michael Ashburner, Judith A. Blake, J. Michael Jerry, Midori Harris, and Suzanna Lewis. A short study on the success of the Gene Ontology. *Journal of Web Semantics*, 1:235–240, 2004.

[6] P.G. Baker, C.A. Goble, S. Bechhofer, N.W. Paton, R. Stevens, and A Brass. An Ontology for Bioinformatics Applications. *Bioinformatics*, 15(6):510–520, 1999.

[7] Sean Bechhofer, Ian Horrocks, and Carole Goble. OilEd: a reason-able ontology editor for the semantic web. In *DL2001, 14th International Workshop on Description Logics*, Stanford, USA, Aug 2001.

[8] Tim BeißBarth and Terence P. Speed. GOstat: find statistically overrepresented gene ontologies within a group of genes. *Bioinformatics*, 20(9):1464–1465, 2004.

[9] T. Berners-Lee and M. Fischetti. *Weaving the web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. Harper, San Francisco, 1999.

[10] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, May 2001.

[11] J.A. Blake. The mouse genome database (MGD): expanding genetic and genomic resources for the laboratory mouse. *Nucleic Acid Research*, 28:108–111, 2000.

[12] Christian Blaschke and Alfonso Valencia. Automatic ontology construction from the literature. *Genome Informatics*, 13:201–213, 2002.

[13] Olivier Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acid Research*, 32:D267–D270, 2004.

[14] Olivier Bodenreider, Joyce A. Mitchell, and Alexa T. McCray. Biomedical ontologies. In *Pacific Symposium on biocomputing (PSB)*, January 2003.

[15] Paolo Bouquet, Fausto Giunchiglia, Frank van Harmelen, Luciano Serafini, and Heiner Stuckenschmidt. Contextualizing ontologies. *Journal of Web Semantics*, 2004.

[16] Karin Koogan Breitman and Julio Cesar Sampaio do Prado Leite. Ontology as a Requirements Engineering Product. In *Proceedings of the 11th IEEE International Requirements Engineering Conference*, pages 309–320, September 2003.

[17] Sydney Brenner. Life sentences: Ontology recapitulates philology. *Genome Biology*, 4:1006.1–1006.2, 2002.

[18] Albert Burger, Duncan Davidson, and Richard Baldock. Formalization of mouse embryo anatomy. *Bioinformatics*, 20:259–267, 2004.

[19] Jeremy S. Caldwell. Ontology recapitulates physiology. *Chemistry and Biology*, 10:784–786, 2003.

[20] Evelyn Camon, Daniel Barrell, Catherine Brooksbank, Michele Magrane, and Rolf Apweiler. The gene ontology annotation (GOA) project - application of GO in SWISS-PROT, TrEMBL, and InterPro. *Comparative and Functional Genomics*, 4:71–74, 2003.

[21] Evelyn Camon, Daniel Barrell, Michele Magrane, Rolf Apweiler, Vivian Lee, Emily Dimmer, John Maslen, David Binns, Nicola Harte, and Rodrigo Lopez. The gene ontology annotation (GOA) database: sharing knowledge in uniprot with gene ontology. *Nucleic Acid Research*, 32:D262–D266, 2004.

[22] Vincent J. Carey. Ontology concepts and tools for statistical genomics. *Journal of Multivariate Analysis*, 90:213–228, 2003.

[23] Jung-Hsien Chiang and Hsu-Chun Yu. MeKE: discovering the functions of gene products from biomedical literature via sentence alignment. *Bioinformatics*, 19:1417–1422, 2003.

[24] Kuo-Chen Chou and Yu-Dong Cai. Prediction of protein subcellular locations by GO-FunD-PseAA predictor. *Biochemical and Biophysical Research Communications*, 320:1236–1239, 2004.

[25] The FlyBase Consortium. The FlyBase database of the drosophila genome projects and community literature. *Nucleic Acid Research*, 27:85–88, 1999.

[26] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

[27] The Gene Ontology Consortium. Creating the Gene Ontology Resource: Design and Implementation. *The Gene Ontology Consortium*, 11:1425–1433, 2001.

[28] The Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32:D258–D261, 2004.

[29] Kirill Degtyarenko and Sergio Contrino. COMe: the ontology of bioinorganic proteins. *BMC Structural Biology*, 4, 2004.

[30] Minghua Deng, Zhidong Tu, Fengzhu Sun, and Ting Chen. Mapping Gene Ontology to proteins based on protein-protein interaction data. *Bioinformatics*, 20:895–902, 2004.

[31] Minghua Deng, Zhidong Tu, Fengzhu Sun, and Ting Chen. An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 20:349–356, 2004.

[32] Damien Devos and Alfonso Valencia. Intrinsic errors in genome annotation. *Trends in genetics*, 17(8):429–431, 2001.

[33] Scott W Doniger, Nathan Salomonis, Kam D Dahlquist, Karen Vranizan, Steven C Lawlor, and Bruce R Conklin. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology*, 4, 2003.

[34] Michael Y. Galperin. The molecular biology database collection: 2004 update. *Nucleic Acids Research*, 32:D3–D22, 2004.

[35] Nicholas Gibbins, Stephen Harris, and Nigel Shadbolt. Agent-based semantic web services. *Journal of Web Semantics*, 1(1):141–154, 2004.

[36] Asunción Gómez-Pérez and Oscar Corcho. Ontology languages for the semantic web. *IEEE intelligent systems*, 17(1):54–60, 2002.

[37] Detlef Groth, Hans Lehrach, and Steffen Hennig. GOblet: a platform for Gene Ontology annotation of anonymous sequence data. *Nucleic Acids Research*, 32:W313–W317, 2004.

[38] T.R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5:199–220, 1993.

[39] Nicola Guarino and Christopher Welty. Ontological analysis of taxonomic relationships. In *Proceedings of ER-2000: The international Conference on Conceptual Modeling*, Lecture Notes in Computer Science. Springer, October 2000.

[40] Amarnath Gupta, Bertran Ludäscher, Jeffrey S. Grethe, and Maryann E. Martone. Towards a formalisation of disease-specific ontologies for neuroinformatics. *Neural Networks*, 16:1277–1292, 2003.

[41] Volker Haarslev and Ralf Möller. Description of the RACER System and its Applications. In *International Workshop on Description Logics (DL-2001)*, August 2001.

[42] Midori A Harris and Helen Parkinson. Conference report: Standards and ontologies for functional genomics: towards unified ontologies for biology and biomedicine. *Comparative and Functional Genomics*, 4:116–120, 2003.

[43] James Hendler. Agents and the semantic web. *IEEE Intelligent Systems Journal*, 16:30–37, 2001.

[44] Steffen Hennig, Detlef Groth, and Hans Lehrach. Automated Gene Ontology annotation for anonymous sequence data. *Nucleic Acid Research*, 31:3712–3715, 2003.

[45] David P. Hill, Judith A. Blake, Joel E Richardson, and Martin Ringwald. Extension and integration of the Gene Ontology (GO): combining vocabularies with external vocabularies. *Genome Research*, 12:1982–1991, 2002.

[46] Mathew Horridge. A practical guide to building OWL ontologies with the Protégé-OWL plugin. *http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf*, 2004.

[47] Ian Horrocks. DAML+OIL: a reason-able web ontology language. In *Proc. of EDBT 2002*, number 2287 in Lecture Notes in Computer Science, pages 2–13. Springer, March 2002.

[48] Ian Horrocks and Peter F. Patel-Schneider. Reducing OWL entailment to description logic satisfiability. *Journal of Web Semantics*, 1(1):1–13, 2004.

[49] Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From SHIQ and RDF to OWL: the making of a web ontology language. *Web Semantics: Science, Services and Agents on the World Wide Web*, 1:7–26, 2003.

[50] Bo Hu, Srinandan Dasmahapatra, Paul Lewis, and Nigel Shadbolt. Ontology-based Medical Image Annotation with Description Logics. In *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, pages 77–82, November 2003.

[51] Lawrence Hunter. Ontologies for programs, not people. *Genome Biology*, 6:1002.1–1002.2, 2002.

[52] Tor-Kristian Jenssen and Eivind Hovig. The semantic web and biology. *Drug Discovery Today*, 7(19):992, 2002.

[53] Cliff Joslyn and Susan Mniszewski. Combinatorial Approaches to Bio-Ontology Management with Large Partially Ordered Sets. In *SIAM Workshop on Combinatorial Scientific Computing (CSC 04)*, February 2004.

[54] Cliff A. Joslyn, Susan M. Mniszewski, Andy Fulmer, and Gary Heaton. The Gene Ontology Categorizer. *Bioinformatics*, 20:i169–i177, 2004.

[55] Glynn Dennis Jr, Brad T Sherman, Douglas A Hosack, Jun Yang, Wei Gao, H Clifford Lane, and Richard A Lempicki. DAVID: database for annotations, visualisation, and integrated discovery. *Genome Biology*, 4, 2003.

[56] Satoshi Kamegai, Kenji Satou, and Akihiko Konagaya. Toward ontology-based knowledge extraction from biomedical literature. *Genome Informatics*, 13:576–577, 2002.

[57] Peter D. Karp. An ontology for biological function based on molecular interactions. *Bioinformatics*, 16:269–285, 2000.

[58] Peter D. Karp. Pathway databases: a case study in computational symbolic theories. *Science*, 293:2040–2044, 2001.

[59] Peter D. Karp, Monica Riley, Milton Saier, Ian T.Paulsen, Julio Collado-Vides, Suzanne M. Paley, Alida Pellegrini-Toole, César Bonavides, and Socorro Gama-Castro. The EcoCyc database. *Nuclei Acid Research*, 30:56–58, 2002.

[60] Salim Khan, Gang Situ, Keith Decker, and Carl J. Schmidt. GeneFigure: Automated Gene Ontology annotation. *Bioinformatics*, 19:2484–2485, 2003.

[61] Purvesh Khatri, Pratik Bhavsar, Gagandeep Bawa, and Sorin Draghici. Onto-tools: an ensemble of web accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nuclei Acids Research*, 32:W449–W456, 2004.

[62] J.-D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii. GENIA corpus - a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19:i180–i182, 2003.

[63] Oliver D. King, Rebecca E. Foulger, Selina S. Dwight, James V. White, and Frederick P. Roth. Predicting gene function from patterns of annotation. *Genome Research*, 13:896–904, 2003.

[64] Jacob Köhler. Integration of life science databases. *Biosilico*, 2:61–69, 2004.

[65] Jacob Köhler, Stephan Philippi, and Matthias Lange. SEMEDA: ontology based semantic integration of biological databases. *Bioinformatics*, 19:2420–2427, 2003.

[66] Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Feierbach B, Fisk DG, Hirschman JE, Hong EL, Issel-Tarver L, Nash R, Sethuraman A, Starr B, Thusfeld CL, Andrada R, Binkley G, Dong Q, Lane C, Schroeder M, Botstein D, and Cherry JM. Saccharomyces Genome Database (GSD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acid Research*, 32:D311–D314, 2004.

[67] Cynthia J. Krieger, Peifen Zhang anbd Lukas A. Mueller, Alfred Wang, Suzanne Paley, Martha Arnaud, John Pick, Seung Y. Rhee, and Peter D. Karp. MetaCyC: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research*, 32:D438–D442, 2004.

[68] Oh Byung Kwon. *I know what you need to buy*: context-aware multimedia-based recommendation system. *Expert Systems with Applications*, 25:387–400, 2003.

[69] Astrid Lægreid, Torgeir R. Hvidsten, Herman Midelfart, and Jan Komorowski. Predicting Gene Onotology Biological Process from temporal gene expression patterns. *Genome Research*, 13:965–979, 2003.

[70] Patrick Lambrix, Manal Habbouche, and Marta Pérez. Evaluation of ontology development tools for bioinformatics. *Bioinformatics*, 19:1564–1571, 2003.

[71] Ning Lan, Gaetano T Montelione, and Mark Gerstein. Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level. *Current opinion in chemical biology*, 7:44–54, 2003.

[72] Ora Lassila and Deborah L. McGuinness. The role of frame-based representation on the semantic web. *Electronic Transactions on Artificial Intelligence*, 5:1403–2031, 2001.

[73] Sung Geun Lee, Jung Uk Hur, and Yang Seok Kim. A graph-theoretic modeling on GO space fro biological interpretation of gene clusters. *Bioinformatics*, 20:381–388, 2004.

[74] Yuefeng Li and Ning Zhong. Web mining model and its applications for information gathering. *Knowledge-Based Systems*, 2004.

[75] Shun-Chieh Lin and Shian-Shyong Tseng. Constructing detection knowledge for DDoS intrusion tolerance. *Expert Systems with applications*, 27:379–390, 2004.

[76] Jane Lomax and Alexa T. McCray. Mapping the Gene Ontology into the Unified Medical Language System. *Comparative and Functional Genomics*, 5:354–361, 2004.

[77] P.W. Lord, R.D. Stevens, A. Brass, and C.A. Goble. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics*, 19:1275–1283, 2003.

[78] Robert Mack and Michael Hehenberger. Text-based knowledge discovery: search and mining of life-sciences documents. *Drug Discovery Today*, 7(11):S89–S98, 2002.

[79] Pierre Maziére, Claude Granier, and Franck Molina. A description scheme of biological processes based on elementary bricks of action. *J. Mol. Biol.*, 339:77–88, 2004.

[80] Peter E. Midford. An overview of machine ontologies for ethologists. *http://www.mesquiteproject.org/ontology/*, 2002.

[81] N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Fergerson, , and M. A. Musen. Creating semantic web contents with Protégé-2000. *IEEE Intelligent Systems*, 16:60–71, 2001.

[82] P.V. Ogren, K.B. Cohen, G.K. Acquaah-Mensah, J.Eberlein, and L. Hunter. The compositional structure of Gene Ontology terms. In *Pac Symp Biocomput.*, pages 214–25, 2004.

[83] Trevor Paterson, Jessie B Kennedy, Martin R. Pullan, Alan Cannon, Kate Armstrong, Mark F. Watson, Cédric Raguenaud, Sarah M. McDonald, and Gordon Russell. A universal character model and ontology of defined terms for taxonomic description. In *Data Integration in the Life Sciences: First International Workshop, DILS*, number 2994 in Lecture Notes in Computer Science, pages 63–78. Springer, March 2004.

[84] Soumya Raychaudhuri, Jeffrey T. Chang, Patrick D. Sutphin, and Russ B. Altman. Associating genes with Gene Ontology codes using a maximum entropy analysis of biomedical literature. *Genome Research*, 12:203–214, 2002.

[85] A.L. Rector, J.E. Rogers, P.E. Zanstra, and E. Van Der Haring. OpenGALEN: Open source medical terminology and tools. In *Proc AMIA Symp.*, page 982, 2003.

[86] Jason Reed. Trends in commercial bioinformatics. *Biotechnology Review*, pages 1–19, 2000.

[87] Stuart C. G. Rison, T. Charles Hodgman, and Janet M. Thornton. Comparison of functional annotation schemes for genomes. *Funct. Integr. Genomics*, 1:56–69, 2000.

[88] Peter N. Robinson, Andreas Wollstein, Ulrike Böhme, and Brad Beattie. NetAffx Gene Ontology Mining Tool: a visual approach for microarray data analysis. *Bioinformatics*, 20:979–981, 2004.

[89] Peter N. Robinson, Andreas Wollstein, Ulrike Böhme, and Brad Beattie. Ontologizing gene-expression microarray data: characterizing clusters with Gene Ontology. *Bioinformatics*, 20:979–981, 2004.

[90] Jonathan Schug, Sharon Diskin, Joan Mazzarelli, Brian P. Bunk, and Jr. Christian J. Stoeckert. Predicting Gene ontology functions from ProDom and CDD protein domains. *Genome Research*, 12:648–655, 2002.

[91] Jonathan Simon, Matthew Fielding, and Barry Smith. Using philosophy to improve the coherence and interoperability of applications ontologies: a field report on the collaboration of IFOMIS and L&C. In *First International Workshop on Philosophy and Informatics (Cologne)*, April 2004.

[92] Barry Smith. *The Blackwell Guide to the Philosophy of Computing and Information (Blackwell Philosophy Guides)*, chapter 17. Oxford:Blackwell, 2003.

[93] Barry Smith, Jacob Köhler, and Anand Kumar. On the application of formal principles to life science data: A case study in the gene ontology. In *International Workshop on Data Integration in the Life Sciences (DILS 2004)*, Lecture Notes in Computer Science. Springer, March 2004.

[94] Steffen Staab. The semantic web - new ways to present and integrate information. *Comparative and Functional Genomics*, 4:98–103, 2003.

[95] R. Stevens, C. Goble, I. Horrocks, and S. Bechhofer. Building a Bioinformatics Ontology Using OIL. *IEEE Transactions on Information Technology and Biomedicine*, 6(2):135–41, Jun 2002.

[96] R. Stevens, C.A. Goble, and S. Bechhofer. Ontology-based Knowledge Representation for Bioinformatics. *Briefings in Bioinformatics*, 1(4):398–416, November 2000.

[97] Robert Stevens, Carole Goble, Norman W. Paton, Sean Bechhofer, Gary Ng, Patricia Baker, and Andy Brass. Complex Query Formulation Over Diverse Information Sources in TAMBIS. In Zoe Lacroix and Terence Critchlow, editors, *Bioinformatics: Managing Scientific Data*. Morgan Kaufmann, May 2003.

[98] Robert Stevens, Chris Wroe, Sean Bechhofer, Phillip Lord, Alan Rector, and Carole Goble. Building ontologies in DAML+OIL. *Comparative and Functional Genomics*, 4:133–141, 2003.

[99] Robert D. Stevens, Alan J. Robinson, and Carole A. Goble. Mygrid: personalised bioinformatics on the information grid. *Bioinformatics*, 19:i302–i304, 2003.

[100] Jeffrey Thomas. Finding an oasis in the desert of bioinformatics. *Biosilico*, 1(2):56–58, 2003.

[101] Stefano Volinia, Rita Evangelisti, Francesca Francioso, Diego Arcelli, Massimo Carella, and Paolo Gasparini. GOAL: automated Gene Ontology analysis of expression profiles. *Nucleic Acids Research*, 32:W492–W499, 2004.

[102] Jennifer Williams and William Andersen. Bringing Ontology to the Gene Ontology. *Comparative and Functional Genomics*, 4:90–93, 2003.

[103] Patrick Henry Winston. *The psychology of computer vision*. MacGraw-Hill, New York, 1975.

[104] C.J. Wroe, R.D. Stevens, C.A. Goble, and M. Ashburner. A Methodology to Migrate the Gene Ontology to a Description Logic Environment Using DAML+OIL. In *8th Pacific Symposium on biocomputing (PSB)*, pages 624–636, 2003.

[105] Hanqing Xie, Alon Wasserman, Zurit Levine, Amit Novik, Vladimir Grebinskiy, Avi Shoshan, and Liat Mintz. Large-scale protein annotation through Gene Ontology. *Genome Research*, 12:785–794, 2002.

[106] Iwei Yeh, Peter D. Karp, Natalya F. Noy, and Russ B. Altman. Knowledge acquisition, consistency checking and concurrency control for Gene Ontology (GO). *Bioinformatics*, 19(2):241–248, 2003.

[107] Barry R Zeeberg, Weimin Feng, Geoffrey Wang, May D Wang, Anthony T Fojo, Margot Sunshine, Sudarshan Narasimhan, David W Kane, William C Reinhold, Samir Lababidi, Kimberly J Bussey, Joseph Riss, J Carl Barret, and John N Weinstein. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biology*, 4, 2003.

[108] Günther Zehetner. OntoBlast function: from sequence similarities directly to potential functional annotations by ontology terms. *Nucleic Acid Research*, 31:3799–3803, 2003.

[109] Bing Zhang, Denise Schmoyer, Stefan Kirov, and Jay Snoddy. GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC bioinformatics*, 5(16), 2004.

[110] Zuo Zhihong and Zhou Mingtian. Web Ontology Language OWL and its Description Logic Foundation. In *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies*, pages 157–160, 2003.

Appendix A

Binding new relationships

New position: Actinin binding (GO:0042805) should be under cytoskeletal protein binding (GO:0008092) instead of protein binding (GO:0005515)
New IsA link: Activin binding (GO:0048185) IsA hormone binding (GO:0042562)
New IsA link: Acyl carrier activity (GO:0000036) IsA acyl binding (GO:0000035)
New Position: ADP-ribosylation factor binding (GO:0030306) should be under GTPase binding (GO:0051020) instead of protein binding (GO:0005515)
New IsA link: Adrenocorticotropin receptor activity (GO:0004978) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: Adrenocorticotropin receptor activity (GO:0004978) IsA hormone binding (GO:0042562)
New IsA link: Angiotensin receptor activity (GO:0001595) IsA neuropeptide receptor activity (GO:0008188)
New position: apolipoprotein E receptor binding (GO:0050749) should be under low-density lipoprotein receptor binding (GO:0050750) instead of receptor binding (GO:0005102)
New IsA link: Aryl hydrocarbon receptor binding (GO:0017162) IsA transcription factor binding (GO:0008134)
New IsA link: bacteriochlorophyll binding (GO:0042314) IsA protein binding (GO:0005515)
New IsA link: beta-endorphin receptor activity (GO:0004979) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: bombesin receptor activity (GO:0004946) IsA neuropeptide receptor activity (GO:0008188)
New IsA link: bradykinin receptor activity (GO:0004947) IsA neuropeptide receptor activity (GO:0008188)
New position: cadherin binding (GO:0045296) should be under cell adhesion molecule binding (GO:0050839) instead of protein binding (GO:0005515)

```
New IsA link: CD4 receptor binding (GO:0042609) IsA antigen binding
(GO:0003823)
New IsA link: Cholecystokinin receptor activity (GO:0004951) IsA
neuropeptide receptor activity (GO:0008188)
New IsA link: Chromatin binding (GO:0003682) IsA protein binding
(GO:0005515)
New IsA link: cyclin binding (GO:0030332) IsA growth factor
binding (GO:0019838)
New IsA link: dynein binding (GO:0045502) IsA enzyme binding
(GO:0019899)
New position: dynein binding (GO:0045502) should be under
cytoskeletal protein
binding (GO:0008092) instead of protein binding (GO:0005515)
New IsA link: ecdysone binding (GO:0035100) IsA steroid binding
(GO:0005496)
New IsA link: epidermal growth factor binding (GO:0048408) IsA
hormone binding (GO:0042562)
New IsA link: estrogen receptor binding (GO:0030331) IsA
transcription factor binding (GO:0008134)
New IsA link: FAD binding (GO:0050660) IsA adenyl nucleotide
binding (GO:0030554)
New IsA link: FMN binding (GO:0010181) IsA nucleotide binding
(GO:0000166)
New IsA link: FAD binding (GO:0050660) IsA adenyl nucleotide
binding (GO:0030554)
New IsA link: FMN binding (GO:0010181) IsA nucleotide binding
(GO:0000166)
New position: Galanin receptor activity (GO:0004966) should be
under neuropeptide receptor activity (GO:0008188) instead
of peptide receptor activity, G-protein coupled (GO:0008528)
New IsA link: Glutamate binding (GO:0016595) IsA
neurotransmitter binding (GO:0042165)
New IsA link: Glycine binding (GO:0016594) IsA neurotransmitter
binding (GO:0042165)
New position: Glycosaminoglican binding (GO:0005539) should be
under polysaccharide binding (GO:0001871) instead of binding
(GO:0005488)
New position: hemoglobin binding (GO:0030492) should be under
protein binding (GO:0005515) instead of binding (GO:0005488)
New position: hydroxyapatite binding (GO:0046848) should be
under phosphate binding (GO:0042301) instead of binding
(GO:0005488)
```

```
New IsA link: interleukin binding (GO:0019965) IsA growth
factor binding (GO:0019838)
New IsA link: interleukin-1 receptor binding (GO:0005149) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-2 receptor binding (GO:0005134) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-3 receptor binding (GO:0005135) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-4 receptor binding (GO:0005136) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-6 receptor binding (GO:0005138) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-7 receptor binding (GO:0005139) IsA
antigen binding (GO:0003823)
New IsA link: interleukin-8 receptor binding (GO:0005153) IsA
antigen binding (GO:0003823)
New position: ISG15 carrier activity (GO:0019793) should be
under protein carrier activity (GO:0008320) instead of protein
binding (GO:0005515)
New IsA link: kinesin binding (GO:0019894) IsA enzyme binding
(GO:00019899)
New IsA link: macrophage colony stimulating factor receptor
binding (GO:0005157) IsA antigen binding (GO:0003823)
New IsA link: melanocyte stimulating hormone receptor activity
(GO:0004980) IsA hormone binding (GO:0042562)
New position: microfibril binding (GO:0050436) should be under
extracellular matrix binding (GO:0050840) instead of binding
(GO:0005488)
New IsA link: myosin binding (GO:0017022) IsA enzyme binding
(GO:0019899)
New IsA link: NADPH binding (GO:0050661) IsA purine nucleotide
binding (GO:0017076)
New IsA link: neuropeptide binding (GO:0042923) IsA
neurotransmitter binding (GO:0042165)
New IsA link: neuropeptide receptor activity (GO:0008188) IsA
neurotransmitter receptor activity (GO:0030594)
New IsA link: peptide YY receptor activity (GO:0001601) IsA
hormone binding (GO:0042562)
New IsA link: peroxisome targeting signal receptor activity
(GO:0005051) IsA peptide receptor activity (GO:0001653)
New IsA link: platelet derived growth factor binding (GO:0048407)
IsA peptide binding (GO:0042277)
```

```
New position: pyridoxal phosphate binding (GO:0030170) should
be under vitamin binding (GO:0019842) instead of binding
(GO:0005488)
New IsA link: Retinoic acid receptor binding (GO:0042974) IsA
transcription factor binding (GO:0008134)
New position: Ribonucleoprotein binding (GO:0043021) should be
under protein binding (GO:0005515) instead of binding (GO:0005488)
New IsA link: snRNA modification guide activity (GO:0030566) IsA
snRNA binding (GO:0017069)
New IsA link: rRNA modification guide activity (GO:0030556) IsA
rRNA binding (GO:0019843)
New IsA link: tRNA modification guide activity (GO:0030557) IsA
tRNA binding (GO:0000049)
New IsA link: somatostatin receptor activity (GO:0004994) IsA
neuropeptide receptor activity (GO:0008188)
New IsA link: somatostatin receptor activity (GO:0004994) IsA
hormone binding (GO:0042562)
New position: tau protein binding (GO:0048156) should be under
cytoskeletal protein binding (GO:0008092) instead of protein
binding (GO:0005515)
New IsA link: thyroid hormone receptor binding (GO:0046966) IsA
transcription factor binding (GO:0008134)
New IsA link: Transforming Growth Factor beta binding (GO:0050431)
IsA growth factor binding (GO:0019838)
New IsA link: Transforming Growth Factor beta binding (GO:0050431)
IsA peptide binding (GO:0042277)
New IsA link: tRNA modification guide activity (GO:0030557) IsA
tRNA binding (GO:0000049)
New IsA link: ubiquinone binding (GO:0048039) IsA coenzyme
binding (GO:0050662)
New IsA link: vasopressin receptor activity (GO:0005000) IsA
hormone binding (GO:0042562)
New IsA link: vitamin D receptor binding (GO:0042809)
transcription factor binding (GO:0008134)
```

Appendix B

Transporter activity new relationships

New IsA link: GO:0015259 : Glutamate channel activity IsA GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0015415 : phosphate-transporting ATPase activity IsA GO:0015320 : phosphate carrier activity
New IsA link: GO:0015649 : 2-keto-3-deoxygluconate:hydrogen symporter activity IsA GO:0015296 : anion:cation symporter activity
New IsA link: GO:0015649 : 2-keto-3-deoxygluconate:hydrogen symporter activity IsA GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0005278 : acetylcholine:hydrogen antiporter activity IsA GO:0008324 : cation transporter activity
New IsA link: GO:0005278 : acetylcholine:hydrogen antiporter activity IsA GO:0015238 : drug transporter activity
New IsA link: GO:0042962 : acridine:proton antiporter activity IsA GO:0015491 : cation:cation antiporter activity
New IsA link: GO:0042962 : acridine:proton antiporter activity IsA GO:0008324 : cation transporter activity
New position: GO:0015217 : ADP transporter activity should be under GO:0000295 : adenine nucleotide transporter activity instead of GO:0015216 : purine nucleotide transporter activity
New position: GO:0005347 : ATP transporter activity should be under GO:0000295 : adenine nucleotide transporter activity instead of GO:0015216 : purine nucleotide transporter activity
New IsA link: GO:0015655 : alanine:sodium symporter activity IsA GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0005352 : alpha-glucoside:hydrogen symporter activity IsA GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0015532 : alpha-ketoglutarate:hydrogen symporter activity

IsA GO:0015296 : anion:cation symporter activity
New IsA link: GO:0005253 : anion channel activity IsA
GO:0008509 : anion transporter activity
New IsA link: GO:0015296 : anion:cation symporter activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0015296 : anion:cation symporter activity IsA
GO:0008509 : anion transporter activity
New IsA link: GO:0015583 : beta-glucoside [arbutin-salicin-cellobiose]
permease activity IsA GO:0015579 : glucose permease activity
New IsA link: GO:0015582 : beta-glucoside permease activity IsA
GO:0015579 : glucose permease activity
New IsA link: GO:0008508 : bile acid:sodium symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0015657 : branched-chain amino acid:sodium
symporter activity IsA GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015525 : carbonyl cyanide m-chlorophenylhydrazone/nalidixic
acid/organomercurials:hydrogen antiporter activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0005261 : cation channel activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0046583 : cation efflux permease activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0019829 : cation-transporting ATPase activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0015531 : citrate:hydrogen symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: cystine:glutamate antiporter activity IsA
GO:0005326 : neurotransmitter transporter activity
New IsA link: cystine:glutamate antiporter activity IsA
GO:0015328 : cystine porter activity
New IsA link: GO:0015519 : D-xylose:hydrogen symporter activity IsA
GO:0005402 : cation:sugar symporter activity
New position: GO:0042936 : dipeptide transporter activity should be under
GO:0015198 : oligopeptide transporter activity instead of
GO:0015197 : peptide transporter activity (*)
New IsA link: GO:0005329 : dopamine transporter activity IsA
GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0015408 : ferric-transporting ATPase activity IsA
GO:0019829 : cation-transporting ATPase activity
New IsA link: GO:0015535 : fucose:hydrogen symporter activity IsA
hexose:hydrogen symporter activity
New IsA link: GO:0015517 : galactose:hydrogen symporter activity IsA

GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0015517 : galactose:hydrogen symporter activity IsA
GO:0009679 : hexose:hydrogen symporter activity
New IsA link: GO:0005331 : gamma-aminobutyric acid transporter activity IsA
GO:0005328 : neurotransmitter:sodium symporter activity
New IsA link: GO:0005331 : gamma-aminobutyric acid transporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015495 : gamma-aminobutyric acid:hydrogen symporter
activity IsA GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0005332 : gamma-aminobutyric acid:sodium symporter activity
IsA GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015501 : glutamate:sodium symporter activity IsA
GO:0005328 : neurotransmitter:sodium symporter activity
New IsA link: GO:0015501 : glutamate:sodium symporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015187 : glycine transporter activity IsA
GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0015375 : glycine:sodium symporter activity IsA
GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0015375 : glycine:sodium symporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015486 : glycoside-pentoside-hexuronide:cation symporter
activity IsA GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0009679 : hexose:hydrogen symporter activity IsA
GO:0015149 : hexose transporter activity
New IsA link: GO:0015398 : high affinity ammonium transporter activity IsA
GO:0008513 : organic cation porter activity
New IsA link: GO:0005316 : high affinity inorganic phosphate:sodium symporter
activity IsA GO:0015370 : solute:sodium symporter activity
New IsA link: GO:0005316 : high affinity inorganic phosphate:sodium symporter
activity IsA GO:0015319 : sodium:inorganic phosphate symporter activity
New IsA link: GO:0005297 : hydrogen:proline symporter activity IsA
GO:0015295 : solute:hydrogen symporter activity
New IsA link: GO:0015522 : hydrophobic uncoupler:hydrogen antiporter activity
IsA GO:0008324 : cation transporter activity
New IsA link: GO:0005315 : inorganic phosphate transporter activity IsA
GO:0008509 : anion transporter activity
New IsA link: GO:0005216 : ion channel activity IsA
GO:0015075 : ion transporter activity
New IsA link: GO:0015612 : L-arabinose porter activity IsA
GO:0015407 : monosaccharide-transporting ATPase activity
New IsA link: GO:0015512 : L-threonine permease activity IsA

GO:0015195 : L-threonine transporter activity
New position: GO:0015400 : low affinity ammonium transporter activity
should be under GO:0008513 : organic cation porter activity instead of
GO:0015291 : porter activity
New IsA link: GO:0009673 : low affinity phosphate transporter activity IsA
GO:0015320 : phosphate carrier activity
New IsA link: GO:0015366 : malate:hydrogen symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0015366 : malate:hydrogen symporter activity IsA
GO:0015140 : malate transporter activity
New IsA link: GO:0015609 : maltooligosaccharide porter activity IsA
GO:0015422 : oligosaccharide-transporting ATPase activity
New IsA link: GO:0015481 : maltose transporting porin activity IsA
GO:0015478 : oligosaccharide transporting porin activity
New IsA link: GO:0005364 : maltose:hydrogen symporter activity IsA
GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0015410 : manganese-transporting ATPase activity IsA
GO:0019829 : cation-transporting ATPase activity
New IsA link: GO:0015311 : monoamine:hydrogen antiporter activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0015311 : monoamine:hydrogen antiporter activity IsA
GO:0015238 : drug transporter activity
New IsA link: GO:0005451 : monovalent cation:proton antiporter activity IsA
GO:0015491 : cation:cation antiporter activity
New IsA link: GO:0005366 : myo-inositol:hydrogen symporter activity IsA
GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0005295 : neutral amino acid:sodium symporter activity IsA
GO:0015175 : neutral amino acid transporter activity
New IsA link: GO:0015413 : nickel-transporting ATPase activity IsA
GO:0019829 : cation-transporting ATPase activity
New IsA link: GO:0030184 : nitric oxide transporter activity IsA
GO:0005326 : neurotransmitter transporter activity
New IsA link: GO:0015394 : nucleoside (uridine) permease activity IsA
GO:0015536 : nucleoside permease activity
New position: GO:0005415 : nucleoside:sodium symporter activity should be
under GO:0015370 : solute:sodium symporter activity instead of
GO:0015294 : solute:cation symporter activity
New IsA link: GO:0015421 : oligopeptide-transporting ATPase activity IsA
GO:0015637 : peptide uptake permease activity
New IsA link: GO:0015315 : organophosphate:inorganic phosphate antiporter
activity IsA GO:0015301 : anion:anion antiporter activity
New IsA link: GO:0015367 : oxoglutarate:malate antiporter activity IsA

GO:0015491 : cation:cation antiporter activity
New IsA link: GO:0015492 : phenylalanine:hydrogen symporter activity IsA
GO:0015494 : aromatic amino acid:hydrogen symporter activity
New IsA link: GO:0015317 : phosphate:hydrogen symporter activity IsA
GO:0015320 : phosphate carrier activity
New IsA link: GO:0015317 : phosphate:hydrogen symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0015312 : polyamine:hydrogen antiporter activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0015312 : polyamine:hydrogen antiporter activity IsA
GO:0015238 : drug transporter activity
New IsA link: GO:0017032 : potassium:amino acid transporter activity IsA
GO:0005416 : cation:amino acid symporter activity
New IsA link: GO:0005298 : proline:sodium symporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015266 : protein channel activity IsA
GO:0008565 : protein transporter activity
New IsA link: GO:0005427 : proton-dependent oligopeptide transporter activity
IsA GO:0015322 : oligopeptide porter activity
New IsA link: GO:0015389 : pyrimidine- and adenine-specific:sodium symporter
activity IsA GO:0015391 : nucleobase:cation symporter activity
New IsA link: GO:0015389 : pyrimidine- and adenine-specific:sodium symporter
activity IsA GO:0005337 : nucleoside transporter activity
New IsA link: GO:0015561 : rhamnose:hydrogen symporter activity IsA
GO:0015153 : rhamnose transporter activity
New IsA link: GO:0015561 : rhamnose:hydrogen symporter activity IsA
GO:0009679 : hexose:hydrogen symporter activity
New IsA link: GO:0015533 : shikimate:hydrogen symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0005283 : sodium:amino acid transporter activity IsA
GO:0005416 : cation:amino acid symporter activity
New IsA link: GO:0005283 : sodium:amino acid transporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0017153 : sodium:dicarboxylate symporter activity IsA
organic acid:sodium symporter activity
New IsA link: GO:0008507 : sodium:iodide symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0005469 : succinate:fumarate antiporter activity IsA
GO:0015138 : fumarate transporter activity
New IsA link: GO:0005469 : succinate:fumarate antiporter activity IsA
GO:0015141 : succinate transporter activity
New IsA link: GO:0008506 : sucrose:hydrogen symporter activity IsA

GO:0005402 : cation:sugar symporter activity
New IsA link: GO:0008512 : sulfate:hydrogen symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0005369 : taurine:sodium symporter activity IsA
GO:0015296 : anion:cation symporter activity
New IsA link: GO:0005369 : taurine:sodium symporter activity IsA
GO:0005343 : organic acid:sodium symporter activity
New IsA link: GO:0015520 : tetracycline:hydrogen antiporter activity IsA
GO:0008324 : cation transporter activity
New IsA link: GO:0015250 : water channel activity IsA
GO:0005372 : water transporter activity
New IsA link: GO:0008559 : xenobiotic-transporting ATPase activity IsA
GO:0042910 : xenobiotic transporter activity
New IsA link: GO:0015341 : zinc efflux permease activity IsA
GO:0046583 : cation efflux permease activity

Appendix C

Metabolism new relationships

New IsA link: peptidyl-leucine racemization GO:0019129 IsA peptidyl-leucine modification GO:0018204
New IsA link: peptidyl-methionine racemization GO:0019123 IsA peptidyl-leucine modification GO:0018204
New IsA link: peptidyl-phenylalanine racemization GO:0019125 IsA peptidyl-phenylalanine modification GO:0018207
New IsA link: peptidyl-serine racemization GO:0019126 IsA peptidyl-serine modification GO:0018209
New IsA link: peptidyl-tryptophan racemization GO:0019128 IsA peptidyl-tryptophan modification GO:0018211
New IsA link: phenylmercury acetate catabolism GO:0019506 IsA aromatic compound catabolism (GO:0019439)
New IsA link: phosphatidylethanolamine biosynthesis GO:0006646 IsA glycerophospholipid biosynthesis GO:0046474
New IsA link: phosphatidylethanolamine catabolism GO:0046338 IsA glycerophospholipid catabolism GO:0046475
New IsA link: phosphatidylethanolamine metabolism GO:0046337 IsA glycerophospholipid metabolism GO:0006650
New IsA link: phospholipid dephosphorylation GO:0046839 IsA phospholipid metabolism GO:0006644
New IsA link: phytochelatin biosynthesis GO:0046938 IsA peptide biosynthesis GO:0043043
New IsA link: poly-N-acetyllactosamine biosynthesis GO:0030311 IsA amine biosynthesis (GO 0009309)
New IsA link: poly-N-acetyllactosamine catabolism GO:0030310 GO:0046230 IsA amine catabolism (GO:0009310)
New IsA link: porphyrin biosynthesis GO:0006779 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: porphyrin catabolism GO:0006787 IsA heterocycle catabolism

(GO:0046700)
New IsA link: protein amino acid myristoylation GO:0018319 IsA protein amino acid lipidation GO:0042050
New IsA link: protein amino acid palmitoleylation GO:0045235 IsA protein amino acid lipidation GO:0042050
New IsA link: protein amino acid palmitoylation GO:0018318 IsA protein amino acid lipidation GO:0042050
New IsA link: protein amino acid prenylation GO:0018346 IsA protein amino acid lipidation GO:0042050
New IsA link: proteoglycan biosynthesis GO:0030166 IsA polysaccharide biosynthesis (GO:0000271)
New IsA link: proteoglycan biosynthesis GO:0030166 IsA glycoprotein biosynthesis (GO:0009101)
New IsA link: proteoglycan catabolism GO:0030167 IsA polysaccharide catabolism (GO:0000272)
New IsA link: proteoglycan metabolism GO:0006029 IsA polysaccharide metabolism GO:0005976
New IsA link: proteoglycan metabolism GO:0006029 IsA glycoprotein metabolism GO:0009100
New IsA link: protocatechuate biosynthesis GO:0046279 IsA carboxylic acid biosynthesis (GO 0046394)
New IsA link: pyridine biosynthesis GO:0046220 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: pyridine catabolism GO:0046221 IsA heterocycle catabolism (GO:0046700)
New IsA link: pyruvate biosynthesis from acetate GO:0019687 IsA pyruvate biosynthesis GO:0042866
New IsA link: pyruvate biosynthesis GO:0042866 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: pyruvate catabolism GO:0042867 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: quinate biosynthesis GO:0042194 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: quinate catabolism GO:0019631 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: retinoid metabolism GO:0001523 IsA terpene metabolism (GO:0042214)
New IsA link: RNA localization GO:0006403 IsA RNA metabolism (GO:0016070)
New IsA link: rRNA transcription GO:0009303 IsA rRNA metabolism GO:0016072
New IsA link: S-adenosylhomocysteine catabolism GO:0019510 IsA homocysteine metabolism GO:0050667

New IsA link: S-adenosylhomocysteine catabolism GO:0019510 IsA amino acid derivative catabolism GO:0042219
New IsA link: S-adenosylmethioninamine biosynthesis GO:0006557 IsA amine biosynthesis (GO 0009309)
New IsA link: S-adenosylmethionine biosynthesis GO:0006556 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: S-adenosylmethionine biosynthesis GO:0006556 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: S-adenosylmethionine biosynthesis GO:0006556 IsA amino acid derivative biosynthesis (GO:0042398)
New IsA link: S-adenosylmethionine catabolism GO:0050843 IsA aromatic compound catabolism (GO:0019439)
New IsA link: S-adenosylmethionine catabolism GO:0050843 IsA heterocycle catabolism (GO:0046700)
New IsA link: S-adenosylmethionine catabolism GO:0050843 IsA amino acid derivative catabolism GO:0042219
New IsA link: s-triazine compound catabolism GO:0042204 IsA aromatic compound catabolism (GO:0019439)
New IsA link: s-triazine compound catabolism GO:0042204 IsA heterocycle catabolism (GO:0046700)
New IsA link: salicylic acid biosynthesis GO:0009697 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: salicylic acid catabolism GO:0046244 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: snRNA transcription GO:0009301 IsA snRNA metabolism GO:0016073
New IsA link: sorbose biosynthesis GO:0042847 IsA alcohol biosynthesis (GO:0046165)
New IsA link: sorbose catabolism GO:0042848 IsA hexose catabolism GO:0019320
New IsA link: sphingosine biosynthesis GO:0046512 IsA amine biosynthesis (GO:0009309)
New IsA link: sterigmatocystin biosynthesis GO:0045461 IsA aromatic compound
biosynthesis (GO:0019438)
New IsA link: sterigmatocystin biosynthesis GO:0045461 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: sterigmatocystin catabolism GO:0045574 IsA aromatic compound catabolism (GO:0019439)
New IsA link: sterigmatocystin catabolism GO:0045574 IsA heterocycle catabolism (GO:0046700)
New IsA link: sterol biosynthesis GO:0016126 IsA alcohol biosynthesis

(GO:0046165)
New IsA link: sterol catabolism GO:0016127 IsA alcohol catabolism
(GO:0046164)
New IsA link: sulfate assimilation via adenylyl sulfate reduction
GO:0010134 IsA sulfate
reduction GO:0019419
New IsA link: teichoic acid biosynthesis GO:0019350 IsA polyol
biosynthesis GO:0046173
New IsA link: teichoic acid biosynthesis GO:0019350 IsA carboxylic acid
biosynthesis (GO 0046394)
New IsA link: teichuronic acid biosynthesis GO:0050845 IsA organic acid
biosynthesis (GO:0016053)
New IsA link: terpenoid indole alkaloid biosynthesis GO:0009709 IsA indole
derivative biosynthesis GO:0042435
New IsA link: thiamin and derivative biosynthesis GO:0042724 IsA
water-soluble vitamin biosynthesis (GO:0042364)
New IsA link: thiamin and derivative catabolism GO:0042725 IsA
water-soluble vitamin catabolism (GO:0042365)
New IsA link: thiocyanate catabolism GO:0046265 IsA organic acid
catabolism (GO:0016054)
New IsA link: toluene-4-sulfonate catabolism GO:0046269 IsA organic acid
catabolism (GO:0016054)
New IsA link: trichloroethylene catabolism GO:0050696 IsA chlorinated
hydrocarbon catabolism (GO:0042205)
New IsA link: triethanolamine catabolism GO:0046267 IsA alcohol catabolism
(GO:0046164)
New IsA link: trisporic acid biosynthesis (GO:0046842) IsA carboxylic acid
biosynthesis (GO:0046394)
New IsA link: UDP-N-acetylgalactosamine biosynthesis GO:0019277 IsA
nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: UDP-N-acetylgalactosamine catabolism GO:0019278 IsA
UDP-glucose catabolism GO:0006258
New IsA link: UDP-N-acetylglucosamine biosynthesis GO:0006048 IsA
nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: UDP-N-acetylglucosamine catabolism GO:0006049 IsA
UDP-glucose catabolism GO:0006258
New IsA link: urate catabolism GO:0019628 IsA heterocycle catabolism
(GO:0046700)
New IsA link: urate catabolism GO:0019628 IsA organic acid catabolism
(GO:0016054)
New IsA link: vanillin biosynthesis GO:0042189 IsA aldehyde biosynthesis
GO:0046184

New IsA link: vanillin catabolism GO:0042190 IsA aldehyde catabolism GO:0046185
New IsA link: vitamin B2 and derivative biosynthesis GO:0042727 IsA water-soluble vitamin biosynthesis (GO:0042364)
New IsA link: vitamin B2 and derivative catabolism GO:0042728 IsA water-soluble vitamin catabolism (GO:0042365)
New IsA link: (R)-4-hydroxymandelate catabolism (GO:0019599) IsA phenol catabolism (GO:0019336)
New IsA link: (R)-mandelate catabolism to benzoate (GO:0019597) IsA xenobiotic catabolism (GO:0042178)
New IsA link: 1,1,1-trichloro-2,2-bis-(4'-chlorophenyl)ethane catabolism (GO:0042188) IsA aromatic compound catabolism (GO:0019439)
New IsA link: 1,3-dichloro-2-propanol biosynthesis (GO:0046309) IsA alcohol biosynthesis (GO:0046165)
New IsA link: 1,3-dichloro-2-propanol catabolism (GO:0046310) IsA alcohol catabolism (GO:0046164)
New IsA link: 1,4-dichlorobenzene catabolism (GO:0019261) IsA aromatic compound catabolism (GO:0019439)
New IsA link: 1-aminocyclopropane-1-carboxylate biosynthesis (GO:0042218) IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: 1-aminocyclopropane-1-carboxylate biosynthesis (GO:0042218) IsA amine biosynthesis (GO:0009309)
New IsA link: 1-aminocyclopropane-1-carboxylate catabolism (GO:0042217) IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 2,4,5-trichlorophenoxyacetic acid catabolism (GO:0046228) IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 2,4-dichlorobenzoate catabolism (GO:0046298) IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 2,4-dichlorobenzoate catabolism (GO:0046298) IsA aromatic compound catabolism (GO:0019439)
New IsA link: 2,4-dichlorophenoxyacetic acid catabolism (GO:0046300) IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 2,4-dichlorophenoxyacetic acid catabolism (GO:0046300) IsA aromatic compound catabolism (GO:0019439)
New IsA link: 2-aminobenzenesulfonate catabolism GO:0046230 IsA organic acid catabolism (GO:0016054)
New IsA link: 2-aminobenzenesulfonate catabolism GO:0046230 IsA amine catabolism (GO:0009310)
New IsA link: 2-aminobenzoate catabolism GO:0019259 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 2-aminobenzoate catabolism GO:0019259 IsA aromatic compound catabolism (GO:0019439)

New IsA link: 2-aminobenzoate catabolism GO:0019259 IsA amine catabolism (GO:0009310)
New IsA link: 2-aminoethylphosphonate catabolism GO:0019635 IsA amine catabolism (GO:0009310)
New IsA link: 2-chloro-N-isopropylacetanilide catabolism GO:0046302 IsA aromatic compound catabolism (GO:0019439)
New IsA link: 3-hydroxyphenylacetate catabolism GO:0019610 IsA phenol catabolism (GO:0019336)
New IsA link: 3-hydroxyphenylacetate catabolism GO:0019610 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 3-methylquinoline catabolism GO:0019335 IsA aromatic compound catabolism (GO:0019439)
New IsA link: 3-methylquinoline catabolism GO:0019335 IsA heterocycle catabolism (GO:0046700)
New IsA link: 3-phenylpropionate catabolism GO:0019380 IsA aromatic compound catabolism (GO:0019439)
New IsA link: 3-phenylpropionate catabolism GO:0019380 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 4-hydroxyproline biosynthesis GO:0019472 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: 4-hydroxyproline catabolism GO:0019470 IsA heterocycle catabolism (GO:0046700)
New IsA link: 4-toluenecarboxylate catabolism GO:0019612 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: 9-cis-retinoic acid biosynthesis GO:0042904 IsA terpene biosynthesis (GO: 0046246)
New IsA link: 9-cis-retinoic acid biosynthesis GO:0042904 IsA hormone biosynthesis (GO:0042446)
New IsA link: 9-cis-retinoic acid biosynthesis GO:0042904 IsA fat-soluble vitamin biosynthesis (GO:0042362)
New IsA link: acetyl-CoA catabolism GO:0046356 IsA coenzyme catabolism (GO:0009109)
New IsA link: 1-aminocyclopropane-1-carboxylate biosynthesis (GO:0042218) IsA amine biosynthesis (GO:0009309)
New IsA link: acyl-CoA biosynthesis GO:0046949 IsA carboxylic acid biosynthesis (GO 0046394)
New IsA link: acylglycerol biosynthesis GO:0046463 IsA glycerol ether biosynthesis (GO: 0046504)
New IsA link: adamantanone catabolism GO:0019263 IsA ketone catabolism (GO:0042182)
New IsA link: aerobic glycerol fermentation GO:0019565 IsA aerobic glycerol catabolism (GO:0019564)

New IsA link: aldonic acid biosynthesis GO:0046175 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: aldonic acid catabolism GO:0046176 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: alginic acid biosynthesis GO:0042121 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: alginic acid catabolism GO:0042122 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: allantoin biosynthesis GO:0019428 IsA amine biosynthesis (GO:0009309)
New IsA link: allantoin catabolism GO:0000256 IsA amine catabolism (GO:0009310)
New IsA link: amino acid biosynthesis GO:0008652 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: anaerobic benzoate catabolism GO:0010128 IsA xenobiotic catabolism GO:0042178
New IsA link: anaerobic benzoate catabolism GO:0010128 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: anaerobic benzoate catabolism GO:0010128 IsA aromatic compound catabolism (GO:0019439)
New IsA link: anaerobic ethylbenzene catabolism GO:0010130 IsA aromatic compound catabolism (GO:0019439)
New IsA link: anaerobic ethylbenzene catabolism GO:0010130 IsA xenobiotic catabolism GO:0042178
New IsA link: anthocyanin catabolism GO:0046284 IsA glycoside catabolism (GO:0016139)
New IsA link: anthocyanin metabolism GO:0046283 IsA glycoside metabolism (GO:0016137)
New IsA link: arsonoacetate catabolism GO:0019501 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: atrazine catabolism GO:0019381 IsA chlorinated hydrocarbon catabolism (GO:0042205)
New IsA link: bacteriochlorophyll metabolism GO:0030493 IsA protein metabolism (GO:0019538)
New IsA link: benzyl isoquinoline alkaloid biosynthesis GO:0009708 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: beta-1,6 glucan catabolism GO:0006079 IsA beta-1,6 glucan metabolism (GO:0006077)
New IsA link: beta-lactam antibiotic biosynthesis GO:0030654 IsA amine biosynthesis (GO:0009309)
New IsA link: beta-lactam antibiotic catabolism GO:0030655 IsA heterocycle catabolism (GO:0046700)

New IsA link: beta-lactam antibiotic catabolism GO:0030655 IsA amine catabolism (GO:0009310)
New IsA link: betaine biosynthesis GO:0006578 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: betaine catabolism GO:0006579 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: biotin biosynthesis GO:0009102 IsA coenzyme biosynthesis (GO:0009108)
New IsA link: biotin biosynthesis GO:0009102 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: biotin biosynthesis GO:0009102 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: biotin catabolism GO:0042367 IsA coenzyme catabolism (GO:0009109)
New IsA link: biotin catabolism GO:0042367 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: biotin catabolism GO:0042367 IsA heterocycle catabolism (GO:0046700)
New IsA link: caprolactam catabolism GO:0019384 IsA heterocycle catabolism (GO:0046700)
New IsA link: carbazole catabolism GO:0046232 IsA heterocycle catabolism (GO:0046700)
New IsA link: carnitine biosynthesis GO:0045329 IsA water-soluble vitamin biosynthesis (GO:0042364)
New IsA link: carnitine biosynthesis GO:0045329 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: carnitine catabolism GO:0042413 IsA water-soluble vitamin catabolism (GO:0042365)
New IsA link: carnitine catabolism GO:0042413 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: catechol biosynthesis GO:0009713 IsA phenol biosynthesis (GO:0046189)
New IsA link: catechol catabolism GO:0019614 IsA phenol catabolism (GO:0019336)
New IsA link: catecholamine biosynthesis GO:0042423 IsA phenol biosynthesis (GO:0046189)
New IsA link: catecholamine catabolism GO:0042424 IsA phenol catabolism (GO:0019336)
New IsA link: choline biosynthesis GO:0042425 IsA ethanolamine biosynthesis (GO:0046335)
New IsA link: cinnamic acid biosynthesis GO:0009800 IsA carboxylic acid biosynthesis (GO:0046394)

New IsA link: cinnamic acid catabolism GO:0046281 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: citrulline biosynthesis GO:0019240 IsA nonprotein amino acid biosynthesis (GO:0019795)
New IsA link: citrulline catabolism GO:0019241 IsA nonprotein amino acid catabolism (GO:0019796)
New IsA link: CMP-N-acetylneuraminate biosynthesis GO:0006055 IsA nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: corrin biosynthesis GO:0046140 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: corrin catabolism GO:0046141 IsA heterocycle catabolism (GO:0046700)
New IsA link: coumarin biosynthesis GO:0009805 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: coumarin catabolism GO:0046226 IsA heterocycle catabolism (GO:0046700)
New IsA link: creatinine catabolism GO:0006602 IsA heterocycle catabolism (GO:0046700)
New IsA link: creatinine catabolism GO:0006602 IsA ketone catabolism (GO:0042182)
New IsA link: creatinine catabolism GO:0006602 IsA amine catabolism (GO:0009310)
New IsA link: cresol catabolism GO:0046199 IsA toluene catabolism (GO:0042203)
New IsA link: cresol catabolism GO:0046199 IsA phenol catabolism (GO:0019336)
New IsA link: CTP deamination GO:0006243 IsA CTP catabolism (GO:0006254)
New IsA link: cyanate biosynthesis GO:0046201 IsA organic acid biosynthesis (GO:0016053)
New IsA link: cyanate catabolism GO:0009440 IsA organic acid catabolism (GO:0016054)
New IsA link: cytokinin biosynthesis GO:0009691 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: cytokinin biosynthesis GO:0009691 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: cytokinin biosynthesis GO:0009691 IsA amine biosynthesis (GO 0009309)
New IsA link: cytokinin catabolism GO:0009823 IsA aromatic compound catabolism (GO:0019439)
New IsA link: cytokinin catabolism GO:0009823 IsA heterocycle catabolism (GO:0046700)

New IsA link: cytokinin catabolism GO:0009823 IsA amine catabolism (GO:0009310)
New IsA link: D-xylose biosynthesis GO:0042842 IsA alcohol biosynthesis (GO:0046165)
New IsA link: D-xylose catabolism GO:0042843 IsA pentose catabolism (GO:0019323)
New IsA link: deoxyguanosine salvage GO:0006180 IsA deoxyguanosine metabolism (GO:0042453)
New IsA link: deoxyinosine salvage GO:0006191 IsA deoxyinosine metabolism (GO:0046094)
New IsA link: dethiobiotin biosynthesis GO:0019351 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: dethiobiotin biosynthesis GO:0019351 IsA amine biosynthesis (GO:0009309)
New IsA link: dethiobiotin biosynthesis GO:0019351 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: diacetyl fermentation GO:0019651 IsA ketone metabolism (GO:0042180)
New IsA link: diaminopimelate biosynthesis GO:0019877 IsA amine biosynthesis (GO:0009309)
New IsA link: dibenzo-p-dioxin catabolism GO:0019341 IsA aromatic compound catabolism (GO:0019439)
New IsA link: dibenzo-p-dioxin catabolism GO:0019341 IsA heterocycle catabolism (GO:0046700)
New IsA link: dibenzofuran catabolism GO:0019340 IsA heterocycle catabolism (GO:0046700)
New IsA link: dibenzothiophene catabolism GO:0018896 IsA aromatic compound catabolism (GO:0019439)
New IsA link: dibenzothiophene catabolism GO:0018896 IsA heterocycle catabolism (GO:0046700)
New IsA link: dimethylsilanediol catabolism GO:0042211 IsA alcohol catabolism (GO:0046164)
New IsA link: dopamine biosynthesis from tyrosine GO:0006585 IsA aromatic amino acid family biosynthesis (GO:0009073)
New IsA link: dTDP-mannose biosynthesis GO:0019308 IsA nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: dTDP-rhamnose biosynthesis GO:0019305 IsA nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: ecdysone modification GO:0006698 IsA lipid modification (GO:0030258)
New IsAlink: ectoine biosynthesis GO:0019491 IsA carboxylic acid biosynthesis (GO:0046394)

New IsA link: ectoine biosynthesis GO:0019491 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: 1,4-dichlorobenzene catabolism (GO:0019261) IsA aromatic compound catabolism (GO:0019439)
New IsA link: 1,3-dichloro-2-propanol biosynthesis (GO:0046309) IsA alcohol biosynthesis (GO:0046165)
New IsA link: ethanolamine catabolism GO:0046336 IsA ethanol catabolism (GO:0006068)
New IsA link: ethanolamine metabolism GO:0006580 IsA ethanol metabolism (GO:0006067)
New IsA link: (R)-4-hydroxymandelate catabolism (GO:0019599) IsA phenol catabolism (GO:0019336)
New IsA link: folic acid and derivative biosynthesis GO:0009396 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: folic acid and derivative biosynthesis GO:0009396 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: folic acid and derivative biosynthesis GO:0009396 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: folic acid and derivative catabolism GO:0009397 IsA aromatic compound catabolism (GO:0019439)
New IsA link: folic acid and derivative catabolism GO:0009397 IsA heterocycle catabolism (GO:0046700)
New IsA link: folic acid and derivative catabolism GO:0009397 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: formaldehyde biosynthesis GO:0046293 IsA one-carbon compound biosynthesis (GO:0019753)
New IsA link: formaldehyde catabolism GO:0046294 IsA one-carbon compound catabolism (GO:0019754)
New IsA link: formate biosynthesis GO:0015943 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: formate catabolism GO:0042183 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: galactosylceramide biosynthesis GO:0006682 IsA glycosphingolipid biosynthesis (GO:0006688)
New IsA link: galactosylceramide catabolism GO:0006683 IsA glycosphingolipid catabolism (GO:0046479)
New IsA link: galactosylceramide metabolism GO:0006681 IsA glycosphingolipid metabolism (GO:0006687)
New IsA link: trisporic acid biosynthesis (GO:0046842) IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: 1-aminocyclopropane-1-carboxylate catabolism (GO:0042217) IsA carboxylic acid catabolism (GO:0046395)

New IsA link: GDP-D-rhamnose biosynthesis GO:0019306 IsA fructose biosynthesis (GO:0046370)
New IsA link: GDP-D-rhamnose biosynthesis GO:0019306 IsA nucleotide-sugar biosynthesis (GO:0009226)
New IsA link: glutamate decarboxylation to succinate GO:0006540 IsA glutamate catabolism to succinate (GO:0019549)
New IsA link: glycosaminoglycan biosynthesis GO:0006024 IsA polysaccharide biosynthesis (GO:0000271)
New IsA link: glycosaminoglycan catabolism GO:0006027 IsA polysaccharide catabolism (GO:0000272)
New IsA link: glycosaminoglycan metabolism GO:0030203 IsA polysaccharide metabolism (GO:005976)
New IsA link: indole biosynthesis GO:0042432 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: indole biosynthesis GO:0042432 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: indole catabolism GO:0042433 IsA aromatic compound catabolism (GO 0019439)
New IsA link: indole catabolism GO:0042433 IsA heterocycle catabolism (GO 0046700)
New IsA link: indole derivative biosynthesis GO:0042435 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: indole derivative biosynthesis GO:0042435 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: indole derivative catabolism GO:0042436 IsA heterocycle catabolism (GO:0046700)
New IsA link: indoleacetic acid biosynthesis via tryptophan GO:0009848 IsA aromatic amino acid family biosynthesis (GO:0009073)
New IsA link: indoleacetic acid biosynthesis via tryptophan GO:0009848 IsA indolalkylamine biosynthesis (GO:0046219)
New IsA link: indoleacetic acid biosynthesis GO:0009684 IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: 1-aminocyclopropane-1-carboxylate catabolism (GO:0042217) IsA carboxylic acid catabolism (GO:0046395)
New IsA link: isocitrate metabolism GO:0006102 IsA citrate metabolism (GO:0006101)
New IsA link: ketone body biosynthesis GO:0046951 IsA ketone biosynthesis (GO:0042181)
New IsA link: ketone body catabolism GO:0046952 IsA ketone catabolism (GO:0042182)
New IsA link: ketone body metabolism GO:0046950 IsA ketone metabolism (GO:0042180)

New IsA link: L-arabinose catabolism to 2-oxoglutarate GO:0019570 IsA carboxylic acid catabolism (GO 0046395)
New IsA link: L-ascorbic acid biosynthesis GO:0019853 IsA carboxylic acid biosynthesis (GO 0046394)
New IsA link: L-ascorbic acid catabolism GO:0019854 IsA carboxylic acid catabolism (GO 0046395)
New IsA link: L-methylmalonyl-CoA biosynthesis GO:0019680 IsA coenzyme biosynthesis (GO:0009108)
New IsA link: lipoic acid biosynthesis GO:0009105 IsA carboxylic acid biosynthesis (GO 0046394)
New IsA link: lipoic acid biosynthesis GO:0009105 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: mandelate biosynthesis GO:0046236 IsA carboxylic acid biosynthesis (GO 0046394)
New IsA link: mandelate catabolism GO:0019596 IsA carboxylic acid catabolism (GO 0046395)
New IsA link: methionine biosynthesis from S-adenosylmethionine GO:0019284 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: methionine biosynthesis from S-adenosylmethionine GO:0019284 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: N-acetylneuraminate biosynthesis IsA carboxylic acid biosynthesis (GO:0046394)
New IsA link: N-acetylneuraminate catabolism GO:0046380 IsA carboxylic acid catabolism (GO:0046395)
New IsA link: N-terminal peptidyl-glutamine methylation GO:0018019 IsA peptidyl-glutamine methylation GO:0018364
New IsA link: nucleobase biosynthesis GO:0046112 IsA heterocycle biosynthesis (GO:0018130)
New IsA link: nucleobase biosynthesis GO:0046112 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: nucleobase catabolism GO:0046113 IsA aromatic compound catabolism (GO 0019439)
New IsA link: nucleobase catabolism GO:0046113 IsA heterocycle catabolism (GO 0046700)
New IsA link: O-glycan processing GO:0016266 IsA protein processing GO:0016485
New IsA link: octamethylcyclotetrasiloxane catabolism GO:0046517 IsA heterocycle catabolism (GO:0046700)
New IsA link: ommochrome biosynthesis GO:0006727 IsA aromatic compound biosynthesis (GO:0019438)
New IsA link: ommochrome biosynthesis GO:0006727 IsA heterocycle biosynthesis (GO:0018130)

New IsA link: orcinol biosynthesis GO:0046197 IsA phenol biosynthesis (GO:0046189)
New IsA link: orcinol catabolism GO:0042209 IsA toluene catabolism (GO:0042203)
New IsA link: orcinol catabolism GO:0042209 IsA phenol catabolism (GO 0019336)
New IsA link: p-cymene catabolism GO:0019334 IsA toluene catabolism (GO:0042203)
New IsA link: para-aminobenzoic acid biosynthesis GO:0008153 IsA amine biosynthesis (GO 0009309)
New IsA link: 1,4-dichlorobenzene catabolism (p1:GO_0019261) IsA aromatic compound catabolism (GO 0019439)
New IsA link: pentaerythritol tetranitrate metabolism GO:0018954 IsA glycol metabolism GO:0042844
New IsA link: peptide biosynthesis GO:0043043 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via (2S,3S,4Xi,6R)-3-methyl-lanthionine sulfoxide GO:0046804 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via (2S,3S,6R)-3-methyl-lanthionine GO:0018156 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 2'-(S-L-cysteinyl)-L-histidine GO:0018233 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 2-(S-L-cysteinyl)-D-allo-threonine GO:0046926 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 2-(S-L-cysteinyl)-D-phenylalanine GO:0046925 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 2-(S-L-cysteinyl)-L-phenylalanine GO:0046924 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 2-imino-glutaminyl-5-imidazolinone glycine GO:0019729 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 3'-(S-L-cysteinyl)-L-tyrosine GO:0018234 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 3-(S-L-cysteinyl)-L-aspartic acid GO:0019928 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 4'-(L-tryptophan)-L-tryptophyl quinone GO:0018069 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 4'-(S-L-cysteinyl)-L-tryptophyl quinone GO:0019927 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 4-(S-L-cysteinyl)-L-glutamic acid GO:0019929 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 4-(S-L-cysteinyl)-L-glutamic acid

GO:0019929 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 5'-(N6-L-lysine)-L-topaquinone
GO:0018124 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via 5-imidazolinone glycine GO:0018253
IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via chondroitin 4-sulfate
glycosaminoglycan GO:0019800 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via L-cysteinyl-L-selenocysteine
GO:0050837 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via L-cystine GO:0018316 IsA peptide
metabolism GO:0006518
New IsA link:peptide cross-linking via L-histidyl-L-tyrosine GO:0018151
IsA peptide metabolism GO:0006518
New IsA link:peptide cross-linking via L-lysinoalanine GO:0018274 IsA
peptide metabolism GO:0006518
New IsA link: peptide cross-linking via N6-(L-isoaspartyl)-L-lysine
GO:0018420 IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via
S-(2-aminovinyl)-3-methyl-D-cysteine GO:0018162
IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via
S-(2-aminovinyl)-3-methyl-D-cysteine GO:0018162
IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via
S-(2-aminovinyl)-3-methyl-D-cysteine GO:0018162
IsA peptide metabolism GO:0006518
New IsA link: peptide cross-linking via
S-[5'-(L-tryptoph-6'-yl)-L-tyrosin-3'-yl]-L-methionin-S-ium GO:0050739 IsA
peptide metabolism GO:0006518
New IsA link: peptide cross-linking via S-glycyl-L-cysteine GO:0018255 IsA
peptide metabolism GO:0006518
New IsA link: peptide cross-linking via the thioethers lanthionine or
3-methyl-lanthionine GO:0018081 IsA peptide metabolism GO:0006518
New IsA link: peptidyl-isoleucine racemization GO:0019124 IsA
peptidyl-isoleucine modification GO:0018203
New IsA link: peptidyl-leucine esterification GO:0018439 IsA
peptidyl-leucine modification GO:0018204