Vavilov Journal of Genetics and Breeding. 2015;19(6):652-660
DOI 10.18699/VJ15.090

Ontologies in bioinformatics and systems biology

N.L. Podkolodnyy[1, 2, 3], O.A. Podkolodnaya[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia

Computer simulation is now becoming a central scientific paradigm of systems biology and the basic tool for the theoretical study and understanding of the complex mechanisms of living systems. The increase in the number and complexity of these models leads to the need for their collaborative development, reuse of models, and their verification, and the description of the computational experiment and its results. Ontological modeling is used to develop formats for knowledge-oriented mathematical modeling of biological systems. In this sense, ontology associated with the entire set of formats, supporting research in systems biology, in particular, computer modeling of biological systems and processes can be regarded as a first approximation to the ontology of systems biology. This review summarizes the features of the subject area (bioinformatics, systems biology, and biomedicine), the main motivation for the development of ontologies and the most important examples of ontological modeling and semantic analysis at different levels of the hierarchy of knowledge: the molecular genetic level, cellular level, tissue levels of organs and the body. Bioinformatics and systems biology is an excellent ground for testing technologies and efficient use of ontological modeling. Several dozens of verified basic reference ontologies now represent a source of knowledge for the integration and development of more complex domain models aimed at addressing specific issues in biomedicine and biotechnology. Further formalization and ontological accumulation of knowledge and the use of formal methods of analysis can take the entire cycle of research in systems biology to a new technological level.

Key words: ontological modeling; bioinformatics; systems biology.

Онтологии в биоинформатике и системной биологии

Н.Л. Подколодный[1, 2, 3], О.А. Подколодная[1]

[1] Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия
[2] Федеральное государственное бюджетное учреждение науки Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Компьютерное моделирование в настоящее время становится центральной научной парадигмой системной биологии и основным инструментом для теоретического исследования и понимания механизмов функционирования сложных живых систем. Увеличение количества и сложности этих моделей приводит к необходимости их коллективной разработки, повторного использования, верификации, описания вычислительного эксперимента и его результатов. При разработке форматов представления знаний для математического моделирования биологических систем активно применяют онтологическое моделирование предметной области. В этом смысле онтологию, связанную со всей совокупностью форматов, обеспечивающих поддержку исследований в системной биологии, в частности компьютерное моделирование биологических систем и процессов, можно считать первым приближением к онтологии системной биологии. В обзоре кратко представлены особенности предметной области (биоинформатика, системная биология, биомедицина), основные мотивации в развитии онтологий и наиболее важные примеры онтологического моделирования и семантического анализа на разных уровнях иерархии знаний: молекулярно-генетическом, клеточном, тканевом, органов и организма. Биоинформатика и системная биология являются прекрасным полигоном для отработки технологий и эффективного использования онтологического моделирования. Создание нескольких десятков базовых ссылочных онтологий и их верификация позволяют использовать эти онтологии в качестве источников знаний для интеграции и построения более сложных моделей предметной области, ориентированных на решение конкретных задач биомедицины и биотехнологии. Дальнейшая формализация и накопление онтологических знаний, а также использование формальных методов их анализа могут поднять весь цикл научных исследований в области системной биологии на новый технологический уровень.

Ключевые слова: онтологическое моделирование; биоинформатика; системная биология.

HOW TO CITE THIS ARTICLE?
Podkolodnyy N.L., Podkolodnaya O.A. Ontologies in bioinformatics and systems biology. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2015;19(6):652-660. DOI 10.18699/VJ15.090

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ?
Подколодный Н.Л., Подколодная О.А. Онтологии в биоинформатике и системной биологии. Вавиловский журнал генетики и селекции. 2015;19(6):652-660. DOI 10.18699/VJ15.090

УДК 577.214, 004.822
Received 28.09.2015
Accepted for publication 29.10.2015
© АВТОРЫ, 2015

e-mail: pnl@bionet.nsc.ru

Появление качественно новых возможностей для проведения исследований, основанных на использовании высокопроизводительных экспериментальных технологий массового параллельного секвенирования ДНК, многолокусного генотипирования, многопараметрического профилирования экспрессии генов с использованием ДНК-чипов, ChiP-on-chip технологий, протеомных и метаболомных технологий и др., привело к накоплению беспрецедентно больших массивов экспериментальных данных и знаний.

Огромный объем молекулярно-биологической информации, ее сложность и наличие большого числа барьеров: технологических, информационных, ресурсных и т. д. затрудняют ее анализ, систематизацию и применение для решения конкретных задач биоинформатики, биотехнологии, фармакологии, персонализированной медицины и др. Чтобы освоить, систематизировать и эффективно использовать такого рода информацию, необходимы новые подходы к обработке больших данных (BIG DATA), в частности автоматические методы семантической интеграции гетерогенных данных, одним из основных этапов которой является согласование понятий предметной области, способов их описания и использования (сопоставление, обработка данных и т. д.). Такое согласованное описание конкретной предметной области называют онтологией.

Разработка онтологии является сложным и затратным процессом. Первый этап этого процесса – онтологический анализ и моделирование предметной области, включая создание словаря терминов, точных их определений и взаимосвязей между ними, правил и ограничений, согласно которым на базе введенной терминологии формируются достоверные утверждения, описывающие состояние изучаемого объекта.

Для чего же нужны онтологии? Онтологии позволяют представить понятия в таком виде, что они становятся пригодными для машинной обработки и вследствие этого используются в качестве посредника между пользователем и информационной системой или между членами научного сообщества при обмене данными. Молекулярному биологу важно иметь возможность описывать молекулярные события, взаимодействующие компоненты, роли, которые играют эти компоненты в молекулярных событиях и процессах, оценивать гипотезы. Биоинформатик заинтересован в интеграции данных, компьютерной аннотации, моделировании процессов и систем. Общезначимыми потребностями является использование онтологий в образовании.

Системная биология как научная дисциплина как раз и возникла с появлением возможностей построения портретных моделей биологических систем и процессов на основе интеграции и совместного компьютерного анализа большого объема такого рода принципиально новых экспериментальных данных, описывающих поведение молекулярно-генетических систем в целом. Предметом исследования в системной биологии являются биологические объекты и формируемые ими сложные, иерархически организованные сети взаимодействий, контролируемые информацией, закодированной в геномах (Kitano, 2002).

В связи с этим в настоящее время онтологический анализ становится одним из основных инструментов биоинформатики и системной биологии, используемым для семантической интеграции экспериментальных данных и знаний с целью построения «единой картины мира» (Подколодный, 2011).

Формальное представление онтологий

В информатике термин «онтология» означает концептуальную модель представления объектов, свойств объектов и отношений между ними (Chandrasekaran et al., 1999). Онтология включает набор понятий (терминов) предметной области, их определений и атрибутов, а также связанное с ними множество аксиом и правил вывода (Gruber, 1995).

Таким образом, формальная модель онтологии – это упорядоченная тройка конечных множеств $O = <T, R, F>$, где T – конечное и непустое множество классов и концептов (понятий, терминов) предметной области как части реального мира, рассматриваемой в пределах заданного контекста (в нашем случае – биоинформатика и системная биология), которую описывает онтология O; R – конечное множество отношений между концептами заданной предметной области; F – конечное множество функций интерпретации, заданных на понятиях и/или отношениях онтологии O или аксиом, используемых для моделирования утверждений, которые всегда являются истинными, что ограничивает интерпретацию и обеспечивает корректное использование понятий.

Одним из наиболее продуктивных подходов к представлению и использованию знаний о предметной области являются дескриптивные логики (ДЛ), определяющие формальный язык для описания понятий (концéпт, класс, категория или сущность) и отношений между ними (называемых ролями), утверждений о фактах и запросах к ним. Кроме этого, в ДЛ входят конструкторы (операции) для понятийных выражений, включающие конъюнкцию, дизъюнкцию и определение отношений.

Базы знаний предметной области с позиции дескриптивной логики подразделяются на общие знания о множестве классов понятий предметной области, свойствах и отношениях между ними (terminological knowledge, или T-Box) и знания об индивидуальных объектах (экземплярах класса), их свойствах и связях с другими объектами (assertional knowledge, или A-Box), т. е. описывают предметную область на уровне конкретных данных (база данных). В базе знаний обе компоненты взаимосвязаны.

В общем случае создание прикладных онтологий, ориентированных на конкретную предметную область, может быть существенно ускорено, если использовать ранее разработанные канонические (ссылочные) онтологии для построения онтологических классов и отношений между ними.

В частности, в качестве такого рода ссылочной онтологии может быть онтология верхнего уровня или онтология базовых знаний, в которой описываются наиболее общие концепты (пространство, время, материя, объект, система, состояние, поведение, событие, процесс, действие, структура, функция и т. п.) и отношения («часть – целое», «общее – частное», «является подклассом», «оказывает воздействие», «является причиной», «приводит к», «регулирует», «связан с», «похоже на», а также простран-

Types of *part_of* relations proposed by (Winston et al, 1987) and their specifying properties

Types of *part_of* relations	Properties		
	«Functional»	«Homeomerous»	«Separable»
Component/Integral-Object	+	–	+
Member/Collection	–	–	+
Portion/Mass	–	+	+
Stuff/Object	–	–	–
Feature/Activity	+	–	–
Place/Area	–	+	–

ственные и темпоральные отношения и т. п.). Эти концепты не зависят от конкретной проблемы или области, поэтому представляется разумным унифицировать их для больших сообществ пользователей.

Проект Open Biological Ontologies

Целью проекта Open Biological Ontologies (OBO) является разработка унифицированных подходов для создания онтологий, методов их интеграции, а также инструментальных средств работы с ними (Bada, Hunter, 2007; Smith et al., 2007). В OBO содержится информация об онтологиях и проектах, которые выполняются в области биологии (http://obofoundry.github.io/).

В настоящее время в OBO описано более 70 онтологий по различным направлениям, включая анатомию, биохимию, биологические процессы, функции и последовательности, заболевания, окружающую среду, экспериментальные доказательства, фенотип, белки, таксономии и др. (Schober et al., 2009).

Для обеспечения совместимости разрабатываемых в биомедицине онтологий в рамках проекта OBO предложены рекомендации по стандартизации, используемым онтологическим отношениям. Задаются формальные свойства отношений, которые можно использовать при логическом выводе новых утверждений. В частности, принимается, что отношения *part_of* и *is_a* транзитивны, рефлексивны, антисимметричны.

Однако на самом деле в зависимости от дальнейшего уточнения семантики отношений и специфики их применения, свойства этих отношений могут не выполняться. Даже такие распространенные отношения, как *part_of* и *is_a*, имеют на практике различные интерпретации. В этом случае свойства транзитивности могут нарушаться. С этим столкнулись разработчики проекта Gene Ontology (GO), когда приступили к формализованной проверке ее онтологии (Smith et al., 2003).

Ниже представлены некоторые проблемы, которые возникли в GO при интерпретации отношения *part_of*:

P1. *A part_of B* означает: *A* иногда является частью *B*, т. е. для каждого *A* в некоторое время t *A* является частью *B*.

Пример: «replication fork» *part_of* «nucleoplasm» («вилка репликации» наблюдается в определенной фазе клеточного цикла).

P2. *A part_of B* означает: *A* может быть частью *B*. Класс *A* является частью класса *B* тогда и только тогда, если существует подкласс $C \subset B$, в котором все экземпляры *A* включены как часть в экземпляры *C* и все экземпляры класса *C* имеют как часть экземпляры класса *A*.

Пример: «flagellum» *part_of* «cell» (некоторые типы клеток включают как часть жгутики).

P3. *A part_of B* означает: *A* всегда является частью *B*.

Пример: «membrane» *part_of* «cell» (мембрана является частью любой клетки).

Для разрешения этих проблем в настоящее время в GO вводятся более специфичные отношения типа: *is_localized_in* или *is_involved_in*.

Аналогично в GO возникли проблемы интерпретации отношения *is_a*. Хорошо структурированная классификация может быть получена путем замены отношения *is_a* на специальные типы, например: *has_role*, *is_dependent_on*, *is_involved_in*, *contributes_to*, *is_located_in*, а также добавления различных категории сущностей: *sites*, *constituents*, *roles*, *functions*, *qualities*.

В общем случае для разрешения этих проблем требуется уточнение семантики этих отношений, используемой при разработке конкретной онтологии.

Существуют различные попытки уточнения семантики отношений *part_of* и *is_a* и их классификации для разрешения типовых конфликтов и нарушений свойств этих отношений. В статье (Winston, et al., 1987) введено 6 различных типов отношения *part_of* (таблица) на основе следующих свойств или критериев:

- «Functional» выполняется, когда части в специфической пространственной или временной локализации выполняют такую же функциональную роль, как и целое.
- «Homeomerous» выполняется, когда части подобны каждая друг другу и целому, которому принадлежит.
- «Separable» выполняется, когда части могут быть физически не связаны и хотя бы в принципе отделены от целого, которое они составляют.

Проблемы с транзитивностью отношения *part_of* возникают, когда комбинируются различные типы отношений *part_of*. В общем случае транзитивность должна предполагаться, по крайней мере, тогда, когда используются отношения *part_of* того же типа, т. е. имеют те же свойства.

В зависимости от особенностей предметной области можно использовать другие наборы свойств, определяющих классы отношения *part_of*, например: configurational, encapsulated, exchangeable, functional, homeomerous, homogeneous, mandatory, canonically necessary, removable, segmental, separable, shareable.

В биоинформатике и системной биологии традиционно широко используются представления онтологий на языке OBO. В последнее время многие онтологии транслируются в представление на языке OWL (Ontology Web Language) (Stevens et al., 2007). Главной проблемой при таких преобразованиях является наличие ошибок, противоречий и нарушений интерпретаций отношений. Использование формальных методов поиска противоречий, неполноты позволяет существенно улучшить качество описаний.

Развитие онтологий в биоинформатике

В настоящее время в области биологии разработано несколько сотен онтологий, которые можно использовать для описания и интеграции знаний, а также вывода новых знаний.

В частности, разработаны и активно используются биоинформационные ресурсы и онтологии, позволяющие описывать молекулярные структуры, функции, процессы и генные сети (GO).

Онтологии MIAPE (Minimum Information About a Proteomics Experiment) (Taylor et al., 2007) и MIMIx (Minimum Information required for reporting a Molecular Interaction eXperiment) (Orchard et al., 2007) предложены рабочей группой Human Proteome Organization (HUPO) для описания протеомных исследований и экспериментов по молекулярным взаимодействиям соответственно.

Небольшая онтология верхнего уровня BFO (Basic Formal Ontology) предназначена для разработки онтологий, ориентированных на поиск и интеграцию научных данных. BFO уже использовалась для разработки более 130 онтологий в разных предметных областях (http://ifomis.uni-saarland.de/bfo/).

База знаний ChEBI (Chemical Entities of Biological Interest) включает онтологию молекулярных объектов – природных соединений или синтетических продуктов, воздействующих на процессы в живых организмах, включая любые уникальные по структуре или изотопному составу атомы, молекулы, ионы, ионные пары, радикалы, ион-радикалы, комплексы, конформеры и т. п. (https://www.ebi.ac.uk/chebi/). ChEBI в настоящее время (release 131) включает 46477 полностью аннотированных молекулярных объектов.

Онтология клеточных типов CL (Cell Type Ontology), по сути, является структурированным контролируемым словарем, включающим описание клеточных типов различных видов организмов – от прокариот до млекопитающих (http://www.obofoundry.org/cgi-bin/detail.cgi?id=cell).

Онтологии, разработанные в рамках KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al., 2004), ориентированы на широкую область молекулярной биологии – от генов и белков до метаболических и генных сетей. База знаний TAMBIS (Transparent Access to Multiple Bioinformatics Information Source, http://www.cs.man.ac.uk/~stevensr/tambis/) предоставляет пользователям-биологам единую точку доступа к мировым источникам биологической информации, которая интегрируется в рамках этой системы с помощью онтологического описания (Stevens et al., 2000). EcoCyc – база научных данных, в которой накапливается информация, полученная в результате аннотирования научных публикаций по геному *E. coli*, регуляции транскрипции ее генов, транспортным и метаболическим путям (Karp et al., 2014).

Sequence Ontology (SO) включает множество понятий и контролируемых словарей, используемых для описания свойств и первичной аннотации нуклеотидных или белковых последовательностей, структурного представления этих аннотаций в геномных базах данных, мутаций в обоих типах последовательностей и на более высоком уровне (Cunningham et al, 2015).

В рамках Mouse Atlas Project развиваются цифровой атлас и база данных об экспрессии генов мыши и клеточных линиях с описанием анатомической локализации клеток (Graham et al., 2015).

Онтология MGED (Microarray and Gene Expression Data) используется для описания экспериментов и данных по экспрессии генов (Whetzel et al., 2006).

MIAME (Minimum Information About a Micro-array Experiment) используется для описания экспрессионных данных (Brazma et al., 2001).

В базе знаний PharmGKB (https://www.pharmgkb.org/) представлена информация по фармакогенетике (Klein et al., 2001).

Целью проекта Cell Cycle Ontology (Antezana et al., 2009) является расширение существующих онтологий, связанных с клеточным циклом, для интеграции и управления знаниями о его компонентах и регуляторных аспектах. В качестве источников этих знаний используются уже существующие ресурсы (GO, UniProt, IntAct, BIND, NCBI taxonomy и др.). Интеграция и комбинация этих знаний позволяют представлять наиболее полную картину процессов деления клетки.

Анатомические и морфологические онтологии для модельных организмов являются важнейшими примерами онтологического моделирования, в котором используется большое разнообразие пространственно-временных отношений.

В частности, основная модель анатомии FMA (Foundational Model of Anatomy ontology) является ссылочной онтологией, включающей понятия и отношения, которые используются для описания структурной организации тела человека на различных уровнях – от макромолекул, клеток, ткани, органа до организма с учетом онтогенеза, и предназначенной для символического компьютерного моделирования анатомических структур (Rosse et al., 2003; Rosse et al., 2007).

Онтологические знания в FMA представлены в виде фреймов и хранятся в реляционной базе данных. FMA (http://sig.biostr.washington.edu/projects/fm/) включает около 75 тыс. анатомических классов, более 130 тыс. уникальных терминов, более 205 тыс. фреймов и 174 уникальных слота, которые используются для представления различных типов отношений, атрибутов и атрибутных отношений. Сеть отношений FMA содержит более 2,5 млн экземпляров отношений, свыше 1 млн экземпляров классов, около 450 тыс. связей между классами.

Gene Ontology

В качестве примера одного из самых успешных проектов создания онтологии можно привести GO (http://www.geneontology.org/).

В состав GO входят 3 раздела:

- Молекулярная функция (Molecular function) – элементарная активность/задача или роль, которую может выполнять ген, продукт гена в каких-либо биологических процессах, например, «catalytic activity» или «Toll receptor binding».
- Биологические процессы (Biological process) описывают серию событий, реализующих одну или более организованных ансамблей молекулярных функций. В отличие от функции процесс должен иметь несколько различающихся этапов. Например: «pyrimidine metabolic process».
- Клеточные компоненты (Cellular component) как часть анатомической структуры, в которой описывается локализация гена или его продукта в организме на уровнях клеточных структур и макромолекулярных комплексов (например, «nucleus», «membrane») или групп продуктов генов (например, «ribosome», «proteasome» или «protein dimer»).

По сути, GO позволяет описывать знания о том, какую функцию выполняет ген или его продукт (РНК, белок) в том или ином биологическом процессе и той или иной клеточной структуре.

GO содержит более 40 тыс. понятий (The Gene Ontology Consortium, 2015), включая:

- Biological process – около 30 тыс.;
- Molecular function – более 10 тыс.;
- Cellular component – 3 758.

На основе GO разработан ресурс GOA (Gene Ontology Annotation, http://www.ebi.ac.uk/GOA), который используется для аннотации белков из UniProtKB (UniProt Knowledgebase). В настоящее время GOA включает 368 млн аннотаций GO для почти 54 млн белков из 480 тыс. таксономических групп (Huntley et al., 2015).

Основные отношения между понятиями, которые используются в GO, это *is_a, part_of* и *regulates*.

- *is_a* – простое отношение «класс – подкласс», где A *is_a* B означает, что A является подклассом B.
- *part_of* – отношение «часть – целое». Выражение A *part_of* B означает, что если A существует, то A всегда является частью B.
- *Regulates*, *positively_regulates* и *negatively_regulates* описывают отношения между биологическими процессами, молекулярными функциями или биологическими свойствами.

В GO описываются иерархические отношения, однако граф отношений не является деревом. Одно понятие может иметь несколько предков. Свойство транзитивности отношений, используемых в GO, позволяет строить решетку отношений между понятиями и выполнять логический вывод о свойствах понятий и их отношениях (Srinivas, 2009).

Онтология системной биологии

Компьютерные модели становятся центральной научной парадигмой системной биологии и основным инструментом для теоретического исследования и понимания механизмов функционирования сложных живых систем. С увеличением числа и размеров этих моделей возрастает необходимость коллективной разработки, повторного использования моделей, их верификации, описания вычислительного эксперимента и его результатов.

При разработке форматов представления знаний для математического моделирования биологических систем активно применяют онтологическое моделирование предметной области. В этом смысле онтологию, связанную со всей совокупностью форматов, обеспечивающих поддержку исследований в системной биологии, в частности компьютерное моделирование биологических систем и процессов, можно считать первым приближением к онтологии системной биологии.

В настоящем разделе мы рассмотрим онтологии системной биологии, которые ориентированы на описание генетических систем и их моделей.

Многие онтологии не разрабатываются с нуля, а интегрируют (или компилируют) уже разработанные ранее частные онтологии, которые используются в качестве базовых источников знаний. К ним относятся формализация структуры математических моделей (SBML), стандартизация описания компонент модели с кинетической и биологической точки зрения (SBO, GO, UniProt). Однако формализация только структуры компьютерной модели недостаточна для поддержки всей технологической цепочки вычислительного эксперимента и компьютерного моделирования в системной биологии. Необходимо формализовать прагматические и динамические аспекты процесса моделирования.

Примером специализированного проекта по созданию онтологии для системной биологии является Systems Biology Ontology (SBO) (http://www.ebi.ac.uk/sbo/). Цель проекта – разработка контролируемых словарей и онтологий, ориентированных на решение задач системной биологии, особенно в контексте компьютерного моделирования.

SBO состоит из 6 ортогональных контролируемых словарей, включающих описание ролей участников реакций (например, «reactant», «product», «modifier»), значения количественных параметров моделей реакций (например, «kinetic constant»), точную классификацию математических выражений, которые описывают систему (например, «mass action rate law»), тип используемой среды моделирования (например, «logical framework», «discrete framework»), типы сущностей системы (например, «macromolecule», «enzyme», «ligand») и взаимодействий в ней (например, «process», «biochemical reaction», «genetic interaction», «relationship»). Количественные параметры и математические выражения описываются на языке MathML 3.0 (http://www.w3.org/TR/MathML3/).

Математическая модель может быть проаннотирована в SBO на любой стадии жизненного цикла (от времени создания до расширения и модификации модели) путем последовательного расширения ее семантики.

Можно привести примеры других форматов представления знаний и связанных с ними онтологий системной биологии:

- Systems Biology Markup language (SBML, http://sbml.org) – формат представления структуры биологических моделей (Hucka et al., 2003).
- Biological Pathway Exchange Language (BioPAX, http://www.biopax.org) – формат для описания и интеграции

информации о молекулярных взаимодействиях и биологических процессах (Demir et al., 2010). Этот подход используется для представления знаний в существующих базах данных (BioCyc, BIND, WIT, aMAZE, KEGG, Reactome и т. д.), важных для описания механизмов экспрессии генов.

- Minimal Information Requested In the Annotation of biochemical Models (MIRIAM) (Le Novère et al., 2005) – формат для стандартизации минимального множества информации, необходимой для аннотации модели и обеспечения возможности коллективной аннотации, курирования и развития, а также повторного использования моделей.
- Simulation Experiment Description Markup Language (SED-ML, http://sed-ml.org/) – формат для описания экспериментов по моделированию и обмену результатами моделирования независимо от использованного языка спецификации модели и среды моделирования (Waltemath et al., 2011).
- TErminology for the Description of DYnamics (TEDDY, http://www.ebi.ac.uk/compneursrv/teddy) (Chelliah, Endler, 2009) – онтология для описания динамического поведения биологической системы или динамического явления, управления элементами биологической модели и системы в системной и синтетической биологии. В частности, используя TEDDY, можно количественно описывать характеристики биологической системы или модели: тип осцилляции (хаотическая, периодическая, квазипериодическая и т. д.), области осцилляции, период, точки покоя или нестабильности, бифуркации, зависящие от параметров, функциональные мотивы (например, отрицательные обратные связи) и т. д.
- Kinetic Simulation Algorithm Ontology (KiSAO, http://biomodels.net/kisao/) – онтология для описания алгоритмов моделирования кинетических процессов (Chelliah, Endler, 2009). Классификация алгоритмов моделирования биологических моделей в KiSAO построена с использованием различных категорий и с учетом версий алгоритмов. Например, детерминированные или стохастические правила, пространственные или непространственные подходы, дискретные или непрерывные переменные, фиксированные или адаптивные временные шаги и т. д.
- Systems Biology Graphical Notation (SBGN, http://www.sbgn.org) – формат для графического представления биологических систем и процессов (Dada et al., 2010).
- SBRML (www.comp-sys-bio.org/tiki-index.php?page=SBRML) – формат для сохранения результатов моделирования (Orchard et al., 2007).
- CellML (www.cellml.org) – формат для описания математических моделей биологических систем и процессов (Lloyd et al., 2004). Для представления математических выражений в CellML используется язык MathML.

Ссылочные онтологии, GO, SO, Chemical Entities of Biological Interest (ChEBI), FMA, FMP, CPRO, PaTO, Pro, RnaO, CARO, описывающие биологические системы на различных уровнях, могут быть интегрированы в единую онтологию для описания объекта исследования системной биологии (организм, орган, ткани, клетки).

Применение онтологий в системной биологии и биоинформатике

В целом в биоинформатике и системной биологии можно выделить следующие задачи, в решении которых применение онтологий дает ощутимый эффект (Bodenreider, Stevens, 2006; Bodenreider, 2008; Beck et al., 2009; Noy et al., 2009; Подколодный, 2011).

1. Интерпретация молекулярно-генетических знаний, семантическая интерпретация методов анализа данных и моделей в системной биологии. В частности, анализ обогащения генов терминами из GO (GO Enrichment Analysis) используется для интерпретации данных (например, функциональное описание множества генов), контроля качества, систематизации и отбора данных.
2. Приоритизация генов, белков, биомаркеров и т. д.
3. Анализ сходства и кластеризация объектов. В качестве примера можно привести анализ уровня экспрессии десятков тысяч генов в различных клеточных ситуациях, при разных состояниях и на различных этапах развития клетки, ткани, органа или организма. После выделения группы генов со схожими паттернами экспрессии (коэкспрессирующиеся гены) возникает задача описания этих групп. Использование GO позволяет описать, в реализации каких функций участвуют гены, входящие в кластер (Khatri, Draghici, 2005). По сути, используя онтологии, можно количественно оценивать семантическое сходство объектов предметной области.
4. Поддержка интероперабельности и обмена знаниями:

унифицированный доступ к множествам гетерогенных источников данных;

поиск релевантной информации в документах. Онтология в этом случае задает структуру для аннотации содержания документа с семантической информацией, а также обеспечивает индексирование и связывание фактов, описанных в базах данных (Shah et al., 2009);

интеграция информации из различных источников и создание больших баз знаний;

комбинирование экспериментальных данных и знаний из онтологии для формирования баз знаний;

интероперабельность, поддержка коммуникации (между людьми и организациями) и обмена знаниями (между людьми и/или системами);

анализ текстов и семантический анализ (Chapman, Cohen, 2009);

приобретение знаний, извлечение знаний, неявных и явных отношений между сущностями в аннотированных источниках, аналитика.

5. Создание новых онтологий на основе повторного использования базовых канонических онтологий и различного типа операции с ними, включая сопоставление (ontology matching), слияние (ontology merging), отображение (ontology mapping), выравнивание (ontology alignment) и т. д.
6. Обеспечение непротиворечивости и корректности представления знаний. Поддержка процесса построения онтологий, включая любые типы автоматического вывода для поиска ошибок и выявления новых отношений. Количество понятий и отношений в современных онтологиях исчисляется сотнями тысяч, поэтому ручная проверка невозможна. Эксперт в этом случае

проверяет противоречия и результаты, полученные путем формального вывода на онтологиях (Livingston et al., 2015).

7. Поддержка индуктивного вывода для извлечения дополнительных знаний из множества фактов и тестирование гипотез. Например, в работе Подколодного с коллегами (2012) представлены подходы к онтологическому моделированию механизмов регуляции транскрипции генов и показаны примеры реконструкции гипотетических механизмов регуляции транскрипции с учетом информации о строении регуляторных районов генов и функциях регуляторных белков, присутствующих в заданных клетках или тканях на определенной стадии развития.
8. Повышение аргументации методов биоинформатики, включая точное описание биомедицинских экспериментальных протоколов, методов анализа данных и моделирования биологических процессов и систем и т. д. (Chen et al., 2007).

Можно привести примеры успешного применения дескриптивной логики и логического вывода для формальных онтологий в области биомедицины. В обзоре Кит с коллегами (Keet et al., 2007) приводится список типовых сценариев, полученных путем анализа и обобщения примеров использования средств логического вывода. При этом отмечено, что стандартные средства Racer, Pellet, FaCT++ не позволяют выполнять многие сценарии на реальных биомедицинских онтологиях из-за их большого объема и сложности. Поэтому весьма актуальным остается развитие новых эффективных программных средств логического вывода и подходов к формализации биологических знаний.

Программные средства для GO

В настоящее время среди всех онтологий в области биоинформатики и системной биологии наиболее широко используется GO. Для этого разработано большое число программных средств:

AMIGO (http://amigo.geneontology.org/amigo), QuickGO (http://www.ebi.ac.uk/QuickGO/), Protein2GO и Ontology Lookup Service (http://www.ebi.ac.uk/ontology-lookup/) – поиск и просмотр GO и аннотаций данных (Carbon et al., 2009).

OBO-Edit – просмотр и редактирование онтологических описаний (Day-Richter et al., 2007).

Blast2GO (https://www.blast2go.com/) – функциональная аннотация неизвестных последовательностей путем поиска гомологий с помощью BLAST и анализа GO аннотаций полученных результатов (Conesa, Gotz, 2008).

GoPubMed – поиск биологических текстов на основе GO и Medical Subject Headings (MeSH, http://www.nlm.nih.gov/mesh/introduction.html). GoPubMed связывает GO с базой рефератов PubMed и отвечает на вопросы: Что? Кто? Где? Когда? (http://www.gopubmed.org/web/gopubmed/).

Onto-Tools (http://vortex.cs.wayne.edu/) – набор сервисов, включающий: Onto-Express, Onto-Compare, Onto-Design, Onto-Translate, Onto-Miner, Pathway-Express, Promoter-Express, nsSNPCounter, TAQ и OE2GO, – профилирование множества генов, сравнение экспрессионных данных, компьютерное сопровождение процесса конструирования ДНК-чипов, включая выбор множества генов на основе их функции, процессов, в которых эти гены участвуют, или клеточных компонент, где эти гены экспрессируются, и т. д.

GOToolBox – анализ результатов ДНК-чиповых экспериментов (http://genome.crg.es/GOToolBox/) (Martin et al., 2004).

Наиболее часто GO используется для анализа обогащения описания генов терминами генной онтологии. Для решения этой задачи разработаны несколько десятков программных систем, среди которых можно выделить наиболее популярные (ftp://ftp.geneontology.org/pub/go/www/GO.tools.microarray.shtml):

- Gorilla (Gene Ontology enRIchment anaLysis and visualizAtion tool) (http://cbl-gorilla.cs.technion.ac.il/) – Web-приложение для идентификации обогащенных терминов GO в ранжированных списках генов.
- PANTHER (http://pantherdb.org/) – классификации белков (и их генов) в соответствии с информацией из GOA, эволюционными взаимосвязями, взаимодействиями, участием в метаболических и генных сетях и т. д. Как часть проекта GO эта система регулярно обновляется и использует наиболее актуальные данные (Thomas et al., 2003; Huaiyu et al., 2003; Mi et al., 2005).
- DAVID (http://david.abcc.ncifcrf.gov/) – Web-сервис для аннотации и анализа экспрессионных и протеомных данных, полученных с помощью высокопроизводительных экспериментальных технологий, а также обогащения описания генов. Основной недостаток DAVID – это использование устаревших данных GO (отставание – 3–4 года).
- Веб-сервис WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) используется в функциональной геномике, протеомике и широкомасштабных генетических исследованиях для функциональной аннотации больших групп генов, например, групп дифференциально экспрессирующихся и коэкспрессирующихся генов и т. д.

Бурное развитие экспериментальных технологий в области молекулярной биологии привело к тому, что онтологическое моделирование становится базовым методом в биоинформатике и системной биологии для интеграции и анализа гетерогенных экспериментальных данных и использования их для построения математических моделей молекулярно-генетических систем и процессов.

Биоинформатика и системная биология являются прекрасным полигоном для отработки технологий и эффективного использования онтологического моделирования. Создание нескольких десятков базовых ссылочных онтологий и их верификация позволяют использовать эти онтологии в качестве источников знаний для интеграции и построения более сложных моделей предметной области, ориентированных на решение конкретных задач биомедицины. В качестве примера можно привести базу знаний KaBOB (Livingston et al., 2015), в которой интегрированы знания из 14 гетерогенных источников – баз данных по генам и их гомологам, белкам, лекарствам, генетическим ассоциациям, регуляторным последовательностям, белок-белковым взаимодействиям и др. Интеграция данных и поиск информации в этой базе знаний основаны на онтологии, объединяющей:

1. Basic Formal Ontology (BFO) (http://purl.obolibrary.org/obo/bfo.owl);
2. BRENDA Tissue / Enzyme Source (BTO) (http://purl.obolibrary.org/obo/bto.owl);
3. ChEBI (http://purl.obolibrary.org/obo/chebi.owl);
4. Cell Type Ontology (CL) (http://purl.obolibrary.org/obo/cl.owl);
5. Gene Ontology (GO) (http://purl.obolibrary.org/obo/go.owl);
6. Information Artifact Ontology (IAO) (http://purl.obolibrary.org/obo/iao.owl);
7. Protein-Protein Interaction Ontology (MI) (http://purl.obolibrary.org/obo/mi.owl) ;
8. Mammalian Phenotype Ontology (MP) (http://purl.obolibrary.org/obo/mp.owl);
9. NCBI Taxonomy (http://purl.obolibrary.org/obo/ncbitaxon.owl);
10. Ontology for Biomedical Investigation (OBI) (http://purl.obolibrary.org/obo/ obi.owl);
11. Protein Modification (MOD) (http://purl.obolibrary.org/obo/mod.owl);
12. Protein Ontology (PR) (http://purl.obolibrary.org/obo/pr.owl);
13. Relation Ontology (RO) (http://purl.obolibrary.org/obo/ro.owl);
14. Sequence Ontology (SO) (http://purl.obolibrary.org/obo/so.owl).

Для представления знаний использовался язык OWL. Описание онтологии KaBOB составляет более 13 млн RDF-триплетов. Для поиска используется язык SPARQL 1.1. Реализованы две версии базы знаний KaBOB только для человека и для человека и ряда модельных организмов. В итоге система позволяет выполнять запросы на множестве гетерогенных данных, сопоставляя знания на различных уровнях описания биологических систем. В качестве примера можно привести следующий запрос: «Какие гены или продукты генов у человека, локализованные в митохондриях, вовлечены в процесс окислительного фосфорилирования и являются мишенями лекарств? Что это за лекарства?». Архитектура системы KaBOB и применяемые информационные технологии позволяют расширять используемые онтологии и базы данных и в перспективе решать сложные биоинформационные задачи.

Таким образом, дальнейшие формализация и накопление онтологических знаний, а также применение формальных методов их анализа могут поднять весь цикл научных исследований в области системной биологии на новый технологический уровень.

Acknowledgments

This work was supported by the Russian Science Foundation, project 14-24-00123.

Conflicts of interest

The authors declare no conflicts of interest.

References

Antezana E., Egaña M., Blondé W., Illarramendi A., Bilbao I., De Baets B., Stevens R., Mironov V., Kuiper M. The Cell Cycle Ontology: an application ontology for the representation and integrated analysis of the cell cycle process. Genome Biol. 2009;10(5):R58. DOI 10.1186/gb-2009-10-5-r58

Bada M., Hunter L. Enrichment of OBO ontologies. J. Biomed. Inform. 2007;40:300-315.

Beck T., Morgan H., Blake A., Wells S., Hancock J.M., Mallon A.-M. Practical application of ontologies to annotate and analyze large-scale raw mouse phenotype data. BMC Bioinformatics. 2009;10(Suppl5): S2. DOI 10.1186/1471-2105-10-S5-S2

Bodenreider O. Biomedical ontologies in action: Role in knowledge management, data integration and decision support. IMIA Yearbook Medical Informatics, 2008.

Bodenreider O., Stevens R. Bio-ontologies: Current trends and future directions. Brief Bioinform. 2006;7(3):256-274.

Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J., Vingron M. Minimum information about a microarray experiment (MIAME)–toward standards for microarray data. Nat. Genet. 2001;29:365-371.

Carbon S., Ireland A., Mungall C.J., Shu S., Marshall B., Lewis S. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25(2):288-289. DOI 10.1093/bioinformatics/btn615

Chandrasekaran B., Josephson J.R., Benjamins V.R. What are ontologies and why do we need them? IEEE Intelligent Systems. 1999;14(1): 20-26. DOI 10.1109/5254.747902

Chapman W.W., Cohen K.B. Current issues in biomedical text mining and natural language processing. J. Biomed. Inform. 2009;42:757-759. DOI 10.1016/j.jbi.2009.09.001

Chelliah V., Endler L., Juty N., Laibe C., Li C., Rodriguez N., Le Novère N. Data integration and semantic enrichment of systems biology models and simulations. Data integration in the life sciences. Lecture Notes in Computer Science. 2009;5647:5-15.

Chen Q., Chen Y.-P.P., Zhang C. Detecting inconsistency in biological molecular databases using ontologies. Data Min. Knowl. Disc. 2007;15:275-296. DOI 10.1007/s10618-007-0071-0

Conesa A., Gotz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics. 2008;2008:619832. DOI 10.1155/2008/619832

Cunningham F., Moore B., Ruiz-Schultz N., Ritchie G.R., Eilbeck K. Improving the Sequence Ontology terminology for genomic variant annotation. J. Biomed Semantics. 2015; 6:32. DOI 10.1186/s13326-015-0030-4

Dada J.O., Spasić I., Paton N.W., Mendes P. SBRML: a markup language for associating systems biology data with models. Bioinformatics. 2010;26(7):932-938. DOI 10.1093/bioinformatics/btq069

Day-Richter J., Harris M.A., Haendel M. Gene ontology OBO-edit working group, Lewis S. OBO-Edit – an ontology editor for biologists. Bioinformatics. 2007;23(16):2198-2200.

Demir E., Cary M.P., Paley S. et al. The BioPAX community standard for pathway data sharing. Nat. Biotechnol. 2010;28:935-942. DOI 10.1038/nbt.1666

Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucl. Acids Res. 2015;43(Database issue):D1049-D1056. DOI 10.1093/nar/gku1179

Graham E., Moss J., Burton N., Armit C., Richardson L., Baldock R.D. The atlas of mouse development eHistology resource. Development. 2015;142:1909-1911. DOI 10.1242/dev.124917

Gruber T.R. Toward principles for the design of ontologies used for knowledge sharing. Int. J. Human-Computer Studies.1995;43(5/6): 907-928.

Huaiyu M.I., Lazareva-Ulitsky B., Loo R., Kejariwal A., Vandergriff J., Rabkin S., Guo N., Muruganujan A., Doremieux O., Campbell M.J., Kitano H., Thomas P.D. The PANTHER database of protein families, subfamilies, functions and pathways. Nucl. Acids Res. 2005; 33(suppl.1):D284-D288. DOI 10.1093/nar/gki078

Hucka M., Finney A., Sauro H.M., Bolouri H., Doyle J.C., Kitano H., Arkin A.P., Bornstein B.J., Bray D., Cornish-Bowden A., Cuellar A.A.,

Dronov S., Gilles E.D., Ginkel M., Gor V., Goryanin I.I., Hedley W.J., Hodgman T.C., Hofmeyr J.H., Hunter P.J., Juty N.S., Kasberger J.L., Kremling A., Kummer U., Le Novère N., Loew L.M., Lucio D., Mendes P., Minch E., Mjolsness E.D., Nakayama Y., Nelson M.R., Nielsen P.F., Sakurada T., Schaff J.C., Shapiro B.E., Shimizu T.S., Spence H.D., Stelling J., Takahashi K., Tomita M., Wagner J., Wang J. SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19(4):524-31.

Huntley R.P., Sawford T., Mutowo-Meullenet P., Shypitsyna A., Bonilla C., Martin M.J., O'Donovan C. The GOA database: Gene Ontology annotation updates for 2015. Nucl. Acids Res. 2015;43(Database issue):D1057-D1063. DOI 10.1093/nar/gku1113

Kanehisa M., Goto S., Kawashima S., Okuno Y., Hattori M. The KEGG resource for deciphering the genome. Nucl. Acids Res. 2004;32 (Database issue):D277-80.

Karp P.D., Weaver D., Paley S., Fulcher C., Kubo A., Kothari A., Krummenacker M., Subhraveti P., Weerasinghe D., Gama-Castro S., Huerta A.M., Muñiz-Rascado L., Bonavides-Martinez C., Weiss V., Peralta-Gil M., Santos-Zavaleta A., Schröder I., Mackie A., Gunsalus R., Collado-Vides J., Keseler I.M., Paulsen I. The EcoCyc Database. Ecosal Plus. 2014;2014. DOI 10.1128/ecosalplus

Keet C.M., Roos M., Marshall M.S. A survey of requirements for automated reasoning services for bio-ontologies in owl. Workshop on OWL: Experiences and Directions. Innsbruck, Austria, 2007.

Khatri P., Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics. 2005; 21(18):3587-3595.

Kitano H. Systems biology: a brief overview. Science. 2002;295:1662-1664.

Klein T.E., Chang J.T., Cho M.K., Easton K.L., Fergerson R., Hewett M., Lin Z., Liu Y., Liu S., Oliver D.E., Rubin D.L., Shafa F., Stuart J.M., Altman R.B. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenomics J. 2001;1: 167-170.

Le Novère N., Finney A., Hucka M., Bhalla U.S., Campagne F., Collado-Vides J., Crampin E.J., Halstead M., Klipp E., Mendes P., Nielsen P., Sauro H., Shapiro B., Snoep J.L., Spence H.D., Wanner B.L. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 2005;23(12):1509-1515.

Livingston K.M., Bada M., Baumgartner W.A. Jr, Hunter L.E. KaBOB: ontology-based semantic integration of biomedical databases. BMC Bioinformatics. 2015;16:126. DOI 10.1186/s12859-015-0559-3

Lloyd C.M., Halstead M.D.B., Nielsen P.F. CellML: its future, present and past. Progr. Biophys. Mol. Biol. 2004;85(2/3):433-450.

Martin D., Brun C., Remy E., Mouren P., Thieffry D., Jacq B. GOToolBox: functional analysis of gene datasets based on Gene Ontology. Genome Biol. 2004;5(12):R101.

Mi H., Lazareva-Ulitsky B., Loo R., Kejariwal A., Vandergriff J., Rabkin S., Guo N., Muruganujan A., Doremieux O., Campbell M.J., Kitano H., Thomas P.D. The PANTHER database of protein families, subfamilies, functions and pathways. Nucl. Acids Res. 2005;1:33 (Database issue):D284-8.

Noy N.F., Shah N.H., Whetzel P.L., Dai B., Dorf M., Griffith N., Jonquet C., Rubin D.L., Storey M.A., Chute C.G., Musen M.A. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucl. Acids Res. 2009;37:W170-W173. DOI 10.1093/nar/gkp440

Orchard S., Salwinski L., Kerrien S., Montecchi-Palazzi L., Oesterheld M., Stümpflen V., Ceol A., Chatr-aryamontri A., Armstrong J., Woollard P., Salama J.J., Moore S., Wojcik J., Bader G.D., Vidal M., Cusick M.E., Gerstein M., Gavin A.C., Superti-Furga G., Greenblatt J., Bader J., Uetz P., Tyers M., Legrain P., Fields S., Mulder N., Gilson M., Niepmann M., Burgoon L., De Las Rivas J., Prieto C., Perreau V.M., Hogue C., Mewes H.W., Apweiler R., Xenarios I., Eisenberg D., Cesareni G., Hermjakob H. The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat. Biotechnol. 2007;25(8):894-898.

Podkolodnyy N.L. Ontological modeling in bioinformatics and systems biology. Trudy Vtorogo simpoziuma "Ontologicheskoe modelirovanie" [Proceedings of the 2nd Symposium "Ontological modeling"]. Moscow, IPI RAN, 2011:233-269.

Podkolodnyy N.L., Ignatieva E.V., Podkolodnaya O.A., Kolchanov N.A. Information suppotr of research on transcriptional regulatory mechanisms: an ontological approach. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2012; 16(4/1):742-755.

Rosse C., Mejino J.L.Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J. Biomed. Inform. 2003;36(6): 478-500.

Rosse C., Mejino J.L.V. The Foundational Model of Anatomy Ontology. Anatomy Ontologies for Bioinformatics: Principles and Practice. Eds A. Burger, D. Davidson, R. Baldock. N.Y.: Springer, 2007.

Schober D., Smith B., Lewis S.E., Kusnierczyk W., Lomax J., Mungall C., Taylor C.F., Rocca-Serra P., Sansone S.A. Survey-based naming conventions for use in OBO foundry ontology development. BMC Bioinformatics. 2009;27;10:125. DOI 10.1186/1471-2105-10-125

Shah N.H., Jonquet C., Chiang A.P., Butte A.J., Chen R., Musen M.A. Ontology-driven indexing of public datasets for translational bioinformatics. BMC Bioinformatics. 2009;10(Suppl. 2):S1. DOI 10.1186/1471-2105-10-S2-S1

Smith B., Williams J., Schulze-Kremer S. The Ontology of the Gene Ontology. AMIA Annual Symp. Proceedings, 2003.

Smith B., Ashburner M., Rosse C., Bard J., Bug W., Ceusters W., Goldberg L.J., Eilbeck K., Ireland A., Mungall C.J., OBI Consortium, Leontis N., Rocca-Serra P., Ruttenberg A., Sansone S.A., Scheuermann R.H., Shah N., Whetzel P.L., Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotech. 2007;25(11):1251-1255.

Srinivas K. OWL Reasoning in the Real World: Searching for Godot. Proc. of the 22nd Intern. Workshop on Description Logics (DL 2009), Oxford, UK, July 27–30, 2009.

Stevens R., Aranguren M.E., Wolstencroft K., Sattler U., Drummond N., Horridge M., Rector A. Using OWL to model biological knowledge. Int. J. Human-Computer Studies http://portal.acm.org/citation.cfm?id=1247774. 2007;65(7):583-594.

Stevens R., Baker P., Bechhofer S., Ng G., Jacoby A., Paton N.W., Goble C.A., Brass A. TAMBIS: transparent access to multiple bioinformatics information sources. Bioinformatics. 2000;16(2):184-185.

Taylor C.F., Paton N.W., Lilley K.S., Binz P.A., Julian R.K.Jr., Jones A.R., Zhu W., Apweiler R., Aebersold R., Deutsch E.W., Dunn M.J., Heck A.J., Leitner A., Macht M., Mann M., Martens L., Neubert T.A., Patterson S.D., Ping P., Seymour S.L., Souda P., Tsugita A., Vandekerckhove J., Vondriska T.M., Whitelegge J.P., Wilkins M.R., Xenarios I., Yates J.R. 3rd, Hermjakob H. The minimum information about a proteomics experiment (MIAPE). Nat. Biotechnol. 2007;25(8):887-893.

Thomas P.D., Campbell M.J., Kejariwal A., Mi H., Karlak B., Daverman R., Diemer K., Muruganujan A., Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129-2141. PubMed PMID:12952881; PubMed Central PMCID: PMC403709.

Waltemath D., Adams R., Bergmann F.T., Hucka M., Kolpakov F., Miller A.K., Moraru I.I., Nickerson D., Snoep J.L., Le Novère, N. Reproducible computational biology experiments with SED-ML – the simulation experiment description markup language. BMC Systems Biol. 2011;5:198.

Whetzel P.L., Parkinson H., Causton H.C., Fan L., Fostel J., Fragoso G., Game L., Heiskanen M., Morrison N., Rocca-Serra P., Sansone S.A., Taylor C., White J., Stoeckert C.J.Jr. The MGED Ontology: a resource for semantics-based description of microarray experiments. Bioinformatics. 2006;22(7):866-73.

Winston M.E., Chaffin R., Herrman D. A taxonomy of part-whole relations. Cognitive Sci. 1987;11:417-444.