BIOINFORMATICS

Ontologies of Cellular Networks

Robert Arp* and Barry Smith

Published 16 December 2008; Volume 1 Issue 50 mr2

A report on the workshop "Ontologies of Cellular Networks," Newark, New Jersey, 27 to 28 March 2008.

As part of a series of workshops on different aspects of biomedical ontology sponsored by the National Center for Biomedical Ontology (NCBO), a workshop titled "Ontologies of Cellular Networks" took place in Newark, New Jersey, on 27 to 28 March 2008. This workshop included more than 30 participants from various backgrounds in biomedicine and bioinformatics. The goal of the workshop was to provide an introduction to the basic tools and methods of ontology, as well as to enhance coordination between groups already working on ontologies of cellular networks. The meeting focused on three questions: What is an ontology? What is a pathway? What is a cellular network?

Introduction

A workshop titled "Ontologies of Cellular Networks" took place in Newark, New Jersey, on 27 to 28 March 2008 as part of a series of workshops on different aspects of biomedical ontology sponsored by the National Center for Biomedical Ontology (NCBO). The workshop, which was organized by Yves Lussier (University of Chicago, Chicago, Illinois), Alan Ruttenberg (Neurocommons, Cambridge, Massachusetts), and Barry Smith (University at Buffalo, Buffalo, New York), involved more than 30 participants from backgrounds in biomedicine and bioinformatics, representing communities such as BioPAX, Reactome, the Integrating Network Objects with Hierarchies Pathway Database, the Semantic Web Health Care and Life Sciences Interest Group, the Open Biomedical Ontologies Foundry, and the National Institutes of Health. There were four main goals of the workshop: (i) provide an introduction to the basic tools and methods of ontology; (ii) foster networking of, and enhance coordination between, groups already working on ontologies of cellular networks; (iii) identify problems that must be solved if ontology methods are to be extended to represent biological mechanisms in greater detail; and (iv) promote further ontology development in this area with the goal of accelerating our ability to understand basic biological phenomena and to leverage experimental data (*1*). Three questions formed the focus of the meeting: What is an ontology? What is a pathway? What is a cellular network?

National Center for Biomedical Ontology, University at Buffalo, Buffalo, NY 14260, USA.

*Corresponding author. E-mail, rarp@buffalo.edu

What Is an Ontology?

In the 1990s, biologists and information scientists began using the term "ontology" to refer to the taxonomic classification of the entities in a particular domain of study (*2*). Today, it is common to hear the term "domain ontology." A domain is a delineated portion of reality corresponding to a discipline, such as cell biology or electron microscopy. A first approximation of a domain ontology is that it is a graph consisting of a backbone taxonomic tree whose nodes represent types of entities in reality. These nodes are connected to the tree by edges representing the "is_a" (= subtype) relation. In examples from the Lipid Ontology being developed by Baker and colleagues (*3*), we have "LC hepoxilin is_a LC eicosanoid" and "lipid is_a small molecule" (Fig. 1). Other edges of the graph relate the nodes to each other through such relations as "part_of" (is part of), "preceded_by" (is preceded by), "has_participant" (has as a participant), or "inheres_in" (inheres in). Examples of the use of these types of relationships are evident in the Infectious Disease Ontology being developed by Lindsay Cowell (Duke University, Durham, North Carolina) and others (*4*) (Fig. 2).

In addition, the ontology can be supplemented with axioms and definitions that enable reasoning on the basis of these relationships—for example, axioms to the effect that "is_a" and "part_of" are transitive (Thus: "If A part_of B and B part_of C, then A part_of C"). In this way, the ontology can be used to infer new information about the underlying instances that comprise the domain of study; for example, a finger can be inferred as part of the body based on the following ontological relationships: "hand part_of body" and "finger part_of hand," thus "finger part_of body."

Examples of domain ontologies in this sense include the Gene Ontology (GO) (http://www.geneontology.org), which is an ontology of three domains—Molecular Function, Biological Process, and Cellular Component; the Foundational Model of Anatomy (FMA) (http://sig.biostr.washington.edu/projects/fm); and the Ontology for Biomedical Investigations (OBI) (http://obi-ontology.org/page/Main_Page), which, together with other ontologies, make up the Open Biomedical Ontologies (OBO) Foundry (*5*).

A massive amount of data and information is produced daily by researchers. The different ways in which researchers collect these data and information have produced a silo effect, whereby the data collected by different research groups or disciplines do not work well together. Even where domain ontologies are used to organize data and information within one specific field, there may still remain the problem of noninteroperability resting on the limited shareability and reusability of data and information deriving from neighboring domains because the ontologies and terminologies used have been created independently and in ad hoc and often unprincipled ways. At the workshop, Peter Lyster (National Institutes of Health, Bethesda, Maryland) made the important observation that "there is a disconnect between modeling, visualizing, analyzing, tagging, and organizing data, and building ontologies that make what is involved in modeling, visualizing, and the like, interoperable."

Ontologies that rest on common principles and enjoy a large degree of acceptance across multiple groups help to facilitate interoperability by providing a controlled, structured representation that can serve as a common framework for the annotation of data and, thereby, make the data more easily searchable by human beings and processable by computers. They provide also a framework for the annotation of literature, which may make available for algorithmic reasoning information hitherto available only in the form of free text descriptions of scientific experiments.

To create an ontology that can serve these purposes, the first goal is to compose a list of relevant terms, codes, or other

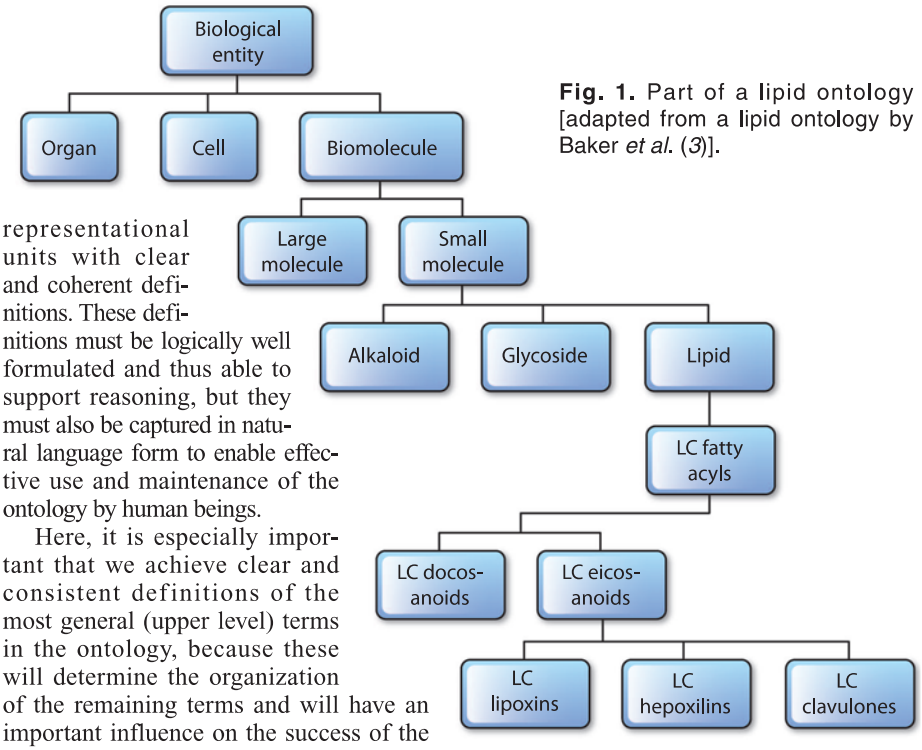

Fig. 1. Part of a lipid ontology [adapted from a lipid ontology by Baker *et al.* (*3*)].

representational units with clear and coherent definitions. These definitions must be logically well formulated and thus able to support reasoning, but they must also be captured in natural language form to enable effective use and maintenance of the ontology by human beings.

Here, it is especially important that we achieve clear and consistent definitions of the most general (upper level) terms in the ontology, because these will determine the organization of the remaining terms and will have an important influence on the success of the ontology in serving integration. Definitions of the most general terms involved in research on cellular networks, including the term "pathway," thus were an important topic of discussion at the meeting.

What Is a Pathway?

Chris Sander's (Memorial Sloan-Kettering Cancer Center, New York, New York) opening talk was titled "Pathways and Networks: An Overview of the Science" and defended one standard view of a cellular network as a collection of pathways. Unfortunately, the term "pathway" is itself used in different ways by researchers, so this term, too, is in need of definition.

Six definitions of "pathway" were offered at the meeting.

- Definition 1: A pathway is a set or series of interactions, often forming a network, which biologists have found useful to group together for organizational, historic, biophysical, or other reasons.

- Definition 2: A pathway is a representation of the processes involved in the cell and between cells.

- Definition 3: A pathway is a series of processes, starting with an input or stimulus and ending with a measured output or phenotype.

- Definition 4: A pathway is a connected sequence of two or more processes happening in the cell and involving distinct participants.

- Definition 5: A pathway is any process happening in the cell whose participants are, or are composed of, chemical substances.

- Definition 6: A pathway is a module containing a collection of processes that form a common function for a common purpose represented as a directed connected subgraph where nodes are cellular components and links represent their regulatory relationships.

Definition 1, which is used by BioPAX (*6*), was presented by Ken Fukuda (National Institute of Advanced Industrial Science and Technology, Tokyo, Japan) in his talk on the Integrating Network Objects with Hierarchies (INOH) Pathway Database. It raises an issue much discussed at the meeting: Are pathways genuine biological entities, there to be discovered? Or, are they merely the reflections of conventional and arbitrary historical groupings made by biologists, analogous to constellations, such as Orion or the Great Bear, or to geographic fiat entities such as Norway or French Polynesia?

Definition 2 was suggested by Emek Demir (Memorial Sloan-Kettering Cancer Center, New York, New York) and raises an analogous issue: Are pathways biological entities out there in the world? Or are they merely representations created by biologists? Certainly representations of pathways exist and are published in pathway databases. But are there also actual pathway-like entities out there in the world, waiting to be studied by biologists and serving as the targets of correct and incorrect representations?

Definition 3, suggested by Chris Sander, introduces the notion of a pathway as a sequence of processes, and thus raised the question as to what are the appropriate participants in such processes. What kinds of interactions link them together? What does it mean to say that interactions form a network? Further, what does it mean to say that the sequence of processes can end with a phenotype?

Definitions 4 and 5 were both from Oliver Ruebenacker (University of Connecticut Health Center, Farmington, Connecticut) and provide two alternative specifications of "pathway" as signifying a process "happening in the cell"—thus leaving aside what seem to be pathways that include many cells; there is not only intra- but also intercellular signaling and signal transduction.

In Definition 6, from Avi Ma'ayan (Mount Sinai School of Medicine, New York, New York), a pathway is defined as a functional module, which conveys the idea that the processes that comprise a pathway must be distinguished from surrounding processes by some kind of boundary and suggests also that the participants inside the module must enjoy some kind of common fate or shared causality in that they contribute to realizing a single function or purpose.

At the meeting, many researchers questioned whether pathways could have clear boundaries. Consider the issue of when a pathway starts and when it ends: Is where we draw these boundaries an arbitrary matter? If yes, does this mean that the pathway we identify by drawing these boundaries does not exist, is a mere fiction? Although it is certainly true that researchers determine the boundaries of pathways when they label them—as in "the MAPK/ERK pathway" or "the Notch signaling pathway" (*7*)—the entities that comprise these pathways are perfectly real, even though, like French Polynesia,

they have arbitrary borders. Further, although it may be difficult to determine exactly and clearly the boundaries of pathways because different researchers may see a larger or a smaller number of entities as being involved in a given pathway, the cells, organelles, enzymes, interactions, signals, and other objects and processes that are described in these pathways are real and so also are the interconnections between them.

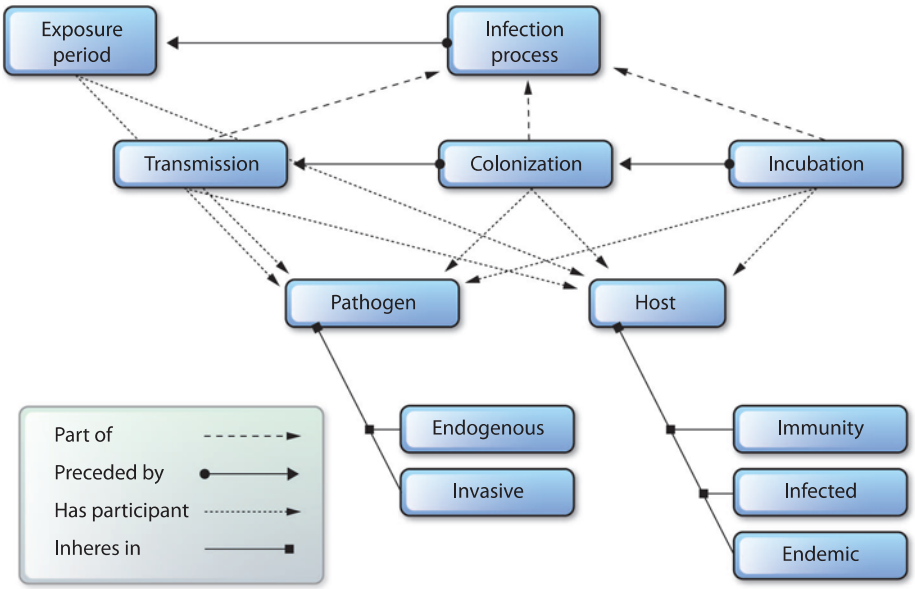

Fig. 2. Part of the Infectious Disease Ontology (*4*), illustrating the different types of relations among terms in the ontology.

What Is a Cellular Network?

Inspired by the ontological work at the meeting, we here attempt to define "cellular network" and include definitions for the terms used. We suggest that a pathway is a connected sequence of two or more processes having a shared causality in that the processes contribute to realizing a common function, whereby the output of one process is the input for the next process in the sequence. An intracellular process is a chemical reaction or collection of chemical reactions taking place within the cell. An intercellular process is a chemical reaction or collection of chemical reactions that take place between and among cells. A cellular pathway is a pathway that is composed of intracellular and intercellular processes. A network is a connected sequence of two or more pathways having a shared causality in that the pathways contribute to realizing a common function, and involving distinct pathways, where the output of one pathway is the input for the next pathway in the sequence. A cellular network, then, is a network that is composed of cellular pathways.

Progress

This workshop opened the door to further research on the ontology of cellular networks and included attempts to establish an acceptable consensus definition of "cellular network." Such a definition will serve as an important first step in developing an ontology of cellular networks that is interoperable with the already established ontologies of the OBO Foundry (*5*, *8*).

References and Notes

1. Ontology of Cellular Networks, http://www.bioontology.org/wiki/index.php/Networks.
2. T. Gruber, Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum. Comput. Stud.* **43**, 907–928 (1995).
3. C. Baker, R. Kanagasabai, W. Ang, A. Veeramani, H.-S. Low, M. Wenk, Towards ontology-driven navigation of the lipid *bibliosphere*. *BMC Bioinformatics* **9** (suppl. 1), S5 (2008).
4. Infectious Disease Ontology, http://www.infectiousdiseaseontology.org/Home.html.
5. Open Biomedical Ontologies Foundry, http://www.obofoundry.org.
6. BioPAX, http://biomserv.univ-lyon1.fr/baobab/parsebionet/JavaDoc/baobab/parseBioNet/bioData/BioPathway.html.
7. R. Orton, O. Sturm, V. Vyshemirsky, M. Calder, D. Gilbert, W. Kolch, Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. *Biochem. J.* **392**, 249–261 (2005).
8. B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, The O.B.I. Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. Sansone, R. Scheuermann, N. Shah, P. Whetzel, S. Lewis, The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251–1255 (2007).
9. Thanks to N. Shah and N. Gough for comments. This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028. Information on the National Centers for Biomedical Computing can be found at http://nihroadmap.nih.gov/bioinformatics.

10.1126/scisignal.150mr2

Citation: R. Arp, B. Smith, Ontologies of cellular networks. *Sci. Signal.* **1**, mr2 (2008).