

Correspondence

Ontologies for programs, not people

Lawrence Hunter

A response to **Life sentences: Ontology recapitulates philology** by Sydney Brenner, *Genome Biology* 2002, **3**:comment1006.1-1006.2.

Address: Center for Computational Pharmacology, Department of Pharmacology, School of Medicine, University of Colorado Health Sciences Center, 4200 E. Ninth Avenue, Denver, CO 80262, USA. E-mail: Larry.Hunter@uchsc.edu

Published: 31 May 2002

Genome Biology 2002, **3**(6):interactions1002.1-1002.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/6/interactions/1002>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

In a recent column [1], the wry and erudite Sydney Brenner expressed his disdain for current efforts to create computational ontologies of molecular biology. In essence, Brenner argued that building a network of names of biological entities is a waste of time. It is the nucleotide sequences or amino-acid conformations of these objects, not their names, "that create the processes that produce outcomes for cells, organs and organism," he says. "Very simply, the network we should be interested in is not the network of names but the network of the objects themselves."

Brenner's article misses the point - several points, actually. First, the essence of the Gene Ontology project, of which he is specifically critical, and of other knowledge-bases of molecular biology, such as EcoCyc [2] or the Unified Medical Language System (UMLS) [3], is not in the list of names they embody, but in the relationships they represent. The names are convenient symbols to which more complex statements can be attached. Without the names, it is impossible to specifically represent relationships such as 'activates' or 'binds to'. Surely that sort of information must be the kind of thing that Brenner means when he says we are interested in the interactions between the objects themselves, rather than their names.

If we are to build useful databases of the interactions that Brenner suggests ought to hold our interest, then there are significant advantages to being able to make statements about various groupings of genes and gene products together, using the terminology that is familiar to molecular biologists. For example, representation of the statement 'the balance between pro- and anti-apoptotic members of the *bcl2* family of genes determines whether apoptosis proceeds' is straightforward if we use an ontology that contains the appropriate abstractions, and painfully difficult if we are limited to expressions of direct interactions between pairs of genes and proteins.

The third important point to consider is to whom Brenner is referring when he uses "we" in his argument. Knowledge-bases are not generally used directly by an end user, but instead by computer programs in order to accomplish complex inference tasks. Many productive and promising approaches to bioinformatics require a computationally manipulable representation of existing biological understanding - incomplete and incorrect as it may be - as a vital prerequisite. For example, inference from gene-expression data using Bayesian networks [4] can take advantage of online sources of information about the likely probabilistic

dependencies among expression levels of various genes. Knowledge-bases built from textbooks, review articles, or even the Oxford Dictionary of Molecular Biology can provide precisely this sort of computationally useful information.

The fourth issue is that if bioinformaticians are to build useful tools for managing the ever-growing onslaught of research publications resulting from high-throughput instrumentation and exacerbated by the collapse of disciplinary distinctions, then they must first create computer programs that recognize references to genes, proteins and other biological entities in texts. Automatically linking references to molecular entities and processes in texts (such as Medline abstracts) to the appropriate entries in molecular databases (such as GenBank) can save enormous amounts of researcher time and facilitate the kind of biology that Brenner holds dear. Such a mapping, however, requires the presence of a well-represented knowledge-base of molecular biological entities - perhaps like the Gene Ontology.

Brenner is, of course, entitled to his opinion about the utility of efforts like the Gene Ontology and the UMLS. Perhaps he doesn't need any of the computational tools for analyzing high-throughput data in light of prior

knowledge, or managing the vast scientific literature, either. For those of us who use bioinformatics software to advance scientific understanding, however, broad community efforts at knowledge representation - like the effort of the Gene Ontology Consortium - are invaluable.

References

1. Brenner, S: **Life sentences: ontology recapitulates philology.** *Genome Biol* 2002, **3**:comment1006.1-1006.2. Also published in *The Scientist* 2002, **16**:12.
2. Karp, PD: **Pathway databases: a case study in computational symbolic theories.** *Science* 2001, **293**:2040-2044.
3. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO: **The Unified Medical Language System: an informatics research collaboration.** *J Am Med Inform Assoc* 1998, **5**:1-11.
4. Segal E, Taskar B, Gasch A, Friedman N, Koller D: **Rich probabilistic models for gene expression.** *Bioinformatics* 2001, **17** Suppl 1:S243-S252.