

Developing a protein-interactions ontology

Esther Ratsch

European Media Laboratory, Heidelberg, Germany

Within a living cell, many processes are regulated by signal transduction pathways, which thus are crucial for life. I want to present here the procedure we followed and some of the problems we met, while developing an ontology for the description of signal transduction pathways.

To get an idea of the concepts involved in the description of signalling pathways, we chose a case example: the Jak-Stat-pathway, which includes several different types of interactions. This pathway starts at the cellular membrane, where a ligand binds to a dimeric receptor, with which Jak proteins are associated. Ligand binding leads to phosphorylation of Stat monomers by activated Jak proteins and subsequent dimerisation of the Stat monomers. The Stat dimers translocate into the nucleus, where they upregulate the transcription of a set of target genes. To model such a pathway, several concepts need to be covered, which in a first step of ontology building need to be identified. In the case of signal transduction, these are for example proteins and their characteristics at different states. The interactions making up a signal transduction pathway can be seen as state transitions of proteins, where a state is described by a protein's localisation, its modifications (e.g. phosphorylation) and other compounds it is bound to. The next steps are the identification of relationships between the different concepts (like a protein being composed of amino acids) and rules governing them. After this, the formalisation of the involved concepts, relationships and rules can begin, which leads to the question of how much formalisation is needed and useful.

The group working on this project is composed of researchers from different disciplines: biology, computer science and computational linguistics. The combination of the modeling and formalisation experience of the computer scientists and the computer linguists, together with the domain knowledge of the biologists, allows the generation of a high quality model. Nevertheless, the interdisciplinarity also bears problems, as it is not always easy to find a compromise between applicability and formalisation. Another difficulty that arises in some occasions is the communication, as these communities have somewhat different vocabularities. This can also be seen in a positive way, as this compels both the domain knowledge people and the modelers to be clear in what they express and represent.

Biographical notes:

Esther Ratsch is a research scientist in the Scientific Databases and Visualization Group at the European Media Laboratory (EML) in Heidelberg, Germany. Her research interest is knowledge modeling in the biological domain with the focus on the representation of experimental data and on the modeling of signal transduction pathways. Before joining the EML in 2001, she worked in a virological laboratory on the regulation of the replication of Hepatitis B Virus. She received her degree in biology from the University of Constance, Germany.