

**Poster F-3**  
**Protein Structure Homology**  
**Modelling assisted by Ontology**



**Authors:**

Kellen Pinagé (*Computing Science Department, Federal University of Amazonas*)  
Virginia Brilhante (*Computing Science Department, Federal University of Amazonas*)

**Short Abstract:** Determining a protein 3D structure is fundamental to knowing its function. The Homology Modelling method applies knowledge reuse to determine protein structures. Improve the method by introducing an ontology to it and demonstrate that an ontology-based knowledge reuse technique is applicable to the protein structure prediction is the goal

**Long Abstract:**

Proteins are chain-shaped molecules formed through amino acid bonds. In their natural environment, proteins fold in several ways giving rise to 3D structures. The specific biological function of a protein in a cell is related with its 3D structure and only in this conformation can a protein carry it out. Therefore, determining a protein structure is fundamental to knowing its function.

In order to find out the structure of proteins, the so-called methods of tertiary structure prediction are applied. Predictions are made from the proteins primary conformation, that is, their amino acid sequences. It is believed that knowing the sequence of amino acids of a protein suffices to predict its 3D conformation. Experiments in laboratory show that when a protein structure is stretched in vitro and then set free, it returns to its original shape after a short time, provided that it does not suffer interference from some other molecule [4].

Some prediction methods are performed by computer, such as the Comparative Modelling of Protein Structure method, also known as Homology Modelling. The method involves knowledge reuse – knowledge used to generate new knowledge – in that information bases about known protein structures are used in the prediction of unknown structures. Homology Modelling can thus be seen as a Knowledge Engineering technique.

An ontology-based knowledge reuse technique has been proposed and applied to the problem of designing data-compliant structures of ecological simulation models [2]. The technique comprises an ontology of ecological data and a modelling mechanism which matches existing model structures against new ontology-annotated data; the matching substructures are then filled in with the data and assembled together to yield new models. The technique has been shown to bring about an efficiency gain to a modelling system in comparison to one that does not reuse model structures.

In general terms, this is a knowledge engineering technique in which: i) data to feed a model structure is annotated with an ontology of the concerned domain; ii) framework structures are abstracted from existing reference models and their characteristics matched, or associated, with the ontological concepts that describe the domain data; iii) in this way, the reference structures are reused to produce new, data compatible model structures [1].

The Homology Modelling method already applies knowledge reuse but does not yet benefit from a domain ontology. We are currently in pursuit of improving the method by introducing an ontology to it and demonstrating that an ontology-based knowledge reuse technique, such as the one aforementioned, is applicable to the problem of protein structure prediction.

To this end, we have adopted the Protein Ontology (PO) [5, 6], an outcome of a still active project of the Sidney Technology University. The construction of PO has been motivated by the need for a structured representation of proteomics data to be applied, for example, to facilitate the sharing of knowledge of the proteomics domain by researchers and their software applications and the integration of multiple protein data bases such as PDB and SWISS-PROT. The ontology provides a structured vocabulary that is suitable for describing protein data including amino acid sequences as well as data on bonds between atoms and protein functions.

As for the tool to be improved through our technique, we are working with Nest, considered to be one of the best Homology Modelling tools [7]. Nest is the central program in the JACKAL package [8] and implements modelling by assembly of rigid bodies, whereby a model is composed from a small number of rigid bodies obtained from the core of aligned sequence regions [3,7]. The program takes as input a file containing the alignment between a query sequence and a template sequence whose structure is known. Information about such structure is then sought for in the PDB repository, on the basis of which Nest predicts the structure of the query sequence.

We will semantically enrich the information concerning sequences and structures used in the Homology Modelling process by annotating it with the Protein Ontology. This will have the effect of narrowing the conceptual gap between the amino acid sequences data and protein structure information, leading to an improved prediction performance since the loss of matching information is bound to be reduced.

## References

1. Brilhante, Virgínia. Ontology-Enabled Reuse of Structural Data-Based Model. In Proceedings of the Workshop on Ontologies and their Applications (WONTO'04) at the 17th Brazilian Symposium on Artificial Intelligence, São Luís, Brazil, 2004.
2. Brilhante, Virgínia. Ontology and Reuse in Model Synthesis. Ph.D. thesis. School of Informatics, University of Edinburgh, 2003.
3. Martí-Renom, Marc A., Ashley C. Stuart, András Fiser, Roberto Sánchez, Francisco Melo and Andrej Sali. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29:291–325, 2000.
4. Meidanes, João and João C. Setúbal. Uma Introdução à Biologia Computacional. IX Escola de Computação, Recife, 1994.
5. Protein Ontology. <http://www.proteinontology.info>. Last accessed on 17 Apr 2006.
6. Sidhu, Amandeep S. and Tharam S. Dillon. Protein Ontology: Vocabulary for Protein Data. In Proceedings of the 3rd International Conference on Information Technology and Applications (ICITA'05). IEEE Computer Society. 2005.
7. Wallner, Björn and Arne Elofsson. All Are Not Equal: a Benchmark of Different Homology

Programs. Protein Science14:1315–1327, 2005.

8. Xiang, Jason Z. JACKAL: A Protein Structure Modeling Package. Columbia University & Howard Hughes Medical Institute. July, 2002.