Master's Thesis Proposal

General Information

Master's Thesis Title:	Bayesian Hierarchical Models for the Detection and Description of Linking Sites in GPCRs
Publication Date:	March 2010
Expiry Date:	June 2010
Orientation:	☐ professional ☒ research
M.Sc. Th. Advisor:	Alfredo Vellido and Jesús Giraldo
M.Sc. Th. Advisor's Dept. & University:	A.V: Lenguatges i Sistemes Informàtics. UPC J.G.: Institut de Neurociències. UAB
M.Sc. Th. Advisor e-mail:	avellido@lsi.upc.edu , Jesus.Giraldo@uab.es
Observations:	
Student's Name: (if already known)	Martha Ivón Cárdenas

M.Sc. Thesis Description

Main issues / Brief Description:

In this thesis, the student will address the problems of detection of linking sites and prediction of their location in GPCR proteins through the definition and use of Bayesian statistical ML models. These will follow a hierarchic approximation that will use not only the properties stemming from the amino-acid sequences, but also the reliable information on the secondary and tertiary structures of the proteins.

Detailed Description:

G protein-coupled receptors (GPCRs) are essential components of signal transduction processes in proteins. These receptors regulate many cell functions and have a relationship with around 3% of the human genes. GPCR links are the estimated target of approximately half of the medicaments currently in clinical use.

Probabilistic modelling and, specifically, machine learning (ML) probabilistic models have only recently begun to be applied to the analysis of GPCR functioning, although their application is expected to generate new insights in this field. Statistical ML techniques are specially suited to deal with some of the common challenges of molecular modelling in proteins, and should be of special interest when the three-dimensional structures of the proteins and receptors remain unknown at large.

The detection of linking sites and the prediction of their location in proteins, and in GPCRs in particular, are tasks of great relevance, both from the point of view of basic biochemical research and as key elements on the discovery of substances of pharmacologic interest. In the proposed thesis, the student will address these problems through the definition and use of statistical ML models using Bayesian techniques. The model will follow a hierarchic approximation that will use not only the properties stemming from the amino-acid sequences, but also the reliable information on the secondary and tertiary structures of the proteins.

References:
[1] I. Olier, A. Vellido and J. Giraldo (2010) Kernel Generative Topographic Mapping. In Procs. of the 18th European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium.
[2] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge Univ. Press, Cambridge, 2004.
[3] Z. R. Yang and R. Thomson, A novel neural network method in mining molecular sequence data, IEEE Trans. on Neural Networks, 16:263–274, 2005.

Minimal Requirements & Previous Knowledge:
Basic Knowledge on medium/high level programming language
Good background in mathematics and, as a plus, in statistics
Master course-level knowledge of data mining / machine learning techniques
Advanced reading/writing English language skills

Other comments:

Location and Date: 22/03/10   Barcelona,

To the Academic Commission of the Master in Artificial Intelligence (CAIMIA)