Postgraduate Student: **Dionisis Markopoulos**

Thesis Title:

Development of a software application for managing protein crystallization data

Abstract:

The new automated methods and the high throughput techniques in the field of macromolecular crystallization lead to the generation of a large volume of data. The increasing need for the management and exploitation of those data for the production of new knowledge has driven the scientific community efforts towards the development of new more effective data management solutions.

This process is summarized in two parts. The first one comprises the systematic recording of both the crystallization trials and the scientific data related to the experiments performed in a research laboratory. It is considered extremely useful for the researcher to monitor the progress of his/her work by keeping a record in an organized way and based on a stable formalism. He/she should also be in position to recover some parts or the entire datasets that is required, using appropriate search queries. The second part is the ability to exploit the recorded data in a way that allows new information to be extracted to be used by the researcher potentially for a new series of successful crystallization experiments.

The objective of the current thesis was the formation of a software application/system for the recording and managing of crystallization trials as well as for the evaluation of their result. For this purpose we have created a relational database coupled with a management system. The application developed is coded in html and thus can be operated through any internet browser. Besides the crystallization experiments there is also the capability of recording sets of commercial kits regularly used in crystallization trials of macromolecular targets, as well as other combinations of chemical solutions prepared in a laboratory.

Additional groups of data that can be recorded in the database are information for every protein sample under investigation as well details on the scale up experiments. Users are able to annotate the experimental results by the insertion of written comments, imported images and score assignments. The database is formed in a way that accepts data from both successful and unsuccessful crystallization trials, accompanied by their proper annotation. This option facilitates further development and upgrade of the new tool for predicting the optimal crystallization conditions, using the available data for the training of machine learning models. The search queries have been formed with the aim to give the researcher the potential to

define the desired criteria and recover all the experiments and relevant information that correspond to particular conditions.

The software application is easy to install and can be used by any member of a crystallization research laboratory. The already recorded records can be transferred in from system to system via exported text files. This software application allows can be easily expanded and customized with changes that can be incorporated either the in database or in the management front-end system to cover future needs.

SUBJECT AREA: Bioinformatics, Protein Crystallization

KEY WORDS: crystallization trials, crystallization conditions, X-ray protein crystallography, laboratory information management system, recording and organizing of experiments, search queries

Examining Committee:

Elias S. Manolakos, Associate Professor, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens (Advisor)

Georgios Th. Tsangkaris, Staff Research Scientist - Professor Level, Biomedical Research Foundation of the Academy of Athens

Evangelia Chrysina, Scientific staff, National Hellenic Research Foundation (co-advisor)