

197b Agml Central: a Proteomic Infrastructure for the Representation, Storage and Analysis of 2-D Gel Electrophoresis Data

Romesh Stanislaus, C Chen, John Arthur, and Jonas S. Almeida

High throughput proteomic initiatives require seamless bioinformatics integration of a range of analytical steps between sample collection and systems modeling immediately accessible to the participants involved in the process. Additionally, these proteomic efforts results in the generation of large amount of data that are in several different data types and formats. In our effort to handle such data in a high throughput fashion we have created infrastructure known as AGML Central. AGML Central is an open-source public proteomic infrastructure for dissemination of 2-D Gel Electrophoresis (2-DE) proteomics data. It is based on the AGML (Annotated Gel Markup Language) an XML standard developed for the annotation of 2-DE data. AGML is an open-source platform independent file format that can store 2-DE data, images and mass spectrometry data as a single XML document. AGML Central infrastructure consists of an input layer which consists of a growing collection of converters from proprietary formats such as those produced by PDQUEST (BioRad), PHORETIX 2-D (Nonlinear Dynamics) and Melanie (GenBio SA). Conversion layer creates an AGML document instance with all the input information and data with or without the raw gel images, mass spectrometry data and stored in a database for future reference. Analysis layer consists of many analytical and search tools. Analysis tools consist of an ever growing list of Matlab toolboxes, Visualization tools and search tools to search the AGML Central database. In order to facilitate automated access a SOAP web service is also provided in the AGML Central infrastructure. This work was supported by the NHLBI Proteomics Initiative through contract N01-HV-28181.