BIOINFORMATICS APPLICATIONS NOTE Vol. 19 no. 16 2003, pages 2149–2151
DOI: 10.1093/bioinformatics/btg281

The UAB Proteomics Database

Aubrey Hill[1,*] and Helen Kim[2]

[1]UAB Comprehensive Cancer Center, [2]Department of Pharmocology and Toxicology, UAB Mass Spectrometry and Proteomics Shared Facility, University of Alabama at Birmingham, 1530 3rd Ave South, Kaul 602A, Birmingham, AL 35294, USA

Received on February 21, 2003; revised on April 25, 2003; accepted on May 8, 2003

ABSTRACT

Summary: The University of Alabama at Birmingham (UAB) Proteomics Database (UPD) (http://www.uab.edu/proteinmenu) was created to provide a repository for the storage and linkage of two-dimensional (2D) gel images and the associated information obtained through mass spectrometry analysis of the proteins excised from the 2D gels in a manner similar to the SWISS-2DPAGE database and the Stanford Microarray Database. This was accomplished through the development of a web interface, a relational database, image maps and hyperlinks stored in the database. In addition to the internally generated data, UPD provides links to the National Center for Biotechnology Information via accession number hyperlinks. UPD currently contains information on 44 individual proteins derived from four experiments conducted by four UAB faculty members. Images of the gels from which each of these proteins was isolated are accessed by hyperlinks embedded in the database.

Availability: The UAB Proteomics Database can be accessed at http://www.uab.edu/proteinmenu

Contact: ahill@uab.edu

INTRODUCTION

Proteomics is an increasingly important area of research with great potential to yield both basic and medically useful information. The separation and enumeration by 2-D gel electrophoresis of proteins associated with various conditions and their subsequent identification by mass spec generates massive amounts of data which must then be processed, stored and made accessible in a manner similar to the SWISS-2DPAGE database (Sanchez *et al*. 1995; Hoogland *et al*. 2000) and the Stanford Microarray Database (Sherlock *et al*. 2001).

The UPD provides a solution to these needs. The design of the UPD assumes that biomedical researchers are not interested in the raw data at its point of acquisition (such as mass spec raw data), but instead expect processed data which is meaningful at a biological level. A second and equally important design principle was to provide a very structured, yet flexible and robust query facility. These two objectives were accomplished by storing proteomics data from a spot/protein perscpective and allowing relational queries to be constructed from any combination of attributes and values. The result of the queries include hyperlinks to both an image map of the original gels as well as a hyperlink to a corresponding mass spec data sheet. The gel image map in turn provides a hyperlink from each protein spot back to the mass spec data.

IMPLEMENTATION

UPD is implemented as a 3-tier software architecture consisting of a web interface, a Java servlet middle tier which dynamically constructs and executes queries, and a relational database. Java servlets are also used to dynamically generate the initial web menus and the web pages that present the query results. The results of the queries are presented in a tabular format.

UPD complies with the five rules of federated 2-DE database construction as described by Appel *et al*. (1996), although currently it provides links only to the National Center for Biotechnology Information (NCBI) protein database.

The primary database entity is a protein/spot on a gel image. The defining attributes of each protein are included and the database is searchable by these attributes/values. These attributes include sample id, protein name, source (tissue, cell line, etc.), apparent molecular mass, theoretical molecular mass, theoretical isoelectric point, measured isoelectric point, gel spot number, investigator name and investigator's department. Items can be retrieved by queries of any of these fields or combinations of these fields. Figure 1 shows the initial menu which is used to construct queries. This web interface is dynamically created from the database each time it is accessed.

Figure 2 presents the results of a query for all samples with an Apparent Molecular Mass greater than 13 300. In this example, the first sample which is returned has a hyperlink, 'JG Spot1' (Data Sheet field), to the mass spectrometry data sheet. Clicking on this field presents the data sheet for this protein.

The NCBI accession number *gi|87303* is included as an item in this data sheet. This hyperlink goes directly to the protein

*To whom correspondence should be addressed.

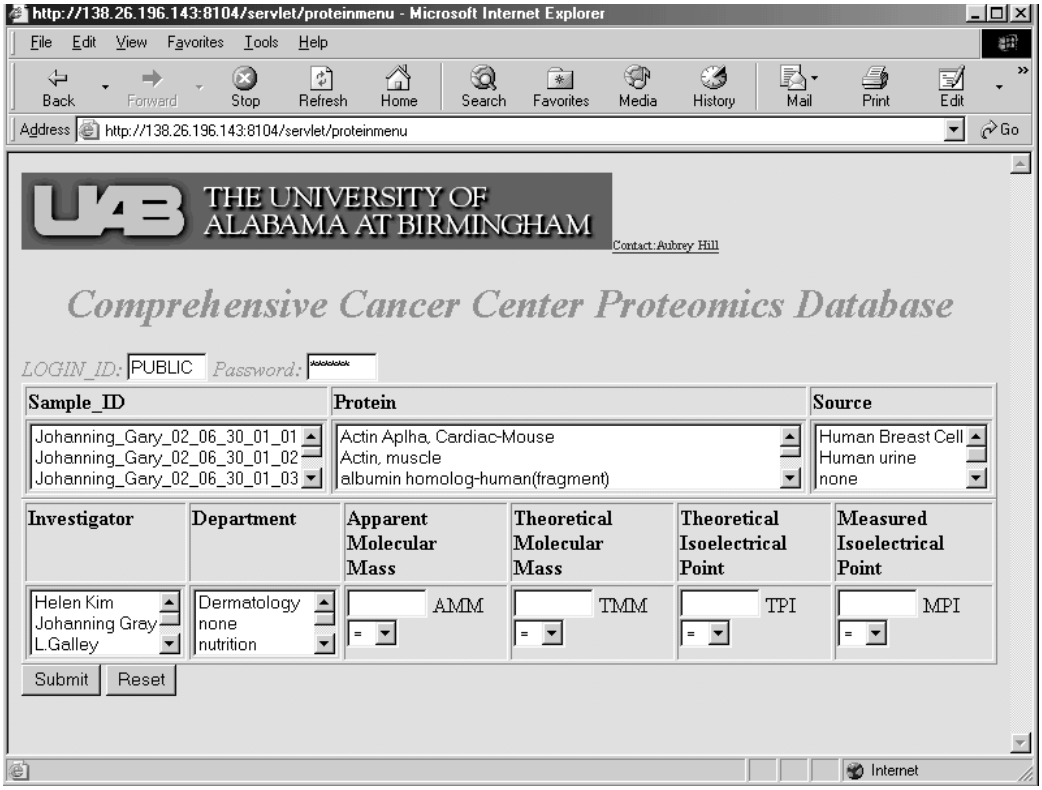
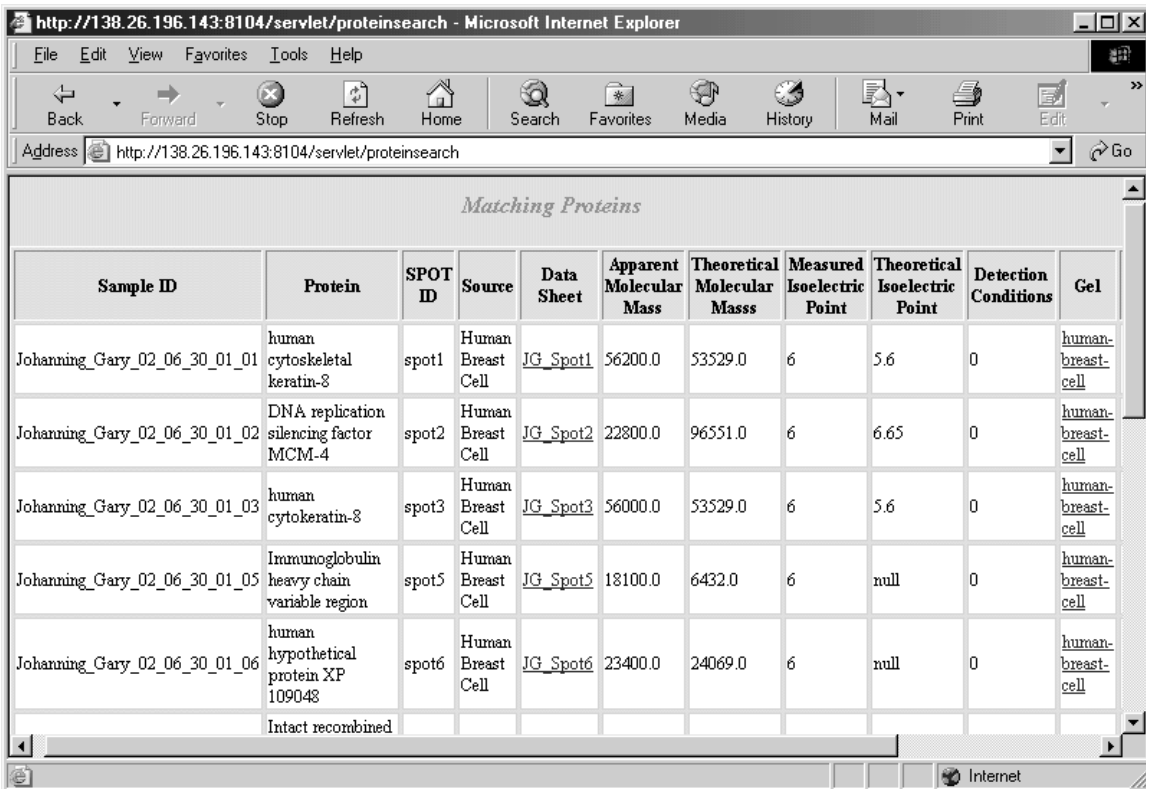


Fig. 1. Query menu.



Sample ID	Protein	SPOT ID	Source	Data Sheet	Apparent Molecular Mass	Theoretical Molecular Masss	Measured Isoelectric Point	Theoretical Isoelectric Point	Detection Conditions	Gel
Johanning_Gary_02_06_30_01_01	human cytoskeletal keratin-8	spot1	Human Breast Cell	JG_Spot1	56200.0	53529.0	6	5.6	0	human-breast-cell
Johanning_Gary_02_06_30_01_02	DNA replication silencing factor MCM-4	spot2	Human Breast Cell	JG_Spot2	22800.0	96551.0	6	6.65	0	human-breast-cell
Johanning_Gary_02_06_30_01_03	human cytokeratin-8	spot3	Human Breast Cell	JG_Spot3	56000.0	53529.0	6	5.6	0	human-breast-cell
Johanning_Gary_02_06_30_01_05	Immunoglobulin heavy chain variable region	spot5	Human Breast Cell	JG_Spot5	18100.0	6432.0	6	null	0	human-breast-cell
Johanning_Gary_02_06_30_01_06	human hypothetical protein XP 109048	spot6	Human Breast Cell	JG_Spot6	23400.0	24069.0	6	null	0	human-breast-cell
	Intact recombined									

Fig. 2. Query result.

record at NCBI. Also shown in Figure 2 is a hyperlink to the original gel from which this protein was isolated. In this case, clicking on *human-breast-cell* displays the gel image, where Spot 1 is human cytoskeletal keratin-8.

FUTURE ENHANCEMENTS

Future enhancements will include a hyperlinked description of the experimental objectives and results for each set of proteins identified from a given set of gels, as well as a means to perform quantitative comparisons between samples using statistical programs such as SAS. Additional hyperlinks will be added to point to other databases such as the Online Mendelian Inheritance in Man.

REFERENCES

Sanchez,J.C., Appel,R.D., Golaz,O., Pasquali,C., Ravier,F., Bairoch,A., and Hochstrasser,D.F. (1995) Inside SWISS-2DPAGE database. *Electrophoresis*, **16**, 1131–1151.

Hoogland,C., Sanchez,J.C., Tonella,L., Binz,P.A., Bairoch,A., Hochstrasser,D.F. and Appel,R.D. (2000) The 1999 SWISS-2DPAGE database update. *Nucleic Acids Res.*, **28**, 286–288.

Sherlock,G., Hernandez-Boussard,T., Kasarskis,A., Binkley,G., Matese,J.C., Dwight,S.S. *et al.* (2001) The Stanford Microarray Database. *Nucleic Acids Res.*, **29**, 152–155.

Appel,R.D., Bairoch,A., Sanchez,J.C., Vargas,J.R., Golaz,O., Pasquali,C., and Hochstrasser,D.F. (1996) Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis*, **17**, 540–546.