Peter F. Lemkin[1]
Geoffrey A. Orr[2]
Mark P. Goldstein[3]
Joseph Creed[4]
Edward Whitley[5]
James E. Myrick[6]
Carl R. Merril[4]

[1]Image Processing Section/LMMB,
Bldg. 469, Room 150,
NCI-FCRDC/NIH, Frederick, MD
[2]PRI/Dyncorp, FCRDC, Frederick,
MD
[3]Monoclonetics International,
Houston, TX
[4]LBG, MIMH/NIH Neuroscience
Center, Washington, DC
[5]NCHGR, Bethesda, MD
[6]NCEH, CDC, Atlanta, GA

The protein disease database

The protein disease database (PDD) correlates quantitative and qualitative protein variations with disease states (C. R. Merril *et al.*, and P. F. Lemkin *et al.* in preparation). The data represented in this database has been extracted from the biomedical literature. Much of the data being collected was determined by many different assay methods and has in general been published in a piecemeal fashion that does not easily allow for searches for correlations between quantitative and qualitative variations and disease states. However, the encapsulation of this type of data in a relational database allows for rapid searches for correlations between disease states and qualitative and quantitative alterations of protein. Quantitative correlations can be performed despite the variations caused by the use of different assay methods by normalizing the protein measurements as fold-changes determined from the ratio of disease-state concentrations/activities divided by the normal-state concentrations/activities. Because many protein concentration changes are nonlinear with disease state, ranges of fold changes can be useful in discrimination between disease states. Ranges of fold changes can be specified in PDD searches. In addition to information on protein variations determined by immunologic, colorimetric, and enzymatic methods reported in the literature, the PDD will also include quantitative and qualitative protein variations observed by high-resolution two-dimensional electrophoresis.

As is increasingly common, we are using the Internet to provide wide user access to this database. The Internet World-Wide Web (WWW) provides the medium and the National Center for Supercomputing Applications (NCSA). Mosaic browser paradigm provides the method for wide distribution and querying of the PDD. The WWW uses a hypertext transfer protocol and an extension, the Common Gateway Interface. The latter makes it possible to build powerful graphical user interfaces supporting data retrieval using images and query-forms. What makes this new delivery mechanism intriguing is that the details of these interactions are totally transparent to the users of these forms. This is because the user only sees the data they are interested in rather than the underlying interactions with the relational database system.

The PDD uses a client-server memory-based structure query language (SQL) relational database engine developed at the National Cancer Institute (NCI). This SQL server in turn is accessible *via* Mosaic over the Internet using the PDD programs residing on the PDD WWW server to provide user access, data entry and database maintenance for all aspects of the database system. Of course security features are in place to protect the integrity of the database. Another aspect of the system is the use of dynamic data entry forms. These are synthesized from current PDD data minimizing the amount of typing required and reducing data entry errors by providing choices from data which already exists in the database. A more detailed description of the PDD system, including biomedical issues and implications of the database, is described (C. R. Merril *et al.*, submitted). Design and implementation issues are discussed (P. F. Lemkin *et al.*, submitted). The PDD group may be contacted through the pdd@ncifcrf.gov E-mail address. The Uniform Resource Locator (URL) http://www-pdd.-ncifcrf.gov/ may be used with Mosaic to attach to the PDD using the WWW*.

Correspondence: Dr. P. F. Lemkin, Image Processing Section, LMMB, Bldg. 469, Rm. 150, NCI-FCRDC/NIH, Frederick, MD 21702, USA

* Parts of this paper were presented at the September 5–7, 1994 conference "Two-dimensional Electrophoresis: From Protein Maps to Genomes" in Siena, Italy