An Attempt to Integrate Various Protein Databases Using an Object-Oriented Database Management System

H. Suzuki
hsuzuki@sci.jip.co.jp

H. Katoh
watson@sci.jip.co.jp

Social and Scientific Department,
Japan Information Processing Service Co., Ltd.
2-4-24 Toyo, Koto-ku, Tokyo, 135 Japan

Abstract

The Human Genome Project activities have been resulting in data accumlation on various databases. Each of them consists of a large amount and variety of data. By using these databases cooperatively, researchers can get information more widely and deeply.

Relational database management systems have been used for one of methods for integration of these databases. These database systems, however, require that data be represented as groups of two-dimensional tables, with the relationships among these tables defined by special columns and separate tables. Because of this restriction, users are forced to contort data into the required form.

By contrast, object-oriented database management systems let users model information in structures that are natural to the data and the application. These systems can manage complex data powerfully and integrate several databases easily.

We have been developed protein variant database collaborated with JIPID, Japan node of PIR-International, and also developed protein database query system. In addition, several novel databases are developed by JIPID in collaboration with National Biomedical Research Foundation (NBRF). All these databases as well as the sequence databases will have interfaced data items, so data that exist in one database may be transferred to another database. These databases are collectively called the database web. In order to utilize these database more effectively, we have attempted to develop an integrated protein database system using an object-oriented database management system. In this system, we treat PIR protein sequence

鈴木博実、加藤博章：日本電子計算株式会社社会科学事業部データベース営業部システム第二課,
〒135 東京都江東区東陽 2-4-24

database, variant database, enzyme database, and protein secondary structure database. Last three databases are developed by JIPID.

This integrated database system has following main characteristics:

1. When users display one of entries, they can pick up corresponding entries from other databases by pushing a button. For example, when users push a button labeled "variant" on the window showing one of PIR entries, this system searchs for variant entries which have same protein and species name with this PIR entry.

2. The data common to plural database is stored in one database, and is automatically transferrable to the others.

3. Color graphic user interface is designed so that researchers who are not familiar with UNIX workstation can use it. Commands are inputed by pushing a button on the window or selecting from pull down menu.

4. The result of searching is listed in a table in a new window. Contents of entries selected in this list are displayed. Users can select more than one entris in the same time.

References

[1] A. Tsugita, A. Ubasawa, C. S. Jone, In: Ikeraha M. (ed) Protein Engineering. Springer, Berlin Heidelberg New York, pp. 273-278, 1990.

[2] A. Ubasawa, F. Okibayashi, C. S. Jone, M. Ikehara, D. G. George, *Protein Seq. Data Anal.*, Vol. 4, pp. 341-347, 1991.

[3] C. S. Jone, A. Tsugita, K. Satake, F. Okibayashi, K. Imai, T. Yagi, K. Takahashi, L. -S. Yeh, *Protein Seq. Data Anal.*, Vol. 4, pp. 367-374, 1991.

[4] H. Suzuki, A. S. Kolaskar, S. L. Samuel, J. Otsuka, A. Tsugita, *Protein Seq. Data Anal.*, Vol. 4, pp. 91-104, 1991.