

3DInSight: An Integrated Relational Database and Search Tool for Molecular Structures and Functional Sites

J. An ¹

ajh@dbl-lab.is.tsukuba.ac.jp

Y. Kubota ²

kubota@human-system.co.jp

T. Nakama ²

nakama@human-system.co.jp

A. Sarai ³

sarai@rtc.riken.go.jp

¹ Institute of Electronics and Information Sciences, University of Tsukuba
1-1 Tannoudai, Ibaraki 305 Japan

² Advanced Technology Institute Inc.,
3-23-15 Jinbo, Kanda, Tokyo 101 Japan

³ RIKEN Life Science Tsukuba Center, The Institute of Physical and
Chemical Research (RIKEN)
3-1-1 Koyadai, Tsukuba, Ibaraki 305 Japan

Although the Protein Data Bank (PDB) contains a large number of three-dimensional structures of biomolecules, conventional search tools provide only simple functions such as keyword searching (entry, header, etc.), because the PDB is in flat format, i.e., text file. It is difficult to understand the relationships between structures and functions from the coordinate data of PDB alone. On the other hand, PROSITE database contains information on functional site of proteins. In order to extract the maximum functional information from three-dimensional structures, we developed a tool called 3DInSight, which consists of the data of PDB and PROSITE in a relational database system SYBASE as well as the tool set for retrieval and display of the information.

The 3DInSight system runs on the World Wide Web (WWW) and provides flexible search functions such as pattern matching of sequences, search by secondary structures, and so on. Furthermore, PDB structural information and PROSITE functional information are cross-linked. The three-dimensional structures and the functional sites can be displayed together by using VRML and more detailed document information can be referred by clicking on the interesting parts of the structures.

The PDB and related data are divided into 14 tables according to the semantic relationships between data items for flexible and efficient searching. For example, there are table of helix, table of sheet, table of properties of amino acids, table of distance information, table of motif position in chains, and so on. Moreover, several subsets of the database such as subsets of homology, subset of DNA, subset of DNA-Protein complex, etc., are created for specific users to search on different categories effectively.

The WWW interface of 3DInSight is of two levels: the form-based interface and SAL-based one. Using the form-based interface users can do most of regular searchings such as searching by keywords, searching by sequence pattern, etc. by simply filling out a form. On the other hand, the SAL-based interface provides powerful expression capability to search information under more complex conditions. However, users have to construct SAL statements by themselves.