PDB Retriever: A Simple and Integrated Browser for Protein Data Bank

Tadashi Mizunuma tmizunum@genes.nig.ac.jp
Sadahiko Misu smisu@genes.nig.ac.jp
Motonori Ota mota@genes.nig.ac.jp
Ken Nishikawa knishika@genes.nig.ac.jp

National Institute of Genetics
1111 Yata, Mishima, Shizuoka, Japan

As the structure determination technique of proteins has been developed, there are many entries accumulated in the structural data base (Protein Data Bank; PDB). The wealth of PDB offers important insights concerning to the principles of the protein architecture. However the content of PDB is a mixture of structures with different qualities (having missing residues, incomplete side chain coordinates, or different resolutions and B-factors) and with different redundancy (there are many mutant structures of lysozyme, etc). In order to make the most of these precious data, elaborate annotations for each structure are required. Especially in the application of the threading method, which provides a plausible candidate for the solution of the structure prediction, we should equip a structural library of high quality, i.e., an ensemble of independent and extensively assessed structures. So far, some researchers prepared their own libraries, but it is a laborious task to construct such a library.

To help this kind of task, we have developed a simple and integrated browser for PDB, named PDB Retriever. We construct two types of sequence data bases from PDB; PDBsh (PDB sequence taken from the header) and PDBsc (PDB sequence from the coordinate data). As their names indicate, PDBsh gathers the sequence data written in the header region of each entry, while PDBsc takes them attached to the coordinate data. Comparing PDBsh and PDBsc, we can count for the missing residues in the entry. Both the data base adopt the FASTA format of one comment line followed by a sequence in one letter representation. In the comment line, the numbers of the missing residues or incomplete side chains are indicated as well as the resolution, chain identifier, length, index of native sequence or mutants, etc. PDB Retriever offers an interface to access both of these data bases. To refer the comment line, we could retrieve the entries under several conditions. We could see the entry retrieved in the form of PDBsh and PDBsc, secondary structure assignment described in the header, and the DSSP output, PDB format itself and three dimensional structure by the RASMOL like Plug-in device (Chemscape). It is available from the WWW site of DDBJ (http://www.ddbj.nig.ac.jp).

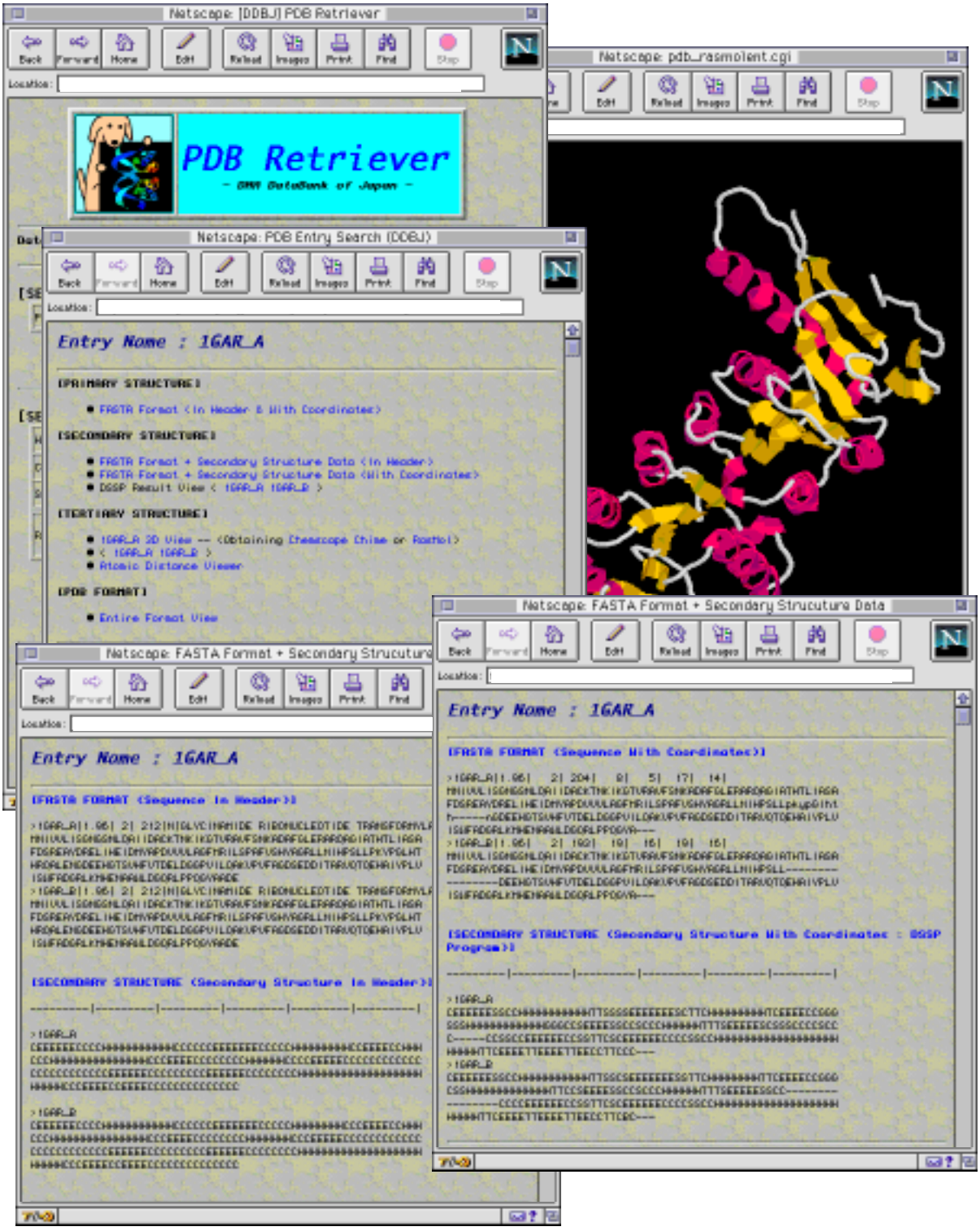

Figure 1: Results of the retrieval by the PDB Retriever (glysinamide ribonucleotide transformylase; 1GAR). This structure was determined with less than 2.0 Å resolution, but has many missing residues. In the page of PDBsc (lower right), they are indicated as "-" in the sequence of the fasta format, and the residues with incomplete side chain are represented by the "small" letters. To compare it with the corresponding PDBsh entry (lower left), we can recognize these dislocations at a glance.