

PDBML: the XML-based Database and its Applications

Haruki Nakamura^a, Nobutoshi Ito^a, Reiko Yamashita^b, Daron Standley^b, Arno Paehler^b, Atsuro Yoshihara^b, ^a*School of Biomedical Science, Tokyo Medical and Dental University*. ^b*PDBj, BIRD, Japan Science and Technology Agency*. ^c*Institute for Protein Research, Osaka University*. E-mail: harukin@protein.osaka-u.ac.jp

A canonical XML for PDB, PDBML [1], has been developed by PDBj (Protein Data Bank Japan) [2] and RCSB. Its structure is defined in XML schema (<http://deposit.pdb.org/pdbML/pdbx.xsd>) and all the contents in the PDBML files are now well validated.

We have built a native XML database with this new format and made it available to public through a web service called xPSSS (<http://www.pdbj.org/xpsss>). In addition to simple keyword search full XPath searches with the SOAP interface are also implemented for complicated searches and large-scale analyses. The contents of the database are also enhanced; additional data such as the biological and biochemical functions and the experimental details extracted from the literatures and other databases are included.

A few applications have also been developed around the database; (i) a molecular graphics viewer, jV, which can parse the PDBML data [3], (ii) electron density maps for evaluation of the structure quality, (iii) the sequence and structure neighbors defined by maximizing the number of equivalent residues (NER) rather than minimizing conventional RMSD [4].

[1] Westbrook J., et al., *Bioinformatics*, 2005, *in press*. [2] Berman H., et al., *Nat. Struct. Biol.*, 2003, **10**, 980. [3] Kinoshita K., Nakamura H., *Bioinformatics*, 2004, **20**, 1329. [4] Standley D., et al., *Proteins*, 2004, **57**, 381.

Keywords: databases, database manipulation, data representation