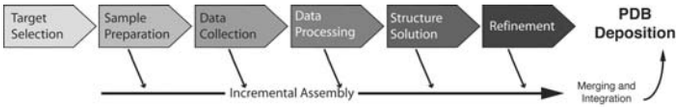

Fig. 5.5.3.5. Schematic diagram of a structure-determination data pipeline.

data file ready for deposition. The *PDB_EXTRACT* program also carrys out this merging operation.

Some steps in the structure-determination pipeline may not be driven by software. For instance, the details of protein production may be held in laboratory databases or within laboratory notebooks. A version of *ADIT* with a data view including all of the structural genomics data extensions has been created for entering these data. This *ADIT* tool can also be used to validate and check the completeness of the final data file.

5.5.4. Access

All of the software tools and libraries described in this chapter are distributed with full source under an open-source licence. Applications are also distributed in binary form for Intel/Linux, Sun/Solaris, SGI/IRIX and Dec Alpha platforms.

The RCSB/PDB is operated by Rutgers, The State University of New Jersey; the San Diego Supercomputer Center at the University of California, San Diego; and the Center for Advanced Research in Biotechnology of the National Institute of Standards and Technology. RCSB/PDB is supported by funds from the National Science Foundation (NSF), the National Institute of General Medical Sciences (NIGMS), the Department of Energy (DOE), the National Library of Medicine (NLM), the National Cancer Institute (NCI), the National Center for Research Resources (NCRR), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the National Institute of Neurological Disorders and Stroke (NINDS).

References

Berman, H. M., Henrick, K. & Nakamura, H. (2003). *Announcing the worldwide Protein Data Bank. Nature Struct. Biol.* **10**, 980.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *The Protein Data Bank. Nucleic Acids Res.* **28**, 235–242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–542.

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *Macromolecular Crystallographic Information File. Methods Enzymol.* **277**, 571–590.

Callaway, J., Cummings, M., Deroski, B., Esposito, P., Forman, A., Langdon, P., Libeson, M., McCarthy, J., Sikora, J., Xue, D., Abola, E., Bernstein, F., Manning, N., Shea, R., Stampf, D. & Sussman, J. (1996). *Protein Data Bank contents guide: Atomic coordinate entry format description.* Brookhaven National Laboratory, New York, USA. Available from http://www.wwpdb.org/documentation/PDB_format_1996.pdf.

Greer, D. S., Westbrook, J. D. & Bourne, P. E. (2002). *An ontology driven architecture for derived representations of macromolecular structure. Bioinformatics*, **18**, 1280–1281.

Kuller, A., Fleri, W., Bluhm, W. F., Smith, J. L., Westbrook, J. & Bourne, P. E. (2002). *A biologist's guide to synchrotron facilities: the BioSync web resource. Trends Biochem. Sci.* **27**, 213–215.

Schirripa, S. & Westbrook, J. D. (1996). *CIFOBJ. A class library of mmCIF access tools.* Reference guide. http://sw-tools.pdb.org/apps/CIFOBJ/cifobj/index.html.

Tosic, O. & Westbrook, J. D. (2000). *CIFParse. A library of access tools for mmCIF.* Reference guide. http://sw-tools.pdb.org/apps/CIFPARSE-OBJ/cifparse/index.html.

Westbrook, J. & Bourne, P. E. (2000). *STAR/mmCIF: an ontology for macromolecular structure. Bioinformatics*, **16**, 159–168.