

24.5. THE PROTEIN DATA BANK, 1999–

according to major consensus. Table 24.5.9.1 indicates how to access these resources.

24.5.10. Conclusion

These are exciting and challenging times to be responsible for the collection, curation and distribution of macromolecular structure data. Since the RCSB assumed responsibility for data deposition in February 1999, the number of depositions has averaged approximately 50 a week. However, with the advent of a number of structure genomics initiatives worldwide, this number is likely to increase. We estimate that the PDB, which at writing contains approximately 10 500 structures, could triple or quadruple in size over the next five years. This presents a challenge of timely distribution while maintaining high quality. The PDB's approach of using modern data-management practices should permit us to accommodate a large data influx.

The maintenance and further development of the PDB are community efforts. The willingness of others to share ideas, software and data provides a depth to the resource not obtainable otherwise. Some of these efforts are acknowledged below. New

input is constantly being sought and the PDB invites comments at any time by e-mail to info@rcsb.org.

Acknowledgements

The continuing support of Ken Breslauer (Rutgers), John Rumble (NIST) and Sid Karim (SDSC) is gratefully acknowledged. Current collaborators contributing to the future development of the PDB are the BioMagResBank, the Cambridge Crystallographic Data Centre, the HIV Protease Database Group, The Institute for Protein Research, Osaka University, The National Center for Biotechnology Information, the ReLiBase developers, the Swiss Institute for Bioinformatics/Glaxo and the European Bioinformatics Institute.

The cooperation of the BNL PDB staff is also gratefully acknowledged.

Parts of this chapter have appeared in *Nucleic Acids Research* (Berman *et al.*, 2000) and are reproduced here with permission of Oxford University Press.

This work is supported by grants from the National Science Foundation, the Office of Biology and Environmental Research at the Department of Energy, and two units of the National Institutes of Health: the National Institute of General Medical Sciences and the National Library of Medicine.

References

24.1

- Abola, E. E. (1994). *PDB-SHELL*. Available at ftp://pdb.bmc.uu.se/pub/databases/pdb/pdb_software/pdbshell/.
- Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Protein Data Bank*. In *Crystallographic databases – information content, software systems, scientific applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107–132. Bonn: International Union of Crystallography.
- Abola, E. E., Sussman, J. L., Prilusky, J. & Manning, N. O. (1997). *Protein Data Bank archives of three-dimensional macromolecular structures*. *Methods Enzymol.* **277**, 556–571.
- Bairoch, A. (1994). *The ENZYME data bank*. *Nucleic Acids Res.* **22**, 3626–3627.
- Bairoch, A. & Boeckmann, B. (1994). *The SWISS-PROT protein sequence data bank: current status*. *Nucleic Acids Res.* **22**, 3578–3580.
- Baker, E. N., Blundell, T. L., Vijayan, M., Dodson, E., Dodson, G., Gilliland, G. L. & Sussman, J. L. (1996). *Crystallographic data deposition*. *Nature (London)*, **379**, 202.
- Bloom, F. E. (1998). *Policy change*. *Science*, **281**, 175.
- Cambell, P. (1998). *New policy for structure data*. *Nature (London)*, **394**, 105.
- Commission on Biological Macromolecules (2000). *Guidelines for the deposition and release of macromolecular coordinate and experimental data*. *Acta Cryst.* **D56**, 2.
- Editorial Board (1998). *New policy on release of structural coordinates*. *Proc. Natl Acad. Sci. USA*, **95**, iii.
- Jiang, J., Abola, E. & Sussman, J. L. (1999). *Deposition of structure factors at the Protein Data Bank*. *Acta Cryst.* **D55**, 4.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody*. *Nature (London)*, **393**, 648–659.
- Lin, D., Manning, N. O., Jiang, J., Abola, E. E., Stampf, D., Prilusky, J. & Sussman, J. L. (2000). *AutoDep: a web-based system for deposition and validation of macromolecular structural information*. *Acta Cryst.* **D56**, 828–841.
- Madden, D. R., Garboczi, D. N. & Wiley, D. C. (1993). *The antigenic identity of peptide–MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2*. *Cell*, **75**, 693–708.

- Peitsch, M. C., Stampf, D. R., Wells, T. N. C. & Sussman, J. L. (1995). *The Swiss 3D-image collection and Brookhaven Protein Data Bank browser on the World-Wide Web*. *Trends Biochem. Sci.* **20**, 82–84.
- Phillips, D. C. (1971). *Protein crystallography 1971: coming of age*. *Cold Spring Harbor Symp. Quant. Biol.* pp. 589–592.
- Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A. & Sodroski, J. (1998). *A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding*. *Science*, **280**, 1949–1953.
- Sayle, R. A. & Milner-White, E. J. (1995). *RASMOL: biomolecular graphics for all*. *Trends Biochem. Sci.* **20**, 374–376.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees*. *Science*, **253**, 1001–1007.
- Seavey, B. R., Farr, E. A., Westler, W. M. & Markley, J. L. (1991). *A relational database for sequence-specific protein NMR data*. *J. Biomol. Nucl. Magn. Reson.* **1**, 217–236.
- Stampf, D. R., Felder, C. E. & Sussman, J. L. (1995). *PDBBrowser – a graphics interface to the Brookhaven Protein Data Bank*. *Nature (London)*, **374**, 572–574.
- Sussman, J. L. (1997). *Bridging the gap*. *Nature Struct. Biol.* **4**, 517.
- Sussman, J. L. (1998). *Protein Data Bank deposits*. *Science*, **282**, 1991.
- Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). *Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules*. *Acta Cryst.* **D54**, 1078–1084.

24.2

- Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information*. *Acta Cryst.* **B35**, 2331–2339.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). *The Nucleic Acid Database – a comprehensive relational database of three-dimensional structures of nucleic acids*. *Biophys. J.* **63**, 751–759.