

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.

24.2 (cont.)

- 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., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 535–542.
- Bourne, P., Berman, H. M., Watenpaugh, K., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *The macromolecular Crystallographic Information File (mmCIF)*. *Methods Enzymol.* **277**, 571–590.
- Brünger, A. T. (1992). *X-PLOR*. Version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven, CT, USA.
- Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. & Berman, H. M. (1996). *Geometric parameters in nucleic acids: nitrogenous bases*. *J. Am. Chem. Soc.* **118**, 509–518.
- Feng, Z., Hsieh, S.-H., Gelbin, A. & Westbrook, J. (1998). *MAXIT: macromolecular exchange and input tool*. NDB-120. Rutgers University, New Brunswick, NJ, USA.
- Feng, Z., Westbrook, J. & Berman, H. M. (1998). *NUCheck*. NDB-407. Rutgers University, New Brunswick, NJ, USA.
- Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K. & Berman, H. M. (1996). *Geometric parameters in nucleic acids: sugar and phosphate constituents*. *J. Am. Chem. Soc.* **118**, 519–528.
- Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). *The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G: the effect of base pair reversal*. *J. Biol. Chem.* **266**, 8861–8883.
- Lavery, R. & Sklenar, H. (1989). *Defining the structure of irregular nucleic acids: conventions and principles*. *J. Biomol. Struct. Dyn.* **6**, 655–667.
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). *New parameters for the refinement of nucleic acid-containing structures*. *Acta Cryst.* **D52**, 57–64.
- Sayle, R. & Milner-White, E. J. (1995). *RasMol: biomolecular graphics for all*. *Trends Biochem. Sci.* **20**, 374.
- Schneider, B., Neidle, S. & Berman, H. M. (1997). *Conformations of the sugar-phosphate backbone in helical DNA crystal structures*. *Biopolymers*, **42**, 113–124.
- Scott, W. G., Finch, J. T. & Klug, A. (1995). *The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage*. *Cell*, **81**, 991–1002.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model*. *Acta Cryst.* **D55**, 191–205.
- Westbrook, J. (1998). *AutoDep input tool*. NDB-406. Rutgers University, New Brunswick, NJ, USA.
- Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *IsoStar: a library of information about nonbonded interactions*. *J. Comput.-Aided Mol. Des.* **11**, 525–537.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *The crystallographic information file (CIF): a new standard archive file for crystallography*. *Acta Cryst.* **A47**, 655–685.
- Hayes, I. C. & Stone, A. J. (1984). *An intermolecular perturbation theory for the region of moderate overlap*. *J. Mol. Phys.* **53**, 83–105.
- Kennard, O. & Allen, F. H. (1993). *3D search and research using the Cambridge Structural Database*. *Chem. Des. Autom. News*, **8**, 1, 31–37.
- RCSB (2000). The Protein Data Bank. Research Collaboratory for Structural Bioinformatics, Department of Chemistry, Rutgers University, Piscataway, NJ, USA (<http://www.rcsb.org>).
- Sayle, R. (1996). *The RASMOL visualiser*. Glaxo Wellcome Research, Stevenage, Hertfordshire, England.

24.4

- Ariyoshi, M., Vassilyev, D. G., Iwasaki, H., Fujishima, A., Shinagawa, H. & Morikawa, K. (1994). *Preliminary crystallographic study of Escherichia coli RuvC protein. An endonuclease specific for Holliday junctions*. *J. Mol. Biol.* **241**, 281–282.
- Athanasiadis, A. & Kokkinidis, M. (1991). *Purification, crystallization and preliminary X-ray diffraction studies of the PvuII endonuclease*. *J. Mol. Biol.* **222**, 451–453.
- Balendiran, K., Bonventre, J., Knott, R., Jack, W., Benner, J., Schildkraut, I. & Anderson, J. E. (1994). *Expression, purification, and crystallization of restriction endonuclease PvuII with DNA containing its recognition site*. *Proteins*, **19**, 77–79.
- Bannikova, G. E., Blagova, E. V., Dementiev, A. A., Morgunova, E. Yu., Mikchailov, A. M., Shlyapnikov, S. V., Varlamov, V. P. & Vainshtein, B. K. (1991). *Two isoforms of Serratia marcescens nuclease. Crystallization and preliminary X-ray investigation of the enzyme*. *Biochem. Int.* **24**, 813–822.
- 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.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography*. New York: Academic Press.
- Bozic, D., Grazulis, S., Siksnys, V. & Huber, R. (1996). *Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 Å resolution*. *J. Mol. Biol.* **255**, 176–186.
- Carter, C. W. Jr & Carter, C. W. (1979). *Protein crystallization using incomplete factorial experiments*. *J. Biol. Chem.* **254**, 12219–12223.
- Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Screening and optimization strategies for macromolecular crystal growth*. *Acta Cryst.* **D50**, 414–423.
- D'Arcy, A., Brown, R. S., Zabeau, M., van Resandt, R. W. & Winkler, F. K. (1985). *Purification and crystallization of the EcoRV restriction endonuclease*. *J. Biol. Chem.* **260**, 1987–1990.
- Gilliland, G. L. (1988). *A biological macromolecule crystallization database: a basis for a crystallization strategy*. *J. Cryst. Growth*, **90**, 51–59.
- Gilliland, G. L. & Bickham, D. (1990). *The Biological Macromolecule Crystallization Database: a tool to assist the development of crystallization strategies*. *Methods Companion Methods Enzymol.* **1**, 6–11.
- Gilliland, G. L. & Davies, D. R. (1984). *Protein crystallization: the growth of large-scale single crystals*. *Methods Enzymol.* **104**, 370–381.
- Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological Macromolecule Crystallization Database, version 3.0: new features, data and the NASA archive for protein crystal growth data*. *Acta Cryst.* **D50**, 408–413.
- Gilliland, G. L., Tung, M. & Ladner, J. (1996). *The Biological Macromolecule Crystallization Database and NASA Protein*