24.2. The Nucleic Acid Database (NDB)

By H. M. Berman, Z. Feng, B. Schneider, J. Westbrook and C. Zardecki

24.2.1. Introduction

The Nucleic Acid Database (NDB) (Berman *et al.*, 1992) was established in 1991 as a resource for specialists in the field of nucleic acid structure. Its purpose was to gather all of the structural information about nucleic acids that had been obtained from X-ray crystallographic experiments and to organize them in such a way that it would be easy to retrieve the coordinates, the information about the experimental conditions used to derive these coordinates, and the structural information that could be derived from these coordinates. Since many NDB users are not crystallographers, the information provided by the database has been presented in such a way as to maximize its utility for various types of modelling and structure prediction.

Since the NDB was founded, many new technologies have presented new challenges and opportunities. The emergence of the World Wide Web has allowed for the creative and powerful dissemination and collection of data and information. The development of a standard interchange format for handling crystallographic data, the macromolecular Crystallographic Information File (mmCIF; Bourne *et al.*, 1997), has made it possible to ensure the integrity and consistency of the data in the archive. The NDB has used these resources to provide both a relational database and an archive of information to a global community.

Table 24.2.2.1. *The information content of the NDB*

(*a*) Primary experimental information stored in the NDB.

Structure summary – descriptor; NDB, PDB and CSD names; coordinate availability; modifications, mismatches and drugs (yes/no)

Structural description – sequence; structure type; descriptions about modifications, mismatches and drugs; description of asymmetric and biological units

Citation – authors, title, journal, volume, pages, year

Crystal data – cell dimensions; space group

Data-collection description – radiation source and wavelength; data-collection device; temperature; resolution range; total and unique number of reflections

Crystallization description – method; temperature; pH value; solution composition

Refinement information – method; program; number of reflections used for refinement; data cutoff; resolution range; R factor; refinement of temperature factors and occupancies

Coordinate information – atomic coordinates, occupancies and temperature factors for asymmetric unit; coordinates for symmetry-related strands; coordinates for unit cell; symmetry-related coordinates; orthogonal or fractional coordinates

(*b*) Derivative information stored in the NDB.

Distances – chemical bond lengths; virtual bonds (involving phosphorus atoms)

Torsions – backbone and side-chain torsion angles; pseudorotational parameters

Angles – valence bond angles, virtual angles (involving phosphorus atoms)

Base morphology – parameters calculated by different algorithms

Nonbonded contacts

Valence geometry r.m.s. deviations from small-molecule standards

Sequence pattern statistics

24.2.2. Information content of the NDB

Structures available in the NDB include RNA and DNA oligonucleotides with two or more bases either alone or complexed with ligands, natural nucleic acids such as tRNA, and protein–nucleic acid complexes. The archive stores both primary and derived information about the structures. The primary data include the crystallographic coordinate data, structure factors and information about the experiments used to determine the structures, such as crystallization information, data collection and refinement statistics. Derived information, such as valence geometry, torsion angles, base-morphology parameters and intermolecular contacts, is calculated and stored in the database. Database entries are further annotated to include information about the overall structural features, including conformational classes, special structural features, biological functions and crystal-packing classifications. Table 24.2.2.1 summarizes the information content of the NDB.

24.2.3. Data processing

Data processing includes data collection, integrity checking and validation of the entries. Once processing is completed, the data are entered into the database. This is accomplished using the integrated system that is illustrated in Fig. 24.2.3.1.

Structures are entered electronically into the NDB after they have been deposited directly by the experimentalist or by the NDB annotators, who scan the literature and the Protein Data Bank (PDB; Bernstein *et al.*, 1977; Berman *et al.*, 2000). The coordinate data may be deposited in any PDB format or in mmCIF format. The entries are transformed into mmCIF format and then annotated using a web-based tool (Westbrook, 1998). This tool operates on top
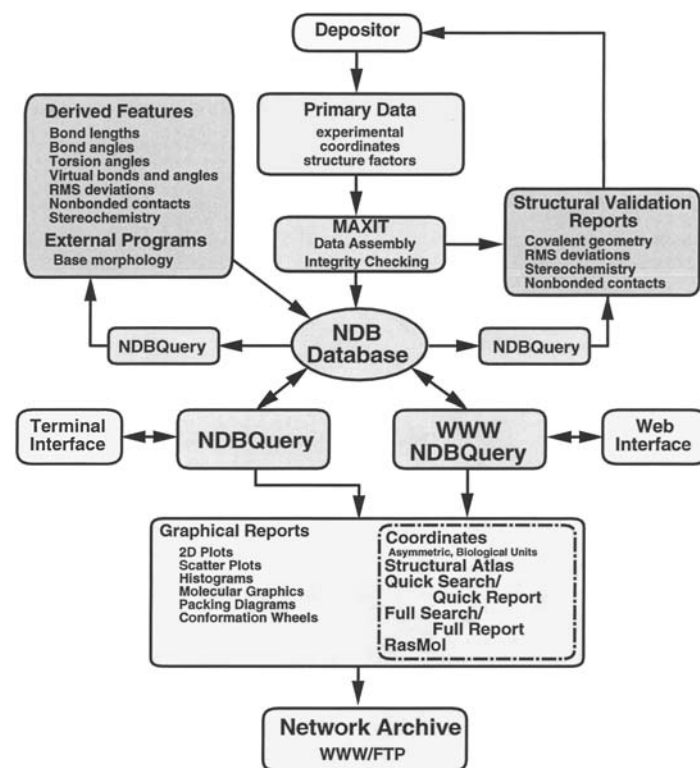

Fig. 24.2.3.1. Flow chart showing the organization of the Nucleic Acid Database Project. The core of this integrated system is the database.
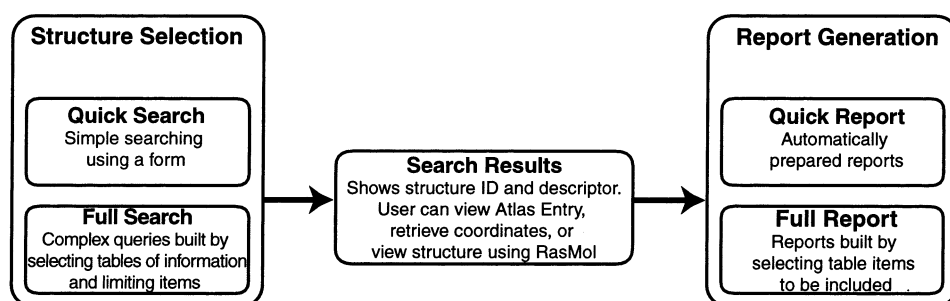
657

Fig. 24.2.4.1. Flow chart demonstrating the two steps involved in searching the NDB: structure selection and report generation.

of the mmCIF dictionary (Bourne *et al.*, 1997) and is used to incorporate experimental information to create a fully populated mmCIF format file. In the next stage of data processing, a program called *MAXIT* (*Macromolecular Exchange and Input Tool*; Feng, Hsieh *et al.*, 1998) checks and corrects atom numbering and ordering as well as the correspondence between the PDB SEQRES record and the residue names in the coordinate files. Once these integrity checks are completed, the structures are validated using a variety of programs.

NUCheck (Feng, Westbrook & Berman, 1998) verifies valence geometry, torsion angles, intermolecular contacts and the chiral centres of the sugars and phosphates. The dictionaries used for checking the structures were developed by the NDB project from analyses (Clowney *et al.*, 1996, Gelbin *et al.*, 1996) of high-resolution small-molecule structures from the Cambridge Structural Database (CSD; Allen *et al.*, 1979). The torsion-angle ranges were derived from an analysis of high-resolution nucleic acid structures (Schneider *et al.*, 1997). One important outgrowth of these
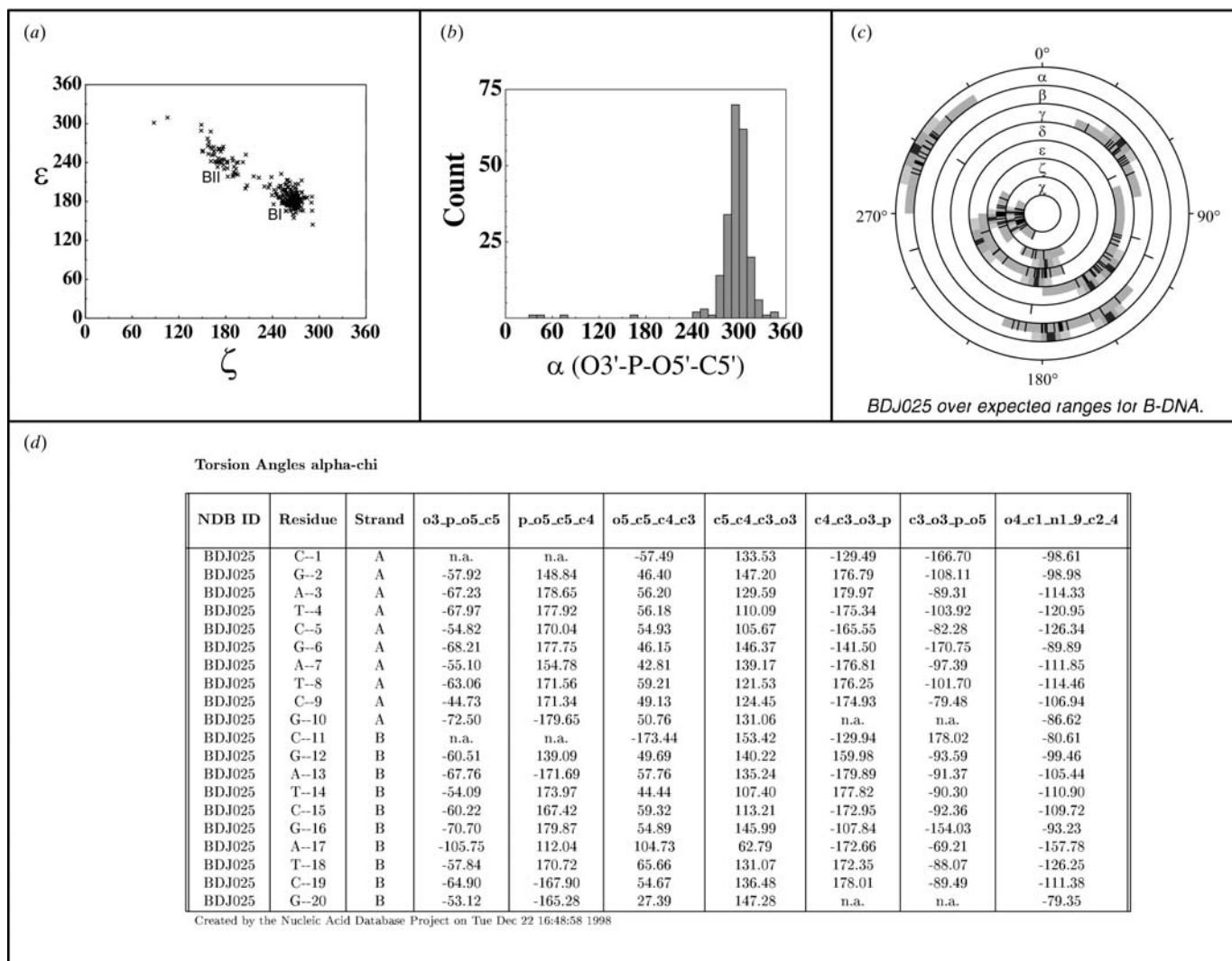
Fig. 24.2.4.2. Examples of reports generated from the NDB about torsion angles. (*a*) A scattergram showing the relationship of ε (C4′—C3′—O3′—P) *versus* ζ (C3′—O3′—P—O5′). The two clusters, BI and BII, are labelled. (*b*) A histogram for α (O3′—P—O5′—C5′) for all B-DNA. (*c*) A conformation wheel showing the torsion angles for structure BDJ025 (Grzeskowiak *et al.*, 1991) over the average values for all B-DNA. (*d*) A torsion-angle report for BDJ025.

Table 24.2.5.1. *Quick reports available from the NDB*

Report name	Contents
NDB status	Processing status information
Cell dimensions	Crystallographic cell constants
Primary citation	Primary bibliographic citations
Structure identifier	Identifiers, descriptor, coordinate availability
Sequence	Sequence
Nucleic acid sequence	Nucleic acid sequence only
Protein sequence	Protein sequence only
Refinement information	R factor, resolution and number of reflections used in refinement
Nucleic acid backbone torsions (NDB)	Sugar–phosphate backbone torsion angles using NDB residue numbers
Nucleic acid backbone torsions (PDB)	Sugar–phosphate backbone torsion angles using PDB residue numbers
Base-pair parameters (global)	Global base-pair parameters calculated using *Curves* 5.1 (Lavery & Sklenar, 1989)
Base-pair step parameters (local)	Local base-pair step parameters calculated using *Curves* 5.1
Groove dimensions	Groove dimensions using Stoffer & Lavery definitions from *Curves* 5.1

validation projects was the creation of the force constants and restraints that are now in common use for crystallographic refinement of nucleic acid structures (Parkinson *et al.*, 1996). The program *SFCHECK* (Vaguine *et al.*, 1999) is used to validate the model against the structure-factor data. The R factor and resolution are verified and the residue-based features are examined with this program. Once an entry has been processed satisfactorily, it is entered into the database.

24.2.4. The database

The core of the NDB project is a relational database in which all of the primary and derived data items are organized into tables. At present, there are over 90 tables in the NDB, with each table containing five to 20 data items. These tables contain both experimental and derived information. Example tables include: the citation table, which contains all the items that are present in literature references; the cell_dimension table, which contains all items related to crystal data; and the refine_parameters table, which contains the items that describe the refinement statistics.

Interaction with the database is a two-step process (Fig. 24.2.4.1). In the first step, the user defines the selection criteria by combining different database items. As an example, the user could select all B-DNA structures whose resolution is better than 2.0 Å, whose R factor is better than 0.17, and which were determined by the authors Dickerson, Kennard, or Rich. Once the structures that meet the constraint criteria have been selected, reports may be written using a combination of table items. For any set of chosen structures, a large variety of reports may be created. For the example set of structures given above, a crystal-data report or a backbone torsion-angle report can be easily generated, or the user could write a report that lists the twist values for all CG steps together with statistics, including mean, median and range of values. The constraints used for the reports do not have to be the same as those used to select the structures. Some examples of reports from the NDB are given in Fig. 24.2.4.2.

24.2.5. Data distribution

Data are made available *via* a variety of mechanisms, such as ftp and the World Wide Web. Coordinate files, reports, software programs and other resources are available *via* the ftp server (ndbserver.rutgers.edu). In addition to links to the ftp server, the web server provides a variety of methods for querying the NDB and

accessing reports prepared from the database (http://ndbserver. rutgers.edu/).

24.2.5.1. *Archives*

The NDB archives, a section of the web site, contain a large variety of information and tables useful for researchers. Prepared reports about the structure identifiers, citations, cell dimensions and structure summaries are available and are sorted according to structure type. The dictionaries of standard geometries of nucleic acids as well as parameter files for *X-PLOR* (Brünger, 1992) are also available. The archives section links to the ftp server, providing coordinates for the asymmetric unit and biological units in PDB and mmCIF formats, structure-factor files, and coordinates for nucleic acid structures determined by NMR.

24.2.5.2. *Atlas*

A very popular and useful report is the NDB Atlas report page. An Atlas page contains summary, crystallographic and experimental information, a molecular view of the biological unit and a crystal-packing picture for a particular structure. Atlas pages are created directly from the NDB database (Fig. 24.2.5.1). The Atlas entries for all structures in the database are organized by structure type on the NDB web site.

24.2.5.3. *NDB searches*

A web interface was designed to make the query capabilities of the NDB as widely accessible as possible. To highlight the special features of NDB, the interface operates in two modes. In the quick search/quick report mode, several items, including structure ID, author, classification and special features, can be limited either by entering text in a box or by selecting an option from the pull-down menu. Any combination of these items may be used to constrain the structure selection. If none are used, the entire database will be selected. After selecting 'Execute Selection', the user will be presented with a list of structure IDs and descriptors that match the desired conditions. Several viewing options for each structure in this list are possible. These include retrieving the coordinate files in either mmCIF or PDB format, retrieving the coordinates for the biological unit, viewing the structure with *RasMol* (Sayle & Milner-White, 1995), or viewing an NDB Atlas page.

Preformatted quick reports can then be generated for the structures in this results list. The user selects a report from a list of 13 report options (Table 24.2.5.1), and the report is created

NDB ID: URX035

Features

RNA, RNA hammerhead ribozyme, catalytic RNA, loop

Compound name

RNA hammerhead ribozyme

Sequence in asymmetric unit

- Chain A: GUGGUCUGAUGAGGCC
- Chain B: GGCCGAAACUCGUAAGAGUCACCAC
- Chain C: GUGGUCUGAUGAGGCC
- Chain D: GGCCGAAACUCGUAAGAGUCACCAC

Citation

W.G. Scott, J.T. Finch, A. Klug
The Crystal Structure of an All–RNA Hammerhead Ribozyme: A Proposed Mechanism for RNA Catalytic Cleavage
Cell, **81**, pp. 991–1002, 1995.

Space group

$P3_121$

Cell constants

```
   A = 64.980      B = 64.980      C = 138.140 (Angstroms)
alpha = 90.00   beta = 90.00   gamma = 120.00 (degrees)
```

Crystallization conditions

- Method: vapor diffusion, sitting drop
- Solution: nucleic acid, water, PEG 6000, glycerol, spermine, magnesium acetate, ammonium acetate, ammonium cacodylate
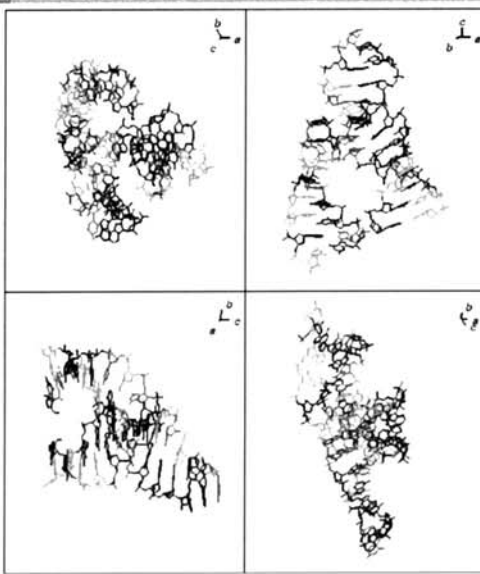
Refinement

The structure was refined using the X–PLOR 3.1 program. The R value is 25.1 for 6193 reflections in the resolution range 15.0 to 3.1 Å with Fobs > 2.0 sigma(Fobs).

Coordinates

The coordinates for the asymmetric unit of this structure are stored in the NDB archive.
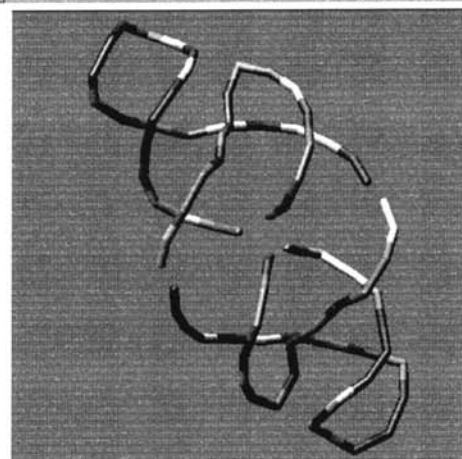
Four views of URX035 (biological unit 1)
The nucleic acid is colored by residue. C, G, A, T, U, and I are yellow, green, red, blue, cyan and dark green, respectively. Other residues and drugs are black.

©1995,1996 The Nucleic Acid Database Project
Rutgers, The State University of New Jersey
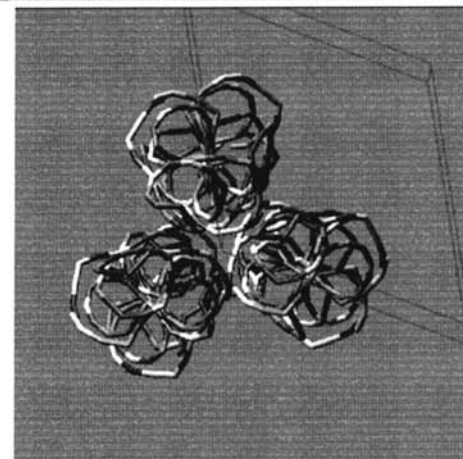
The asymmetric unit of URX035
The nucleic acid is colored by residue.

©1995,1996 The Nucleic Acid Database Project
Rutgers, The State University of New Jersey

A crystal packing picture of URX035

©1995,1996 The Nucleic Acid Database Project
Rutgers, The State University of New Jersey

Fig. 24.2.5.1. NDB Atlas page for URX035 (Scott *et al.*, 1995) that highlights structural information that is contained in the database and provides images of the biological unit, asymmetric unit and crystal packing of the structure.

Top-left panel — NDB Quick Report Display (base-pair parameters):

Structure_ID	pair_number	pair_name	sx	sy	sz	prop_twist	buckle	opening
UHX026	1	GC	-0.010	-0.020	0.080	-16.740	-1.130	2.840
UHX026	2	GC	-0.100	-0.340	0.620	-18.830	-17.680	-2.140
UHX026	3	CG	-0.410	-0.520	0.200	-11.110	-10.940	-2.870
UHX026	4	GC	0.260	-0.140	0.240	-15.710	-5.080	0.680
UHX026	5	AT	-0.120	-0.520	0.100	-14.360	0.690	-2.140
UHX026	6	CG	0.300	0.040	0.080	-6.670	1.240	4.540
UHX026	7	CG	1.140	0.590	-0.080	-13.380	2.900	10.170
UHX026	8	AT	-0.170	-0.190	0.180	-12.260	19.100	-36.110
UHX026	9	CG	0.290	0.460	-0.030	-19.390	-4.210	4.710
UHX026	10	CG	-0.170	-0.200	-0.320	-11.620	2.780	-0.730
UHX026	11	GC	-0.150	-0.430	0.200	-8.310	-1.020	-0.850
UHX026	12	UA	0.920	0.510	0.630	-2.350	1.060	10.930
UHX026	13	GC	0.080	0.030	-0.210	-1.420	-5.620	0.500
UHX026	14	GC	0.040	0.180	-0.240	-13.280	0.220	6.120
UHX026	15	CG	0.270	-0.090	-0.080	-3.320	-1.840	2.930
UHX026	16	CG	0.070	-0.160	-0.050	6.500	2.520	1.630
URX035	1	GC	-0.740	-0.090	-0.050	-12.870	-2.580	-5.930
URX035	2	UA	-0.070	1.300	-0.700	-27.570	12.310	11.520
URX035	3	GC	0.780	0.030	-0.760	-9.080	-2.300	-2.220
URX035	4	GC	-0.170	0.720	-0.760	-3.970	-2.580	17.400
URX035	5	UA	0.060	0.480	-0.640	0.080	6.130	8.840
URX035	6	GC	-1.000	-0.060	0.570	-7.600	-7.780	-1.700
URX035	7	GC	1.230	-0.200	0.350	10.300	7.420	8.010

Top-right panel — NDB Quick Report Display (nucleic acid backbone torsions):

Structure_ID	chain_id	residue_no	residue_name	o3_p_o5_c5	p_o5_c5_c4	o5_c5_c4_c3	c5_c
UHX026	A	1	G	n.a.	-171.101	47.737	8
UHX026	A	2	G	-58.093	179.061	50.907	8
UHX026	A	3	C	-63.706	-178.431	53.546	7
UHX026	A	4	G	-67.894	157.393	69.264	7
UHX026	A	5	A	-65.069	170.547	57.717	8
UHX026	A	6	C	-75.770	162.552	72.037	8
UHX026	A	7	C	-67.265	170.302	60.858	8
UHX026	A	8	C	-61.620	176.075	51.317	8
UHX026	A	9	U	-68.667	173.155	55.497	8
UHX026	A	10	G	147.843	-174.729	64.071	9
UHX026	A	11	A	-78.532	170.064	56.171	9
UHX026	A	12	U	-107.914	171.613	48.388	8
UHX026	A	13	G	-62.051	-178.004	51.875	15
UHX026	A	14	A	-84.608	171.925	41.037	8
UHX026	A	15	G	-69.001	165.307	58.355	7
UHX026	A	16	G	-68.871	176.296	63.239	8
UHX026	A	17	C	-80.633	-168.209	55.190	8
UHX026	A	18	C	-56.489	-175.186	34.742	8
UHX026	A	19	G	-87.854	-178.156	45.743	7
UHX026	A	20	A	157.115	167.168	57.207	9
UHX026	A	21	A	-66.629	154.809	65.115	8
UHX026	A	22	A	-93.254	175.458	62.711	9

Bottom-left panel — NDB Quick Report Display (citation report):

Structure_ID	Authors	Title	Publication	V
PD0001	K.Luger, A.W.Maeder, R.K.Richmond, D.F.Sargent, T.J.Richmond	Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution	Nature	3
PD0002	K.E.Flick, M.S.Jurica, R.J.Monnat Junior, B.L.Stoddard	DNA Binding and Cleavage by the Nuclear Intron-Encoded Homing Endonuclease I- PpoI	Nature	3
PD0003	F.Guo, D.N.Gopaul, G.D.Van Duyne	Structure of Cre Recombinase Complexed with DNA in a Site-Specific Recombination Synapse	Nature	3
PD0004	A.White, X.Ding, J.C.Vanderspek, J.R.Murphy, D.Ringe	Structure of the Metal-Ion-Activated Diphtheria Toxin Repressor/Tox Operator Complex	Nature	3
PD0005	J.Ding, K.Das, H.Yu, S.G.Sarafianos, A.D.Clark Junior, A.Jacobo-Molina, C.Tantillo, S.H.Hughes, E.Arnold	Structure and Functional Implications of the Polymerase Active Site Region in a Complex of HIV-1 RT with a Double-Stranded DNA and an Antibody Fab Fragment at 2.8 Angstroms Resolution	J.Mol.Biol.	2
PD0006	J.R.Horton, H.G.Nastri, P.D.Riggs, X.Cheng	Asp34 of PvuII Endonuclease is Directly Involved in DNA Minor Groove Recognition and Indirectly Involved in Catalysis	J.Mol.Biol.	2
PD0007	E.Fraenkel, C.O.Pabo	Comparison of X-Ray and NMR Structures for the Antennapedia Homeodomain-DNA Complex	Nat.Struct.Biol.	5
	Q.Zhao			

Bottom-right panel — NDB Quick Report Display (structure identifier report):

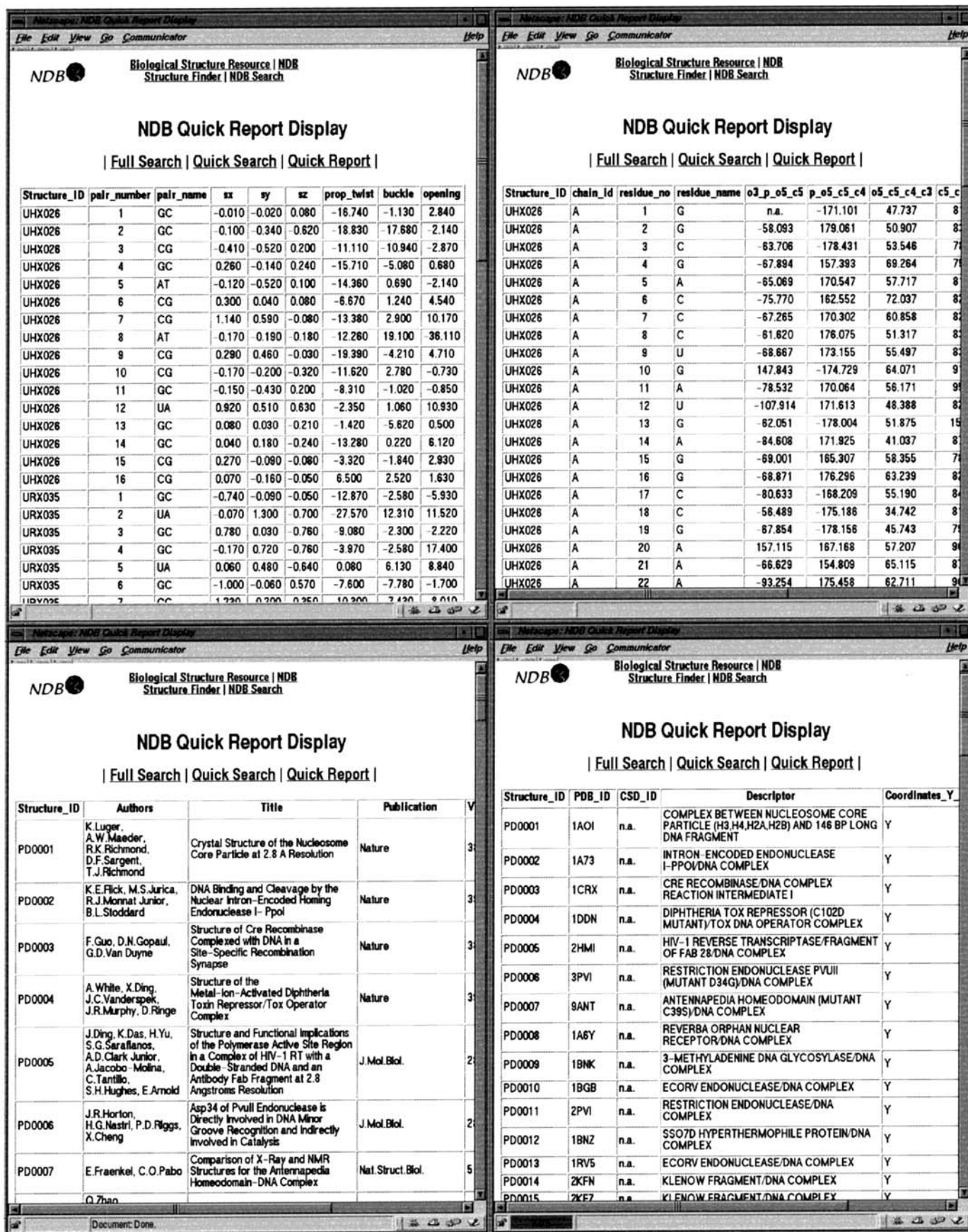| Structure_ID	PDB_ID	CSD_ID	Descriptor	Coordinates_Y
PD0001	1AOI	n.a.	COMPLEX BETWEEN NUCLEOSOME CORE PARTICLE (H3,H4,H2A,H2B) AND 146 BP LONG DNA FRAGMENT	Y
PD0002	1A73	n.a.	INTRON-ENCODED ENDONUCLEASE I-PPOI/DNA COMPLEX	Y
PD0003	1CRX	n.a.	CRE RECOMBINASE/DNA COMPLEX REACTION INTERMEDIATE I	Y
PD0004	1DDN	n.a.	DIPHTHERIA TOX REPRESSOR (C102D MUTANT/TOX DNA OPERATOR COMPLEX	Y
PD0005	2HMI	n.a.	HIV-1 REVERSE TRANSCRIPTASE/FRAGMENT OF FAB 28/DNA COMPLEX	Y
PD0006	3PVI	n.a.	RESTRICTION ENDONUCLEASE PVUII (MUTANT D34G)/DNA COMPLEX	Y
PD0007	9ANT	n.a.	ANTENNAPEDIA HOMEODOMAIN (MUTANT C39S)/DNA COMPLEX	Y
PD0008	1A6Y	n.a.	REVERBA ORPHAN NUCLEAR RECEPTOR/DNA COMPLEX	Y
PD0009	1BNK	n.a.	3-METHYLADENINE DNA GLYCOSYLASE/DNA COMPLEX	Y
PD0010	1BGB	n.a.	ECORV ENDONUCLEASE/DNA COMPLEX	Y
PD0011	2PVI	n.a.	RESTRICTION ENDONUCLEASE/DNA COMPLEX	Y
PD0012	1BNZ	n.a.	SSO7D HYPERTHERMOPHILE PROTEIN/DNA COMPLEX	Y
PD0013	1RV5	n.a.	ECORV ENDONUCLEASE/DNA COMPLEX	Y
PD0014	2KFN	n.a.	KLENOW FRAGMENT/DNA COMPLEX	Y
PD0015	2KFZ	n.a.	KLENOW FRAGMENT/DNA COMPLEX	Y

Fig. 24.2.5.2. Examples of quick reports. Clockwise from top left: base-pair parameters [global; calculated using *Curves* 5.1 (Lavery & Sklenar, 1989)] report for ribozyme structures; nucleic acid backbone torsions (NDB) report for ribozyme structures; structure identifier report for protein–DNA structures; citation report for protein–DNA structures.

automatically. Multiple reports can be easily generated. These reports are particularly convenient for being able to produce reports quickly based on derived features, such as torsion angles and base morphology (Fig. 24.2.5.2).

In the full search/full report mode, it is possible to access most of the tables in the NDB to build more complex queries. Instead of limiting items that are listed on a single page, the user builds a search by selecting the tables and then the items that contain the desired features. These queries can use Boolean and logical operators to make complex queries.

After selecting structures using the full search, a variety of reports can be written. The report columns are selected from a variety of database tables, and then the full report is automatically generated. Multiple reports can be generated for the same group of selected structures; for example, reports on crystallization, base modification, or a combination of these reports can be generated for a particular group of structures.

24.2.5.4. *Mirror sites*

The NDB is based at Rutgers University (http://ndbserver. rutgers.edu/) and is currently mirrored at three other sites: the Institute of Cancer Research (ICR) in London, England (http://www.ndb.icr.ac.uk), the San Diego Supercomputer Center in San Diego, USA (http://ndb.sdsc.edu/NDB/) and the Structural Biology Centre in Tsukuba, Japan (http://ndbserver.nibh.go.jp/NDB/). These mirror sites are updated daily, are fully synchronous, and contain the ftp directories, the web site and the full database.

24.2.6. Outreach

The NDB has worked closely with the community of researchers to ensure that their needs are met. A newsletter is published electronically four times a year and provides information about the newest features of the system. Questions and very complex queries can be handled by the staff in response to user requests *via* e-mail to ndbadmin@ndbserver.rutgers.edu.

Acknowledgements

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.

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.* D**56**, 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.* D**55**, 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.* D**56**, 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.* D**54**, 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.* B**35**, 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.* D**52**, 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.* D**55**, 191–205.

Westbrook, J. (1998). *AutoDep input tool*. NDB-406. Rutgers University, New Brunswick, NJ, USA.

24.3

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.

Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). *The development of versions 3 and 4 of the Cambridge Structural Database system*. *J. Chem. Inf. Comput. Sci.* **31**, 187–204.

Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1987). *Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds*. *J. Chem. Soc. Perkin Trans.* 2, pp. S1–S19.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1972). *The Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 535–547.

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.* A**47**, 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., Vassylyev, 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.* D**50**, 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.* D**50**, 408–413.

Gilliland, G. L., Tung, M. & Ladner, J. (1996). *The Biological Macromolecule Crystallization Database and NASA Protein*