The Nucleic Acid Database

A comprehensive relational database of three-dimensional structures of nucleic acids

Helen M. Berman, Wilma K. Olson, David L. Beveridge,* John Westbrook, Anke Gelbin, Tamas Demeny, Shu-Hsin Hsieh, A. R. Srinivasan, and Bohdan Schneider

Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903; and
*Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459 USA

INTRODUCTION

The last several years have seen an information explosion in molecular biology as the number of available nucleic acid sequences has grown at a very rapid rate (1, 2). An extensive effort has already been made to develop programs to analyze the literal patterns in these sequences, and computer packages now exist to compare sequences, to look for restriction sites, to search for palindromes, etc. (3). At the same time, information about the three-dimensional structures of nucleic acids has also grown (see references 4 and 5, and reviews on some topics of nucleic acid three-dimensional structure can be found in *Current Opinion in Structural Biology,* 1 (2) (1991), and 2 (3) (1992)). It is clear that if one hopes to truly understand the properties of nucleic acids and the mechanisms by which they are recognized by proteins, one must understand the details of their fine structures and be able to couple the sequences with the structural data.

Although nucleic acid structures are contained within the Protein Data Bank (PDB) (6) and the Cambridge Structural Database (7), no single database is devoted entirely to nucleic acid structures. Towards that end we have initiated the Nucleic Acid Database Project, whose goals are to continually develop computational tools for the extraction of information about nucleic acid structures and to distribute both the software and the data. This report describes our progress towards achieving these goals.

CONTENTS OF THE NUCLEIC ACID DATABASE

As of May 1992, the Nucleic Acid Database (NDB) has stored information on 192 DNA and RNA crystal structures, including 130 entries having atomic coordinates. Table 1 outlines the holdings. Coordinates are derived from both the Protein Data Bank and the literature.

Using procedures described below, a relational database that contains tables of primary and derivative information has been developed. The overall scheme for maintenance and distribution of the database, detailed in a flowchart, is shown in Fig. 1.

The *primary* experimental information stored in NDB is shown in Table 2 *a*. These include atomic coordinates, crystal data, bibliographic references, crystallization conditions, data collection methods, refinement information, and various structural descriptions. The coordinates of atoms other than those of the nucleic acid are also stored and designated as to whether they belong to drugs, ions, or water, or if they modify the DNA/RNA molecule.

The *derivative* information (Table 2 *b*) is calculated from the atomic coordinates and includes chemical bond lengths and angles, torsion angles, virtual bond lengths and angles involving phosphorus atoms, and base morphology parameters calculated according to various algorithms (8–12).

SOFTWARE

A collection of applications has been developed to facilitate many of the tasks involved in building the Nucleic Acid Database. These tasks include the encoding of primary experimental data, the organization of primary data and derived structural parameters in a relational database management system, and the simplified selection and display of all of the stored information.

Data preparation

The experimental data in the NDB have been collected from a variety of sources. The most common means of data accumulation have been the manual entry of data from published literature, the collection of information from one of the standard crystallographic archive file types (13, 14), the extraction of information from a crystallographic application program, and the collection of data from some combination of these sources.

All of the primary experimental data collected for the Nucleic Acid Database are first encoded in an ASCII NDB format file (15). This format is a superset of the Brookhaven PDB format (13). The direct entry of experimental data from the literature is performed using an interactive editor (16) which is sensitive to this NDB file format. The editor program automatically enforces the syntactic conventions of the NDB format and can consequently be used productively by inexperienced users.

To permit the extraction of information from existing crystallographic file formats, a suite of filter programs (17) has been developed to perform format translations. These programs translate between the NDB format and the Brookhaven PDB format, and between the NDB format and the Crystallographic Information File (CIF) (14) format. New terms specific to nucleic acid crystallography have been locally added to the existing CIF dictionary specification. An electronic template based on this expanded CIF dictionary has been developed to permit

0006-3495/92/09/751/09 $2.00

TABLE 1 **NDB holdings as of May, 1992**

Structure type	Number	Coordinates available	Coordinates in preparation
A-DNA	30	11	5
A-RNA	6	5	—
B-DNA	34	19	1
Z-DNA	30	16	1
Unusual DNA	9	7	—
Unusual RNA	8	5	—
DNA/RNA hybrid	2	—	—
DNA-intercalated drug	26	19	—
DNA-groove binding drug	12	11	—
DNA-other binding drug	6	4	—
RNA-intercalated drug	16	15	—
RNA-other binding drug	3	3	—
tRNA	10	8	—
Total	192	123	7

direct submission of all information currently collected for the NDB database in a single format.

The preparation of experimental crystallographic data for submission to a crystallographic archive involves time consuming and tedious reformatting of the output of refinement programs in a manner required by the archive. To simplify this process for the commonly used NUCLSQ refinement package (18), a program (19) has been developed to extract coordinate and refinement information from the output files of NUCLSQ and to reformat this information in the NDB format. Extraction and reformatting tools for other refinement programs are under development.

Database management

Information stored in the NDB format files is reorganized into the table structure of a relational database management system. The program that performs this task (20) creates the appropriate ANSI SQL (21) commands to build the schema for all of the primary data in the database. Although a proprietary relational database system (22) has been employed for the preliminary phase of this project, no use has been made of any feature of the database system beyond the features described in the ANSI standard.[1] Hence, the database can be easily implemented under any standard database management system.

The primary means of interaction with information stored in the relational database is through an interactive menu-driven application program (23). This application permits a user to construct a query for the selection of stored structures using any primary or derived structural features as selection criteria. Having created a structure selection, the user may then construct a formatted report incorporating any primary or derived structural features for this selection, or in the case of angular structural features, the results may be displayed in the form of conformation rings (24). The program permits the user to save selection and report queries in a natural language format for subsequent reuse or modification.

In addition to providing a user interface to the NDB database, this program also manages the calculation of derived structural parameters. Simple structural features which can be defined unambiguously are calculated internally. The symmetry related strand is also generated

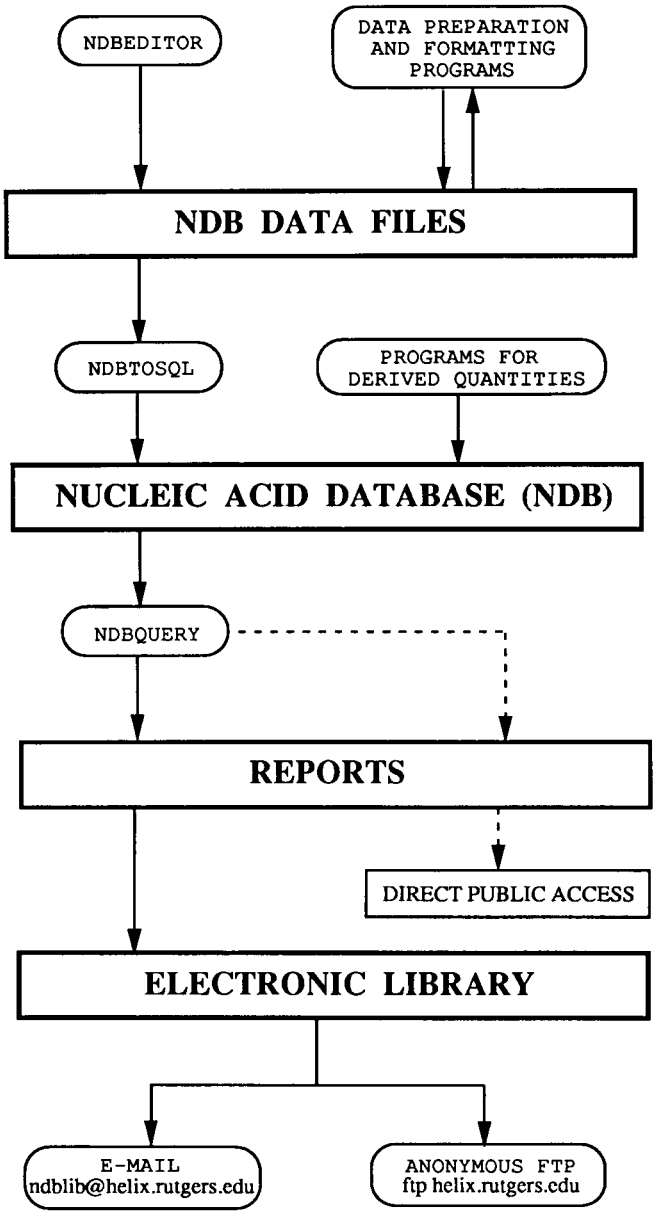

FIGURE 1 Flow chart illustrating the Nucleic Acid Database maintenance and distribution.

[1] There are certain features concerning the distribution between multiple platforms of a database maintained by the SYBASE DBMS which may not be generic to all such products. These features are managed by the SYBASE DBMS server in a manner completely transparent to an application program. We have been able to take considerable advantage of these distributed features without introducing DBMS or platform specific code.

TABLE 2a **Primary experimental information stored in NDB**

Structure Summary	Citation
descriptor	authors
NDB-, PDB-, and CSD names	title
coordinates available (yes/no)	journal
modifiers (yes/no)	volume
mismatches (yes/no)	pages
drugs (yes/no)	year
Structural Description	Crystal Data
sequence	cell dimensions (a, b, c, α, β, γ)
structure type (A/B/Z/unusual)	space group
description of modifiers of base, phosphate, and sugar	Data Collection Description
description of base mismatch	source of radiation
name and binding type of drug	data collection device
description of base pairing	radiation
description of contents of asymmetric unit	temperature
molecular weight of asymmetric unit	total resolution range
	number of reflections at specified sigma value
Crystallization Description	Coordinate Information
method	atomic coordinates, occupancies, and temperature factors grouped by DNA/RNA atoms, modifiers, drugs, ions, solvents
temperature	
pH value	
composition of solvents	coordinates for symmetry related second strand
Refinement Information	matrix for transformation of orthogonal to fractional coordinates
method	
program	
number of reflections used for refinement	
R-factor	
refinement of temperature factors and occupancies	

TABLE 2b **Derivative information stored in NDB**

Distances	Angles
chemical bond lengths	valence bond angles
virtual bonds (involving phosphorus atoms)	virtual angles (involving phosphorus atoms)
Torsions	torsion angles
backbone and side chain torsion angles	Base Morphology
	parameters calculated by different algorithms

for those structures that contain only one strand of the double helix in the asymmetric unit. Other structural features, for which there is demonstrated algorithmic sensitivity, are calculated using existing programs (8–12) minimally modified to facilitate the automated processing of all of the structures in the database.

A number of other application programs have also been interfaced to our database software. These include conformational analysis by Dials and Windows (11, 12) and a locally developed three-dimensional molecular display package (25). A filter program designed to produce Kinemage (26) format graphics files for each structure in the database is also under development.

INFORMATION RETRIEVAL FROM THE NUCLEIC ACID DATABASE

The creation of reports has two independent parts: first, the user defines structures that are searched for, and then, in the second part, the database is asked to release specified information on the chosen structures. In the second part of the query, any information can be released on the selected structures and the *output selection* is chosen without regard to the *constraints* by which the included structures were picked.

Structure selection

The *constraints* are chosen by selecting any of the stored properties of structures, and defining a condition this property must meet. The condition may be any numeric operator, such as equal to, not equal to, less than, greater than, and so on for numerical data, and string searches for character fields. Several conditions may be joined together by using logical AND and OR operators. (See below for sample query.)

Report preparation

The *output selection* involves selecting any of the stored properties of structures for inclusion in the reports generated. Then the format of the report can be defined by choosing font styles, sizes, and data arrangements in the columns. The reports can also include simple statistics such as means and standard deviations.

Coordinate files are written out in the PDB format (13). For these files, either the orthogonal or the crystallographic fractional coordinate system may be chosen. For structures with only one strand per asymmetric unit, the symmetry-generated second strand may also be chosen as an option. Coordinate files do not have to include all the atoms of the selected structure but the output atoms can be restricted to certain atom types, residues, strands, etc.

The following two examples demonstrate the two steps involved in creating reports with the program **ndbquery** (23).

Example 1

(Fig. 2.) Give (*a*) descriptors, (*b*) full citations, and (*c*) cell dimensions for each B-DNA containing the residue sequence 'CGCG'. Include only structures without base modifiers, mismatches, and drugs.

(a) Structure selection

The desired B-DNAs have to be chosen by specifying the following constraints.

Property	*Operator*	*Operand*	*Logical*
structure_type	=	**B**	**AND**
residue_sequence	**contains**	**CGCG**	**AND**
base_modifier (y/n)	=	**no**	**AND**
mismatch (y/n)	=	**no**	**AND**
drug (y/n)	=	**no**	

(a) **Descriptors**

NDB ID	Descriptor
BDFP24	5'-D(RP*GP(S)*CP*GP(S)*CP*GP(S)*C)-3'
BDL001	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3', 290 K
BDL002	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3', 16 K
BDL005	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3', 290 K, USING ANISOTROPIC THERMAL MOTION MODEL
BDL020	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3', RE-REFINEMENT
BDL021	5'-D(*CP*GP*CP*GP*AP*AP*AP*AP*CP*GP*CP*G)-3'/ 5'-D(*CP*GP*CP*GP*TP*T)-3'+5'-D(*TP*TP*CP*GP*CP*G)-3'

Page 1 created by the Nucleic Acid Database Project on Thu Apr 30 13:43:55 1992

(b) **Citations**

NDB ID	Citation
BDFP24	W.B.T.Cruse, S.A.Salisbury, T.Brown, R.Cosstick, F.Eckstein, O.Kennard Chiral Phosphorothioate Analogues of B-DNA: The Crystal Structure of RP-d(Gp(S)CpGp(S)CpGp(S)C) *J.Mol.Biol.*, **192**, 891-905, 1986.
BDL001	H.R.Drew, R.M.Wing, T.Takano, C.Broka, S.Tanaka, K.Itakura, R.E.Dickerson Structure of a B-DNA Dodecamer. Conformation and Dynamics *Proc.Natl.Acad.Sci.U.S.A.*, **78**, 2179-2183, 1981.
BDL002	H.R.Drew, S.Samson, R.E.Dickerson Structure of a B-DNA Dodecamer at 16 Kelvin *Proc.Natl.Acad.Sci.U.S.A.*, **79**, 4040-4044, 1982.
BDL005	S.R.Holbrook, R.E.Dickerson, S.-H.Kim Anisotropic Thermal-Parameter Refinement of the DNA Dodecamer CGC-GAATTCGCG by the Segmented Rigid-Body Method *Acta Cryst.*, **B41**, 255-262, 1985.
BDL020	E.Westhof Re-Refinement of the B-Dodecamer d(CGCGAATTCGCG) with a Comparative Analysis of the Solvent in It and in the Z-Hexamer d(5BrCG5BrCG5BrCG) *J.Biomol.Struct.Dyn.*, **5**, 581-600, 1987.
BDL021	J.Aymami, M.Coll, G.A.Van Der Marel, J.H.Van Boom, A.H.-J.Wang, A.Rich Molecular Structure of Nicked DNA: A Substrate for DNA Repair Enzymes *Proc.Natl.Acad.Sci.U.S.A.*, **87**, 2526-2530, 1990.

Page 1 created by the Nucleic Acid Database Project on Tue May 5 14:50:19 1992

(c) **Sequences, cell dimensions, space groups**

NDB ID	Sequence	a b c	α β γ	Space Group
BDFP24	GCGCGC	34.9, 39.1, 20.6	90.0, 90.0, 90.0	P 21 21 21
BDL001	CGCGAATTCGCG	24.9, 40.4, 66.2	90.0, 90.0, 90.0	P 21 21 21
BDL002	CGCGAATTCGCG	23.4, 39.3, 65.3	90.0, 90.0, 90.0	P 21 21 21
BDL005	CGCGAATTCGCG	24.9, 40.4, 66.2	90.0, 90.0, 90.0	P 21 21 21
BDL020	CGCGAATTCGCG	24.9, 40.4, 66.2	90.0, 90.0, 90.0	P 21 21 21
BDL021	CGCGAAAACGCG	26.0, 44.0, 66.6	90.0, 90.0, 90.0	P 21 21 21

Page 1 created by the Nucleic Acid Database Project on Thu Apr 30 13:45:19 1992

FIGURE 2 Three reports describing different properties of B-DNAs which contain the residue sequence 'CGCG' and do not have base modifiers, mismatches, or drugs. (*a*) Descriptors. (*b*) Citations. (*c*) Sequences, cell dimensions, and space groups.

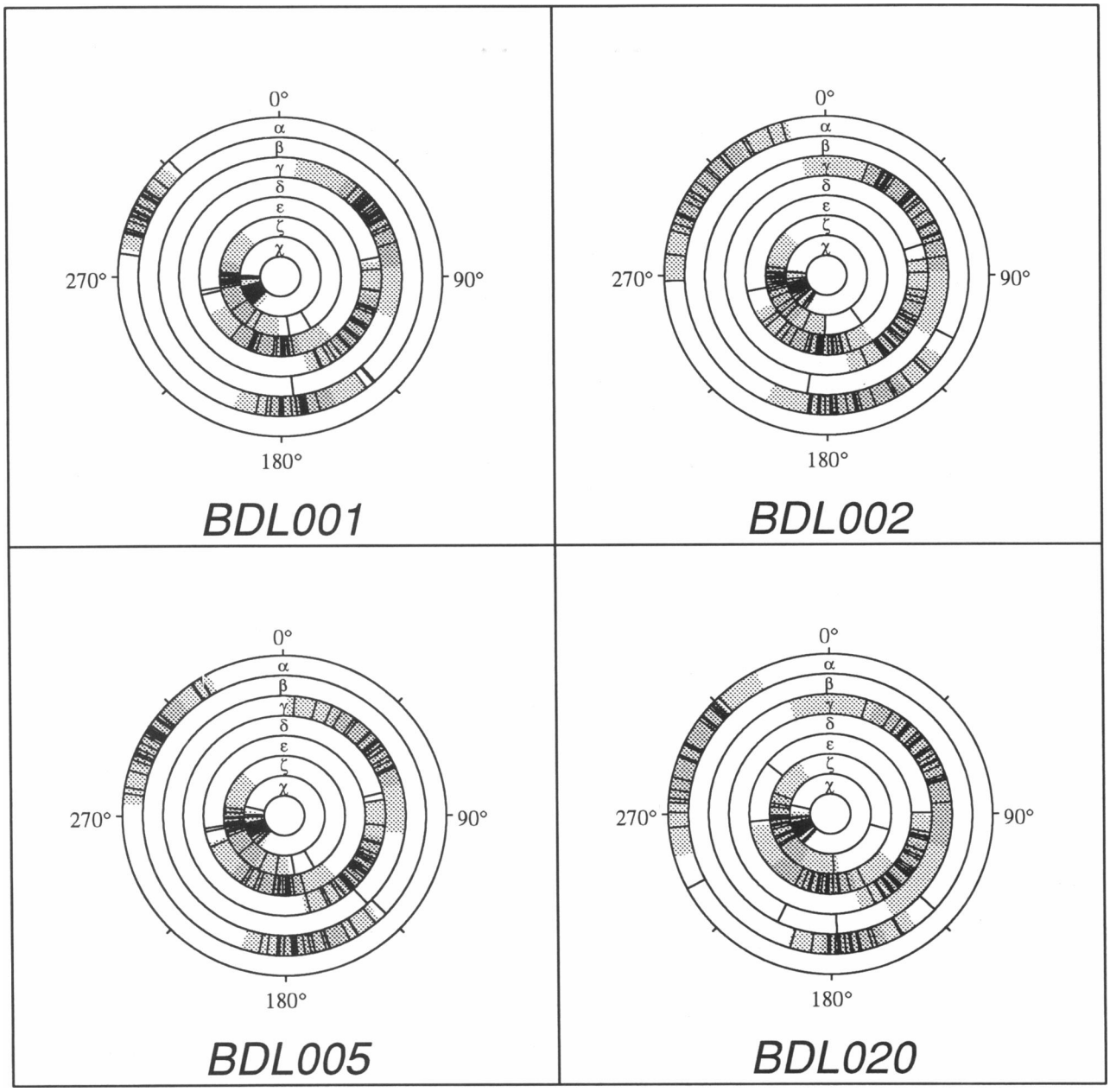

FIGURE 3 Conformation rings depicting the occurrences of backbone and side group torsion angles {P-O5′ α (ω), O5′-C5′ β (ϕ), C5′-C4′ γ (ψ), C4′-C3′ δ (ψ'), C3′-O3′ ϵ (ϕ'), O3′-P ζ (ω'), C1′-N χ} corresponding to BDL001, BDL002, BDL005, and BDL020. Each ring is labeled from 0° to 360° and the occurrence of a particular torsion angle in the corresponding DNA structure is noted by a radial line in the appropriate angular location. Multiple lines in a given ring represent the various values of a particular torsion of the selected structure. Dark and light shadings are, respectively, used to highlight states within one and two standard deviation units of the mean. More details regarding the conformation rings are described in reference 24 and the NDB newsletter of April, 1992. (See the section on distribution and the index for information to obtain the newsletter.)

Thus, each constraint is defined by choosing the four attributes: property, operator, operand, and logical operator. First, the property is selected from the menus. To find the property of interest easily, the menus are arranged in a hierarchical fashion. Then, an operator is chosen for the selected property from the next menu which lists all possible operators. The operand, as shown in the third column, is typed from the keyboard after the operator is chosen. The logical connective is also specified if there are multiple constraints; in most cases, the logical connective is AND.

In this example, the first property, structure type, is restricted to the value *B*, specifying that only structures of type *B* are to be chosen. Next, the residue sequence is defined to contain CGCG. The operator **contains** is an example of an available string search operator. The last three properties can take on the values of **yes** and **no** only, and specify the applicability of the given informa-

Citations, descriptors, refinement information

(a)

NDB ID	Descriptor and Citation	No. of Reflections	Resolution (Å)	R Factor (%)
ZDF028	5'-D(*CP*GP*CP*GP*CP*G)-3', COPPER(II) CHLORIDE SOAKED T.F.Kagawa, B.H.Geierstanger, A.H.-J.Wang, P.S.Ho Covalent Modification of Guanine Bases in Double Stranded DNA: The 1.2 Angstroms Z-DNA Structure of d(CGCGCG) in the Presence of CuCl2 *J.Biol.Chem.*, **266**, 20175-20184, 1991.	4107	1.2	19.8
ZDFB10	5'-D(*(M)CP*GP*UP*AP*(M)CP*G)-3', COPPER(II) CHLORIDE SOAKED B.H.Geierstanger, T.F.Kagawa, S.-L.Chen, G.J.Quigley, P.S.Ho Base Specific Binding of Copper(II) to Z-DNA: The 1.3 Angstroms Single Crystal Structure of d(m5CGUAm5CG) in the Presence of CuCl2 *J.Biol.Chem.*, **266**, 20185-20191, 1991.	2587	1.3	20.9
ZDFB11	5'-D(*CP*(NH2)AP*CP*GP*TP*G)-3' M.Coll, A.H.-J.Wang, G.A.Van Der Marel, J.H.Van Boom, A.Rich Crystal Structure of a Z-DNA Fragment Containing Thymine/2-Aminoadenine Base Pairs *J.Biomol.Struct.Dyn.*, **4**, 157-172, 1986.	3739	1.3	21.7
ZDFB21	5'-D(*CP*GP*CP*(M)GP*CP*G)-3' S.L.Ginell, S.Kuzmich, R.A.Jones, H.M.Berman Crystal and Molecular Structure of a DNA Duplex Containing the Carcinogenic Lesion O6-Methylguanine *Biochemistry*, **29**, 10461-10465, 1990.	1217	1.9	19.0
ZDFB24	5'-D(*(M)CP*GP*UP*AP*(M)CP*G)-3' G.Zhou, P.S.Ho Stabilization of Z-DNA by Demethylation of Thymine Bases: 1.3 Angstroms Single-Crystal Structure of d(m5CGUAm5CG) *Biochemistry*, **29**, 7229-7236, 1990.	2870	1.3	20.8
ZDFB31	5'-D(*CP*GP*UP*(NH2)AP*CP*G)-3' B. Schneider, S. L. Ginell, R. Jones, B. Gaffney, H. M. Berman The Crystal and Molecular Structure of a DNA Fragment Containing a 2-Amino Adenine Modification: d(CGUA'CG)2 *PREPRINT*, *, *-*, *.	4588	1.3	13.8

Page 1 created by the Nucleic Acid Database Project on Thu Apr 30 14:21:46 1992

Torsion angle (ζ = C3'-O3'-P-O5') for residues 3, 4, 5, 9, 10, 11 (in degrees)

(b)

NDB ID	Res 3	Res 4	Res 5	Res 9	Res 10	Res 11
ZDF028	79.2	69.3	73.9	67.2	336.9	49.6
ZDFB10	73.0	28.5	81.8	81.5	310.3	75.1
ZDFB11	77.4	54.2	82.0	62.8	293.7	75.5
ZDFB21	78.5	49.3	58.0	76.1	282.5	75.9
ZDFB24	90.8	43.0	65.2	80.5	306.4	94.6
ZDFB31	76.0	64.8	76.4	74.2	296.3	69.1

Page 1 created by the Nucleic Acid Database Project on Thu Apr 30 13:49:21 1992

FIGURE 4 Two reports describing Z-DNAs with torsion angle (ζ = C3'-O3'-P-O5') at residue 4 between 20° and 120°. Only structures which are chemically modified, have no drugs, and have solvent coordinates deposited are included. (*a*) Citations, descriptors, and refinement information. (*b*) Torsion angle ζ at residues 3, 4, 5, 9, 10, and 11 (in degrees).

tion for the structure. These constraints pick six structures out of the database. The next step is to define the output information to be reported on these structures.

(b) Report preparation

Once the desired structures are found, any kind of report can be written about them. Selection of output fields is done by simply going through the menus listing stored properties, and picking out those of interest. The format of the report is defined after the output columns are chosen. Fig. 2 contains three possible outputs with different information about the selected structures. In Fig. 2 *a*, the descriptor for each structure is picked, in order to identify the chosen structures uniquely. Fig. 2 *b* gives the complete citations for the same structures. Here, more sophisticated formatting options have been defined by the user. A third table, reporting on the sequence, cell dimensions, and space group is given in Fig. 2 *c*.

Conformation rings (24) describing the occurrences of the backbone torsion angles and the glycosyl rotation are an optional output format of the **ndbquery** program. The corresponding rings for four of the six selected B-DNAs are shown in Fig. 3.

Example 2

(Fig. 4.) Find Z-DNA structures that have values of the torsion angle ζ at residue 4 between 20° and 120° (the angle being defined as the C3′4-O3′4-P5-O5′5 atom sequence). Select only structures that have no drugs, are chemically modified, and have solvent coordinates deposited. For these structures, give the complete reference, and also list the *R* factor, resolution, and number of reflections used in refinement as well as the values of ζ at residues 3, 4, 5, 9, 10, and 11.

(a) Structure selection

In order to find the desired Z DNAs, the user defines the following constraints.

Property	*Operator*	*Operand*	*Logical*
structure_type	=	**Z**	**AND**
base_modifier (y/n)	=	**yes**	**AND**
solvent_coordinates_deposited (y/n)	=	**yes**	**AND**
drug (y/n)	=	**no**	**AND**
torsion_angle_residue_number	=	**4**	**AND**
torsion_angle_C3′_O3′_P5_O5′	≤	**120**	**AND**
torsion_angle_C3′_O3′_P5_O5′	≥	**20**	

In this example, the torsion angle residue number is restricted to 4, in order to restrict the torsion values only for this residue. This does not mean that only these torsion angles can be added to the output. The restriction is only applicable to the structure search. The last two constraints define the desired range of the selected torsion angle.

These constraints pick six structures out of the current Nucleic Acid Database holdings. The next step is to define the relevant information to be reported on these structures.

(b) Report preparation

For these structures, two reports are shown. Fig. 4 *a* gives the descriptors and the complete citations, along with the crystallographic resolution, number of reflections used in the refinement, and residual index *R*. The values of the torsion angle ζ at residues 3, 4, 5, 9, 10, and 11 are reported in Fig. 4 *b*.

DISTRIBUTION

Currently, tables produced by the Nucleic Acid Database Project are placed into an electronic library for public access, maintained at NDB. The tables are updated regularly to include information about all newly released structures. The regularly released NDB newsletter contains information on the current contents of the library and detailed access instructions. Subscription to the newsletter may be initiated by sending the following electronic mail message:

To: **ndblib@helix.rutgers.edu**
Subject: **subscribe**

The subject line must contain this one word, as it is read automatically by the mail server. All subscribers will receive regularly updated information on the NDB project. For more information on communicating with the library, refer to the Appendix. At present, the library contains the following types of files. (*a*) Summary reports on all structures, describing structural properties (for example, structure types, cell dimensions, and modifications) and complete references. (*b*) Coordinate files for all structures where coordinates have been deposited. (*c*) Tables of derived information such as torsion angles and base morphology. (*d*) A regularly released *NDB Newsletter* containing updated information about the library.

Users with a SYBASE license may access the database as it is maintained at Rutgers University. In the future, access to the database will be available via the INTERNET.

APPENDIX

Using the nucleic acid database electronic library

The currently available information is in the Nucleic Acid Database Library (ndblib). This electronic library is divided into the following directories:

newsletter, reports_ascii, reports_ps, coordinates.

Files from any of these directories are available to the general public, and can be downloaded either via the mail server or the file transfer program (anonymous ftp).

The directory *newsletter* contains the current and all previous newsletters. *Reports_ascii* and *reports_ps* contain the summary reports about the structures in ascii and PostScript versions, respectively. The directory *coordinates* contains coordinate files for the structures.

To obtain files through the mail server, the user must send a mail message to the library at this address:

ndblib@helix.rutgers.edu

The mail server will read the subject line of the message sent to this address, and automatically reply by posting the requested file to the user's directory. Therefore, the subject line has to be in the standard format to unambiguously identify the file desired. No other message needs to be sent.

The subject line must be in this format:

Subject: send ⟨filename⟩ from ⟨directory⟩

where ⟨directory⟩ is replaced by any one of the directories listed above, and ⟨filename⟩ is replaced by the name of a file in that directory. The word **Subject** is normally a system prompt. Refer to your system administrator for details on how to fill in the subject line on your mailing system. To obtain listings of files in each directory, the file called *index* should be requested.

For example, to get a listing of all newsletters the subject line must look like this:

Subject: send index from newsletter

To get the current newsletter, the subject line must look like this:

Subject: send news.apr92 from newsletter

where *news.apr92* is the name of the file containing the newsletter of April, 1992. This information can be obtained from the index as well.

To get the list of reports available in PostScript format, send the following subject line:

Subject: send index from reports_ps

This index describes all reports available from the reports_ps directory. To obtain one of these reports (for example, the one called *names.ps*) send this line:

Subject: send names.ps from reports_ps

To obtain coordinate files, the same kind of subject line can be sent, with *coordinates* as directory name, and the name of the structure as file name. Again, the *index* file from coordinates will give a listing of all structures in this directory. The following line will request the coordinate file for the structure with NDB ID *addb01:*

Subject: send addb01 from coordinates

The Nucleic Acid Database Library can also be accessed using anonymous ftp. In this case, the user can copy files directly from the library's directories into his/her own directory.

To start the ftp program, the following command has to be issued:

ftp helix.rutgers.edu

Once a proper connection to helix is established, the following command has to be issued in order to log in:

user anonymous

Then, ftp will request a password, to which your own name should be typed.

After getting into the proper session, it is necessary to type **cd pub** as the first command, and then to enter one of the following commands to choose the directory with the files to be transferred:

cd newsletter, cd reports_ascii, cd reports_ps, cd coordinates

(The command **ls** will list the files in the current directory.)

To obtain any of the files in the current directory, issue a get statement, of the form:

get ⟨filename⟩

A session with ftp ends with the **quit** command. For more detailed instructions, please refer to any of our newsletters.

Inquiries about the file server should be directed to:

ndbadmin@helix.rutgers.edu

The continued interest and advice from Frank Manion of the Fox Chase Cancer Center is very much appreciated. We thank Donna Horton for her help and assistance in many aspects of the present effort.

This project is supported by grant DIR N012772 from the National Science Foundation.

Received for publication and in final form 15 May 1992.

REFERENCES

1. Bilofsky, H. S., C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, C. D. Swindell, and S. C. Tung. 1986. The GenBank genetic sequence databank. *Nucleic Acids Res.* 14:1–4.
2. Hamm, G. H., and G. N. Cameron. 1986. The EMBL data library. *Nucleic Acids Res.* 14:5–10.
3. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.
4. Dickerson, R. E., H. K. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka. 1982. The anatomy of A-, B-, and Z-DNA. *Science (Wash. DC).* 216:475–485.
5. Shakked, Z., and D. Rabinovich. 1986. The effect of the base sequence on the fine structure of the DNA double helix. *Prog. Biophys. Molec. Biol.* 47:159–195.
6. Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.
7. Allen, F. H., S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson. 1979. The Cambridge crystallographic data centre: computer-based search, retrieval, analysis and display of information. *Acta Cryst.* B35:2331–2339.

8. Babcock, M. S., and W. K. Olson. 1992. A new program for the analysis of nucleic acid structure: implications for nucleic acid structure interpretation. *In* Computation of Biomolecular Structures: Achievements, Problems, and Perspectives. Springer-Verlag, Heidelberg. In press.
9. Grzeskowiak, K., K. Yanagi, G. G. Prive, and R. E. Dickerson. 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.
10. Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* 6:63–91, 655–667.
11. Ravishanker, G. 1991. MD Toolchest. Wesleyan University, Middletown, CT.
12. Ravishanker, G., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: "curves", dials, and windows. *J. Biomol. Struct. Dynam.* 6:669–699.
13. Protein Data Bank. 1992. Atomic Coordinate and Bibliographic Entry Format Description. Brookhaven National Laboratory, Brookhaven, NY.
14. Hall, S. R., A. H. Allen, and I. D. Brown. 1991. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Cryst.* A47:655–685.
15. Berman, H. M., A. Gelbin, J. Westbrook, and T. Demeny. 1991. The Nucleic Acid Database File Format. Rutgers University, New Brunswick, NJ.
16. Westbrook, J. 1991. Ndbeditor. An Interactive Editor for Nucleic Acid Crystallographic Data. Rutgers University, New Brunswick, NJ.
17. Hsieh, S.-H. 1992. Ndbfilter. A Suite of Translator Programs for Nucleic Acid Database Crystallographic Archive File Format. Rutgers University, New Brunswick, NJ.
18. Westhof, E., P. Dumas, and D. Moras. 1985. Crystallographic Refinement of Yeast Aspartic Acid Transfer RNA. *J. Mol. Biol.* 184:119–145.
19. Hsieh, S.-H. 1992. Ndbextract. A Suite of Programs to Extract and Reformat Crystallographic Refinement Program Output. Rutgers University, New Brunswick, NJ.
20. Westbrook, J., and S.-H. Hsieh. 1992. Ndbtosql. A Program to Generate ANSI SQL Tables from NDB Format Files. Rutgers University, New Brunswick, NJ.
21. American National Standards Institute. 1986. Database Language SQL, Document ANSI X3.135–1986.
22. Sybase, Inc. 1989. SYBASE SQL Server Release 4.0. Emeryville, CA.
23. Westbrook, J., T. Demeny, and S.-H. Hsieh. 1992. Ndbquery. A Simplified User Interface to the Nucleic Acid Database. Rutgers University, New Brunswick, NJ.
24. Srinivasan, A. R. and W. K. Olson. 1980. Yeast $tRNA^{Phe}$ conformation wheels: a novel probe of the monoclinic and orthorombic models. *Nucleic Acid Res.* 8:2307–2329.
25. Macskassy, S. 1992. Ndbview. A Specialized 3-Dimensional Display Program for Crystallographic Structures of Nucleic Acids. Rutgers University, New Brunswick, NJ.
26. Richardson, D. C., and J. S. Richardson. 1992. The kinemage: a tool for scientific communication. *Protein Science* 1:3–9.