

The Network Interface to the Nucleic Acid Database

Helen M. Berman

Lester Clowney

Anke Gelbin

Christine Zardecki

John D. Westbrook

Nucleic Acid Database Project

Department of Chemistry

Rutgers, The State University of New Jersey

ndb.rutgers.edu

Abstract

The Nucleic Acid Database (NDB) [1] is a resource and repository for researchers in the field of nucleic acid structure. This tutorial will describe the contents of the NDB and the query interface that is available to access this information on the World Wide Web.

1 Introduction

The Nucleic Acid Database Project [1] was created in 1991 as a resource for researchers in the field of nucleic acid structure. Beginning in January 1996, the NDB also became a deposition site for oligonucleotide crystal structures. The purpose of the NDB is to gather all structural information about oligonucleotides obtained from x-ray crystallographic experiments, and to organize this information in a manner that maximizes the utility of this information for the diverse community of consumers and producers of nucleic acid structure data.

Table 1 shows the distribution of the 440 crystal structures in the NDB as of July 1996. The NDB includes coordinate data as well as information describing the crystallographic experiment and refinement. Structural features derived from coordinate data such as bond distances, angles, torsions and base morphology are calculated for each structure and included in the NDB. The primary and derived data in the NDB are summarized in Tables 2 and 3.

Table 1 NDB holdings as of July 1996

Structure Type	Number
A-DNA	57
B-DNA	76
Z-DNA	52
DNA-Drug Complexes	98
DNA/RNA Hybrid	15
A-RNA	13
RNA-Drug Complexes	19
t-RNA	10
DNA with unusual features	21
RNA with unusual feature	16
Protein-DNA complexes	67
Protein-RNA complexes	8

Table 2 Primary experimental data stored in the NDB

Structure features	Descriptor
	NDB, PDB and CSD identifiers
	Availability of coordinates
Structure description	Availability of structure factors
	Sequence
	Conformation type
	Description of modifiers of base, phosphate and sugar
	Mismatched base-pairs
	Name and binding type for drug
	Description of base pairing
Citation	Description of asymmetric unit
	Description of the biological unit
	Authors
	Title
	Journal
	Volume
	Pages
	Year
Crystal data	Cell dimensions
Data collection description	Space group
	Source of radiation
	Data collection device
	Radiation wavelength
	Temperature
Crystallization description	Resolution range
	Total number of unique reflections
	Method
	Temperature
	pH value
Refinement information	Composition of solutions
	BMCD crystal identifier
	Method
	Program
	Number of reflections used for refinement
Coordinate information	Data cutoff
	Resolution range
	R-factor
	Refinement of temperature factors and occupancies
	Atomic coordinates, occupancies and temperature factors for the asymmetric unit
	Symmetry operations to generate the unit cell
	Symmetry operations to generate the biological unit

Table 3 Derivative data stored in the NDB

Distances	Covalent bond lengths
	Nonbonded contacts
Torsions	Virtual bonds involving phosphorus atoms
	Backbone and side chain torsion angles
Angles	Pseudorotation parameters
	Valence bond angles
Base morphology	Virtual angles involving phosphorus atoms
RMS deviations	Parameters calculated by a variety of algorithms
	Deviations and RMS deviations from small molecule standard values for covalent bond distances and angles.

2 Data Flow

The flow of data through the NDB is illustrated in Figure 1. The NDB abstracts experimental data from the primary literature and from author depositions, and records this information in mmCIF template files. The template files are ordinary mmCIF data files which have added comments containing definitions and enumerations that assist in data processing and curation. After validation processing, the primary data are loaded into the NDB database system. Once the primary data has been loaded into the database, the derivative features in Table 3 are calculated from coordinate data and loaded into the database. The calculation of derivative features is managed by the application, NDBQuery, which is also responsible for providing local query access and for generating database reports. Some of the features of NDBQuery are discussed in the next section.

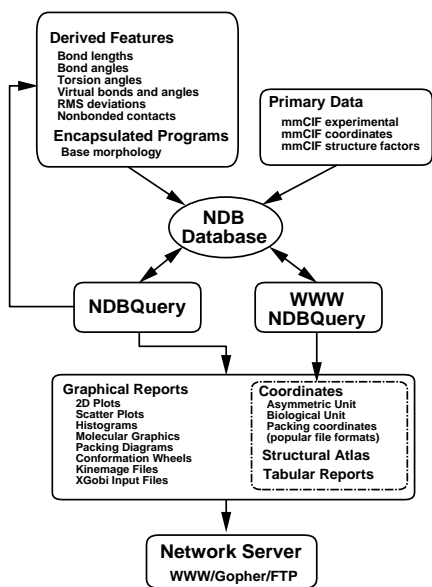


Figure 1 Functional diagram of the data management scheme used by the NDB. The figure illustrates the generation of derived structural features by the NDBQuery program using both internal and encapsulated methods. The collections of graphical and tabular reports created by the NDBQuery application are accessible via the NDB network server.

3 Query Tools

The NDBQuery application provides a query interface to the NDB. This application guides the user in the creating complex selection queries on the NDB through a series of interactive menus. The application is designed to help the user select a collection of structures satisfying a set of conditions which may

be placed on either primary or derived structural features. The application then guides the user in retrieving information related to the group of selected structures.

NDBQuery provides a variety of output options. The form of output is simply the retrieval of coordinates for selected structures in a format such as PDB or mmCIF. Tabular reports can be created in formatted ASCII text files and also using the more sophisticated formatting features provided by L^AT_EX. Some example reports produced by the NDBQuery program are shown in Figures 2 - 4. NDBQuery creates graphical reports in a variety of different formats. Some examples of these graphical reports include the histogram charts shown in Figure 5, the scatter charts shown in Figure 7, the pie charts Figure 6, the torsion ring plots shown in Figure 8, and molecular graphics shown in Figures 9 and 10. All of the query and report generation operations performed by the NDBQuery program can be saved in script files enabling very complicated queries and reports to be processed in a single command, without the aid of the user interface.

Cell Dimensions for Structures Containing the Sequence C A A T T

NDB ID	Descriptor/a b c Alpha Beta Gamma	SpcGrp	Coord
BDJ069	5'-D(*CP*GP*CP*AP*AP*TP*TP*GP*CP*G)-3' 63.63 24.70 38.29 90.00 110.27 90.00	C 2	Y
PDT017	PPR1-DNA COMPLEX 53.80 92.60 174.70 90.00 90.00 90.00	I 2 2 2	Y
UDJ031	5'-D(*CP*GP*CP*AP*AP*TP*TP*GP*CP*G)-3' 27.01 39.26 54.06 90.00 90.00 90.00	I 21 21 21	Y
UDJ049	5'-D(*GP*GP*CP*CP*AP*AP*TP*TP*GP*G)-3' 26.25 36.82 53.23 90.00 90.00 90.00	P 21 21 21	P

Created by the Nucleic Acid Database Project on Sat Jul 27 13:36:42 1996

Figure 2 Example tabular report produced by the NDBQuery program. This report displays the structure descriptor, cell constants, and coordinate status for NDB structures containing the sequence fragment CAATT.

Citations for Structures by Author S. Neidle with Intercalated Drugs

NDB ID	Citation
DDB009	H.-S.Shieh, H.M.Berman, M.Dabrow, S.Neidle The Structure of a Drug-Deoxydinucleoside Phosphate Complex; Generalized Conformational Behavior of Intercalation Complexes with RNA and DNA Fragments <i>Nucleic Acids Res.</i> , 8 , 85-97, 1980.
DRB008	H.M.Berman, W.Stallings, H.L.Carrell, J.P.Glusker, S.Neidle, G.Taylor, A.Achari Molecular and Crystal Structure of an Intercalation Complex: Proflavin-Cytidylyl-(3',5')-Guanosine <i>Biopolymers</i> , 18 , 2405-2429, 1979.
DRD004	A.Agarwal, S.A.Islam, R.Kuroda, S.Neidle X-Ray Crystallographic Analysis of a Ternary Intercalation Complex between Proflavine and the Dinucleoside Monophosphates CpA and UpG <i>Biopolymers</i> , 23 , 1025-1041, 1984.

Created by the Nucleic Acid Database Project on Sat Jul 27 13:24:16 1996

Figure 3 Example tabular report produced by the NDBQuery program. This report displays the citations for structures reported by S. Neidle which contain intercalating drugs.

Mismatched Base-Pairs for Structures With a G-G Mismatch

NDB ID	Strand 1	Strand 2	Descriptor
AHIB53	C--5	U--13	5'-R(*GP*CP*UP*UP*CP*GP*GP*CP*)-D(*(BR)U)-3'
AHIB53	G--6	G--12	5'-R(*GP*CP*UP*UP*CP*GP*GP*CP*)-D(*(BR)U)-3'
AHIB53	U--3	G--15	5'-R(*GP*CP*UP*UP*CP*GP*GP*CP*)-D(*(BR)U)-3'
AHIB53	U--4	C--14	5'-R(*GP*CP*UP*UP*CP*GP*GP*CP*)-D(*(BR)U)-3'
BDL046	G--4	G--9	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*GP*GP*CP*G)-3'
BDL046	G--9	G--4	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*GP*GP*CP*G)-3'

Created by the Nucleic Acid Database Project on Sat Jul 27 14:02:22 1996

Figure 4 Example tabular report produced by the NDBQuery program. This report displays the mismatched base pairs and the structure descriptor for NDB structures containing at least one G-G mismatch.

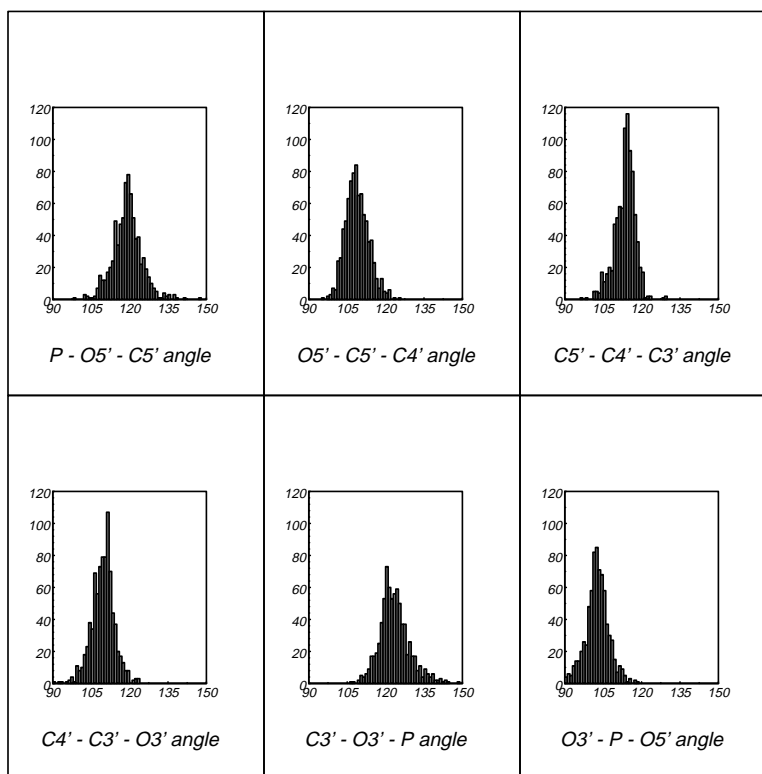


Figure 5 Example histogram charts produced by the NDBQuery program. The distributions of backbone covalent bond angles are shown for B-DNA structures without mismatched base pairs, modifications, or bound drugs.

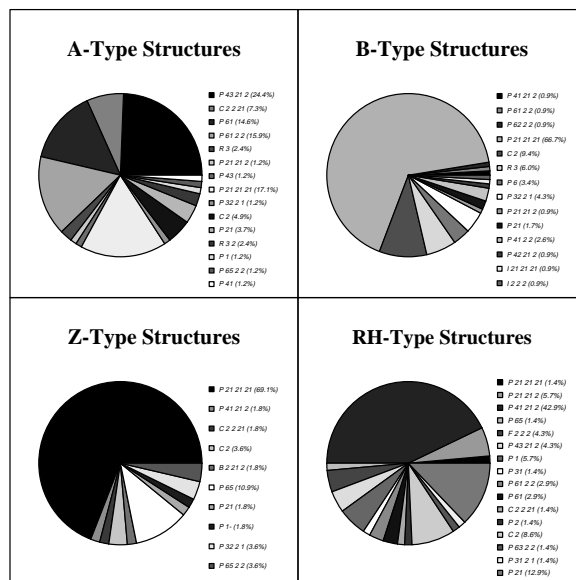


Figure 6 Example pie charts produced by the NDBQuery program. The distributions of space groups are shown for four nucleic acid conformation types.

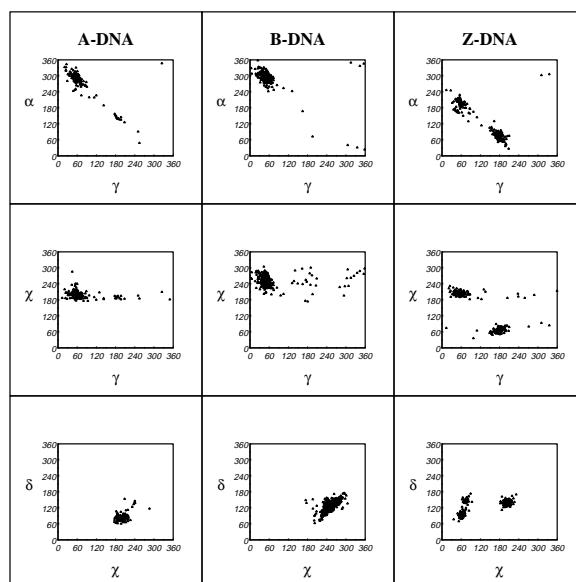


Figure 7 Example scatter charts produced by the NDBQuery program. The pair-wise distributions of selected backbone torsions are shown for DNA structures obtained with resolutions better than 1.9Å for three DNA conformation types.

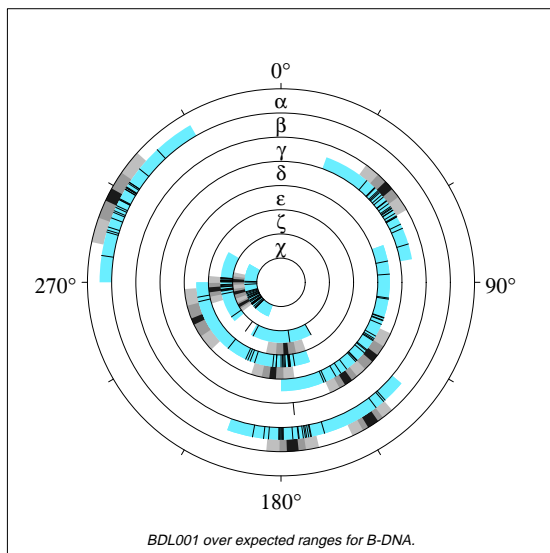


Figure 8 Example torsion ring chart produced by the NDBQuery program. The observed backbone torsions for the Dickerson dodecamer, BDL001 [2] are displayed over a solid cyan band indicating the range of values observed for all B-DNAs in the NDB. The concentric variable shaded regions indicate the one, two and three sigma regions for the distributions of backbone torsions for all B-DNAs in the NDB.

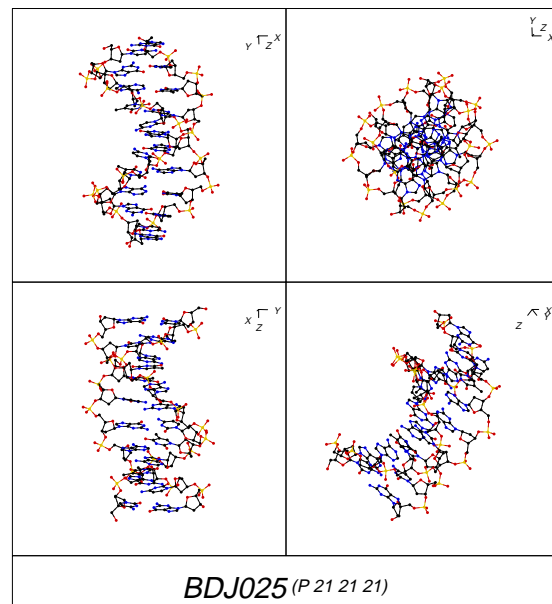


Figure 9 Example molecular graphic produced by the NDBQuery program. Four views of the biological unit of the self-complementary duplex d(CGATCGATCG)₂, BDJ025[3].

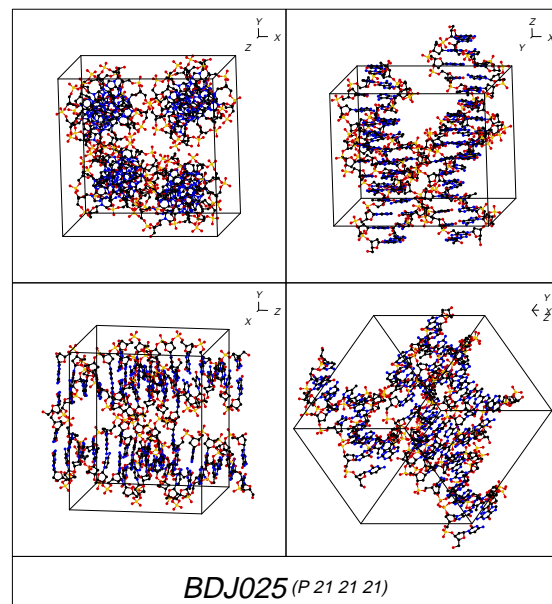


Figure 10 Example molecular graphic produced by the NDBQuery program. Four views of packing in the unit cell for the self-complementary duplex d(CGATCGATCG)₂, BDJ025[3].

4 Distributed Access to the NDB

The NDB can be reached on the World Wide Web at the address <http://ndbserver.rutgers.edu>, and at our mirror site at the European Bioinformatics Institute at address <http://www.ebi.ac.uk/NDB/>.

An overview of the NDB resources available on the Web are summarized on the NDB home page in Figure 11. These resources include: several query interfaces; an atlas of nucleic acid structures; an archive containing coordinate data, standard tabular and graphical reports describing contents of the database; dictionaries of standard geometries and parameter files; information and software related to macromolecular CIF and DDL projects; documentation and newsletters.



Figure 11 The Nucleic Acid Database home page.

4.1 Network Query Interface

The NDB network query interface (Figure 1) provides a subset of the capabilities of the NDBQuery program on the World Wide Web. The menu of available query interfaces is shown in Figure 12. The first interface option provides the most robust access to the NDB. The second interface provides a selection query interface using a simple one page form containing some of the most popular database attributes. The third interface performs a structure selection query and creates a status report for the group of

selected structures.

The control menu for the robust query interface is shown in Figure 13. The selection query and report creation facilities provided by this interface are similar to those provided by the NDBQuery program. This interface allows the user to navigate through the database and to identify the database attributes which will be constrained in the selection query or appear in an output report.

In building a selection query, database features of interest are chosen from the table and column menus shown in Figures (Figure 14) and (Figure 15). Conditions are placed on the selected features using the form shown in Figure 16. This form permits the specification of a target value for the condition and the type of comparison to be made (e.g. equal to, greater than, containing the string, etc.). Multiple conditions are applied using the logical operator (AND/OR) specified with each condition. The result of a selection query is a menu of output options similar to those shown in Figure 17. From this menu it is possible to display/retrieve the coordinates for the asymmetric and biological units for the structure, view the NDB Atlas entry for the structure, or view the structure using the RASMOL viewer [4]. An example of output from a selection query is shown in Figure 18.

In addition to these output options, the interface can be used to build tabular reports for the subset of structures that are returned by the selection query. From the main interface menu shown in Figure 13, a user can enter the report generation section of the interface and select the database features that will populate an output report. Reports are generated by the interface and returned to the network browser in the form of HTML tables (Figure 19).

The second query interface provides a simple alternative to the menu driven selection query. The simple selection form interface is shown in Figure 20. This selection interface presents a subset of the NDB database attributes in a one page form. This selection interface produces the same menu of output options as are provided by the more complicated menu driven selection interface.

The status query selection form is shown in Figure 21. This interface permits a selection query on the restricted set of features: NDB structure identifier, author, citation title, and status code. These features are typically available for all structures throughout the data curation process. The result of the status query is a status report similar to that shown in Figure 22.



Figure 12 The NDB top level query options page. From this menu three query options may be selected: the full WWW NDB query interface, the quick form-based query interface, and the status query Interface. A tutorial describing the use of these query interfaces may also be selected.



Figure 14 The table selection menu provides a list of the table organized in categories according to content. Clicking on a table name from this screen will produce the list of columns in the selected table.



Figure 13 The main menu of options for the full NDB query interface. This menu coordinates the structure selection queries and tabular report generation.



Figure 15 The column selection menu provides a list of the columns in a selected table. Each column is displayed with a description and a few examples.

STRUCTURE SELECTION CONDITION FORM

Use the following form to place a condition on the selected item. Select the appropriate logical and comparison operators using the pull-down menus, and enter the desired constraining value in the text box.

| Main Menu | Tables | Columns | Reset Selection | Make Selection |

The current item selection is: Space_Group

Place the constraining value for Space_Group in space below.
(e.g. P212121, C2, P2 (note spacing))

Value:

Select the comparison operator for this condition from the following menu.

Comparison Operator:

If multiple conditions are selected, select the logical operation that will be used to join the conditions.

Logical Operator:

Current Query

```
SELECT DISTINCT ndbV2.structure_summary.Structure_ID FROM
ndbV2.structure_summary, ndbV2.cell_dimensions WHERE
ndbV2.structure_summary.Structure_ID = ndbV2.cell_dimensions.Structure_ID AND (
ndbV2.cell_dimensions.a < 32.5 AND ndbV2.cell_dimensions.a > 25.8) ORDER BY
ndbV2.structure_summary.Structure_ID
```

Figure 16 The condition form allows the user to place a constraint on the value of the column attribute. In this form, the selection condition that the space group be $P2_12_12_1$ is added to the selection query. This condition on the space group will be appended to the existing selection conditions on the a cell constant which are displayed at the bottom of the form.

NDB STRUCTURE SELECTION

| Main Menu | Tables | Reset Selection |

15 structure(s) were found that satisfied your selection criteria. Select one of the structures from the following list, and then choose the desired display format.

Structure ID
4HJ052
BOL012
BOL021
BOL032
BOL066
BOLB10
BOLB13
DDF064
GDL014

Display option:

Your display:

For dynamic render:

Public Acid Database Project
Release 14 (March 2004) - Download at: <http://www.rcsb.org/>

Figure 17 This form displays the list of the structures which have been selected by the current selection query. A list of options which can be selected from this form include: the coordinates for the asymmetric unit in mmCIF, PDB and NDB formats, the coordinates for the biological unit in PDB format, the full NDB format entry, the structure of the biological unit viewed using a RAS-MOL viewer [4], and the coordinates of the biological unit returned to the user with the *chemical/x-pdb* mime type.

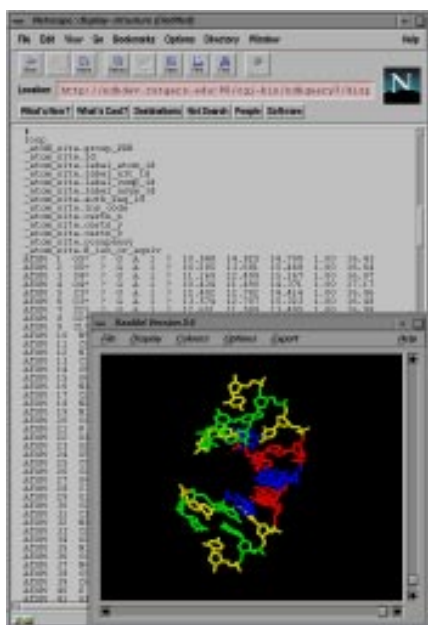


Figure 18 Example display from a structure selection query for NDB structure AHJ043[5] showing the coordinates in mmCIF format and the structure viewed using the RASMOL viewer [4].



Figure 19 An example report created from NDB report preparation interface following a structure selection query. The selection query picked B-DNA structures in P 21 21 21 space group with an *a* cell dimension in the range, 23.5 to 24.5 Å. This example report displays the sequences and cell dimensions for the selected structures.

The screenshot shows the Oracle SQL Developer application. At the top, there's a menu bar with 'File', 'Edit', 'View', 'Go', 'Bookmarks', 'Options', 'Structure', and 'Window'. Below the menu bar is a toolbar with icons for file operations and database navigation. The main window has a title bar 'Oracle SQL Developer' and a menu bar 'File', 'Edit', 'View', 'Go', 'Bookmarks', 'Options', 'Structure', 'Window'. The 'Structure' menu is open, showing a list of structures in the NDB catalog. The list has columns for 'Structure ID', 'Author', 'Structure Description', 'Structure Type', 'Status', 'Group', 'Modifiers', 'Revision', and 'Open Group'. The 'Structure Description' column is highlighted. The 'Structure Description' column is highlighted. The 'Structure Description' column is highlighted.

Figure 20 An example illustrating the use of the NDB Quick Query Form to select B-DNA structures determined by author Dickerson, which contains no modifications nor mismatched base pairs, and were determined in space group $P2_12_12_1$ with resolution better than 1.5 Å.

WebProject - NDB Quick Status Report

File Edit View Go Bookmarks Options Summary Window Help

Location: C:\MSDEV\1\000-NEW\237443-NEW\1274433\NDB\STAT\STAT-STATUS.RPT

Print View Copy Paste Undo Redo

NDB QUICK STATUS REPORT

The following form can be used to obtain a report about the processing status of structures in the NDB Audit Database.

Structure ID: [dropdown]
 Author: [dropdown]
 Title: [dropdown]
 Status Code: [dropdown]
 PAGE: [dropdown]

Status Code	Description
RC1	Coordination mode is not OK (warning)
EOL	Entry Management: add invalid entries (NDB action)
WLD	Secondary data management: verify the status
WLD	Secondary data management: verify the status
WLD	Existing in development mode: action
PNC	Account duplicated after NDB is implemented

Initial Selection Reset Selection

The NDB Audit System is Project
 Department of Computer Systems University of New Brunswick
<http://www.unb.ca/nbs/audit/>

Figure 21 An example illustrating the use of the NDB status query form to select structures that in being processed at the NDB and are waiting for materials or approval from authors.



Figure 24 An atlas entry for the B-DNA dodecamer structure, BDL001 [2]. This page of the atlas gives the structure type, compound name, sequence, citation, space group, cell constants, crystallization conditions and refinement data. Links are provided to coordinate data, images of the biological unit, and images of packing in the unit cell for this structure.

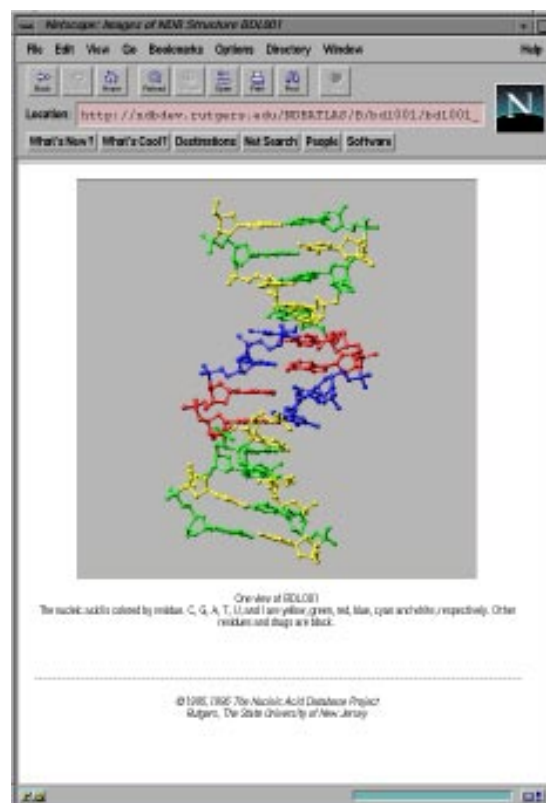


Figure 25 A molecular graphic from the NDB Atlas entry for the B-DNA dodecamer structure, BDL001 [2]. This ball and stick representation is color coded by sequence.

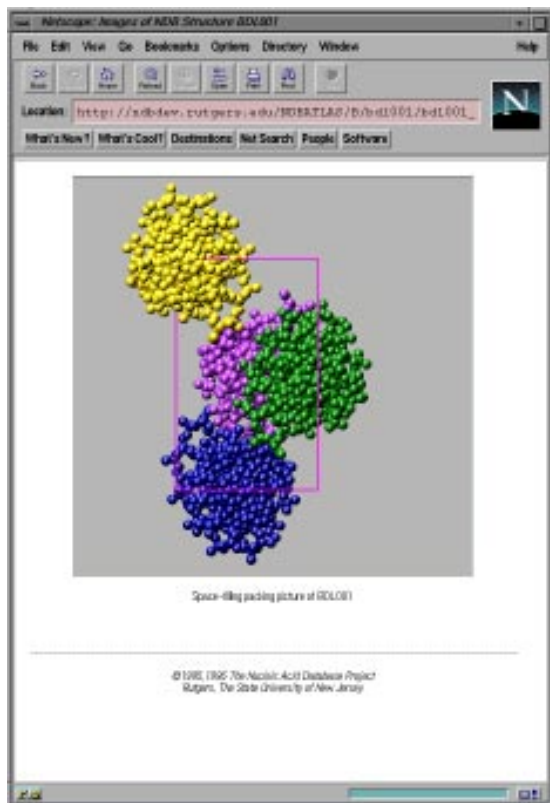


Figure 26 A molecular graphic from from the NDB Atlas entry for the B-DNA dodecamer structure, BDL001 [2]. This space filling representation of the unit is color coded in terms of the symmetry related molecules.



Figure 27 The Archives page provides links to a variety of information that is routinely updated by the NDB. This includes collections of standard reports describing the current contents of the NDB, dictionaries of nucleic acid standard geometries and parameter files, on-line documentation, project newsletters, and links to the NDB FTP archive.



Figure 28 The macromolecular CIF (mmCIF) home page. The information which can be accessed from this page includes: the history of the development of the mmCIF dictionary and Dictionary Description Language (DDL), the text and HTML versions of the mmCIF and DDL dictionaries, descriptions of mmCIF software tools, and related mailing lists.

5 Acknowledgment

This work was supported by the National Science Foundation (BIR9305135 & BIR9510703) and the Department of Energy. The collaboration with Wilma Olson and David Beveridge on nucleic acid analysis projects using the NDB is gratefully acknowledged.

References

- [1] H. M. Berman, W. K. Olson, D. L. Beveridge, J. D. Westbrook, A. Gelben, T. Demeny, S-H. Hsieh, A. R. Srinivasan, and B. Schneider. Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. *Biophys. J.*, 63:751, 1992.
- [2] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. Structure of a B-DNA Dodecamer. Conformation and Dynamics. *Proc. Nat. Acad. Sci. USA*, 78:2179, 1981.
- [3] K. Grzeskowiak, K. Yanagi, G. G. Prive, and R. E. Dickerson. 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 Reversals. *J. Biol. Chem.*, 266:8861, 1991.
- [4] Roger Sayle. RasMol 2.5: Molecular Graphics Visualisation Tool. BioMolecular Structures Group, Glaxo Research and Development, Greenford, Middlesex, UK., 1994.
- [5] M. Egli, N. Usman, and A. Rich. Conformational Influence of the Ribose 2'-Hydroxyl Group: Crystal Structures of DNA-RNA Chimeric Duplexes. *Biochemistry*, 32:3221, 1993.