International Tables for Crystallography (2006). Vol. F, Table 24.2.2.1, p. 657.

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

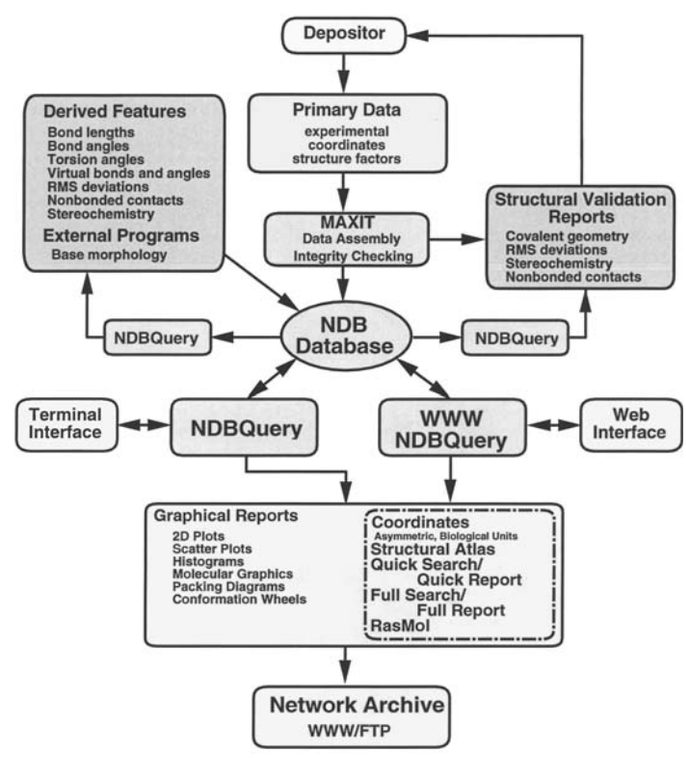

Fig. 24.2.3.1. Flow chart showing the organization of the Nucleic Acid Database Project. The core of this integrated system is the database.

references