NUVIEW: software for display and interactive manipulation of nucleic acid models

M.Bansal, D.Bhattacharyya[1] and S.Vijaylakshmi

Abstract

The NUVIEW software package allows skeletal models of any double helical nucleic acid molecule to be displayed on a graphics monitor and to apply various rotations, translations and scaling transformations interactively, through the keyboard. The skeletal model is generated by connecting any pair of representative points, one from each of the bases in the basepair. In addition to the above mentioned manipulations, the base residues can be identified by using a locator and the distance between any pair of residues can be obtained. A sequence based color coded display allows easy identification of sequence repeats, such as runs of Adenines. The real time interactive manipulation of such skeletal models for large DNA/RNA double helices, can be used to trace the path of the nucleic acid chain in three dimensions and hence get a better idea of its topology, location of linear or curved regions, distances between far off regions in the sequence etc. A physical picture of these features will assist in understanding the relationship between base sequence, structure and biological function in nucleic acids.

Introduction

Sequence directed variability of oligonucleotide structures and intrinsic curvature of polynucleotides have become accepted facts recently. Methods of analyzing these sequence dependent features have been described earlier (Bhattacharyya and Bansal 1988, 1989, Bansal and Bhattacharyya, 1990) and the software package to calculate various local parameters of oligonucleotide structures, (NUPARM) developed by us is now available on the program tape (PGMTP) of Brookhaven Protein Data Bank (PDB) or directly from us. A related program (NUCGEN) has also been developed to generate nucleic acid structures using the local step parameters tilt, roll, twist, shift, slide, rise, as well as the intra basepair parameters propeller twist, buckle and opening, for any given base sequence and obtain some structural parameters describing the shape/curvature of the generated

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

[1] *Present address: Division of Computer Research & Technology, Building 12A, NIH, Bethesda, MD 20892, USA*

model structure (Bhattacharyya and Bansal 1988, 1989—see also accompanying paper). However since a visual representation of the structure facilitates in better understanding the three dimensional shape of the molecule, we have recently developed an interactive graphics program 'NUVIEW' to display and manipulate large DNA molecules.

Since a polynucleotide structure with n basepairs contains about $(n \times 40)$ atoms, it is difficult to see the large number of atoms constituting a polymer with $n > 100$, using the all-atom display packages. Furthermore so many atoms are not really required to visualize the molecule, if one is interested in looking at the tertiary fold of the molecule. Hence, NUVIEW has been developed to show the global fold of the molecule, of length up to 1200 basepairs, by displaying a skeletal model generated using only one atom per nucleotide. The program supports real-time rotation, translation and scaling of the molecule, as well as location of any particular base on the screen, to pinpoint interesting features. It also allows distance calculation between any two bases in the molecule. The other useful features are representation of the DNA double helix in a sequence based color coded form and generation of plotfiles.

Systems and methods

NUVIEW has been written in BORLAND TURBO C++ for the IBM PC computer or compatibles, running under DOS 5.0 (or higher) operating system environment and equipped with a graphics (mono VGA, CGA, EGA or VGA) monitor. All operations, including rotation, translation, scaling, coloring, *etc.* are performed through the various keys of a standard keyboard with 101 keys. The plotfiles are created in a HPGL compatible format, to enable output either on a Hewlett-Packard pen-plotter or a PCL5 compatible laser printer (it has been tested on HP LaserJet III).

Algorithm

The package uses simple graphical routines to connect the coordinates corresponding to the given pair of atoms in each basepair (by dotted line), as well as along each of the polynucleotide chains (by solid line). The DNA molecule

is displayed in a color coded representation, with backbone in pink, A:T basepairs in blue and G:C basepairs in yellow. The 5' end of the first strand is indicated by a circle. The line representing every tenth basepair is slightly elongated to enable the relative phasing of every turn of DNA to be visualized. An important feature of this software, which is very useful in locating precisely the bending locus of a sequence, is the identification of each base along the strand by using the function keys.

The molecule is treated as a rigid body for the purpose of rotation and translation operations, which are applied about three orthogonal axes, X and Y being in the plane of the screen and the Z axis being normal to the screen. The 'append' function allows two molecules or fragments of a large sequence to be merged and viewed, as well as rotated, together.

Implementation

Input file

NUVIEW, requires an input coordinate file, containing the residue identification and one atom from each nucleotide (we generally choose the $C1'$ atom, but any other pair of atoms such as P or C6/C8 can be taken). The input coordinates can be given in one of three ways. (i) The program can read a Brookhaven PDB format file and selectively pick up the $C1'$ atoms. (ii) The output file (***.HLX) of the generation program 'NUCGEN' (described in the accompanying paper). The $C1'$ coordinates are in standard PDB format, but the pair of atoms within a basepair should occur sequentially, unlike the PDB file convention, wherein all residues in one chain of the duplex occur sequentially. (iii) It can also accept a file containing coordinates and residue names in the format [3F10.3,3X,A1], with the basepairs occurring sequentially.

```
"Press Esc to exit NUVIEW"
"Press S to save the reoriented coordinates"
"Press U to translate the molecule UP on the screen"
"Press D to translate the molecule DOWN on the screen"
"Press L to translate the molecule LEFT on the screen"
"Press R to translate the molecule RIGHT on the screen"
"Press X or Y or Z for continuous rotation about the X,Y,or Z axes"
"Press Tab to bring the molecule to the Center"
"Press right arrow to rotate clockwise along Y axis "
"Press left arrow to rotate counter clockwise along Y axis"
"Press up arrow to rotate clockwise along X axis"
"Press down arrow to rotate counter clockwise along X axis"
"Press PgUp to rotate clockwise along Z axis "
"Press PgDn to rotate counter clockwise along Z axis "
"Press + to scale up  "
"Press - to scale down"
"Press C to display molecule with thickline and title"
"Press T to display the title"
"Press Space bar to enter DOS command mode"
"Press N to display a new molecule"
"Press A to add/superpose another molecule"
"Press Enter for the locator mode option to identify bases"
"Press = to calculate distance between any two bases"
"Press . to create an HPGL file for plotting the model"

        "Press any key to continue"
```

Fig. 1. A summary of the command options available in NUVIEW, as displayed on screen in the on-line HELP menu.
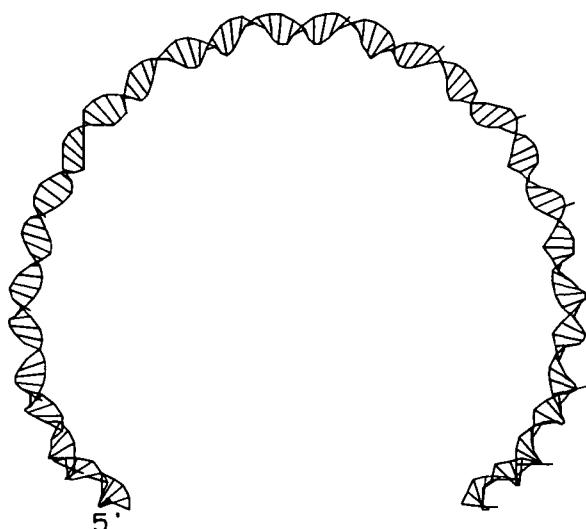

Fig. 2. A line plot of the model structure for $(C_5A_5)_{15}$ molecule showing smooth curvature (Nagaich et al., 1994). The line representing every tenth basepair is slightly extended towards the second strand, as defined by the EMBO Workshop (1989) nomenclature.

The input file type has to to be indicated before specifying the file name.

A limited amount of on-line help is available in the program, while the details are included in the form of application notes. The various operations possible in the NUVIEW software are summarised in Figure 1.

Plot file generation

HPGL plot files can be created, to obtain skeletal drawings of the DNA molecules on an Hewlett-Packard, or compatible, pen-plotter or laser printer. The plot file can be assigned a title and more than one image can be merged to create a composite plot. The file has to be subsequently transferred to the plotter/laser printer using a normal DOS print command.

Examples and discussion

Figure 2 shows the plot of a model structure, corresponding to the repeat sequence $(C_5A_5)_{15}$ which is predicted to be smoothly curved ((Nagaich et al., 1994), while Figure 3(a) shows the screen image for a promoter sequence DNA molecule, in the color coded mode. The models have been generated using the NUCGEN software and the dinucleotide inter-basepair parameters obtained from an analysis of crystal structures of B-DNA oligonucleotides (accompanying paper, also Nagaich et al, 1994). Figure 3(b) shows the locator mode option with a line connecting two residues and the residue identification and distance between them (in Ångströms), being indicated at the top. The NUVIEW software readily allows a large double
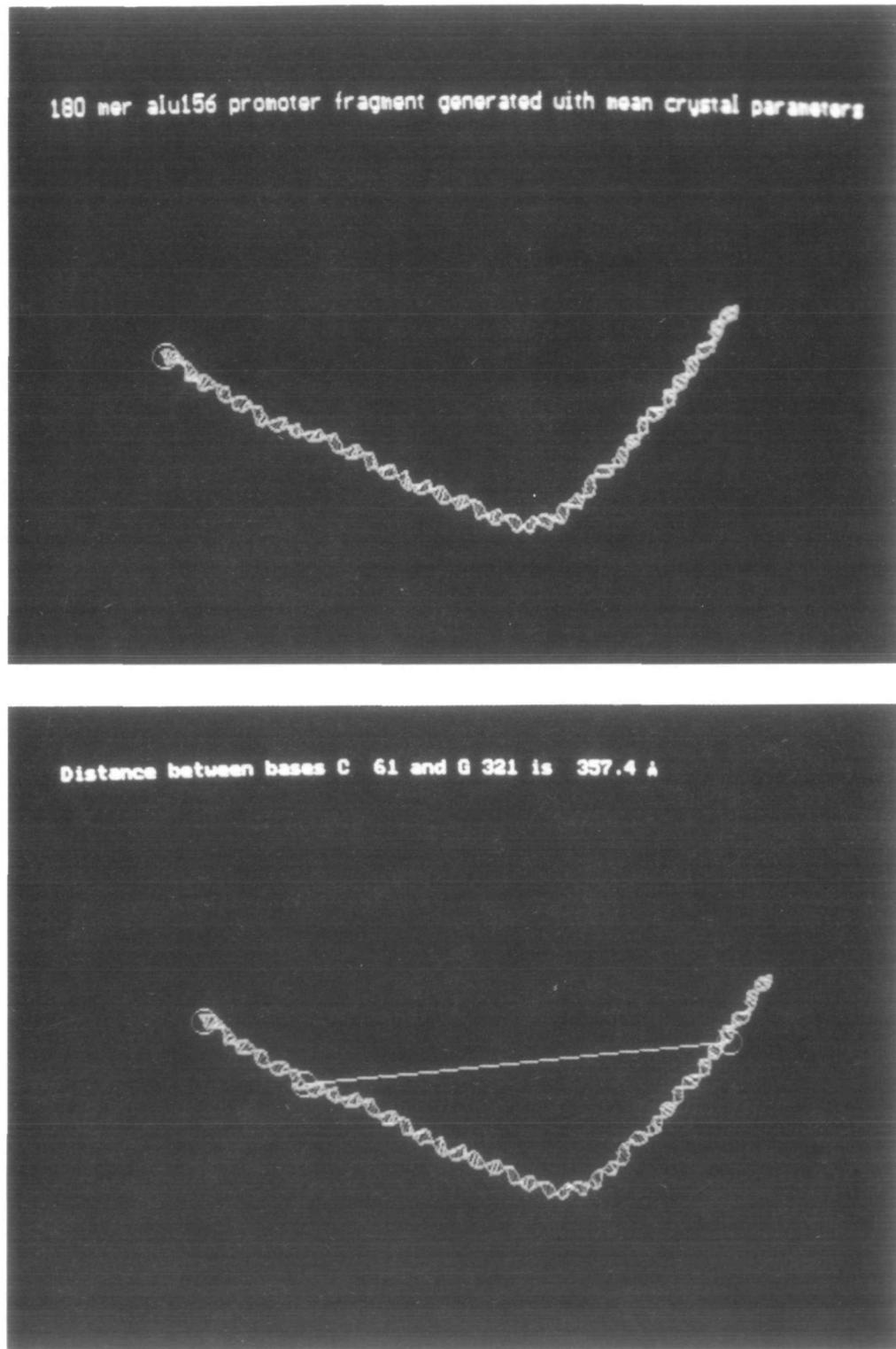
Fig. 3. The predicted structure for a 180 bp long Alu156 promoter sequence (−150 to +30) from *Bacillus subtilis* (McAllister and Achberger, 1988). (a) The color coded screen image is shown along with the title information. (b) The distance between two residues (joined by a magenta line) is also displayed on screen.
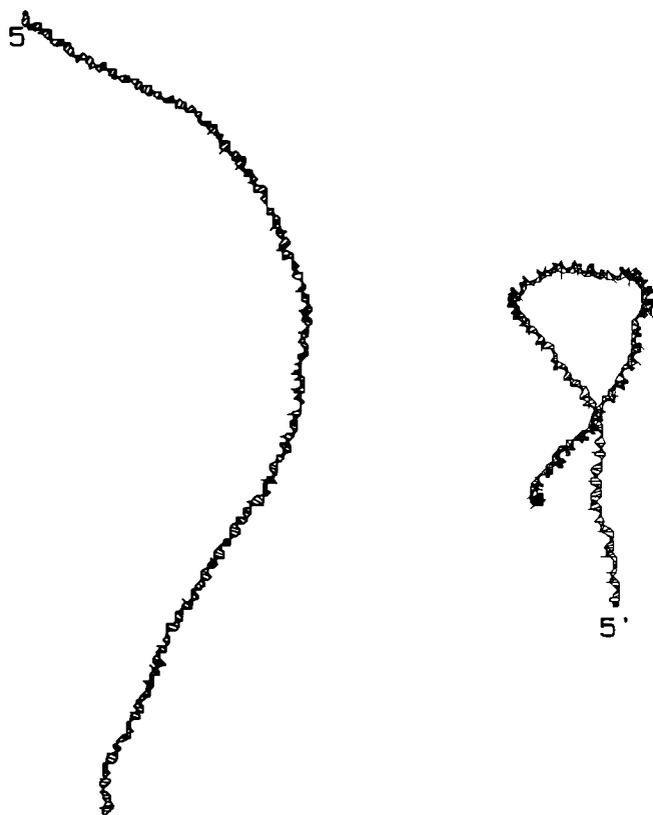
Fig. 4. Two dimensional plots showing mutually orthogonal views of a model structure for the 414 residue long fragment of kinetoplast DNA from *Leishmania tarentolae*, which has been characterised as being highly curved (Diekmann and Wang, 1985).

stranded nucleic acid molecule to be viewed in several different orientations to get a better idea of its three dimensional topology. This is illustrated in Figure 4, where two mutually orthogonal views of the 414 basepairs long curved kinetoplast DNA, from *Leishmania tarento-lae* (Diekmann and Wang, 1985) are shown. The two views together clearly indicate that this molecule has a super-helical type of curved structure, which is not obvious from a single projection.

It is possible to simultaneously display two molecules in either superposed or side by side mode and operate on them together. This helps to visualize the effect of mutations on DNA structure. Thus the NUVIEW software provides a very useful tool to understand the topology of large nucleic acid fragments. A different version of this software is being implemented on Silicon Graphics workstations and will be available shortly. This version, which is menu based and with mouse interface, will enable even larger molecules (up to 5000 bp) to be displayed and manipulated, easily and rapidly. However in the PC version we found the keyboard based manipulation very useful, since it enables the user to control the step size of rotation and translation operations

interactively. While there are several software packages which have superior display features, to our knowledge there is at present no software available, either for PCs or Silicon Graphics workstations, which allows display and manipulation of multiple, large, DNA molecules and the NUVIEW package fills a very useful niche in the rapidly expanding field of structural analysis of large DNA molecules, generated with any of the simulation packages.

Acknowledgements

References

Bansal,M. and Bhattacharyya,D. (1990) Real and apparent curvature in oligonucleotides, in Sarma,R.H and Sarma,M.H. (eds), *Structure & Methods*, Volume 3: *DNA & RNA*, Adenine Press, New York, pp. 139–153.

Bhattacharyya,D. and Bansal,M. (1988) A general procedure for generation of curved DNA molecules, *J. Biomol. Str. Dyn.*, **6**, 93–104.

Bhattacharyya,D. and Bansal,M. (1989) A self-consistent formulation for analysis and generation of non-uniform DNA structures, *J. Biomol. Str. Dyn.*, **6**, 635–653.

Diekmann,S. and Wang,J.C. (1985) On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities, *J. Mol. Biol*, **186**, 1–11.

EMBO Workshop (1989) Definitions and nomenclature of nucleic acid structure parameters, *EMBO J.*, **8**, 1–4, *J. Mol Biol.*, **205**, 787–791.

McAllister,C.F. and Achberger,E.C. (1988) Effect of polyadenine-containing curved DNA on promoter utilization in *Bacillus subtilis*, *J. Biol. Chem.*, **263**, 11743–11749.

Nagaich,A.K. Bhattacharyya,D Brahmachari,S.K. and Bansal,M. (1994) CA/TG sequence at the 5′ end of oligo(A)-tracts strongly modulates DNA curvature, *J. Biol. Chem.*, **269**, 7824–7833.