

SOFTWARE PRESENTATION

NAExplor: A Software Tool for Managing and Analyzing Nucleic Acid Structure Files

Michael Assmann and Hartmut Fritzsche

Institute of Molecular Biology, Department of Biophysical Chemistry, Friedrich-Schiller-University of Jena, Winzerlaer Strasse 10, D-07745 Jena, Germany. Tel.: +49-3641-657516. E-mail: ass@imb-jena.de

Received: 7 April 1998 / Accepted: 30 April 1998 / Published: 10 June 1998

Abstract NAExplor is a software tool for converting coordinates files between the software packages AMBER, CHARMM, and XPLOR. In addition, it manages the conversion of NMR-derived distance restraints information from the MARDIGRAS program into the appropriate file formats used for input in AMBER, CHARMM, and XPLOR. Analyses of H-H distances in nucleic acid structures and calculations of torsion angles for nucleic acid backbone and riboses are also possible.

Keywords Molecular dynamics, Nucleic acid structure, Format conversion, NMR restraints, PDB files

Introduction

For the analyses of nucleic acid fine structure by molecular dynamics simulations including NMR-derived distance restraints [1-3] we use the software packages AMBER [4-5], CHARMM [6-7], and XPLOR [8]. These programs use slightly different formats of their coordinates and distance restraints files. The format frequently used for cartesian coordinates is the Protein Data Bank (pdb) format. When using a structure file in pdb format produced from one of these programs as input for another program, one has to make some changes to the files. As an example, we list an atom in AMBER pdb file format: (' ' indicates space)

```
ATOM.....1.H5T..C5.....1.....-0.657..-9.692..-0.341
```

and the same atom in CHARMM pdb file format:

```
ATOM.....1..H5T.CYT.....1.....-0.657..-9.692..-0.341..  
0.00..0.00.....A
```

XPLOR uses the same format as CHARMM, but needs separate files for each segment (strands, ions). For calculations of distance restraints from 2D NOESY NMR experiments, we use the complete relaxation matrix approach to correct for spin diffusion, as implemented in the program MARDIGRAS [9-10]; the final results of one or more NMR experiments are calculated by the modules AVGMARDI or, in a later version, AVGBNDS, and are written to an output file. Our software tool will help to manage these changes in the pdb files, it will generate distance restraints files for these program packages and, in addition, provide facilities for analyzing the H-H distances in a given structure in accordance with the distance boundaries calculated by the AVGMARDI or AVGBNDS modules from MARDIGRAS. Calculation of torsion angles for nucleic acid backbone and riboses is also possible.

Correspondence to: M. Assmann

Implementation

The program NAEExplor can be obtained via WWW from <http://www.uni-jena.de/molebio/bpc.html>

NAEExplor is written in C and should be easily compiled on any UNIX system. The source and header files `naexplor.c`, `naexplor.h`, `nambcal.c`, `npdb_cvt.c`, `nhdiscal.c`, `ntorsdna.c` are provided in a compressed tar file (`NAEExplor.tar.Z`).

Compilation on UNIX

A simple Makefile is included in the package. You may edit this and change the name of the compiler (by default this is 'cc'). Once this is done, type: `make`

The program will then compile and you will have an executable program: **NAEExplor**

Running NAEExplor

On UNIX machines NAEExplor can be run by typing: **NAEExplor**

Description of the tool NAEExplor

The desired modules can be selected from a menu by typing the corresponding number.

1. Conversion of AMBER pdb format to CHARMM pdb format

As input the filename of a structure in AMBER pdb format is required. The converted CHARMM pdb output file carries the prefix of the input file, ending with '_ac'. For the CHARMM format it has to be accounted for a special naming convention. The first strand of a nucleic acid double helix is named 'A', the second is named 'B', and ions are named 'I'. This convention has to be applied accordingly in the GENERATE statement of the CHARMM input script.

2. Conversion of AMBER pdb format to XPLOR pdb format

As input the filename of a structure in AMBER pdb format is required which is converted to several files (one for each segment and the ions of the molecule) read in by XPLOR. The output files have the prefix of the input pdb file extended by '_ax_n' (where n is 1, 2, (3)).

3. Conversion of CHARMM pdb format to AMBER pdb format

As input, a structure file generated by CHARMM is used, and an output file with the prefix of the input pdb file extended by '_ca' is written.

4. Conversion of XPLOR pdb format to AMBER pdb format

This requires a structure generated by XPLOR, the prefix of the input pdb file is extended by '_xa'.

5. Generate restraints files from MARDIGRAS AVGMARDI or AVGBNDS output files

The desired output file can be selected, then the filename of the AVGMARDI or AVGBNDS output file as well as of the corresponding structure coordinates file in AMBER pdb format are required. When generating an AMBER distance restraints file, force constants `rk2` and `rk3` are asked for, the prefix of the input filename is extended by '_amb.con'. For a CHARMM distance restraints file, force constants `kmin`, `kmax`, and `fmax` are required, the extension of the output file is '_mkcha.inp' (which then has to be incorporated into a CHARMM script to generate and write out the final restraints format). The XPLOR output filename has the extension '_xpl.con'.

The thymine methyl groups are calculated as pseudoatoms 'M7' as the average of the three methyl hydrogen atoms and are written together with the other coordinates in a separate pdb file. When using these methyl pseudoatoms in CHARMM, a patch residue (PRES PSEUDO, for example) has to be generated in the topology file and this residue 'PSEUDO' then has to be patched in the input script.

6. Calculate distance restraints and statistics

The desired output file can be selected, then the program reads the output filename generated from the MARDIGRAS modules AVGMARDI or AVGBNDS and the corresponding structure coordinates file in AMBER pdb format. Data are output in a file extended by '.dat', statistics in '.sta'. Here LOWER and UPPER design the lower and upper distance boundaries from the AVGMARDI/AVGBNDS output file, SDT is the standard deviation, FOUND means the calculated H-H distances in the structure under consideration, and DIFF is the difference between the distance found and the upper or lower boundary, respectively. The statistics file contains the differences DIFF grouped into several ranges, and the RMS (root mean square deviation) value for the restraints set.

7. Calculation of all H-H distances

A structure filename (AMBER pdb format) is asked for. Then one has to enter the width in Å for the region where no restraining force will be applied (for example, 0.4). All hydrogen atoms and methyl pseudoatoms 'M7' are written to a file ending on '_hyd.pdb'. All the H-H distances < 5 C5 are in a file with extension '.dat'. Herein, FOUND means the calculated distances, LOWER and UPPER are the distances FOUND plus or minus half the value SDT, respectively (given as input for the region without any restraining force applied). Distance restraints files can be created with those data as an additional option.

8. Calculation of torsion angles of nucleic acid backbone and riboses

As input, a structure filename has to be defined (AMBER pdb format). An output file with the extension '.tor' is created.

9. Exit the program

Acknowledgements This work was supported by the Thuringian Ministry of Science, Research and Culture and by the Fonds der Chemischen Industrie (FCI).

References

1. Schmitz, U. and James, T. L. *Meth. Enzymol.* **1995**, 261, 3.
2. Nilsson, L.; Clore, G. M.; Gronenborn, A. M.; Brünger, A. T. and Karplus, M. *J. Mol. Biol.* **1986**, 188, 455.
3. Mujeeb, A.; Kerwin, S. M.; Kenyon, G. L. and James, T. L. *Biochem.* **1993**, 32, 13419.
4. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; and Case, D. A. *J. Comp. Chem.* **1986**, 7, 230.
5. Pearlman, D. A.; Case, D. A.; Caldwell, J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K. and Kollman, P. A. *AMBER 4.0 (UCSF)*. University of California, San Francisco 1991.
6. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. and Karplus, M. *J. Comp. Chem.* **1983**, 4, 187.
7. Nilsson, L. and Karplus, M. *J. Comp. Chem.* **1986**, 7, 591.
8. Brünger, A. T. *X-PLOR Version 3.1. A system for X-ray crystallography and NMR*. Yale University Press 1992, New Haven.
9. Borgias, B. A. and James, T. L. *Meth. Enzymol.* **1989**, 176, 169.
10. Borgias, B. A. and James, T. L. *J. Magn. Reson.* **1990**, 87, 475.