**NUPARM**

Authors: Dhananjay Bhattacharyya1, Shayantani Mukherjee1, Debasish Mukherjee1, Sukanya Halder1, Anirban Ghosh2 and Manju Bansal2

1 Saha Institute of Nuclear Physics

Kolkata 700064, INDIA

2 Molecular Biophysics Unit

Indian Institute of Science

Bangalore 560012, INDIA

E-mail: [dhananjay.bhattacharyya@saha.ac.in](mailto:dhananjay.bhattacharyya@saha.ac.in) [bhattasinp@gmail.com](mailto:bhattasinp@gmail.com) mb@mbu.iisc.ac.in

The program calculates base pair orientation parameters in DNA or RNA along with useful torsion angles and sugar pucker pseudo-rotation phase angles in nucleic acids. It has been described in several publications along with further developments. The main ones are the following:

1. M. Bansal, D. Bhattacharyya and B. Ravi (1995) NUPARM and NUCGEN: Software for Analysis and Generation of Sequence Dependent Nucleic Acid Structures *CABIOS* **11:** 281-287.

2. S. Mukherjee, M. Bansal and D. Bhattacharyya (2006) Conformational specificity of non-canonical base pairs 4 and higher order structures in nucleic acids: crystal 5 structure database analysis, *J. Comp. Aided Mol. Des.* **20:** 629-45.

3. P.K. Pingali, S. Halder, D. Mukherjee, S. Basu, R. Banerjee, D. Choudhury and D. Bhattacharyya (2014) Analysis of stacking overlap in nucleic acid structures: Algorithm and application. *J. Comp. Aided Mol. Des* **28**: 851-867.

4. D. Bhattacharyya, S. Halder, S. Basu, D. Mukherjee, P. Kumar, M. Bansal (2017) RNAHelix: Computational Modeling of Nucleic Acid Structures with Watson-Crick and Non-canonical Base Pairs. *J. Comp. Aided. Mol Des.* **31:** 219-235.

The program expects coordinates in PDB formatted file and base pairing information between residues from another file. Few files (AdeVariants.name, GuaVariants.name, CytVariants.namd, UraVariants.name and surface.xyz), supplied along with this package, needs to be placed in specific location and those need to be indicated in the code. Please make minor modification in the code at lines 1165, 1175, 1184, 1193 and 6043 to indicate location of the files in your system, recompile using gfortran or some other good FORTRAN compiler and run. The Different run-time options are the following:

NUPARM can be used in two ways:

1. For a signle coordinate file in PDB format and

2. For a trajectory file in PDB format as created by VMD (for example).

In the first case:

NUPARM [-options] coordinate\_file\_name

In the ssecond case:

NUPARM [-options] trj\_file\_name -trj -param name -output out\_file\_name

where the [-options] are as used in case 1.

Options: -notdouble [default=double]

-notpdb [default=pdb]

-bpinf file\_name\_for\_BasePairing\_Information

-lsfit [default=cross-product method]

-c1c1 [default=C6-C8 line as Y-axis]

-cg [default=C6-C8 midpoint as BP center]

-single [default=Single Strand Parameter not required

-orient "atom selection" [default=Not required]

-pp [default=Not required]

-axis [default=Not required]

-ww (to consider all base pairs as Watson-Crick type) [default=No]

In the implementation for PDBe, the trajectory file reading is not considered and few options have not been implemented. As coordinates may not be available in PDB formatted file, the pre-processing of the mmCIF file is done by BPFIND. That generates two files specifically for use by NUPARM, namely PDB\_ID\_updated.out and PDB\_ID\_updated.cor. The basic run command of NUPARM would thus be:

**nuparm PDB\_ID\_updated.cor -BPINF PDB\_ID\_updated.out**

The program produces three output files: PDB\_ID\_updated\_basepair.csv, PDB\_ID\_updated\_local.cvf and PDB\_ID\_updated\_torsion.csv, which can be loaded into the database. Each of the files contain four letter PDB\_ID as the first column, residue serial number as the second column, residue number as in \_atom\_site.auth\_seq\_id in the third column, single character residue name in the fourth column, \_atom\_site.pdbx\_PDB\_ins\_code in the fifth column.

In addition to these common informations about the residues, the PDB\_ID\_updated\_basepair.csv gives values of the parameters useful to describe orientation of the two bases, of a base pair, with respect to each other as suggested by IUPAC-IUB parameters. The 7th to 11th columns would have residue information of the base which is paired to the residue mentioned in 2nd to 6th columns. These would be followed by the actual parameters: Buckle (12th column), Open (13th column), Propeller (14th column), Stagger (15th column), Shear (16th column) and Stretch (17th column). Most of these parameters are also calculated by few other software, such as X3DNA or CURVES and the values calculated by these three programs are very similar. The values of Stretch calculated by NUPARM, however, depict true separation between the two paired bases, unlike the other programs. Hence, these values are around 3A by NUPARM while are around 0 by the other programs. This file also contain three other parameters, which are useful to detect Isostericity of a base pair with respect to standard Watson-Crick base pairs. These are given in 18th column, which is distance between C1' atoms of the two paired residues; angle between glycosidic bond (C1'--N9/N1) and the C1'...C1' pseudo-bond (whose length is given in the 18th column) and similar angle for the paired base in the 19th and 20th columns. The 21st column contains base pairing type, which can be W:W, W:H, etc, as detected by BPFIND program. The 22nd column represents orientation of the two base pairs, i.e. *Cis* or *Trans*. In cases where a residue is not paired to any other base, the 7th through 11th columns would not contain any information, would have “.” and the 12th to 20th columns would have zero values.

The file PDB\_ID\_local.csv gives values of local doublet parameters describing orientation of two successive base pairs (or bases or bases and a base pair) with respect to each other. The parameter values in each line represents orientation of the base pair (or base) indicated in the current line and the base pair (or base) indicated in the next line. The first few columns, 1st to 12th are identical to the PDB\_ID\_basepair.csv file. The 13th column gives values of Tilt, 14th column gives values of Roll, 15th column gives values of Twist, 16th column gives values of Shift, 17th column gives values of Slide, 18th column gives values of Rise. The 19th column gives values of difference between the Buckle values of the present base pair and the next base pair, which is often related to measurement of Rise. The 20th column gives measure of stacking overlap between the current base pair and the next base pair in A2 unit. The base pairing type etc. are mentioned in the 21st and 22nd columns, as in PDB\_ID\_updated\_basepair.csv file.

The file PDB\_ID\_updated\_torsion.csv also contain residue information similar to the others in first 6 columns. The 7th to 12th columns give main chain torsion angles, a, b, g, d, e and z. The 13th column represents glycosidic torsion angle x. The 15th and 16th columns represent two pseudo-torsion angles, n and q, as defined by Duarte and Pyle (1998) *J. Mol. Biol.* **284**:1465-1478. The 17th column gives pseudo-rotation phase angle, as defined by Altona and Sundaralingam (1972) J. Am. Chem. Soc., 94: 8205–8212. The 18th column represent description of the phase angle in standard definition.