CURVES+ USER GUIDE (V2.6, 4/ 2014)

Compiling

Curves+ is distributed as Fortran code. The distribution includes a Makefile that uses the gfortran compiler. If you have this compiler on your system, you can simply type "make" to generate the executable file cur+, otherwise you will have to modify the Makefile. The main vector and matrix dimensions of the code are contained in curves_data.inc. The program is currently set up to treat a maximum of 500 nucleotides, 15,000 heavy atoms and 100 ions (or ligands). You may need to change these limits in some cases (e.g. analyzing solvent water molecules). In this case, change the Makefile to compile with checks for overflowing dimensions (exchange the comment symbol "#" on the FFLAG lines of the Makefile)

Input data

The example below shows a simple input file. In the example, the compiled program is in the directory /Users/RL/Code. The notation <<! allows the input data to be placed immediately after the line which calls the program. The input is then ended by the explanation mark.
The example analyzes a duplex DNA from the file 1bna.pdb.
The initial input is in namelist format (beginning &inp and ending &end - namelist lines must begin on column 2 or beyond. See below for possible namelist variables. In the example, we give the input file name (the extension .pdb is assumed by default), the output file (r+bdna.lis, in this case it will be created in the /Code subdirectory) and the library file lib (base and backbone reference geometries will be taken from the files standard_b.lib and standard_s.lib, again in the /Code subdirectory).
The next line gives the number of strands, followed by the number of nucleotides in each strand (up to four strands can be analyzed). For each strand, the sign indicates whether the nucleotides are in the direction 5'-3' (positive) or 3'-5' (negative).
The following two lines give the numbers corresponding to the order of the subunits containing the nucleotides which constitute each strand (1=1st, 2=2nd, …). Note proteins, water and HETATM are automatically removed at input. Contiguous numbers can be indicated using a colon (I:J).
Note that bases can be excluded from axis calculations by including them in the I/P lines as negative numbers (useful for flipped out bases or abasic nucleotides) and gaps are indicated by zeros. Nucleotide numbers in each strands should be organized so that paired bases occur in identical positions.

The main options of Curves+ are controlled with namelist input. All namelist parameters have default parameters (see below) that will be used unless new values are given in the namelist section of the input (i.e. between "&inp" and "&end"). Remember that namelist input lines should not start on column 1 (leave at least one blank space).

```
rm r+bdna*.*
/Users/RL/Code/Cur+ <<!
  &inp file=1bna, lis=r+bdna,
  lib=/Users/RL/Code/standard, &end
2 1 -1 0 0
1:12
24:13
!
```

Curves+ namelist variables

CHARACTER (strings without quotes, maximum length 128 characters):

file: file name for input structure (.pdb and .mac extensions need not be given in input, use .trj for MD trajectories)

ftop: name (with extension) of file with topological data (AMBER format) for MD trajectory analysis

lis: root file name for all output (.lis, .cda, .cdi, .cdl, .fra, _X.pdb, _B.pdb, _C.pdb)

lib: root file name for base (_b.lib) and backbone (_s.lib) geometry files

lig: name (.lig extension assumed) for reference geometry of ligand

ibld: name (.cdi extension assumed) of ion coordinates for reconstruction

sol: name of solute molecule in input .pdb file (or Amber topology file) to be analyzed if ions=.t.

back (P): atom used to define backbone. Different backbone can use different atoms if necessary (e.g. P/C5* - a slash is used to separate input names).

REAL DATA (default value of each variable is given):

wback (2.9Å) radius of backbone around spline through phosphate atoms for groove width calculations

wbase (3.5Å) half-width of base pairs for groove depth calculations

INTEGER DATA (default value of each variable is given):

isym (1): symmetry repeat for generating helical axis (1 = mononucleotide, 2 = dinucleotide e.g. ATATAT..., any positive value is allowed)

naxlim(0): if naxliim > 0, calculates overall bend skipping naxlim base levels at the 5'- and 3'-ends

itst (0): first snapshot to analyze (if reading MD trajectory)

itnd (0): last snapshot to analyze (if reading MD trajectory)

itdel (1): spacing of snapshots to analyze

n.b. if itst=n and itnd ≤ itst, analyze only snapshot number "n"

LOGICAL SWITCHES (enter as .t. or .f., the default value of each variable is given):

circ (.f.): .t. implies a closed circular nucleic acid

line (.f.): if .t. find the best linear helical axis

zaxe (.f.): if .t. use the Cartesian Z-axis as the helical axis

fit (.t.): if .t. fit a standard bases to the input coordinates (important for MD snapshots to avoid base distortions leading to noisy helical parameters)

test (.f.): if .t. provide addition output in .lis file on fitting and axis generation

ions (.f.): if .t. helicoidal analysis of ions (or solvent molecules) around solute is carried out

refo (.f.): if .t. use the old Curves convention for defining the reference frame of each base, rather than the more recent Tsukuba convention

axfrm (.f.): if .t. generates closely spaced helical axis frames as input for Canal and Canion

frames (.f.): if .t. outputs base frames in the *lis*.fra file (as nearest integer to value x 1000)

Input files

STRUCTURE
Curves+ can analyze single structures from .pdb or .mac filest (the latter is only of interest for users of JUMNA) or a series of snapshots from an AMBER MD simulation (a .trj file). Analyzing a trajectory also requires reading a topology files that defines the atom types and the order of atoms that will be found in the .trj file. Note that only ATOM lines are used in .pdb files (unless solute molecules are analyzed, in which case HETATM lines are also used, e.g. by putting solute=H2O in the namelist input).

Please note that PDB unit numbers are NOT used in Curves+ input:
- the numbers in the input refer to order in which subunits appear in the .pdb file (i.e. 1,2,3 implies 1st, 2nd, and 3rd subunits encountered in the input file)

LIBRARY FILES (note that all .lib files can be edited by the user to suit individual needs)
standard_b.lib contains the standard geometry for the bases which can ben analyzed. Below is the data for cytosine:

```
C Y 7 'Cytosine'
 1.94866 -1.45161 -0.19029 'C1*'
 0.74385 -0.58352 -0.07229 'N1'
 0.93020  0.79520 -0.12929 'C2'
-0.15547  1.60399 -0.02329 'N3'
-1.37969  1.08626  0.13371 'C4'
-1.59254 -0.32937  0.19471 'C5'
-0.49500 -1.12092  0.08671 'C6'
```

The first line gives the one-letter code for the base, R/Y specifying purine or pyrimidine, the number of ring atoms (plus the sugar C1*) and the full base name (≤ 8 charas.). The following lines give the x,y,z coordinates of each atom and the atom name. The first three atoms must be C1*, the base atom bound to C1* and the atom used to define the base normal using the vector product (1-2)x(3-2). This data is used for least squares fitting of standard base geometries. The base atoms can be specified in any order.

standard_s.lib contains the description of the backbone geometry to be analyzed. Standard lines start with a blank and define a single torsion by giving 4 atom names (≤ 4 chars) followed by the name of the torsion name (≤ 6 chars). Lines starting with 'B' define torsions which depend on the type of base: atoms 1-4 are used with purines and atoms 5-8 are used with pyrimidines

Lines starting with 'S' define sugar rings and give the names of the 5 ring atoms in the order used for pseudorotation calculations. the data below is for standard DNA. Non-standard bases or backbones can be analyzed by modifying the .lib files.

```
  '-O3*'  'P'   'O5*' 'C5*'  'Alpha'
   'P'    'O5*' 'C5*' 'C4*'  'Beta'
   'O5*'  'C5*' 'C4*' 'C3*'  'Gamma'
   'C5*'  'C4*' 'C3*' 'O3*'  'Delta'
   'C4*'  'C3*' 'O3*' '+P'   'Epsil'
   'C3*'  'O3*' '+P'  '+O5*' 'Zeta'
B 'O4*'  'C1*' 'N9'  'C4'  'O4*'  'C1*'
                ...  'N1'  'C2'   'Chi'
S 'C1*' 'C2*' 'C3*'  'C4*' 'O4*'
```

standard_i.lib contains the description of the ions or solute molecules (one atom only) to be analyzed. Each line gives the name of the ion or atom in single quotes followed by its charge (which can be used in Canion for choosing positive, neutral or negative ions/solute atoms).

```
'Na+' 1
'K+' 1
'P' -1
```

Output files

The .lis output from Curves+ lists the input parameters, the base sequence of each strand, then: (A) base pair-axis parameters; (B) intra-base pair parameters; (C) inter-base pair parameters: (D) backbone parameters and (E) groove parameters. In section (C), the first six parameters, including rise and twist correspond to the transformation between successive base pairs. H-Ris and H-Twi correspond to the translation and rotation of successive base pairs along and around the helical axis. With the base pair-axis parameters, H-Ris and H-Twi are useful for understanding the overall helical structure of the fragment analyzed.
Graphic output is a _X.pdb file containing the helical axis and _B.pdb file showing the backbone splines and the vectors defining groove widths. Analyzing snapshots from an MD trajectory (which requires both trajectory and topology I/P files) suppresses the parameters in the .lis output, and either Canion (for structure, using the .cda file) or Canal (for ions/solute molecules, using the .cdi file) should be used for parameter analysis.

Test output for 1bna.pdb

```
    ************************************              **************
    **** CURVES+ Version 2.6  04/2014 ****              *  15 Apr 14 *
    ************************************              **************


 FILE : 1bna.pdb                          ftop :
 LIS  : r+bdna                            LIB  : standard
 lig  :                                   ibld :
 sol  :                                   back : P

 wback :  2.90  wbase :  3.50

 isym  :     1  itst  :     0  itnd  :     0  itdel :     1  naxlim:     0

 circ  :     F  line  :     F  zaxe  :     F  fit   :     T  test  :     F
 ions  :     F  refo  :     F  axfrm :     F  frames:     F


 LS fitting of standard bases ...RMS max =  0.051

 Strands =    2 Atoms =   486 Units =    24 H removed =     0

 Combined strands have   12 levels ...

 Strand  1 has  12 bases (5'-3'): CGCGAATTCGCG
 Strand  2 has  12 bases (3'-5'): GCGCTTAAGCGC

 (A) BP-Axis         Xdisp   Ydisp   Inclin    Tip  Ax-bend

   1) C   1-G  24    0.15    0.08     6.2     1.2     ---
   2) G   2-C  23    0.02   -0.01     4.4     3.1     0.4
   3) C   3-G  22    0.43   -0.36     4.7    -5.9     0.5
   4) G   4-C  21   -0.18    0.02     6.4     0.2     0.5
   5) A   5-T  20   -0.33   -0.02     1.8    -1.3     1.0
   6) A   6-T  19   -0.37    0.06    -1.0    -1.3     1.0
   7) T   7-A  18   -0.02    0.02    -1.7    -4.2     1.0
   8) T   8-A  17   -0.07    0.26    -1.3    -2.6     1.0
   9) C   9-G  16    0.36    0.17    -2.0    -1.7     0.7
  10) G  10-C  15    1.46    0.64    -3.9     5.6     0.7
  11) C  11-G  14    0.52    0.34    -8.3    -3.7     0.3
  12) G  12-C  13    1.25    0.09    -6.6    -1.9     0.3

      Average:       0.27    0.11    -0.1    -1.0  Total bend =     6.8 (  1 to  12)
```

(B) Intra-BP parameters

Strands 1-2			Shear	Stretch	Stagger	Buckle	Propel	Opening
1)	C	1-G 24	-0.61	-0.20	0.09	3.8	-14.8	-2.7
2)	G	2-C 23	0.11	-0.18	0.27	-4.5	-10.9	-2.8
3)	C	3-G 22	-0.15	-0.17	0.23	-7.6	-4.1	-0.9
4)	G	4-C 21	-0.22	-0.35	-0.12	10.1	-11.7	0.3
5)	A	5-T 20	0.28	-0.21	0.14	4.9	-18.4	3.7
6)	A	6-T 19	-0.09	-0.03	0.28	3.5	-20.2	7.4
7)	T	7-A 18	0.32	-0.10	0.23	0.7	-19.3	10.2
8)	T	8-A 17	0.23	-0.20	0.02	-1.8	-19.9	3.0
9)	C	9-G 16	-0.17	-0.16	0.05	-10.8	-19.3	0.5
10)	G	10-C 15	0.23	-0.20	0.32	2.4	-6.3	0.8
11)	C	11-G 14	-0.06	-0.21	0.68	-4.0	-19.8	-4.7
12)	G	12-C 13	-0.36	-0.02	0.32	7.0	0.5	-2.9
	Average:		-0.04	-0.17	0.21	0.3	-13.7	1.0

(C) Inter-BP

			Shift	Slide	Rise	Tilt	Roll	Twist	H-Ris	H-Twi
1)	C	1/G 2	-0.39	0.27	3.54	-3.5	6.2	42.5	3.57	42.9
2)	G	2/C 3	0.52	0.18	3.54	1.0	-5.3	36.1	3.55	35.9
3)	C	3/G 4	-0.32	0.78	3.00	3.2	9.1	26.7	3.04	27.2
4)	G	4/A 5	0.00	0.07	3.38	-3.3	2.1	40.0	3.37	40.1
5)	A	5/A 6	0.11	-0.31	3.31	-0.8	0.4	35.3	3.31	35.3
6)	A	6/T 7	0.35	-0.61	3.34	2.0	-3.5	34.4	3.36	34.6
7)	T	7/T 8	-0.28	-0.17	3.33	3.2	-0.0	35.2	3.31	35.3
8)	T	8/C 9	0.02	-0.07	3.39	0.9	-0.9	38.6	3.39	38.6
9)	C	9/G 10	0.39	0.93	3.21	-2.9	4.8	31.6	3.12	31.5
10)	G	10/C 11	-1.36	0.31	3.71	-5.0	-13.5	38.5	3.67	39.2
11)	C	11/G 12	0.79	0.09	3.22	3.1	-2.9	34.6	3.21	34.8
	Average:		-0.02	0.14	3.36	-0.2	-0.3	35.8	3.35	36.0

(D) Backbone Parameters

Strand 1			Alpha	Beta	Gamma	Delta	Epsil	Zeta	Chi	Phase	Ampli	Puckr
1)	C	1	----	----	174.2	156.8	-141.3	-143.9	-105.0	161.6	56.6	C2'en
2)	G	2	-65.6	169.8	40.1	128.1	174.2	-97.8	-110.5	139.8	42.1	C1'ex
3)	C	3	-62.6	171.8	58.8	98.3	-176.7	-87.6	-135.1	92.8	38.5	O1'en
4)	G	4	-62.9	-179.9	57.2	155.7	-155.3	-152.5	-93.4	166.6	49.8	C2'en
5)	A	5	-43.0	142.8	52.4	119.6	179.9	-92.2	-126.3	128.8	46.8	C1'ex
6)	A	6	-73.3	179.7	66.0	121.1	173.7	-88.5	-122.2	127.3	50.2	C1'ex
7)	T	7	-56.6	-179.2	52.2	98.9	173.6	-85.9	-127.3	101.5	47.6	O1'en
8)	T	8	-59.2	173.4	64.1	108.9	170.6	-89.3	-125.7	115.9	49.7	C1'ex
9)	C	9	-58.5	-179.5	60.5	128.7	-156.9	-94.0	-119.5	140.7	46.9	C1'ex
10)	G	10	-67.3	169.1	47.2	142.9	-103.3	150.2	-89.6	146.5	54.7	C2'en
11)	C	11	-73.9	139.3	56.3	135.7	-161.8	-89.6	-125.1	147.7	47.7	C2'en
12)	G	12	-81.5	175.7	57.2	110.7	----	----	-112.0	114.1	52.1	C1'ex

Strand 2			Alpha	Beta	Gamma	Delta	Epsil	Zeta	Chi	Phase	Ampli	Puckr
1)	G	24	-65.0	170.6	46.6	78.7	----	----	-135.2	34.2	46.4	C3'en
2)	C	23	-72.2	138.5	44.6	112.8	-174.4	-96.8	-125.3	117.3	44.6	C1'ex
3)	G	22	-66.8	179.1	50.2	149.7	-100.1	171.6	-88.4	156.8	52.3	C2'en
4)	C	21	-59.1	-175.4	45.0	110.3	-176.7	-86.5	-114.3	113.9	43.3	C1'ex
5)	T	20	-58.6	179.5	55.3	122.4	178.5	-94.5	-120.5	129.9	50.5	C1'ex
6)	T	19	-58.3	173.6	60.0	109.2	178.8	-88.3	-131.3	116.7	47.9	C1'ex
7)	A	18	-57.1	-173.6	47.7	130.2	174.4	-101.3	-108.3	147.6	43.0	C2'en
8)	A	17	-56.6	-169.5	53.8	146.6	176.9	-97.1	-106.4	169.4	43.1	C2'en
9)	G	16	-69.2	171.1	73.2	135.9	174.1	-98.4	-114.8	149.5	41.5	C2'en
10)	C	15	-63.0	168.8	60.4	85.7	174.8	-85.5	-133.8	67.5	44.1	C4'ex
11)	G	14	-51.3	163.9	49.0	121.9	177.7	-93.0	-116.4	128.5	45.4	C1'ex
12)	C	13	----	----	55.9	136.7	-158.6	-124.9	-127.6	153.5	43.7	C2'en

(E) Groove parameters

Level			W12	D12	W21	D21
1.5						
2.0	G	2				
2.5						
3.0	C	3				
3.5			8.1	5.1		
4.0	G	4	7.2	4.9	11.4	5.0
4.5			6.2	5.1	11.5	5.1
5.0	A	5	5.2	5.1	11.5	5.2
5.5			4.6	5.1	11.2	5.7
6.0	A	6	4.2	4.9	10.5	5.4
6.5			4.1	5.2	10.4	5.0
7.0	T	7	4.0	5.4	10.6	4.9
7.5			3.3	5.6	11.6	4.5
8.0	T	8	3.1	5.6	12.1	3.5
8.5			4.1	5.5	12.4	1.8
9.0	C	9	5.2	5.7	12.4	2.9
9.5						
10.0	G	10				
10.5						
11.0	C	11				
11.5						

Curves+ program subunits

aacur	main program
axis	generate helical axis frame at each base level
axref	output axis frames if test = .t.
backbo	calculate backbone torsions and sugar puckers
bisection	bisection algorithm used in determining ion positions
dotdelta	dot product used by bisection algorithm
eigen	eigenvalue calculation used for fitting standard base geometries
findaxis	locate screw axis to transform one helical axis frame to the next
input	read input conformation (.pdb or .mac)
intaxe	use screw axis to generate helical axis frames between base levels
intop	reads Amber topology file
ionbld	reconstructs ion positions using a .cdi file as input
ionpar	calculate helicoidal ion coordinates
ligloc	find atoms of ligand necessary to fit reference geometry
ligpar	calculate ligand helicoidal parameters
locate	find base and backbone atoms and generate base reference frames
lsfit	least-squares fit of bases to standard geometries
lslig	least-squares fit of ligand to reference geometry .lig file
manta	groove geometry calculation
nml	parsing namelist input
params	calculation of helical parameters
pdbout	output of .pdb format files
screw	apply a screw axis transformation on a helical axis reference frame
setup	read library files for bases (and optionally, ions and ligands)
smooth	polynomial smoothing of helical axis frames at base levels
torp	calculate torsion angles