

User guide Twist-DNA

Daniel Jost

Laboratoire de Physique, École Normale Supérieure de Lyon, CNRS UMR 5672, Lyon, France.

Twist-DNA is a freely-available open source code that allows the computations of base-pair and bubble opening probabilities for any DNA sequences at a given temperature, a given superhelical stress and a given salt concentration.

We need your feedback! Send your comments, suggestions, and questions to daniel.jost@ens-lyon.fr

I. INSTALLATION

Download the Twist-DNA package at

<http://www.cbp.ens-lyon.fr/doku.php?id=developpement:productions:logiciels:twistdna>

and unzip it in your favorite folder. Open the command line (or Terminal), go to the corresponding folder and type 'make' to compile. The main program and subroutines are written in Fortran 90, so before compiling be sure that a fortran compiler is installed on your computer [you can download freely gfortran from the GNU Compiler Collection at <http://gcc.gnu.org/fortran/>] and that you have added the name of your favorite fortran compiler in the Makefile after the tag 'COMP=' (default is gfortran).

After the installation, check up that Twist-DNA works correctly on your computer by typing 'make test' in the command line to perform a test on a predefined DNA sequence. This step is crucial to validate the good installation of Twist-DNA.

The installation has been tested to work on Unix-based operating systems (Linux and Mac OS X) using the free GNU compiler gfortran or the Intel compiler fort.

To delete auxiliary and executable files, type 'make clean' in the command line.

II. UTILIZATION

A. Input parameters

Twist-DNA has several inputs that can be changed in the input file:

- the temperature T (in Celsius);
- the superhelical density Σ ;
- the salt concentration CNa (in mole);
- the range of bubble sizes to investigate: from $bsize_{min}$ to $bsize_{max}$ every $bstep$;
- the threshold btw : bubble whose opening probability is higher than btw (in $\log_{10}$ -unit) are saved.

Open the file *input.dat* and change the parameters if needed.

Important remark on CPU time. If you are only interested in the base-pair opening probabilities, assign 0 to $bsize_{max}$ in the input file, it will save a significant amount of computing time.

B. Computation

To run Twist-DNA, type `./TwistDNA < sequence.fasta` in the command line, where *sequence.fasta* is the FASTA file containing the sequence to study.

Remark on FASTA file. The first line of the file starts with the symbol `>` and contains a description of the file. The subsequent lines describe the DNA sequence and must contain only the standard nucleotide symbols (`'A'/'a'`, `'G'/'g'`, `'T'/'t'`, `'C'/'c'`), see for example *EcoliK12-MG1655.fasta*. Note that each line must contain less than 200 characters. If larger, open *TwistDNAmain.f90*, augment the length of the variable `'line'` to the corresponding maximal line length, and re-compile (`'make'`) the code.

C. Outputs

After running, Twist-DNA generates two files in the BED format (bedGraph track) for visualization in your favorite Genome Browser:

- *openproba.bed* that contains the individual base-pair opening probabilities (in $\log_{10}$ -unit).
- *bubblethresh.bed* that contains the list of bubbles (start, end and opening probability in $\log_{10}$ -unit) whose opening probabilities are higher than the threshold *btw*.

Remark. By default, the program prints `'chr'` as the chromosome name in the BED files. To change to another name, modify the corresponding lines in *TwistDNAprop.f90* and re-compile (`'make'`) the code.

III. EXAMPLE

A. Analysis of the genome of *Escherichia coli*

Type `./TwistDNA < EcoliK12-MG1655.fasta` in the command line to perform the analysis of the genome of *E. coli*. This will print the following lines in the command line:

```

Reading sequence...
Sequence length: 4639675
Reading input file...
Inputs:
Temperature (oC): 37.00
Superhelical density: -0.06
Salt concentration (M): 0.10
Self-consistent linearization...
Computing base-pair opening probabilities...
Computing bubble opening probabilities...
Done
```

and will generate output files.

B. Visualization with the UCSC Genome Browser

Open your favorite web browser and go to the web page of the UCSC Microbial Genome Browser (<http://microbes.ucsc.edu/>). Load the genome of *E. coli* (tag `'Genome'`, clade `'Bacteria-Proteobacteria-Gamma'`, genome `'Escherichia coli K12'`). In the Genome Browser Gateway, click on `'add custom tracks'` and load the file *openproba.bed*.

IV. CITATION

Users of Twist-DNA, please cite:

1. D. Jost (2013) Twist-DNA: computing base-pair and bubble opening probabilities in genomic superhelical DNA. *Bioinformatics*, in submission.
2. D. Jost, A. Zubair and R. Everaers (2011) Bubble statistics and positioning in superhelically stressed DNA. *Phys. Rev. E*, **84**, 031912.