SerraNA
VERSION 1.0

by Victor Velasco and Agnes Noy

York, UK, 2019

Introduction

SerraNA constitutes a software for analysing elastic and structural properties of nucleic acids using ensembles obtained by molecular dynamics (MD) or montecarlo (MC) simulations. By analysing all sub-fragment lengths, the program allows to infer global elastic constants describing fragment's overall flexibility (Figure 1). It is composed by three executables, ***SerraNA, Analysis*** and ***Extract***. The workflow is summarized in Figure 2.

Installation

The only requirement for running *SerraNA* is a FORTRAN compiler. The program can be compiled on a terminal by typing:

```
$ make all
```

This will produce the three executables. They can also be compiled separately:

```
$ make SerraNA
$ make Analysis
$ make Extract
```

SerraNA

It is the main program that processes the DNA trajectory and calculates structural and elastic parameters at all sub-fragment lengths. It will run by typing:

```
$./SerraNA < s_NA.in
```

A trajectory file and a topology file in AMBER style format is needed (10F8.3 for the trajectory). The files can contain ions or other residues and *SerraNA* will ignore them.

s_NA.in is the input file that indicates:

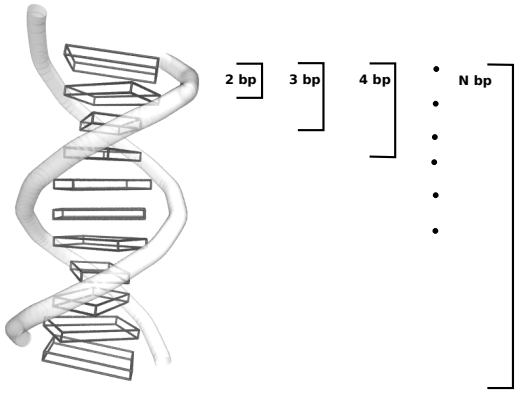

Figure 1: Parameters are calculated between every two bps comprising an oligomer whose length ranges from 2 bp to N, being N the total number of bps of the oligomer. If molecule is linear, two bps for each end are discarded

- The path for topology and trajectory.
- If the structure is double-stranded (typing "2") or if it is single-stranded (typing "1").
- If the structure is linear ("1") or closed ("2"). For linear nucleic acids, *SerraNA* ignores the two base-pairs at each end for avoiding end effects.

The program generates four outputs:

1. **BPP.out** contains the six base-pair parameters: shear, stretch, stagger, buckle, propeller and opening as they are calculated in 3DNA. It has averages and standard deviations over the whole MD or MC ensemble. The output is only written for double-stranded DNA.
2. **BSP.out** contains the six base-step parameter (shift, slide, rise, tilt, roll and twist) plus bending angle. It has averages and standard deviations over the whole MD or MC ensemble. Values should be directly comparable to the ones obtained by 3DNA.
3. **structural_parameters.out** which have variables describing the geometry of the DNA molecule for all possible sub-fragments using an extension of CEHS algorithm as it is explained on Figure 3 and on [1]. It has averages and standard deviations over the whole MD or MC ensemble:
    - Twist and bending angles, roll and tilt, which denote bending towards the major groove and backbone, respectively, at the mid-point of the specified fragment
    - Added shift, added slide, added rise, which are the counterparts of the translational bp step parameters for longer lengths and are defined simply by the addition of values at 2bp level.
    - End-to-end distance and contour length
    - $< \theta >$ ("Bending" as it's labelled on the output), $< \theta^2 >$ ("Bending**2"), and $< \cos\theta >$ ("D correlation")

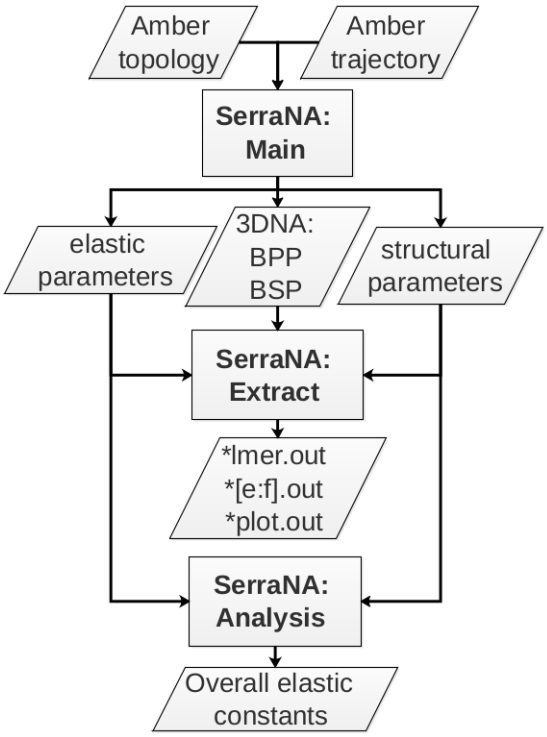

Figure 2: General workflow of **SerraNA**

- From averaged structure: $< \theta_s >$ ("AVSTR B"), $< \theta_s^2 >$ ("AVSTR B**2"), and $< \cos\theta_s >$ ("AVSTR D C"), where θ_s is the static curvature. Average structure is built with mean values of base-step parameters at 2 bp level

Translation are in Å and rotations are in degrees. Note that some of the variables for 2-mers will be directly compatible with those printed in the **BSP.out** output and with 3DNA.

4. **elastic_parameters.out**, which containts the following parameters for all sub-fragments:

   - elastic constants for stretch (pN), twist (nm), roll (nm), tilt (nm), as well as their couplings (nm). These are the terms of elastic matrix $F = k_B T b N V^{-1}$, where V is the corresponding covariance matrix, b is average rise and N is the number of bp-steps.
   - Dynamic persistence lengths defined via $A_d^{-1} = 1/2(A_{tilt}^{-1} + A_{roll}^{-1})$.
   - Variance and partial variance for end-to-end distance (in Å^2) as they are relevant for the calculation of the global stretch modulus.

   This file is only written if the trajectory has more than 1 snapshot.

structural_parameters.out and **elastic_parameters.out** have (N-1)! values for each variable, being N the total number of bps, if it is circular DNA, and the total number of bps minus two for each end, if it is linear DNA.

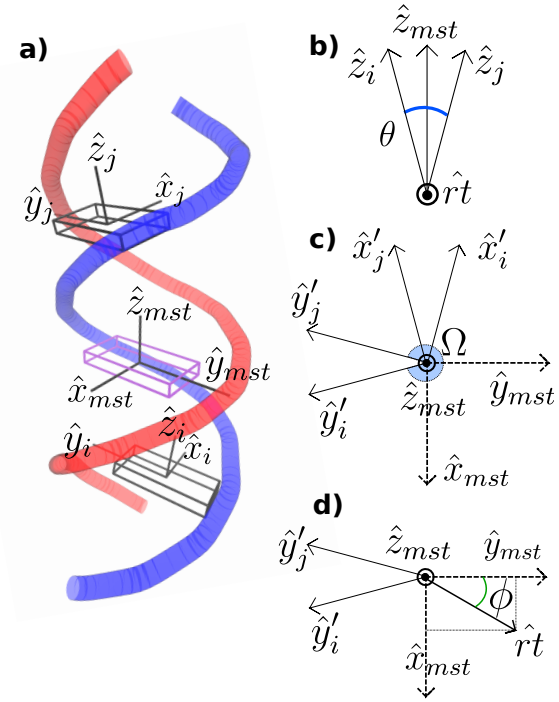

Figure 3: Schematic diagrams of SerraNA's method. (a) Bp-triads and mid-base triad are defined as in 3DNA using bending angle (θ) and roll-tilt axis ($\hat{rt}$) (b). Co-planar vectors $\hat{y}'_i$, $\hat{y}'_j$, $\hat{x}_{mst}$ and $\hat{y}_{mst}$ define twist angle Ω (c) and roll and tilt bending angles (d) with the help of auxiliary angle ϕ.

Analysis

This is the program that calculates the elastic constants at a more global level describing the whole DNA fragment. For execution, simply type:

```
$./Analysis < ov_NA.in
```

ov_NA.in is the input file that indicates:

1. The path to **elastic_parameters.out** and **structural_parameters.out**
2. The part of the molecule that will be used to calculate each of the global elastic constants. Two ranges should be provided:
   - The first one defines the region of the molecule used (from bp "a" to bp "b").
   - DEFAULT OPTION a=b=0 considers the whole fragment except for the stretch modulus where only the central 18-mer is taken to avoid long end-effects.
   - The second indicates the range of sub-lengths analyzed, being from "c" to "d" bp-steps. Note that c > 0 and d <= b-a.
   - DEFAULT OPTION c=d=0 applies the recommended methodology described in [1]:
     - For twist, roll, tilt and dynamic persistence length, c=10 for avoiding local irregularities and d=N-10 to have at least ten different values for each sub-length.

– For stretch modulus, c=8 for avoiding short-ranged stacking effects and d=17 due to only central 18-mer is used

Analysis outputs information on screen regarding the global elastic constants (for more information see [1]):

1. Total, static and dynamic persistence lengths obtained through the linear fitting of the corresponding directional correlation decays (labelled as A^a, A^a_s and A^a_d, respectively). Note that the fitting always uses sub-lengths ranging from 1 bp-step to N-10 bp-steps.
2. Total persistence length recalculated through $1/A = 1/A_s + 1/A_d$ using A^a_s and A^a_d and being labelled as A^b. A^a and A^b should be almost equal.
3. Torsion modulus, together with global values of tilt, roll and the associated A^c_d obtained through averages of the mean values by length. Standard errors are printed for these variables.
   Note that for sufficiently long molecules containing a few DNA turns, roll and tilt converges with A^c_d as the imbalance between directions facing towards grooves and backbone dissipates.
   Also note that $A^c_d > A^a_d$ as A^c_d is based on partial variances. These are the reciprocal of V^-1 diagonal terms and are the residual variance left after removing the influence from other variables
4. Total persistence length recalculated through $1/A^d = 1/A^a_s + 1/A^c_d$.
   Note that A^d should be bigger than all other estimations of A due to the use of partial variances on A^c_d
5. Stretch modulus calculated through the linear fitting of end-to-end partial variances using the central 18-mer for avoiding long end-effects

The interval of confidence are calculated as [1] for the variables obtained through linear fits.

Extract

Extract program process *SerraNA* ouputs, **elastic_parameters.out** and **structural_parameters.out**, creating simple files ready to plot. You can filter a particular sublength to produce plots similar to Figure 5 of Velasco *et al.* or you can extract averages and standard deviations as a function of length from a particular region to obtain plots similar to Figure 3 from [1]

The program can also process **BPP.out** and **BSP.out** for easier plotting. For running it type:

```
$./Extract < ex_NA.in
```

ex_NA.in is the input file that indicates:

1. Path to either BPP, BSP, structural or elastic parameters output file. If you selected to extract BPP.out or BSP.out, then all other inputs will be ignored.
2. Type "0" for extracting a sub-length or "1" for getting avg+-sd as a function of length
3. The following entry is used to indicate:
   - The length (l) you want to process, which should be $0 < l < N$, if you typed "0" before
   - The region (e,f) from which you want to extract avg+-sd as a function of length, if you typed "1" before:
     – If it is linear DNA, then 0 < e < f < N
     – If it is circular DNA, then both e < f or f < e, are valid
     – DEFAULT OPTION, e=f=0, consider the whole fragment.

The program creates different types of outputs:

1. **BPP_plot.out**, which presents parameters with the following order of columns, being averages (first) and standard deviations (second) calculated over all the ensemble. The order of variables is the same as the processed **BPP.out** file:

1	base-pair i
2,3	Shear
4,5	Stretch
6,7	Stagger
8,9	Buckle
10,11	Propeller
12,13	Opening

2. **BSP_plot.out**, as previously, the order of variables is the same as the processed **BSP.out** file:

1	Medium position of bp-step
2,3	Shift
4,5	Slide
6,7	Rise
8,9	Tilt
10,11	Roll
12,13	Twist
14,15	Bending

3. **structural_lmer.out** to extract parameters for a particular sub-length l in the same order as the processed **structural_parameters.out** file:

1	Medium position along the sub-fragment
2,3	Added shift
4,5	Added slide
6,7	Added rise
8,9	End-to-End L
10,11	Contour L
12,13	Twist
14,15	Roll
16,17	Tilt
18,19	Bending
20,21	Bending**2
22,23	D correlation
24	AVSTR B
25	AVSTR B**2
26	AVSTR D C

4. **structural_[e:f].out** to extract length-dependence for a particular molecular part:

1	Sub-length (in bps)

2,3	Added shift
4,5	Added slide
6,7	Added rise
8,9	End-to-End L
10,11	Contour L
12,13	Twist
14,15	Roll
16,17	Tilt
18,19	Bending
20,21	Bending**2
22,23	D correlation
24,25	AVSTR B
26,27	AVSTR B**2
28,29	AVSTR D C

5. **structural_plot.out** if DEFAULT OPTION e=f=0.
6. **elastic_lmer.out** to extract parameters for a particular sub-length l following the same order as the processed **elastic_parameters.out** file:

1	Medium position along the sub-fragment
2	Stretch
3	Twist
4	Roll
5	Tilt
6	Stretch-Twist
7	Stretch-Roll
8	Stretch-Tilt
9	Twist-Roll
10	Twist-Tilt
11	Tilt-Roll
12	Dynamic Persistence Length
13	Variance End-End
14	Partial variance End-End

7. **elastic_[e:f].out** to extract length-dependence for a particular molecular part:

1	Sub-length (in bps)
2,3	Stretch
4,5	Twist
6,7	Roll
8,9	Tilt
10,11	Stretch-Twist
12,13	Stretch-Roll
14,15	Stretch-Tilt
16,17	Twist-Roll
18,19	Twist-Tilt
20,21	Tilt-Roll
22,23	Dynamic Persistence Length
24,25	Variance End-End
26,27	Partial variance End-End

8. **elastic_plot.out**, if DEFAULT OPTION e=f=0.

When a particular sub-length is choosen, then the program places items of x-axis (first column) in the medium position along the fragment. For example:

- At sublength = 1 bp-step, parameters between residue 1 and 2, will be positioned at 1.5, between residue 2 and 3 at 2.5 etc
- At sublength = 2 bp-steps, parameters between residues 1 and 3 will be at 2, between residues 2 and 4 at 3 etc.
- And so on

Note that **a, b, c** and **d** bps speficied in the executable *Analysis* are totally independent from **e** and **f** bps specified here in *Extract*, since the first program has the goal to calculate global elastic constants and the second just aims to become an utility for plotting data.

References

[1] Victor Velasco, Matthew Burman, Jack W. Shepherd, Mark C. Leake, Ramin Golestanian and Agnes Noy, "*SerraNA*: a program to infer elastic constants from local to global using nucleic acids simulation data" in BioRxiv, 2019.