cgDNA+ model for DNA minicircles

Marius BEAUD

July 12, 2021

Goal of the thesis

- The scientific goal of the thesis is to compute discrete shapes for DNA minicircles.
- This has been done in the past with cgDNAmin and Non-continuous closure of the loop [Manning, 2017].[1]
- We adapt cgDNAmin
  - to the more accurate cgDNA+ model [Patelli, 2019] [1]
  - to add periodic closure assumption

[1] All references are detailed in the last slide

Plan for today

I will present:

- Basics of cgDNA [Gonzalez, 2013][Petkevičiūtė, 2014]
- Original cgDNAmin [Manning, 2017]
- cgDNA+ [Patelli, 2019]
- Periodic cgDNA+ [Glowacki, 2017]
- cgDNA+min adaptation
- Example of results
- MATLAB scripts & Demo

cgDNA

[Gonzalez, 2013] [Petkevičiūtė, 2014]

cgDNA basics

- Rigid base model: each base is represented by a frame in $SE(3)$.
- Gaussian pdf:

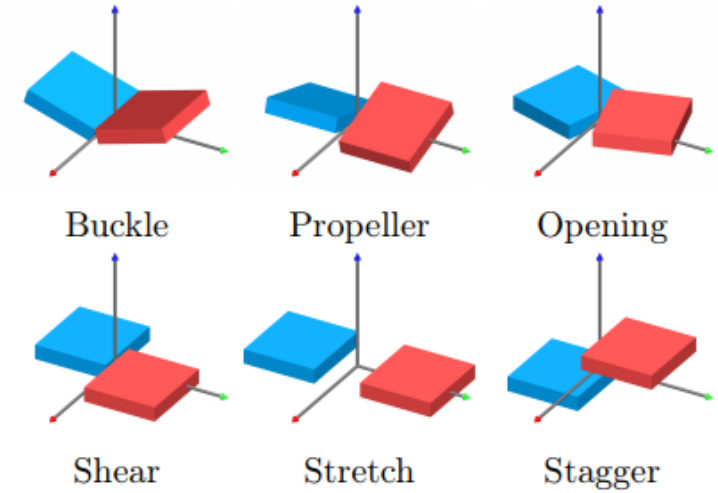
$$\rho(w; S, \mathcal{P}) = \frac{1}{Z} e^{-\beta U(w;S,\mathcal{P})},$$
$$U(w; S, \mathcal{P}) = \frac{1}{2}\big(w - \mu(S,\mathcal{P})\big)^T K(S,\mathcal{P})\big(w - \mu(S,\mathcal{P})\big).$$

- Coordinates:

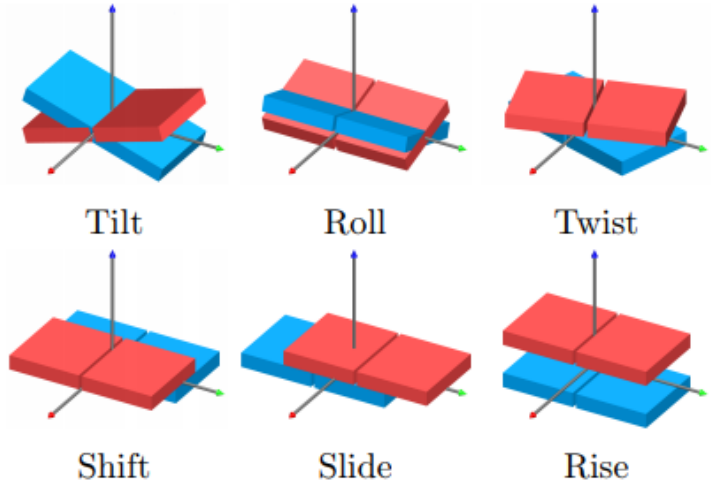
$$w = (x_1, y_1, x_2, y_2, \ldots, x_{n-1}, y_{n-1}, x_n),$$

$x_i \in \mathbb{R}^6$: *intras*, relative displacement between base frames,
$y_i \in \mathbb{R}^6$: *inters*, relative displacement between base pair frames.

cgDNA coordinates

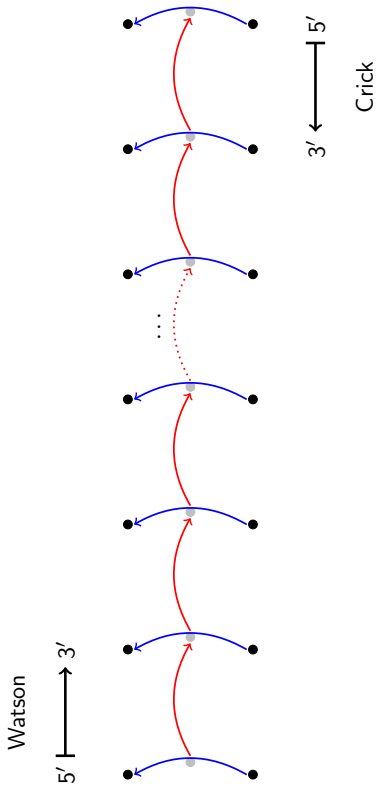


(a) Intras

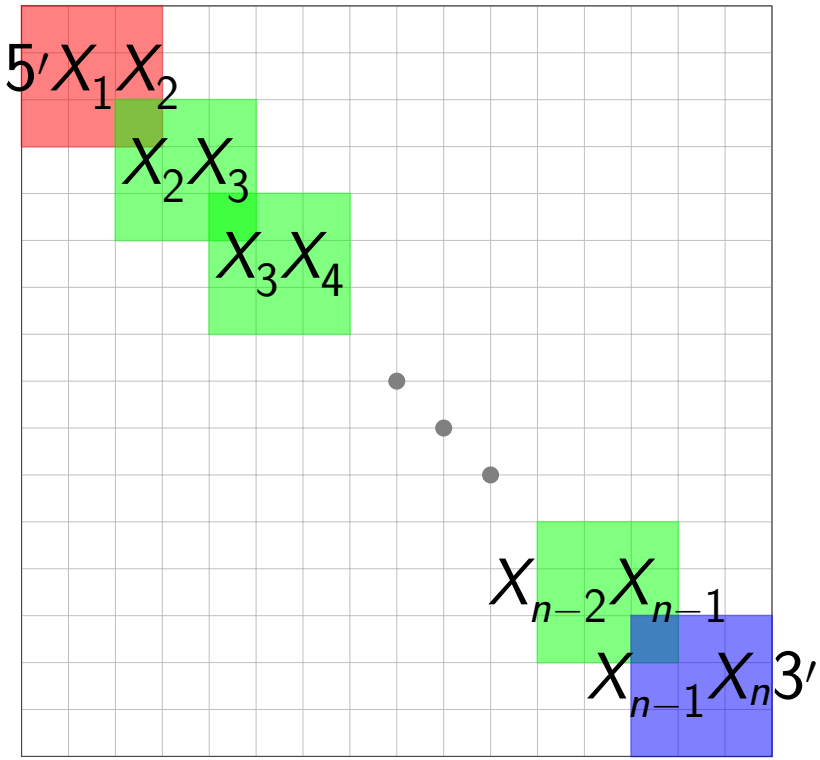


(b) Inters

cgDNA coordinates



(a) Chain structure of cgDNA coordinates.



(b) Construction of cgDNA stiffness matrix.

cgDNAmin

[Manning, 2017]

cgDNAmin: Pipeline

Two steps pipeline:

- Continuum equilibria
  - MATLAB pre-processing to obtain continuum coefficients from cgDNA parameters
  - bBDNA software to compute continuum equilibria [Glowacki, 2017]
- cgDNA energy minimization
  - Uses quaternions
  - New coordinates vector

cgDNAmin: Continuum equilibria

- MATLAB pre-processing scripts
  - Parameters: Sequence, cgDNA parameter set
  - Output: continuum coefficients for birod model
- bBDNA [Glowacki, 2017]
  - Input: Continuum coefficients for birod model
  - Output: Bifurcation diagram of families equilibrium configurations, including closed loops of different length.

cgDNAmin: Continuum equilibria

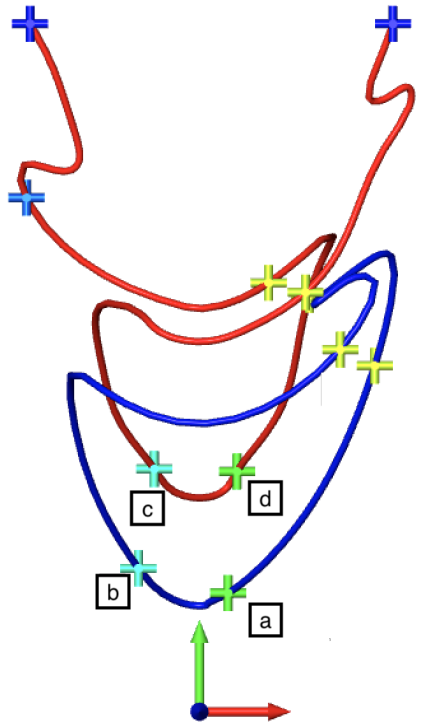


Figure: Example of bifurcation diagram outputed by bBDNA. Plot of the Energy versus the applied torque.

cgDNAmin: Discrete energy minimization

- Uses quaternions
  - 1 quaternion & 1 vector in $\mathbb{R}^3$ for each base pair
  - Benefits: simple expression for the closure assumption
  - Cost: extra transformation to recover inters
- New coordinate vector
  - $z = (x_1, o_1, q_1, x_2, o_2, q_2, \ldots, x_{n-1}, o_{n-1}, q_{n-1}, x_n)$
  - Locally recover original cgDNA inter coordinates:

$$y_i := f(o_i, q_i, o_{i+1}, q_{i+1}) = (\theta_i^1, \theta_i^2, \theta_i^3, \zeta_i^1, \zeta_i^2, \zeta_i^3),$$

$$\theta_i^a = \frac{10 q_{i+1}^T B_a q_i}{q_{i+1}^T q_i}, \ a = 1, 2, 3, \qquad \zeta_i^T = (o_{i+1} - o_i) R(q_{i+1} + q_i).$$

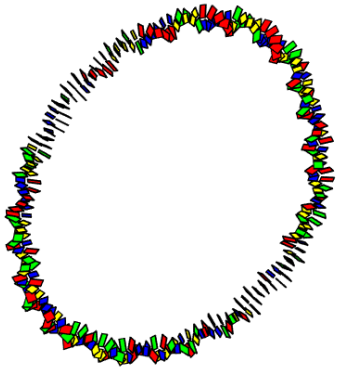
$R(q)$ is the rotation matrix induced by q and B_a are matrices used to form an orthonormal base for the quaternions.

  - global transformation $w = F(z)$,
    Energy $U(z) = \frac{1}{2}(F(z) - \mu)^T K (F(z) - \mu)$

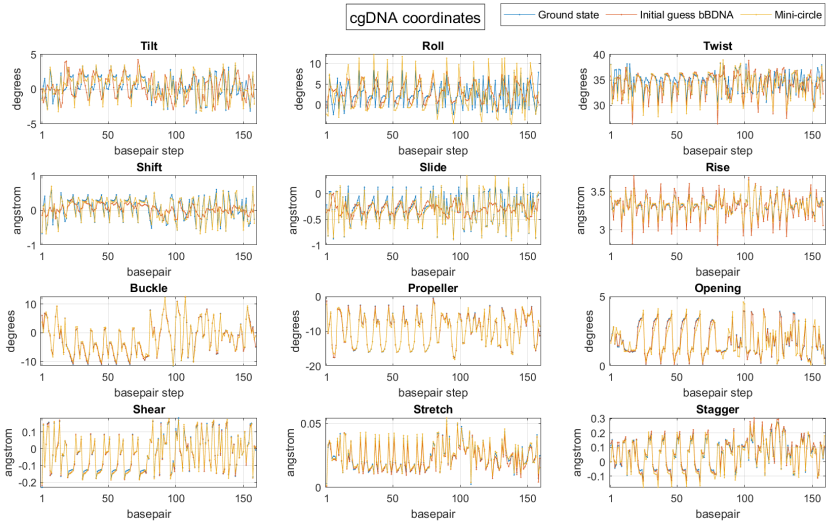
cgDNAmin routines

- Compute cgDNA stiffness & groundstate,
- Read discretized continuum solution,
- Call the `fminunc` MATLAB routine to minimize the energy $U(z)$,
- Provide explicit gradient and Hessian of $U(z)$,
- Recover original cgDNA coordinates, visualize and save the solution.

cgDNAmin example



(a) 3D view of a solution



(b) 2D coordinates plots

cgDNA+ model

[Patelli, 2019]

cgDNA+ model

- Add rigid phosphate groups
  - Each phosphate is associated to a base
  - Phosphate coordinates: relative displacement between the base and the phosphate frames

- New base pair level coordinates

$$\tilde{x}_i = [p^+, x_i, p^-] \in \mathbb{R}^{18}.$$

Exeption for first and last base pair: no external phosphates

cgDNA+ model

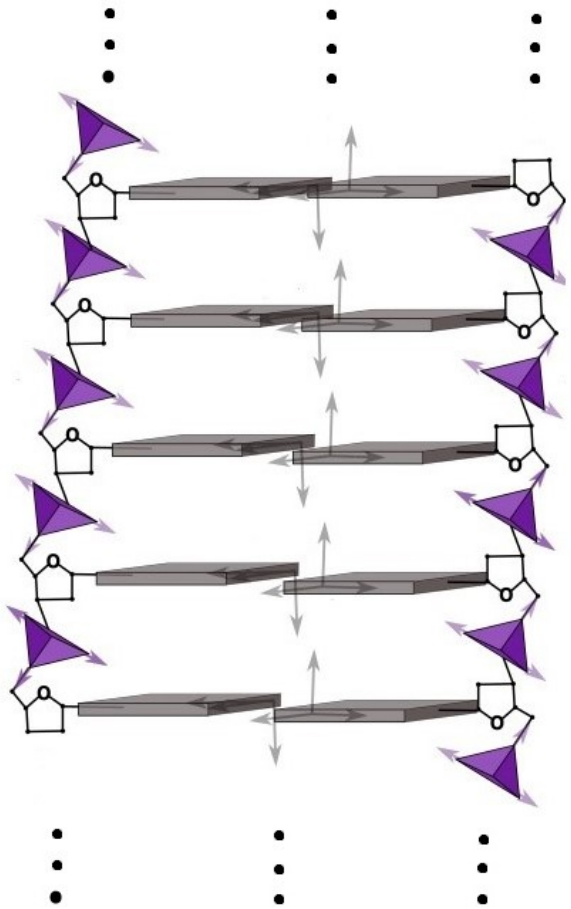
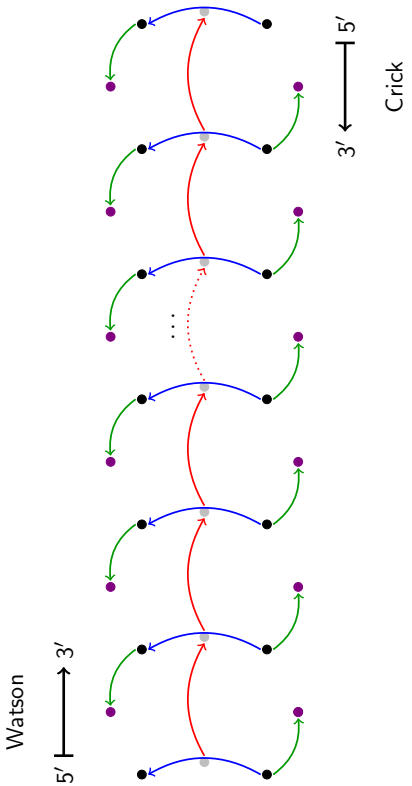
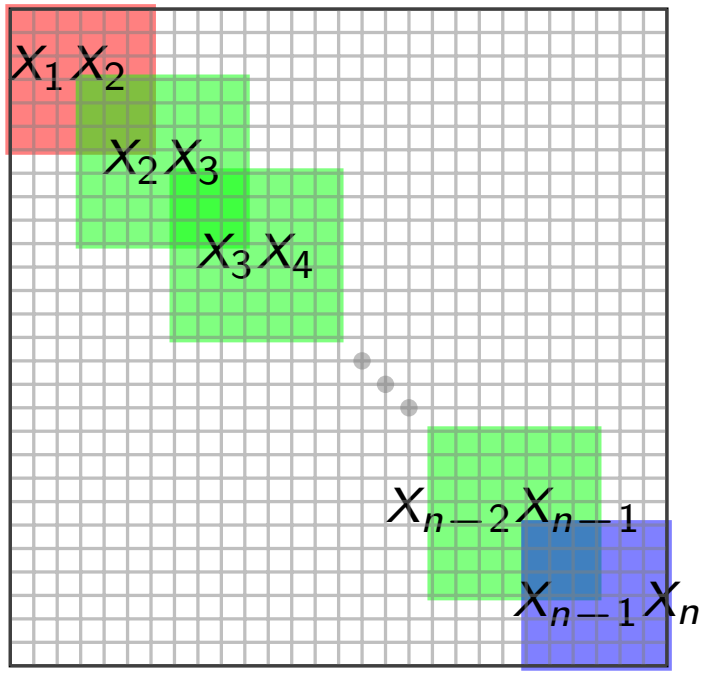


Figure: cgDNA+ structure

cgDNA+ model



(a) Chain structure of cgDNA+ coordinates.



(b) Construction of cgDNA+ stiffness matrix.

Periodic cgDNA+

- Introduced to model repeats of a DNA base sequence [Glowacki, 2017].
- Periodic cgDNA coordinates $\not\Rightarrow$ closed loop !
- Extra set of inter coordinates: inter between base pairs n and 1
- Extra block split in four corners: interactions between bases n and 1

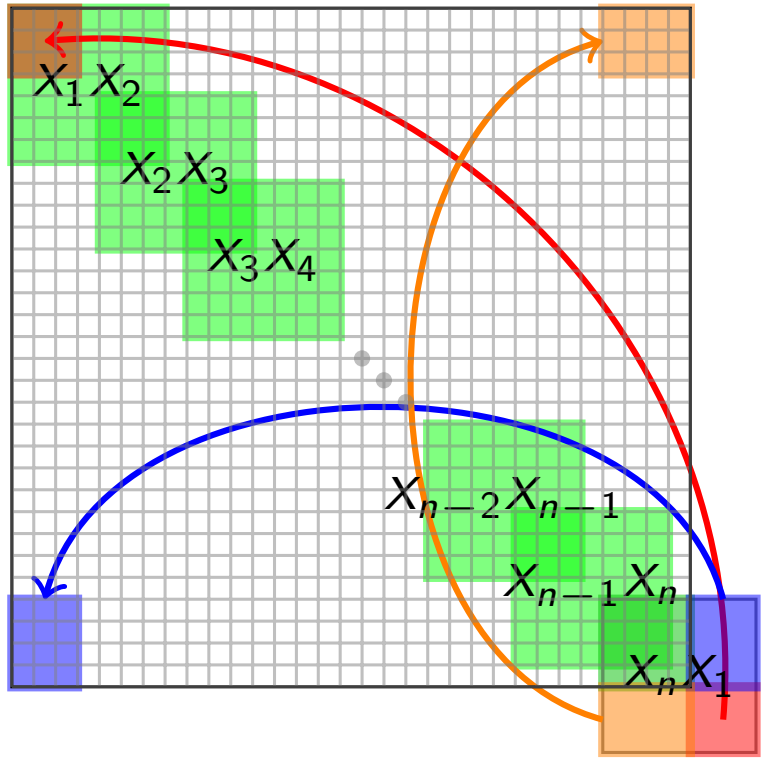


Figure: Periodic cgDNA+ stiffness matrix,
the cgDNA case is analogous.

cgDNA+min

cgDNA+min

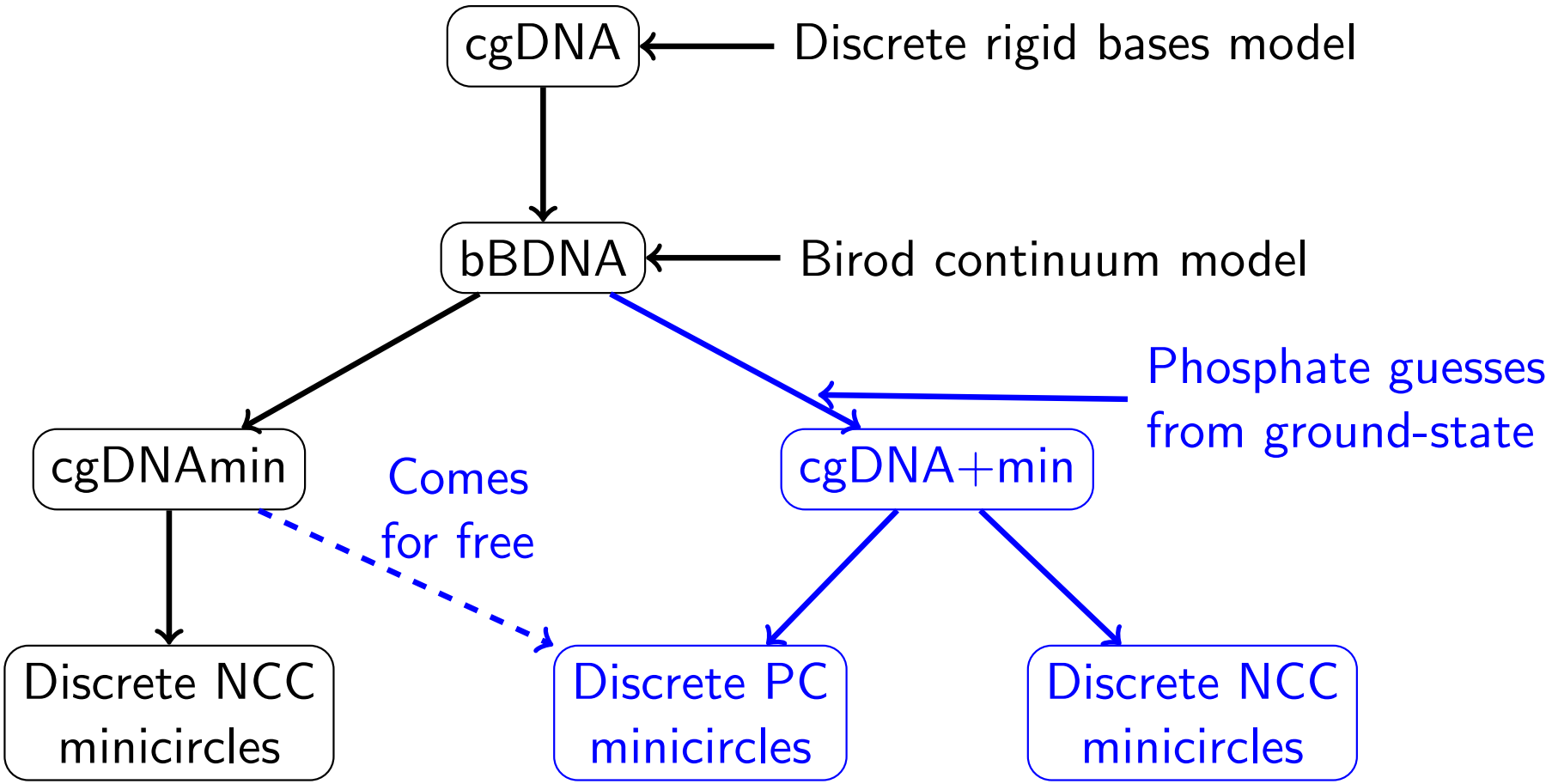
- Base algorithm from cgDNAmin
- Change in coordinates
  - cgDNA+ base pair level coordinates
  - New discrete energy
  - Adapt gradient & Hessian
- Initial guess for cgDNA+ coordinates
  - cgDNA coordinates: from birod model obtained with cgDNA parameters (same as in cgDNAmin)
  - Phosphate coordinates: use ground-state coordinates.

Non-Continuous vs Periodic closures

Two different types of closure:

- Non-continuous closure (NCC):
  Original closure from R. Manning
  - Models **formation** of DNA minicircles
  - Use standard stiffness & ground-state
- Periodic closure (PC)
  - Models **fully formed** DNA minicircles
  - Use periodic stiffness & ground-state from Glowacki
  - Adapt gradient & Hessian
- Both closure assumptions **do not** model the same physical configuration!

Model flowchart



Example of results

Many examples can be found on the webpage:
5 sequences coming from the experimental literature with length ranging form 94 to 399 base pairs.

- Here: Kahn & Crothers sequence, also studied by Manning, Maddocks and Kahn (1996).
- Containing phased A-tracts $\Longrightarrow$ strong intrinsic bend

Example of results

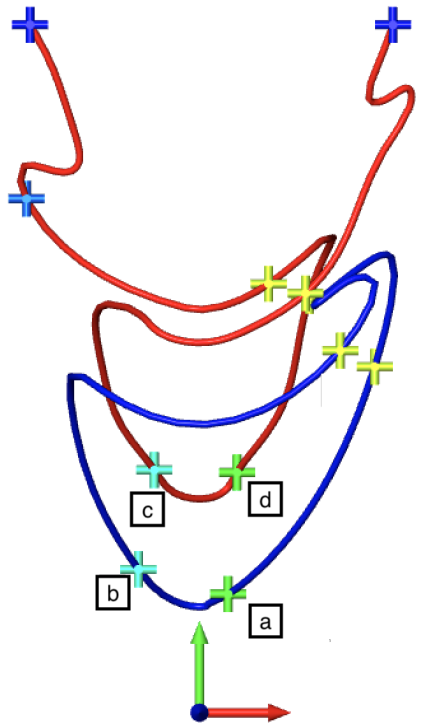
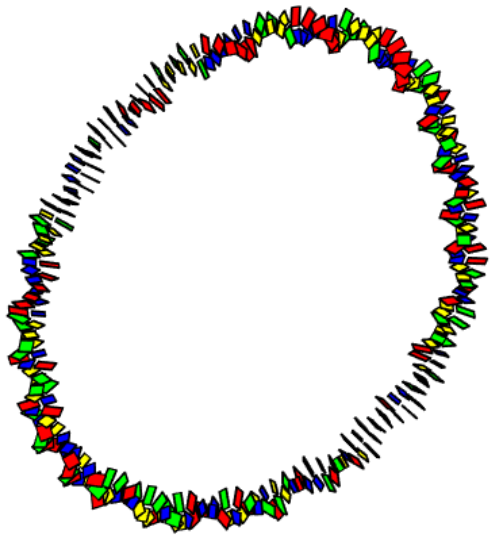
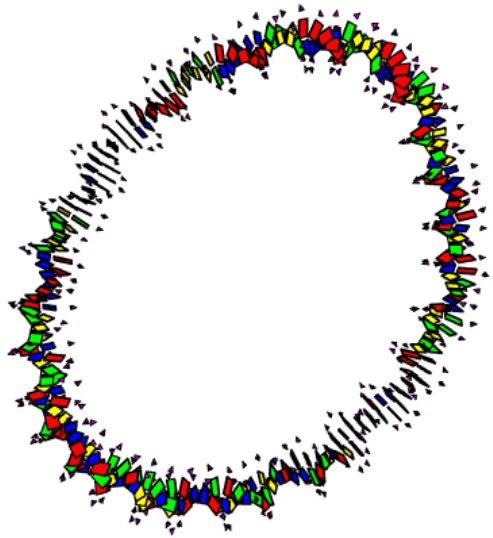


Figure: Bifurcation diagram for the Kahn & Crothers sequence.
On the next slides, we show 3D configurations of results for guesses a) and c).

Results: Kahn & Crothers, guess a)

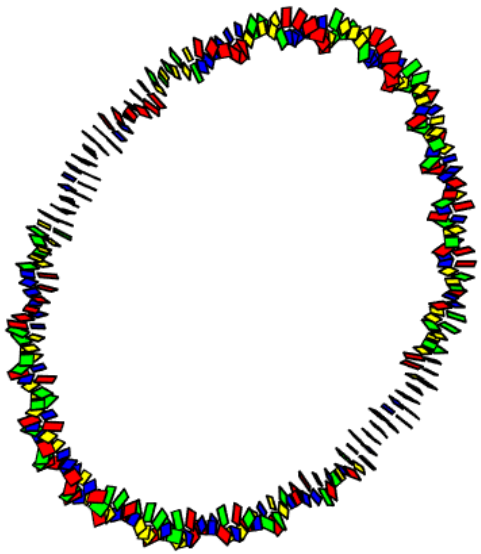


(a) NCC cgDNAmin
Link: 15, Energy: 35

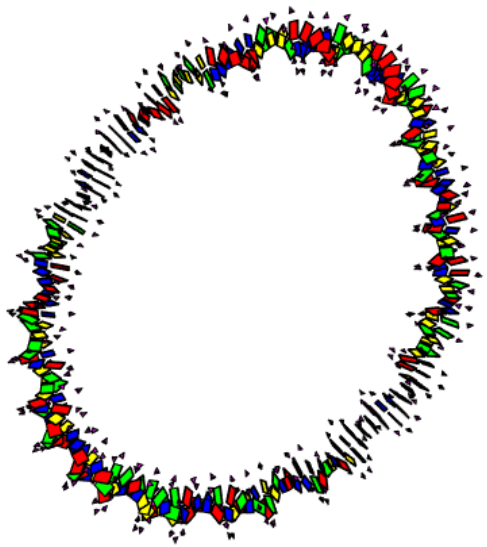


(b) NCC cgDNA+min
Link: 15, Energy: 45

Results: Kahn & Crothers, guess a)



(a) PC cgDNAmin
Link: 15, Energy: 35



(b) PC cgDNA+min
Link: 15, Energy: 47

Comparison NCC-PC, Kahn & Crothers, guess a)

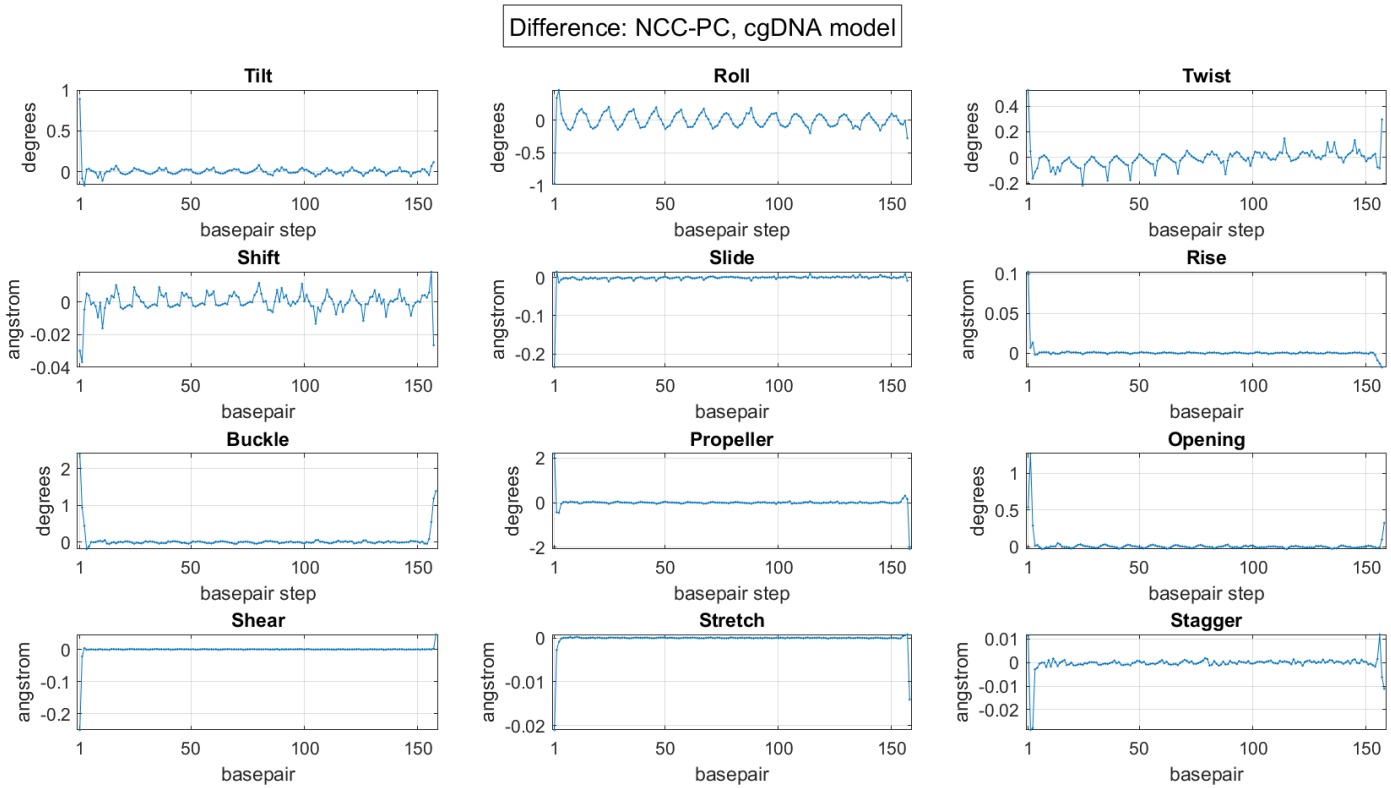


Figure: cgDNA coordinate difference, NCC - PC. Guess a)

Comparison NCC-PC, Kahn & Crothers, guess a)

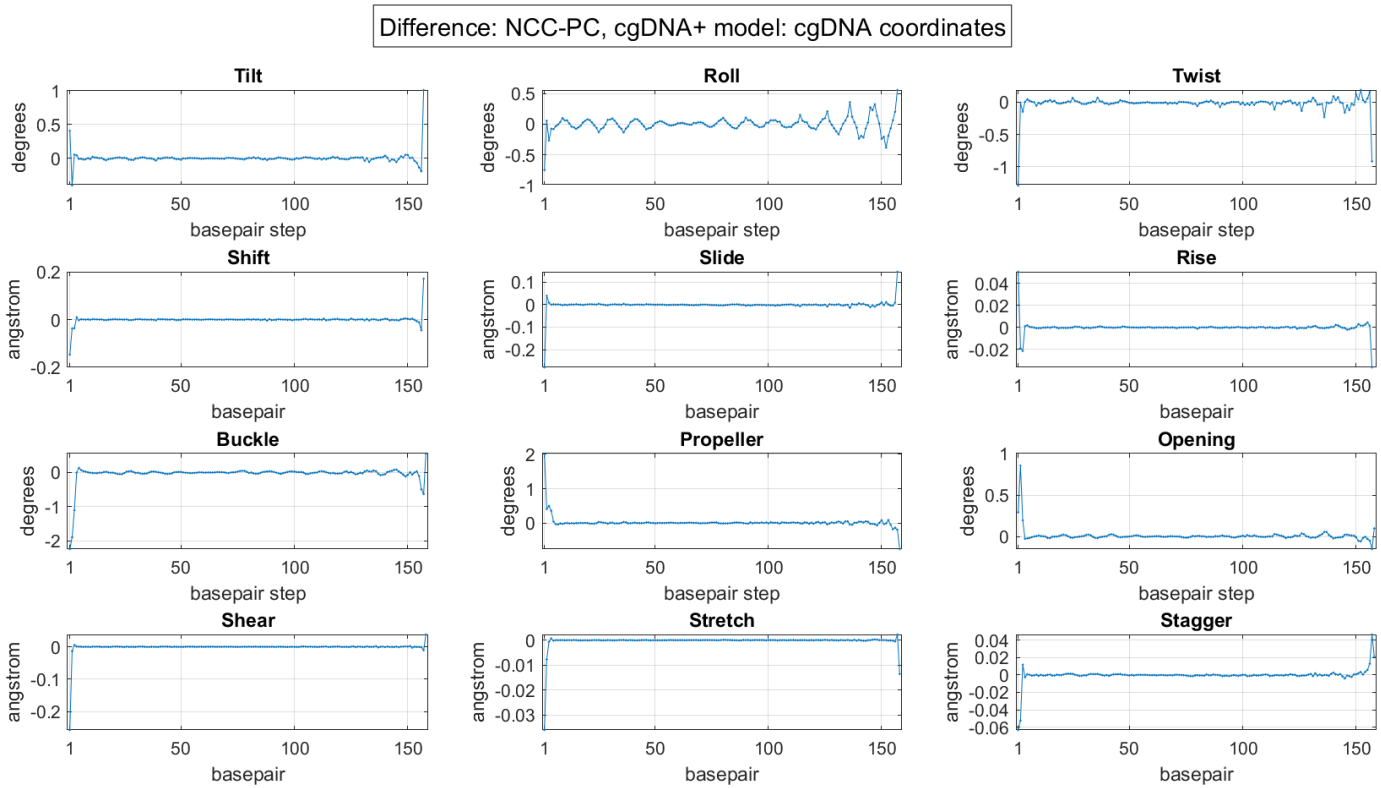


Figure: cgDNA+ coordinate difference, NCC - PC. Guess a)
We do not show the difference in Phosphate coordinates as it is very close to zero in all dimensions.

comparison cgDNA-cgDNA+

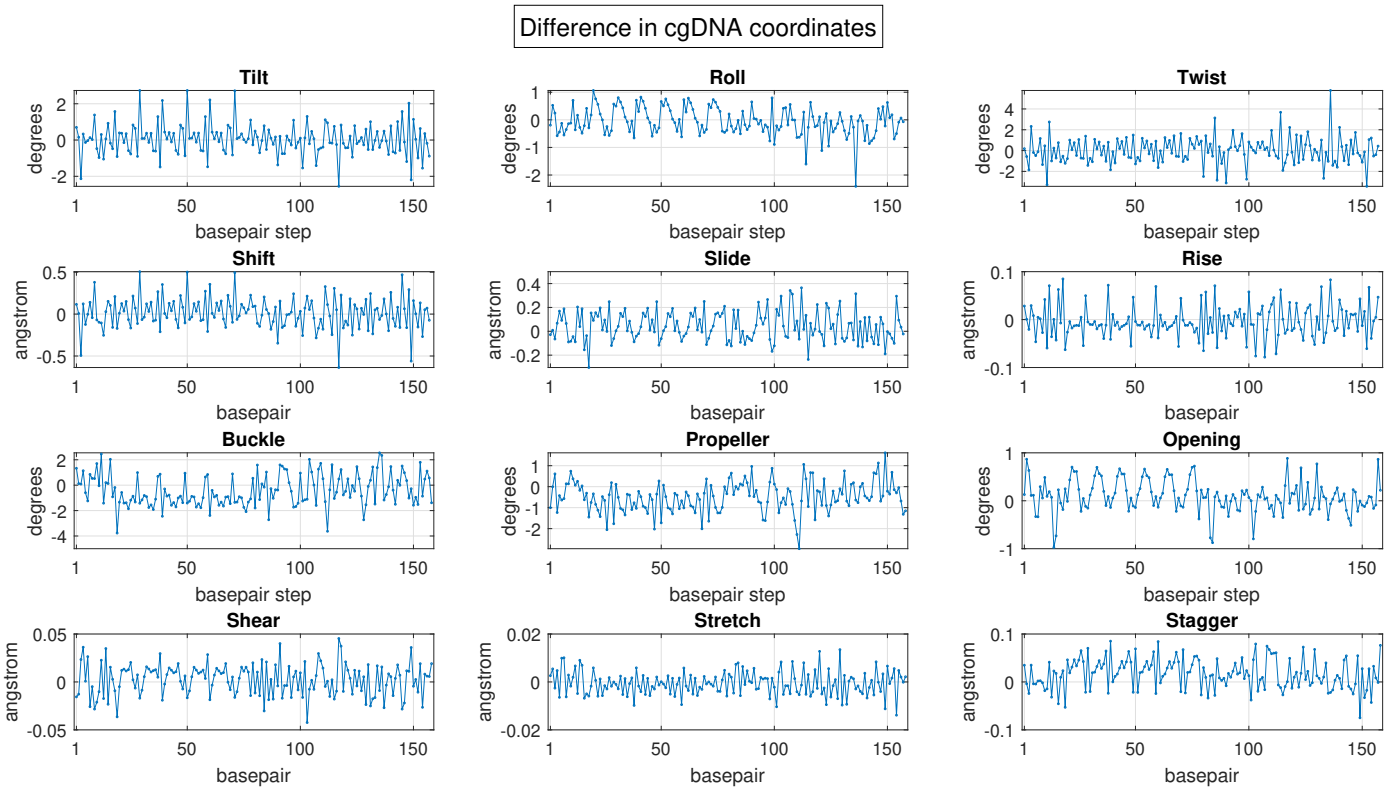
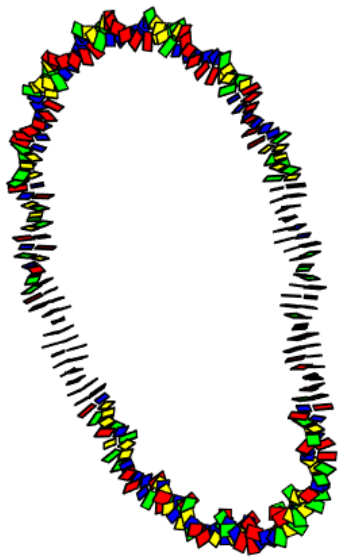
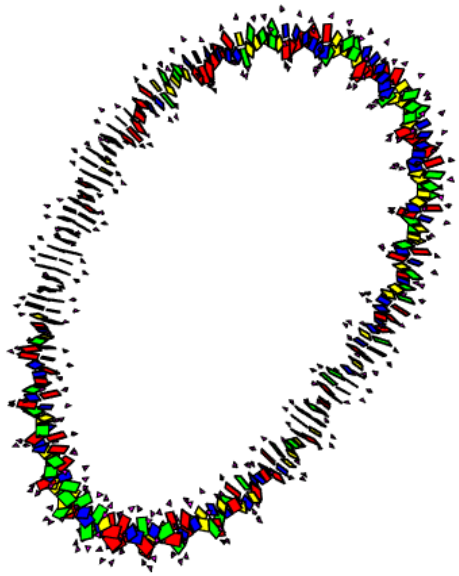


Figure: Comparison of Non-continuous (NCC) cgDNA and cgDNA+ models. Difference in the original cgDNA coordinates, Kahn & Crothers c11t15 sequence, guess a).

Results: Kahn & Crothers, guess c)

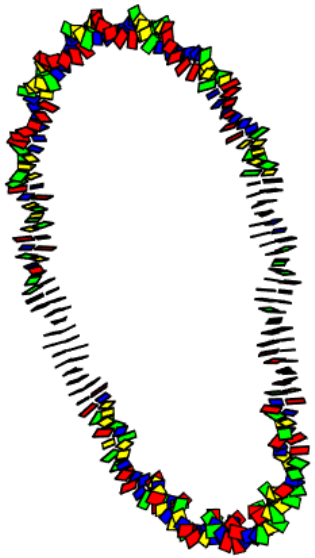


(a) NCC cgDNAmin
Link: 14, Energy: 84

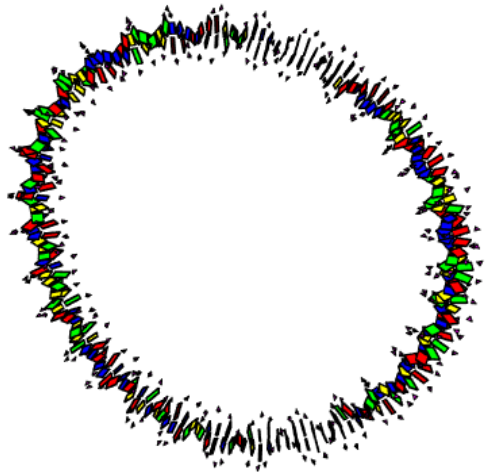


(b) NCC cgDNA+min
Link: 16, Energy: 110

Results: Kahn & Crothers, guess c)



(a) PC cgDNAmin
Link: 14, Energy: 84



(b) PC cgDNA+min
Link: 16, Energy: 117

Comparison NCC-PC, Kahn & Crothers, guess c)

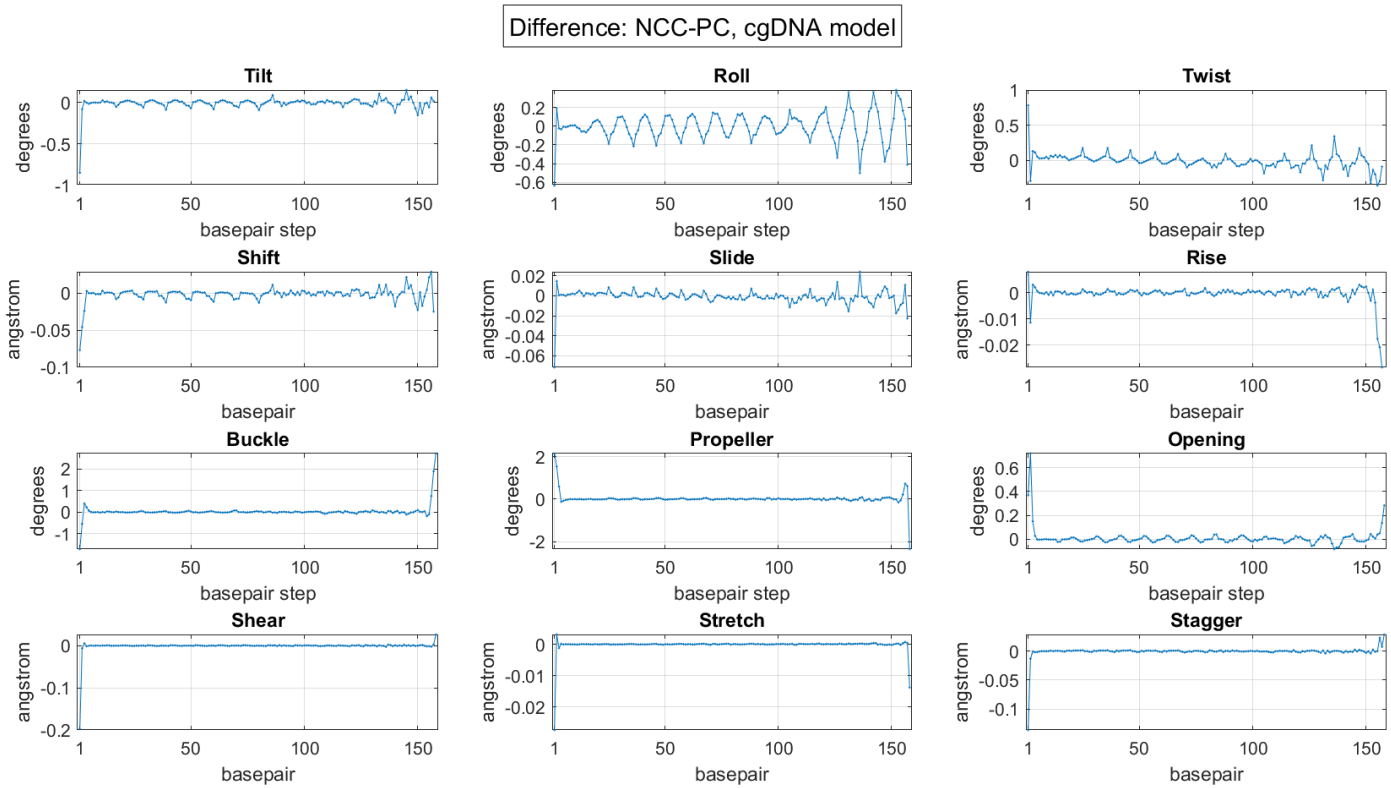


Figure: cgDNA coordinate difference, NCC - PC. Guess c)

Comparison NCC-PC, Kahn & Crothers, guess c)

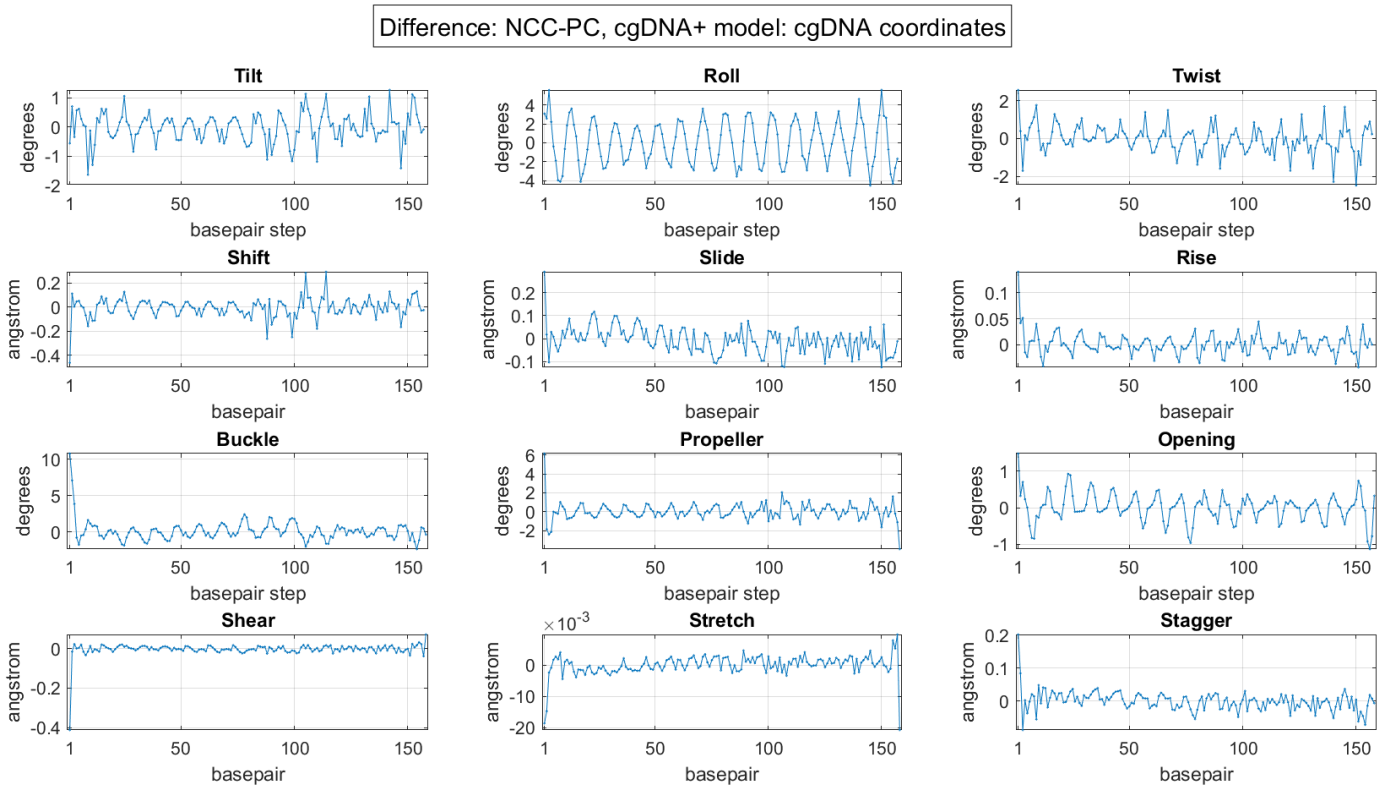


Figure: cgDNA+ coordinate difference, NCC - PC. Guess c).

Comparison NCC-PC, Kahn & Crothers, guess c)

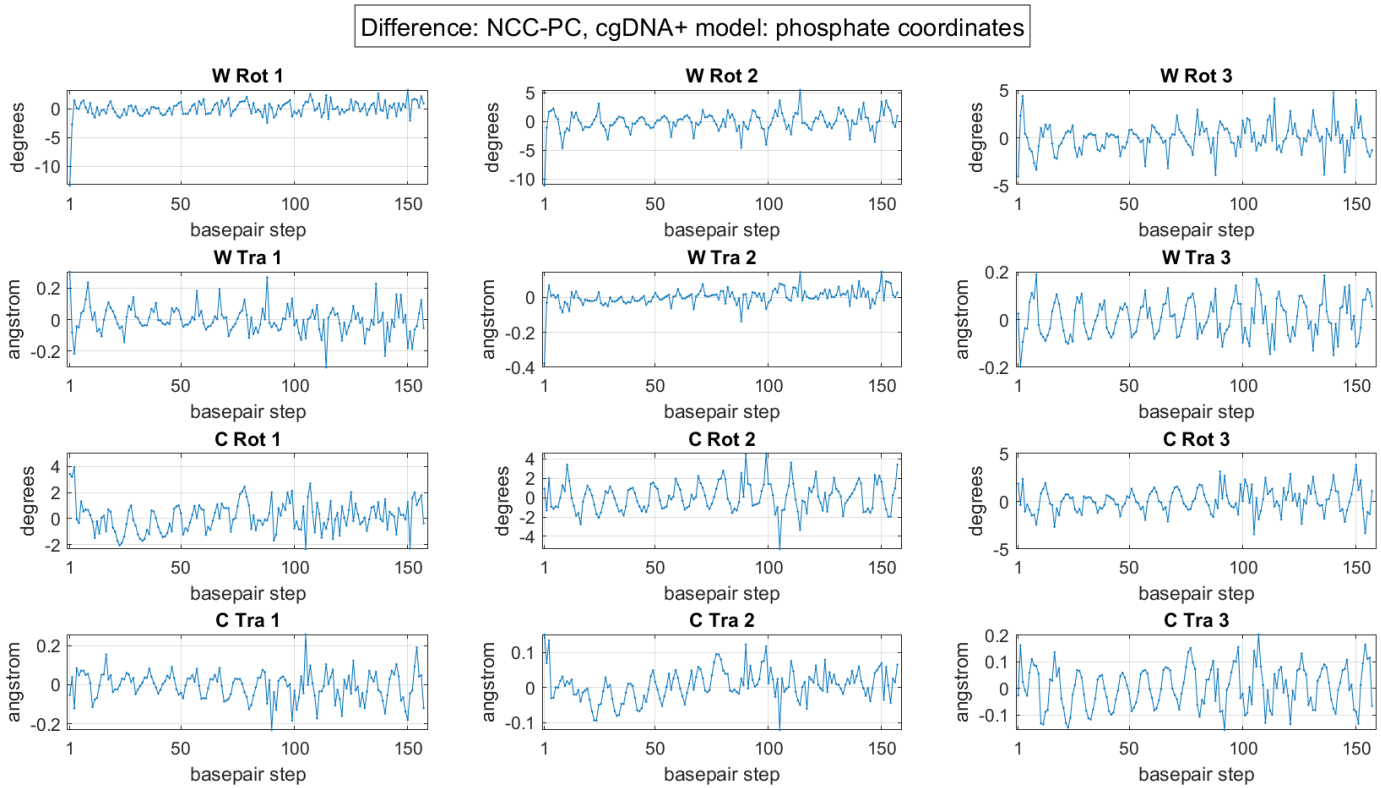


Figure: cgDNA+ phosphate coordinate difference, NCC - PC. Guess c).

comparison cgDNA-cgDNA+

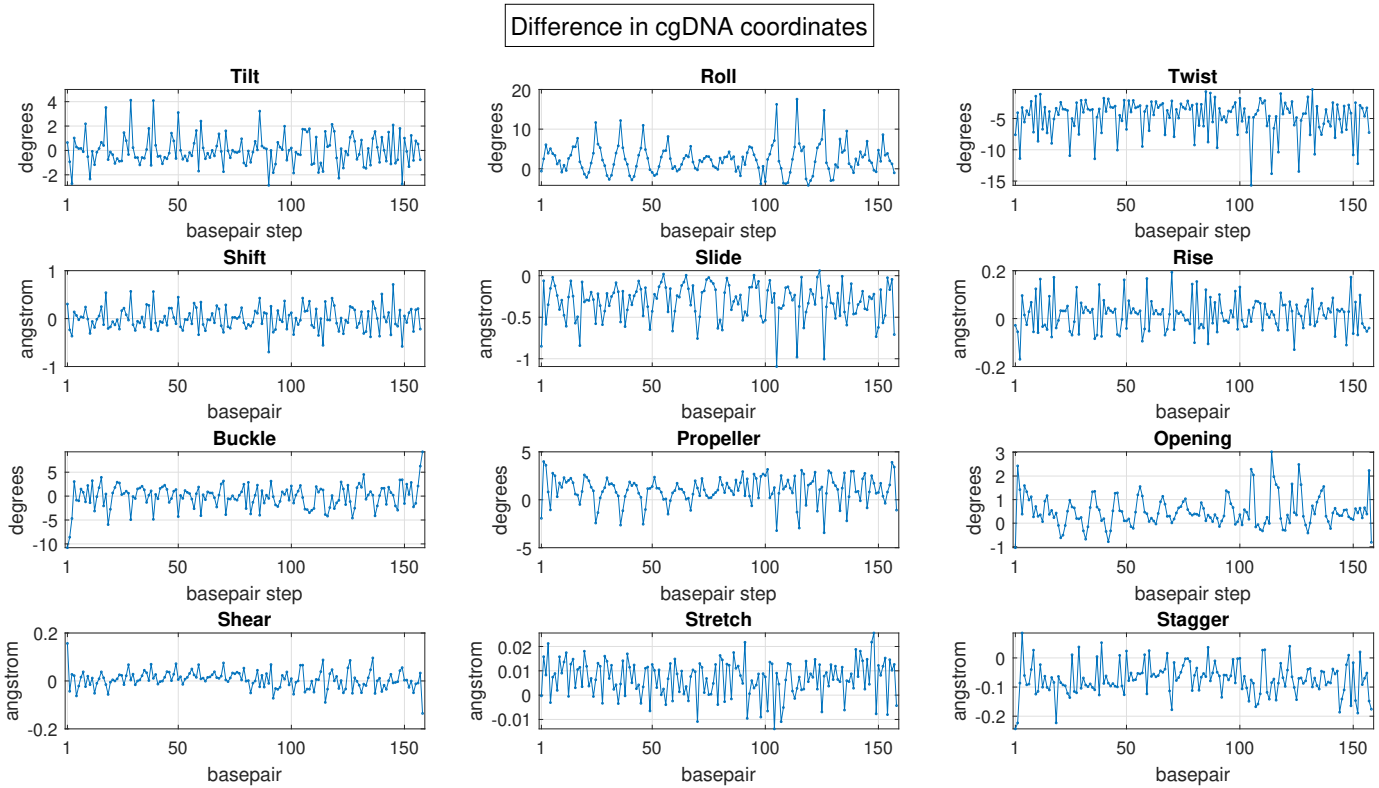


Figure: Comparison of Non-continuous (NCC) cgDNA and cgDNA+ models. Difference in the original cgDNA coordinates, Kahn & Crothers c11t15 sequence, guess c).

Observations

- NCC vs PC
  - Similar in most cases
  - relative difference of the norms of the coordinates vector in the 5% range
  - relative difference in energy in the 1-2% range.
- cgDNA vs cgDNA+
  - Some cases have very different solutions
  - We suppose it is due to difference in stability:
    Unstable equilibrium from bBDNA may not converge to the same stable discrete configurations.
- General observations
  - Link is not conserved between initial guess and final solution.
  - Some cases show negative smallest eigenvalue of the Hessian $\rightarrow$ this need investigation.

MATLAB scripts

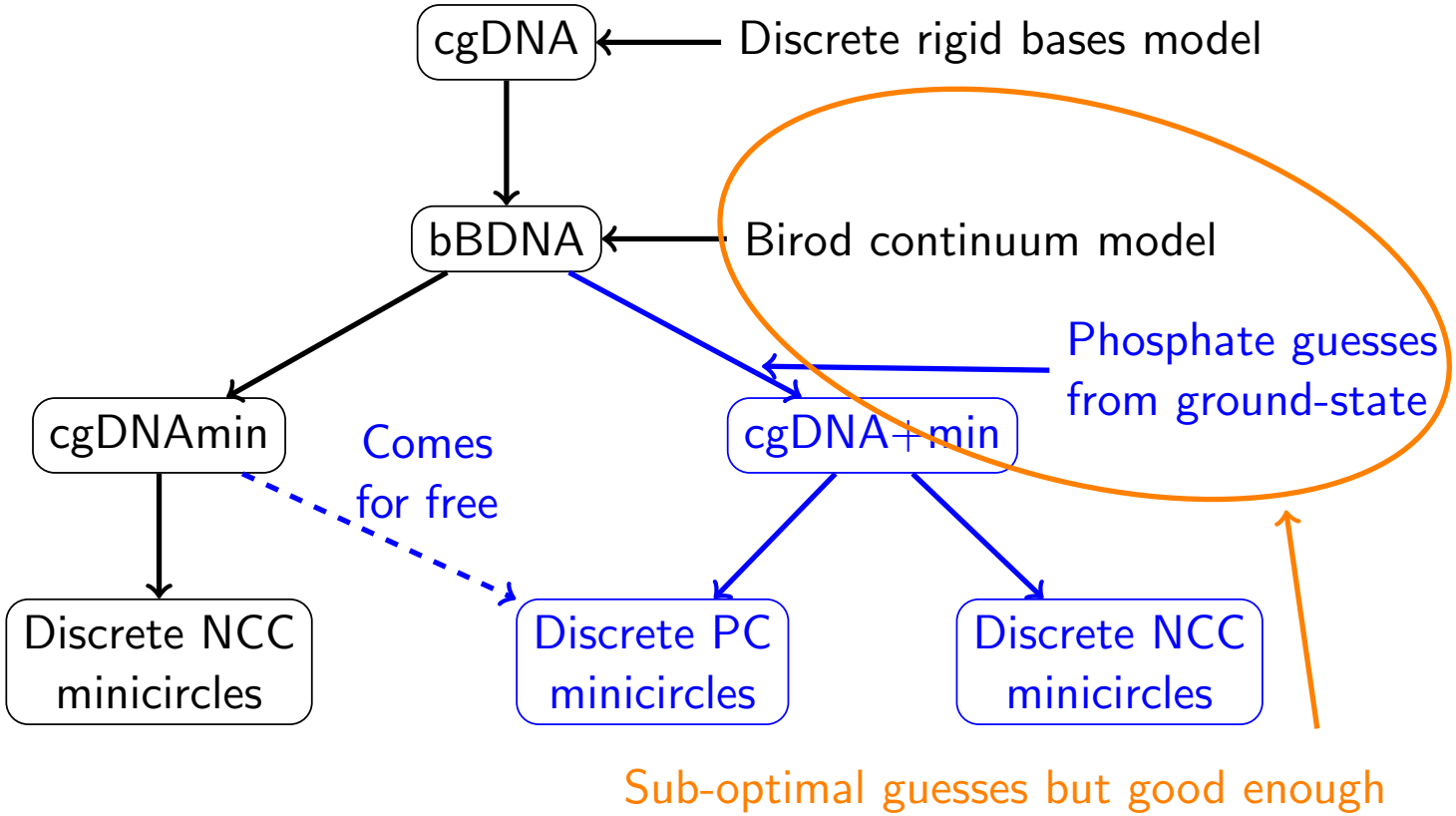
- Energy_min.m
  - Choose sequence
  - Choose model (cgDNA or cgDNA+, NCC or PC)
  - Input initial guess
  - Outputs Discrete energy minimization results
- Visualize_solution.m
  - Choose sequence
  - Choose model (cgDNA or cgDNA+, NCC or PC)
  - Input solution from Energy_min.m
  - Outputs figures shown in the webpage

DEMO

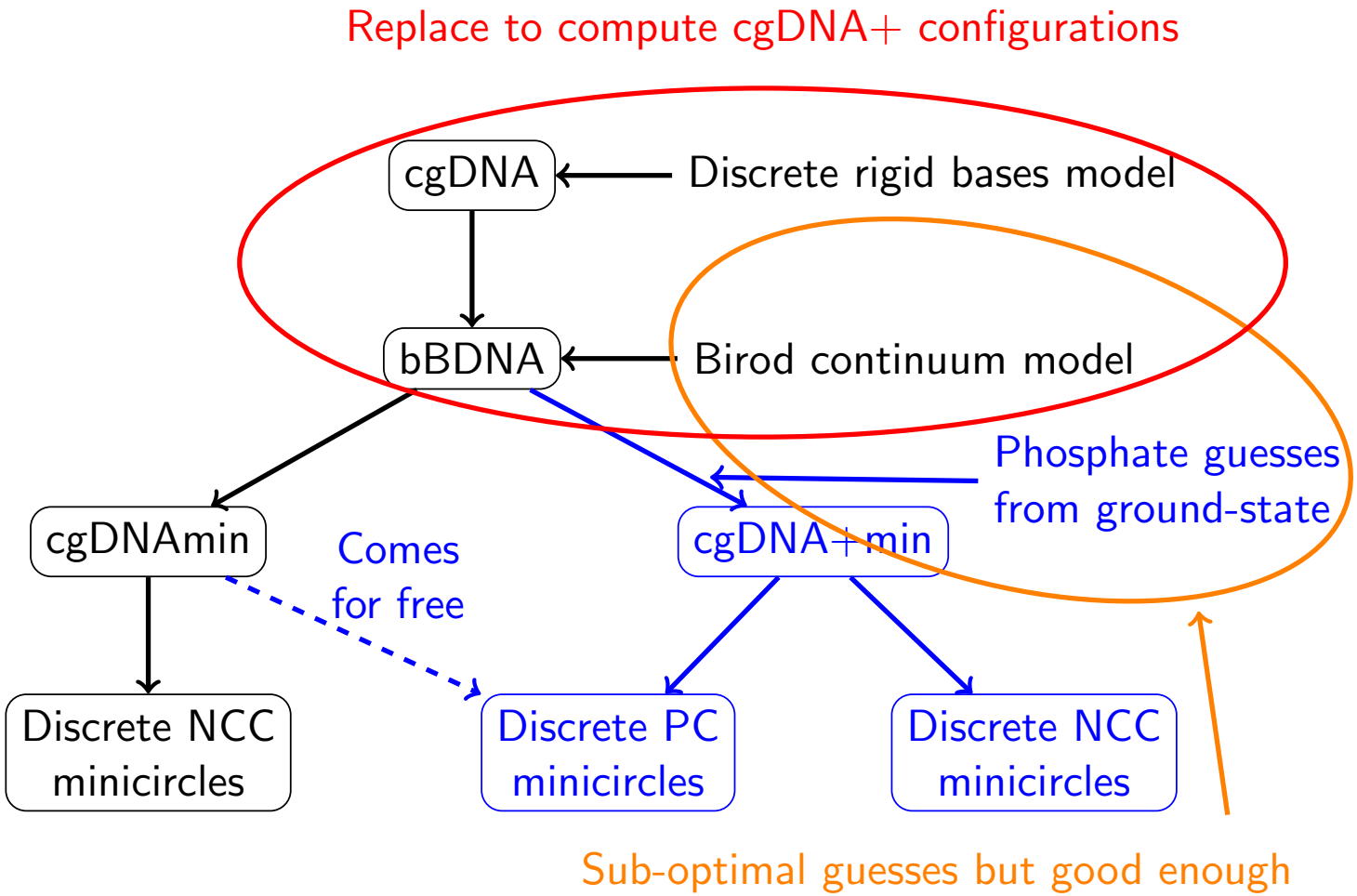
Conclusion

- We have a model for discrete DNA minicircles.
- Input: Sequence, cgDNA+ parameter set, cgDNA parameter set.
- Output: minicircle configurations.

Point of improvement



Point of improvement



Point of improvement

We tried to improve the initial guesses:

- Continuum coefficient obtained from cgDNA parameters
- Try to use cgDNA+ parameters
- Issue: need to recover separate 18×18 positive definite blocks from the stiffness.
  After marginalizing, unable to get consistent positive definite blocks

The End

Thanks for listening!

Any questions?

Main Resources

- Manning, R. *"Notes on cgDNAmin, Discrete-birod DNA Cyclization"*. unpublished. 2017.
- Patelli, A. S. *"A sequence-dependent coarse-grain model of B-DNA with explicit description of bases and phosphate groups parametrised from large scale Molecular Dynamics simulations"*. PhD thesis. EPFL, 2019.
- Glowacki, J. *"Computation and Visualization in Multiscale Modelling of DNA Mechanics"*. PhD thesis. EPFL, 2016.
- Gonzalez, O. et al. *"A sequence-dependent rigid-base model of DNA"*. Journal of Chemical Physics 138, no. 5 (2013), p. 055122 1-28.
- Petkevičiūtė, D. et al. *"cgDNA: a software package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA"*. Nucleic Acids Research 42, no. 20 (2014), p. e153

- The details of all other resources can be found on the thesis provided on the webpage:
  https://lcvmwww.epfl.ch/research/cgDNA/beaud/index.html