Computations with the cgDNA+ coarse-grain model of DNA

Federica PADOVANO

June 1° 2022

EPFL

LCVM²

Goal of the semester project

bBDNA software:

- significant **computational cost**
- necessary for the initial guess

Find a **procedure to find a good initial guess** for the energy minimization without using bBDNA software

Previous works

- cgDNA+ → dsDNA structure
- cgDNA+min → dsDNA minicircle structure

cgDNA+

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

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

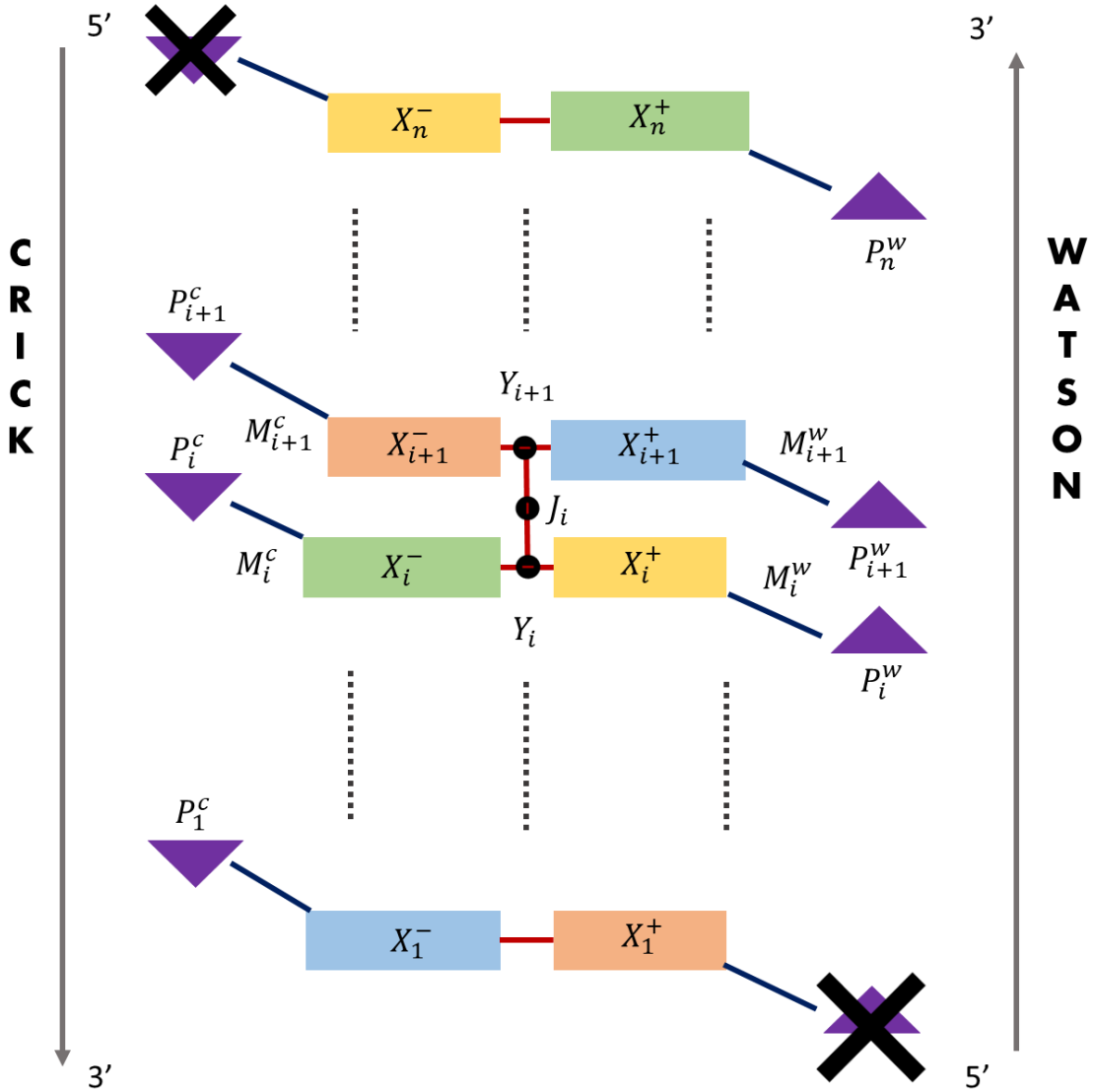
- Coordinates:

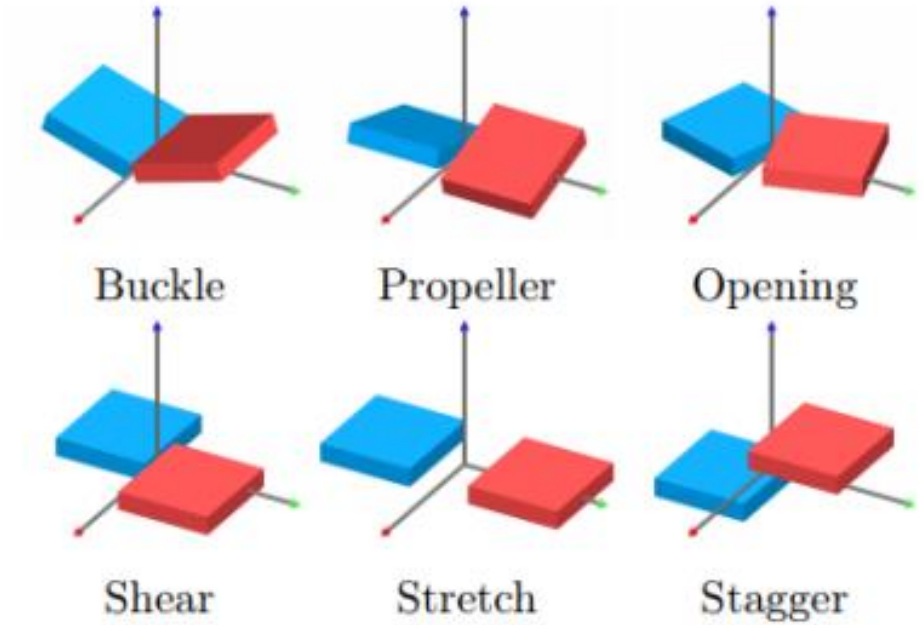
$$w = (x_1, x_1^c, y_1, x_2^w, x_2, x_2^c, y_2, ..., x_i^w, x_i, x_i^c, y_i, ..., y_{n-1}, x_n^w, x_n) \in \mathbb{R}^{24n-18}$$

$$where \quad x_i, x_i^c, x_i^w, y_i \in \mathbb{R}^6,$$

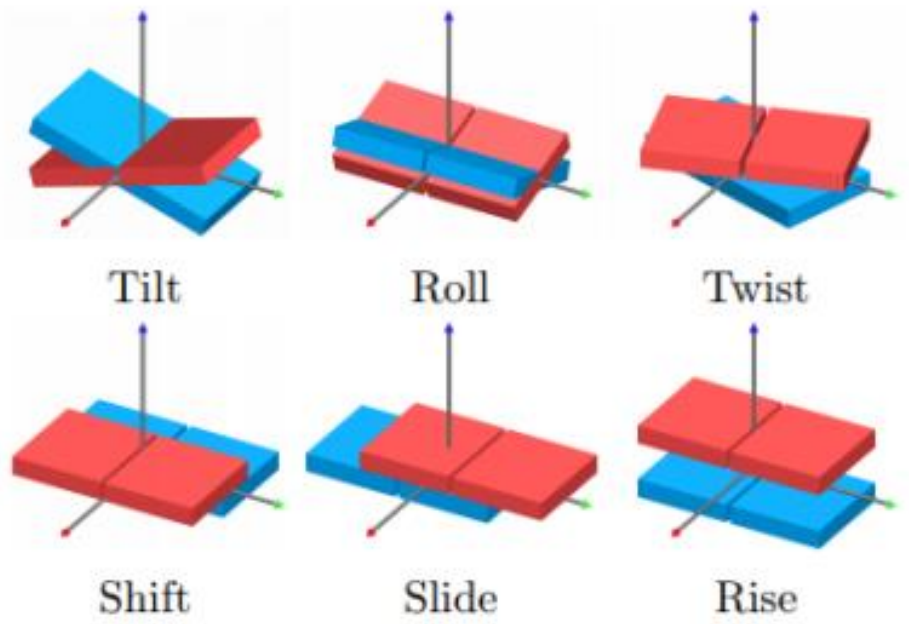
cgDNA+

- Rigid bodies
- Relative coordinates w.r.t. a midway frame
- Cayley vector + Translation vector
- Base + Base pair + Phosphate





(a) Intras

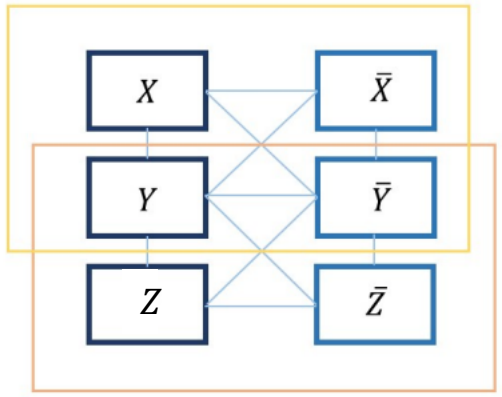


(b) Inters

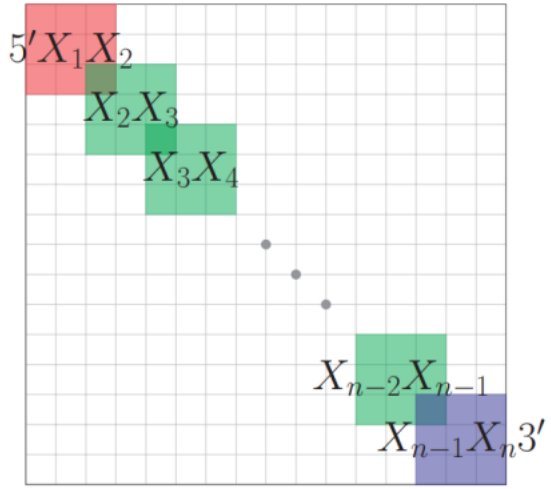
cgDNA+

Three assumptions made on the energy:

- has a **shifted quadratic form**
- the total energy is a **sum over level junctions energies**
- The coefficient in the local junction energy **depends on the local dimer** sequence step



(a) Interactions among the neighbours.



(b) Stiffness matrix construction.

cgDNA+ Periodicity

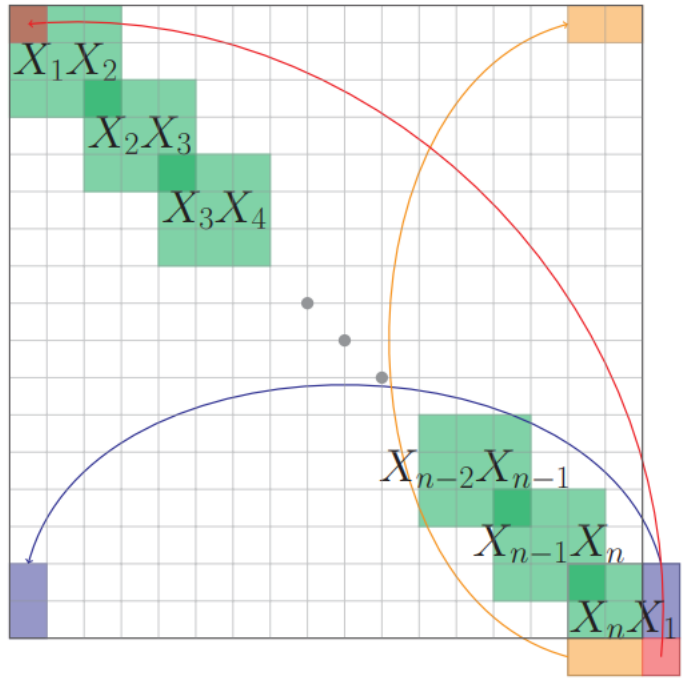
1 2 N

$X_1 \dots X_n$ $X_1 \dots X_n \dots \dots$ $X_1 \dots X_n$

BASE SEQUENCE

- 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

STIFFNESS MATRIX



cgDNA+min

- 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 $U(z) = \frac{1}{2}(F(z) - \mu)^T K (F(z) - \mu)$

cgDNA+min

- Construct periodic stiffness matrix and periodic ground-state (parameter set)
- Run bBDNA software to obtain a continuum initial guess
- Discretize the initial guess
- Run the energy minimization starting in the initial guess:
  - Assumptions for the continuum model are slightly different than the ones of the discrete case
  - During the discretization process, errors are induced.
- Convert quaternions into inter variables

cgDNA+min

Assumptions on energy and quaternions:

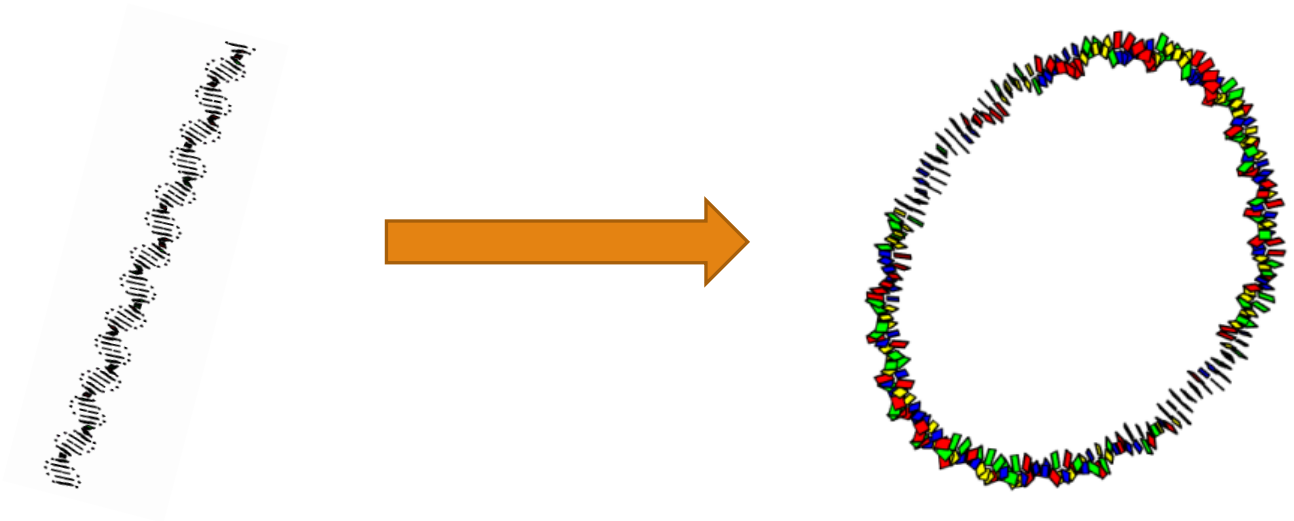
- has a **shifted quadratic form**
- the total energy is a **sum over level junctions energies**
- The coefficient in the local junction energy **depends on the local dimer** sequence step
- Contraint: $\|q_i\|^2 = 1 \; \forall i = 1 \ldots, n$

$$\mathcal{L}(z) = \frac{1}{2}(F(z) - \mu)^T K(F(z) - \mu) + \lambda \sum_{i=1}^{n} (\|q_i\|^2 - 1)^2$$

Construct an initial guess

Two main steps:

- generate special **helicoidal configuration** that have specific integer link number **m**
- **deform** the helicoidal equilibrium into a **twisted circle**



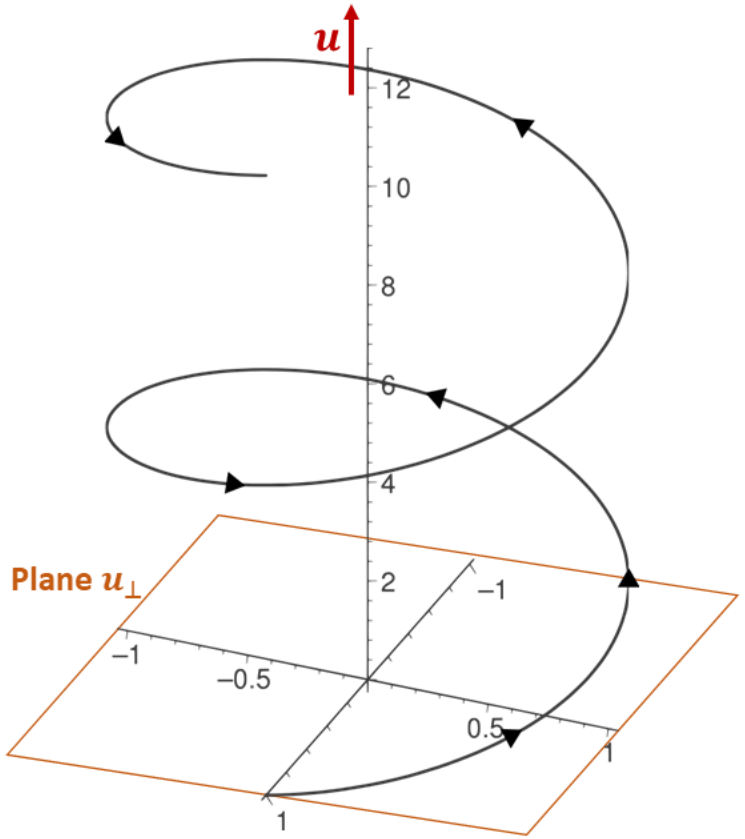
Helicoidal configuration

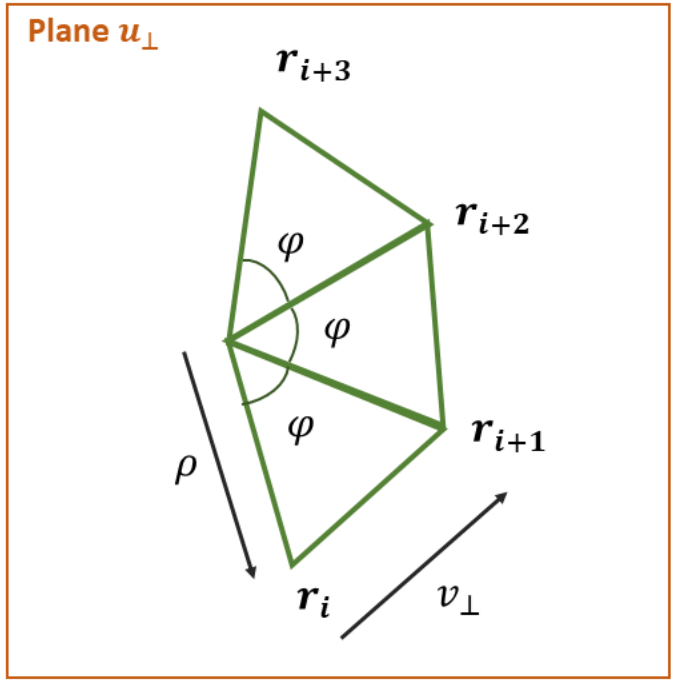
$$w = (x_1, u, v, x_2, u, v, \ldots, x_n, u, v) \in \mathbb{R}^{24n}$$

- the vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$ (***inter*** *variables*) are always the same
- uniformity of ***inter*** *variables*
- rotation axis ($\boldsymbol{u}$) is parallel to the helix axis
- each base pair frame origin has the same distance ($\boldsymbol{\rho}$) from the axis center-line
- relative distance between two consecutive base pairs is the same

Helicoidal configuration

- base pairs frame axis rotate through $\boldsymbol{u}$
- $\boldsymbol{v}_{\perp}$ = translation in the plane $u_{\perp}$
- $\boldsymbol{v}_{\parallel}$ = translation along direction $\boldsymbol{u}$
- $\boldsymbol{\varphi}$ = angle of rotation
- $\rho = \frac{1}{2} \frac{v_{\perp}}{\sin(\varphi/2)}$





Helicoidal configuration

$$z = (x_1, x_2, \ldots, x_n, u, v) \in \mathbb{R}^{18n+6}$$

$$Pz = w$$

$$P = \begin{bmatrix} I_{18} & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & I_6 \\ 0 & I_{18} & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & I_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & I_{18} & 0 \\ 0 & 0 & \ldots & 0 & I_6 \end{bmatrix} \in \mathbb{R}^{24n \times 18n+6}$$

ENERGY:

$$U^*(z) = \frac{1}{2}(Pz - \mu)^\top K(Pz - \mu)$$

$$\nabla U^*(z) = 0 \iff P^\top KPz = PK\mu$$

Solve this equation to find the global minimum

Helicoidal configuration

- the number of links between the two ends must be an integer number **m**

$$\varphi n = 2\pi m$$

$$\|u\| = 10 \tan \frac{\varphi^*}{10}$$

$$\|u\| = 10 \tan \frac{\pi m}{n}$$

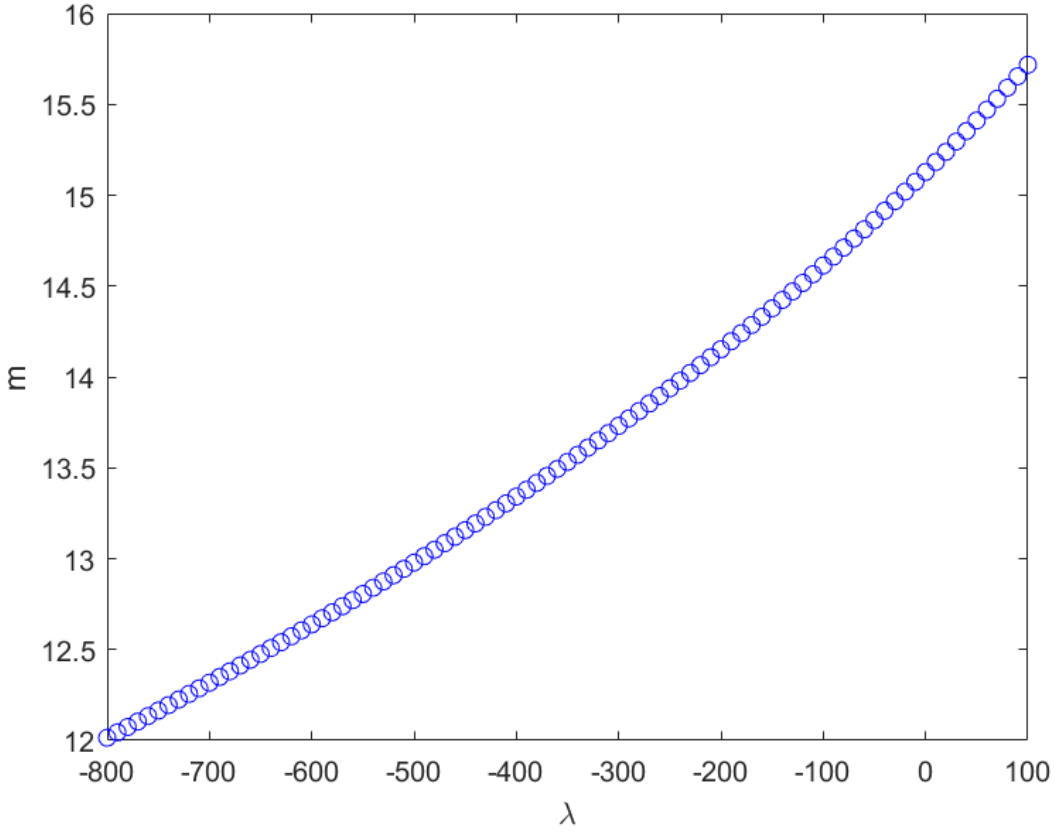
$$z^T E z = \|u\|^2 \qquad E = \begin{bmatrix} 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & I_3 & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{18n+6 \times 18n+6}$$

$$\mathcal{L}(z;\lambda) = U^*(z) + \lambda h(z)$$

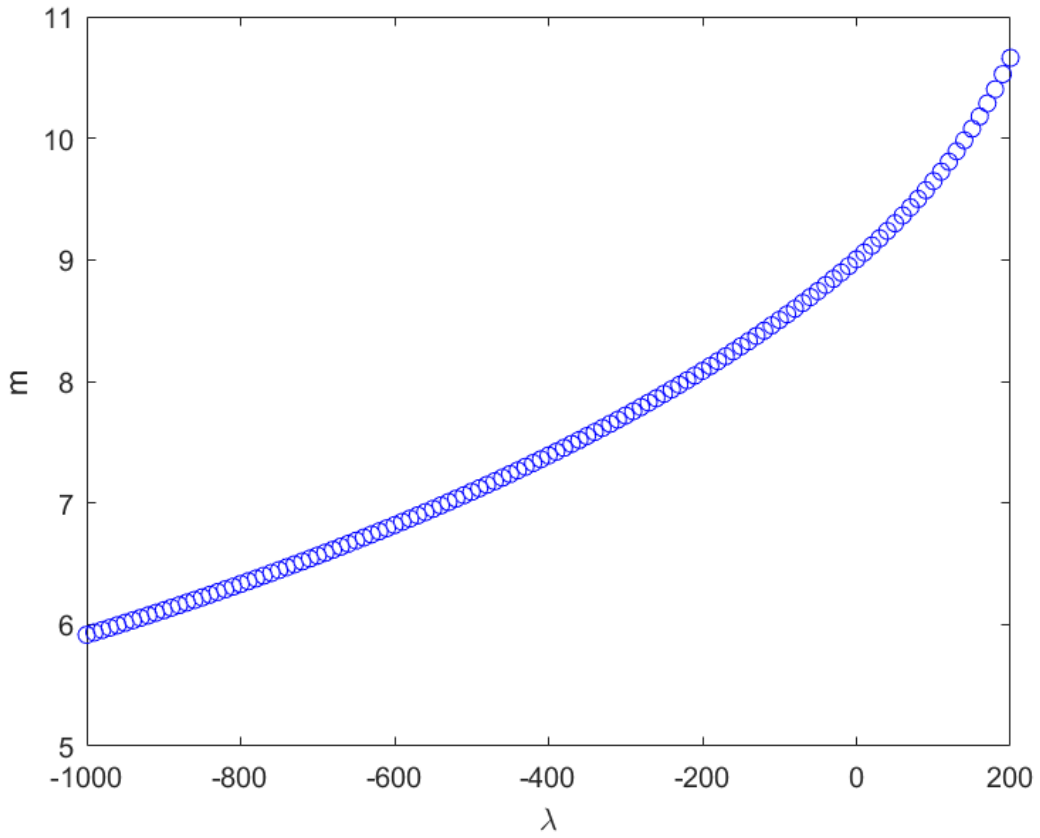
$$\nabla \mathcal{L}(z,\lambda) = 0 \iff (P^T K P - \lambda E) z = P K \mu$$

Solve this equation to find the minimum

Kahn-Crothers



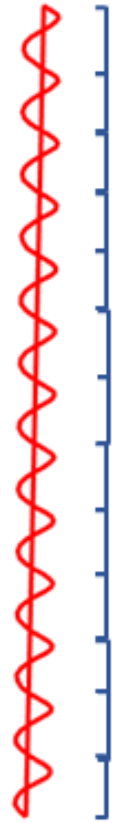
Widom 601



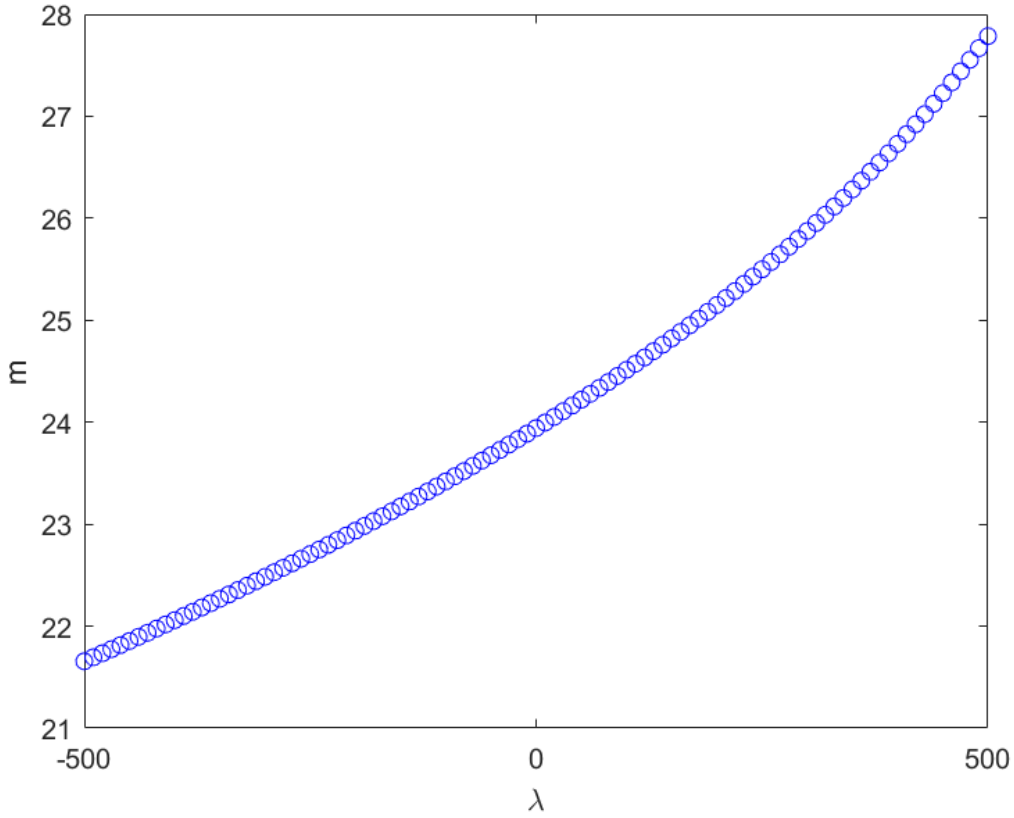
Ex. Kahn-Crothers

$\lambda = -493.91$

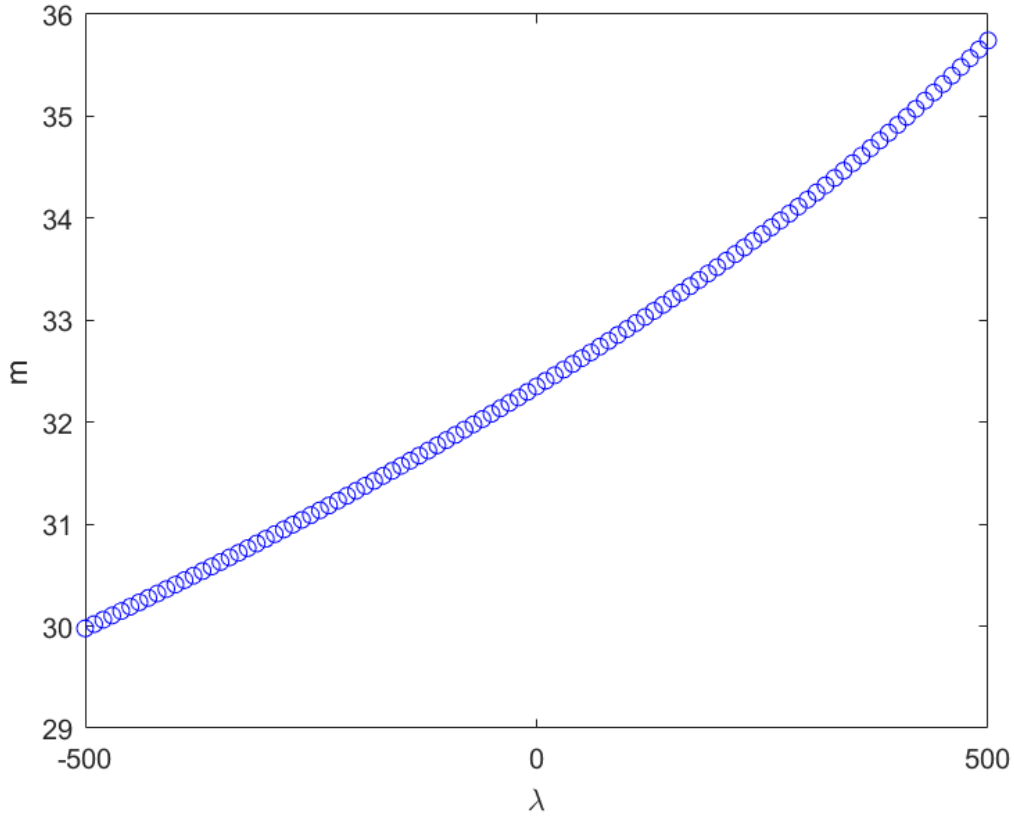
$m = 13$



Pyne 251bp

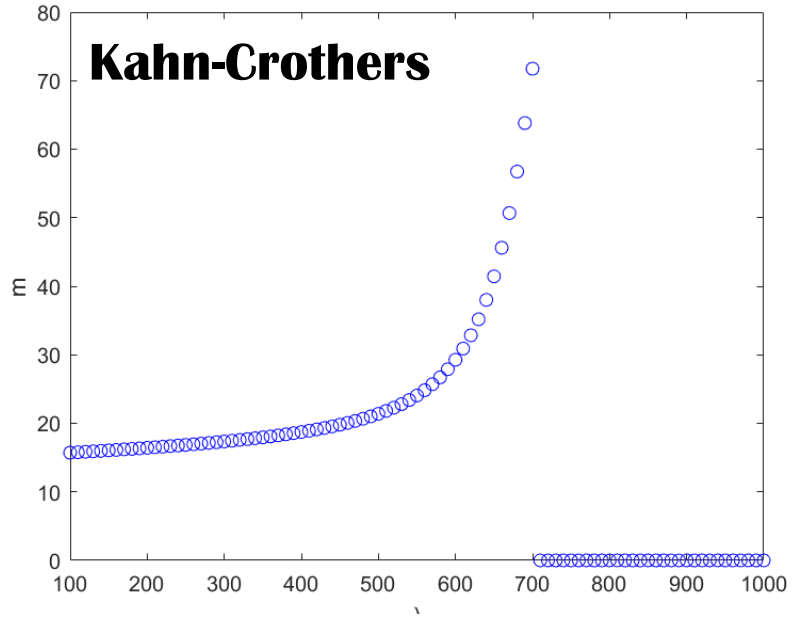


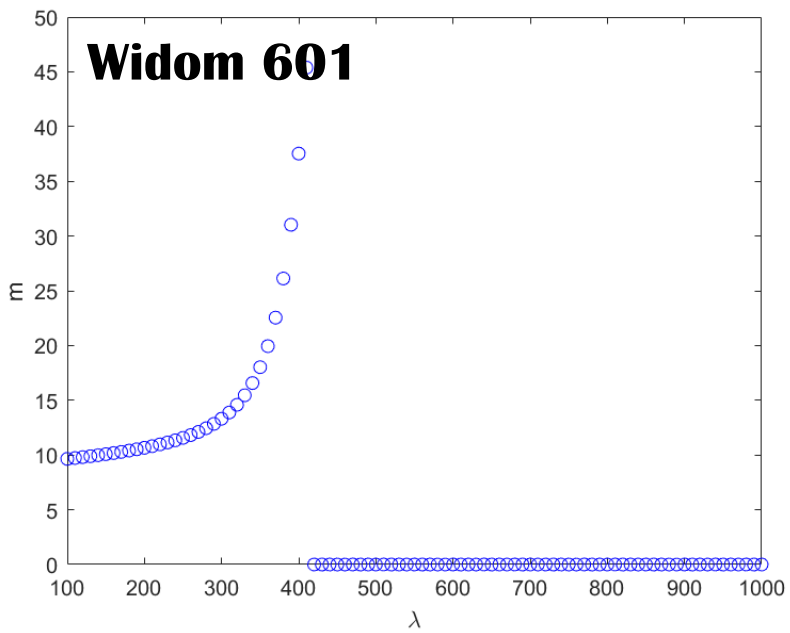
Pyne 339bp

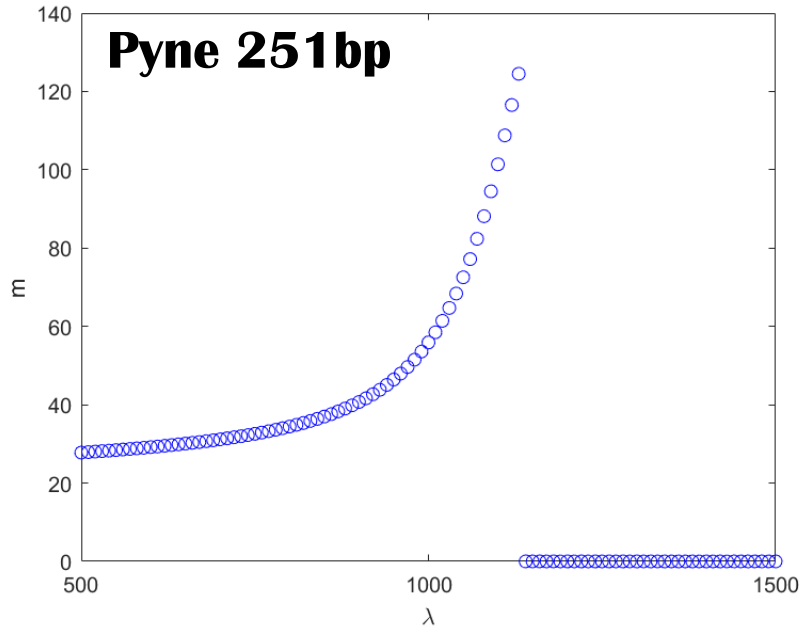


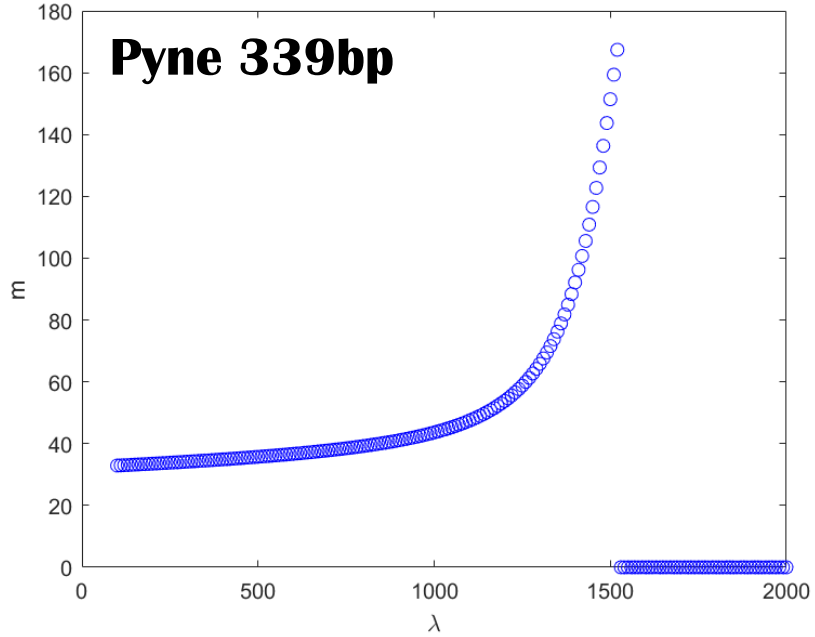
In general we should check that the **matrix** is **symmetric** and **positive definite**

$$(P^T KP - \lambda E)z = PK\mu$$









Analysis on initial guesses

*If we **change inter variables of the periodic ground-state**, while intras and phosphate are left as they are, we can find a good initial guess for the energy optimization*

Consider two different initial guesses for the energy optimization:

- configuration obtained with the **bBDNA software**
- configuration in which the **intra** and **phosphate** coordinates are the same of the **periodic ground-state**, while the **inter** coordinates are the ones of the configuration obtained with the **bBDNA software**

Analysis on initial guesses

Sequences:

- Kahn-Crothers
- Widom 601
- Pyne 251bp
- Pyne 339bp

MatLab fminunc:

- Trust Region
- Quasi-Newton

Results

- *Kahn-Crothers* (4 i.g.): they all converge to the same solution, and they are all local minima → *Abs. difference:* ***1.0e-3*** *Negative eigenvalues:* ***0***

- *Pyne 251* (4 i.g.): they all converge to the same solution, and they are all local minima → *Abs. difference:* ***1.0e-2*** *Negative eigenvalues:* ***0***

Results

- *Widom 601* (4 i.g.):
  - GUESS 1: absolute difference 1.0e-3, but one has a negative eigenvalue (-1.3285e-05) and the other one has all positive eigenvalues (minimum is 3.4011e-05)

    → *Abs. difference:* ***1.0e-3***
    *Negative eigenvalues: {**1 ,0**}*
  - GUESS 2: converge to the same point which is not a minimum (*Quasi-Newton + Trust Region*)

    → *Abs. difference:* ***1.0e-3***
    Negative eigenvalues: ***2***
  - GUESS 3: converge to the same saddle point

    → *Abs. difference:* ***1.0e-3***
    Negative eigenvalues: ***1***
  - GUESS 4: they all converge to the same solution, and they are all local minima (*Quasi-Newton + Trust Region*)

    → *Abs. difference:* ***1.0e-3***
    Negative eigenvalues: ***0***

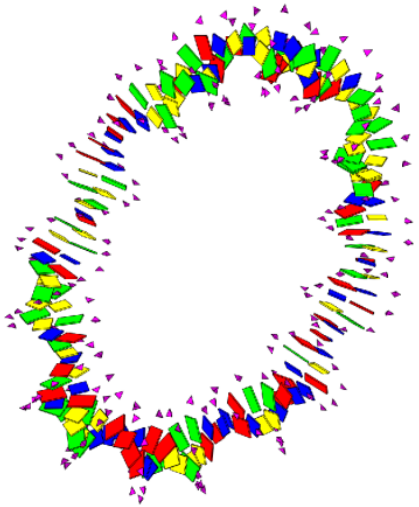
Guess 4, Widom 601bp

- *If we run **Trust Region** method we find **two different solutions**:*
  - *one equal to M. Beaud one* **Energy: 213**
  - *one closer to the initial guess* **Energy: 280**

- *If we run **Quasi-Newton + Trust Region** method they **both converge** to the one Beaud found*

*Abs. difference: **1.0e-4***

Negative eigenvalues: 0

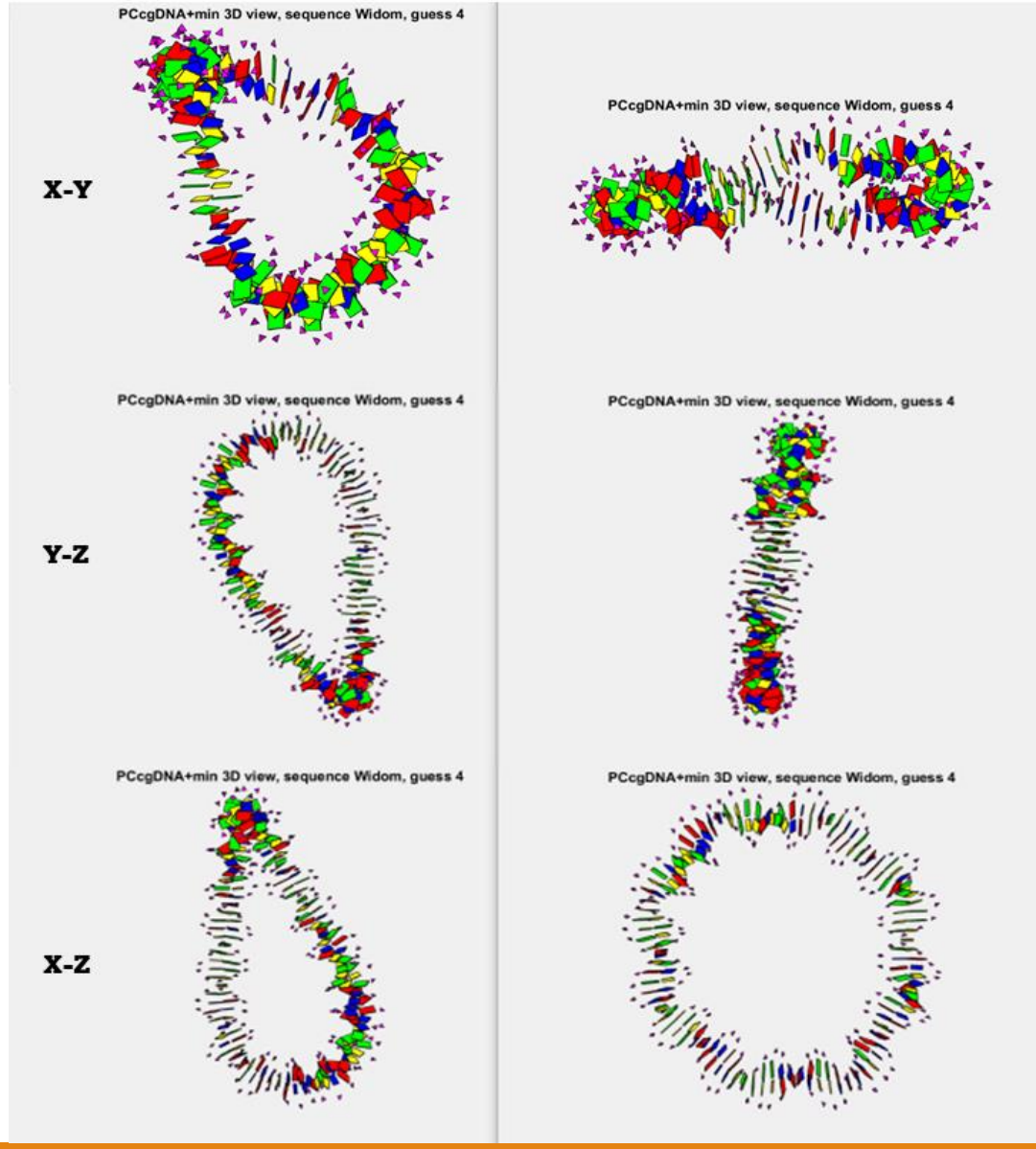


(d) Link: 8, Energy: 4293

bBDNA Initial guess

Trust Region

bBDNA | **bBDNA with gd-intras**



Results

- *Pyne 339* (4 i.g.):
  - GUESS 1: they all converge to the same solution, and they are all local minima
    (*Quasi-Newton + Trust Region*)

    → *Abs. difference:* ***1.0e-4***
    Negative eigenvalues: ***0***
  - GUESS 3: they all converge to the same solution, and they are all local minima

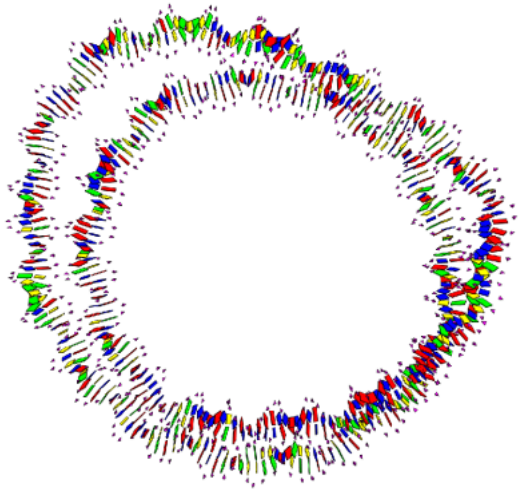
    → *Abs. difference:* ***1.0e-2***
    Negative eigenvalues: ***0***
  - GUESS 6: they all converge to the same solution, and they are all local minima

    → *Abs. difference:* ***1.0e-2***
    Negative eigenvalues: ***0***
  - GUESS 9: converge to the same point which is not a minimum

    → *Abs. difference:* ***1.0e-4***
    Negative eigenvalues: ***+5***

Guess 9, Pyne 339bp

- Trust Region method is really slow
- ***fminunc*** stops because step-size is too short
- Might be that in the starting point the function is too different from a quadratic so the algorithm need a really small region for the model to be adequate
- Using Quasi-Newton we obtain a configuration closer to the initial guess but with much lower energy

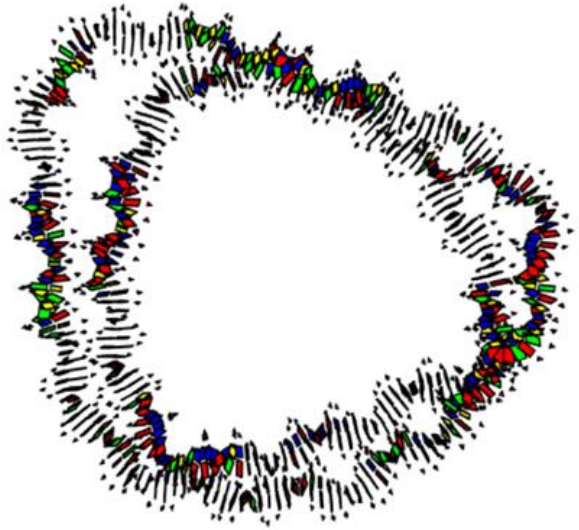


bBDNA Initial guess

Energy: 65140

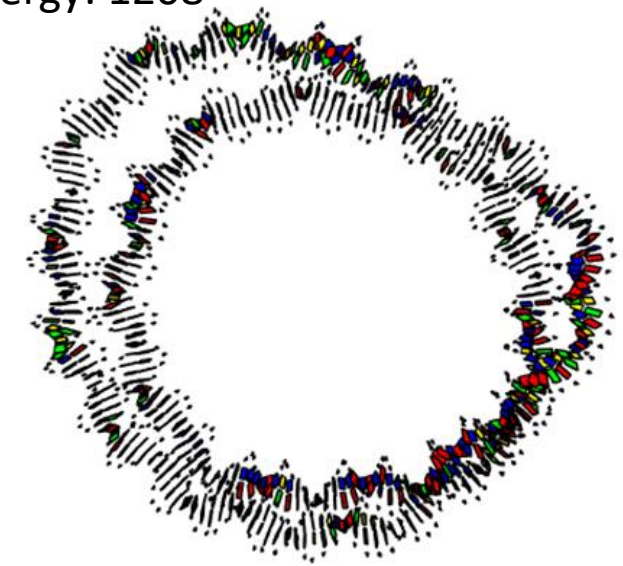
M. Beaud solution

Energy: 1206



Quasi-Newton solution

Energy: 1208



Conclusions

- ***Minimum eigenvalue of the Hessian*** always really small (1.0e-08 ~ 1.0e-04)
- ***Maximum eigenvalue of the Hessian*** always really big (~ 1.0e+4)
- ***Condition number of the Hessian*** really big (1.0e+09 ~ 1.0e+14)
- This leads to oscillations and might be the reason why the convergence to the exact solution is really slow
- The solutions with many negative eigenvalues also have gradients that are not close to the zero value
- All energy optimization converge to the same configuration with the two different initial points

Thank you for listening!

DO YOU HAVE ANY QUESTION?