1 cgDNAweb+

The cgDNAweb+ is an online interface which main purpose is the visualisation of sequence dependent minimum energy coarse grain configurations of B–form DNA predicted by the cgDNA and cgDNA+ models. The user can enter any standard DNA sequence ($\{A, T, G, C\}$ alphabet), and visualize the predicted minimum energy configuration in two different ways:

1. 3D view of the structure
2. 2D plots of the values of the helical parameters (see Figure (1) and Phosphate coordinates (see Figure (2)).

The cgDNAweb+ is accessible via the link `http://cgdnaweb.epfl.ch`. Read the instructions and reconstruct your favourite DNA sequence. Notice that you can visualize multiple sequence at a time.

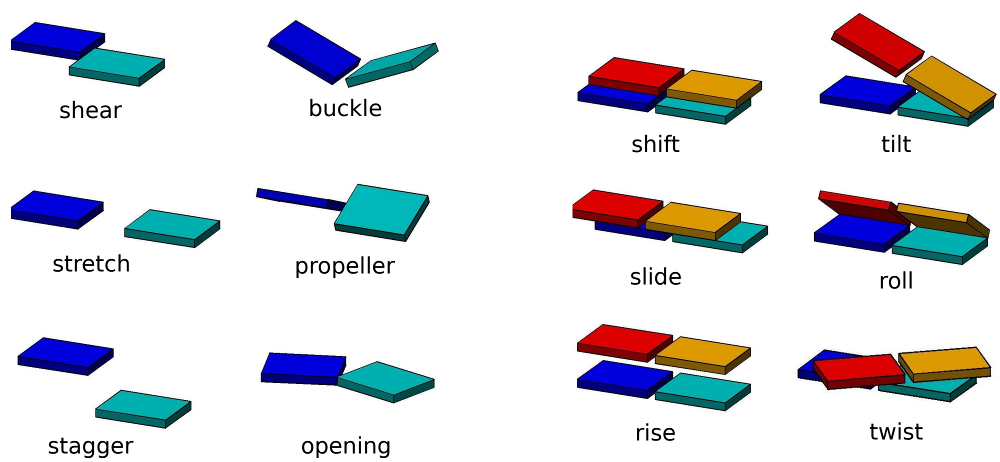

Figure 1:*Standard base pair helical parameters. Intras on the left, inters on the right, translation in columns 1 and 3, rotation (i.e. Cayley vectors) in columns 2 and 4.*

2 The cgDNA+ matlab package

2.1 Download and Run the package

The cgDNA+ MATLAB package can be downloaded from `https://lcvmwww.epfl.ch/teaching/modelling_dna/protected_files/codes_exercises/cgDNA+_matlab_package.zip`.

1. Start by reading the *README* file and by testing the cgDNA+ package on your version of MATLAB.
2. Run `main.m` with *cgDNA+ps1.mat* and be sure no error is reported.
3. Understand the inputs/outputs of the function `main.m`.

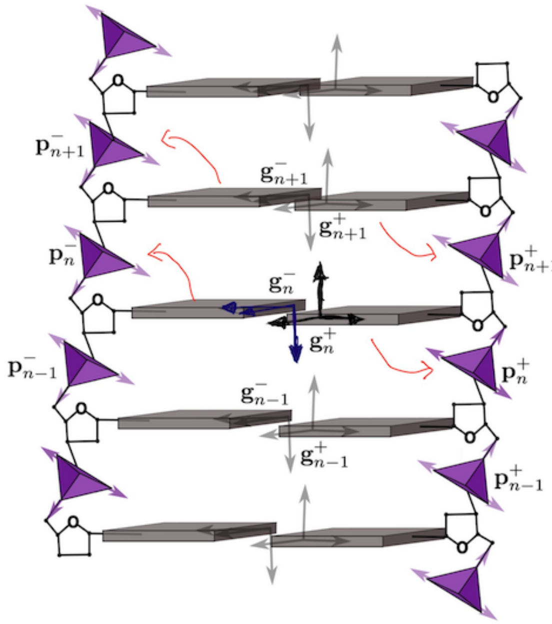

Figure 2: *Rotational (i.e. Cayley vectors) and translational coordinates of Phosphates are defined with respect to its corresponding base, ie for n^{th} Phosphate on Watson strand (strand with superscript +) relative coordinates are defined with respect to n^{th} base on Watson strand and similarly for Phosphates on Crick strand (strand with superscript -). Note that Phosphate coordinates are defined before the base flipping (see the direction of frames fitted in the bases).*

3 Optimisation in SE(3)

i) The first order necessary conditions for minimizing a function $I(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ are $\nabla I(\mathbf{x}) = 0$. This can be proved by considering a curve $\mathbf{x}(\varepsilon) \in \mathbb{R}^n$. The Taylor expansion along this curve is $\mathbf{x}(\varepsilon) = \mathbf{x}(0) + \varepsilon \mathbf{h} + O(\varepsilon^2)$ where $\mathbf{h} = \mathbf{x}'(0)$ is the tangent vector which can be any vector $\mathbf{h} \in \mathbb{R}^n$. Then,

$$\frac{d}{d\varepsilon} I(\mathbf{x}(\varepsilon))\bigg|_{\varepsilon=0} = 0 \Leftrightarrow \mathbf{h} \cdot \nabla I(\mathbf{x}) = 0, \forall \mathbf{h} \tag{1}$$

where

$$\nabla I(\mathbf{x}) \in \mathbb{R}^n = \left[\frac{\partial I}{\partial \mathbf{x}_1} \cdots \frac{\partial I}{\partial \mathbf{x}_n} \right]^T. \tag{2}$$

Then, as $\mathbf{h}$ is arbitrary $\mathbf{h} \cdot \nabla I(\mathbf{x}) = 0, \forall \mathbf{h} \Leftrightarrow \nabla I(\mathbf{x}) = 0 \in \mathbb{R}^n$

ii) Show that if $I(M) \colon \mathbb{R}^{n \times n} \to \mathbb{R}$, then the first order necessary conditions are $H \colon \nabla I(M) = 0$ $\forall H \in \mathbb{R}^{n \times n}$ or $\nabla I(M) = 0 \in \mathbb{R}^{n \times n}$, where $\{\nabla I\}_{ij} = \dfrac{\partial I}{\partial m_{ij}}(I(M))$ and $:$ is the Frobenius matrix inner product $A \colon B = \sum_i \sum_j a_{ij} b_{ij} (= Tr(A^T B))$.

iii) Show that if $I(R) \colon SO(n) \to \mathbb{R}$ then the first-order necessary conditions are $S \colon R^T \nabla I(R) = 0$ $\forall\ S = -S^T \in \mathbb{R}^{n \times n}$, or $R^T \nabla I(R)$ must be symmetric.

iv) Use parts i) & iii) to compute the optimal values of $\mathbf{r} \in \mathbb{R}^3$ and $R \in SO(3)$ where

$$I(\mathbf{r}, R) := \sum_{i=1}^{M} ||\mathbf{r} + R\boldsymbol{\alpha}_i - \mathbf{p}_i||^2, \tag{3}$$

and $\boldsymbol{\alpha}_i \in \mathbb{R}^3 \& \mathbf{p}_i \in \mathbb{R}^3$ are given vectors. The values of $\mathbf{r}\&R$ are the best fit frame to a given set of atomic coordinates of atoms $\mathbf{p}_i$, $i = 1, \dots, M$, whose idealized coordinates in the frame $(\mathbf{r}, R)$ are known to be $\boldsymbol{\alpha}_i$, $i = 1, \cdots, M$. In an exercise of a later series we will apply this result to actual MD data where the list of atom coordinates $\mathbf{p}_i$ make up a base, or make up a phosphate group.

4 Connection between the Cayley transform and the matrix exponential

In exercise 1.2 of Serie 3 we have shown that when a matrix $M \in \mathbb{R}^{n\times n}$ is defined by a Cayley transform of $N \in \mathbb{R}^{n\times n}$ such that $(I - N)$ is invertible, via

$$M = CayTra(N) := (I + N)(I - N)^{-1}, \tag{4}$$

then $M \in SO(3)$ if and only if $N = -N^T \in \mathbb{R}^{3\times 3}$. The function $CayTra : \mathbb{R}^{3\times 3} \to \mathbb{R}^{3\times 3}$ restricts to $Sk \to SO(3)$, $Q = CayTra([\mathbf{u}\times])$ where Sk is all 3×3 skew matrices $[\mathbf{u}\times]$ and Q is a proper rotation matrix. Moreover in Qu1.2 we showed that any rotation $Q \in SO(3)$ is a Cayley transform except for rotations through π.
Remark: Do not confuse the notations $CayTra$ and Cay (introduced in Qu1 Serie 4). Cay is a mapping from a rotation matrix to a vector. Basically, $\mathbf{u} = Cay(Q)$ gives the vector $\mathbf{u}$ corresponding to the skew matrix $[\mathbf{u}\times]$ that is the inverse Cayley transform of the rotation matrix Q. The inverse Cayley transform of Q is defined explicitly as $CayTra^{-1}(Q)$ so that $[\mathbf{u}\times] = CayTra^{-1}(Q)$ if and only if $\mathbf{u} = Cay(Q)$.

We showed in Qu 1.2 Serie 3 that, for $N = [\mathbf{u}\times]$, we have

$$Q := CayTra([\mathbf{u}\times]) := I + \frac{2}{1 + \|\mathbf{u}\|^2}\left([\mathbf{u}\times] + [\mathbf{u}\times]^2\right) \quad \text{[see equation 27 of Corr 3]} \tag{5}$$

$$[\mathbf{u}\times] = CayTra^{-1}(Q) := \frac{1}{1 + \mathrm{tr}Q}\left(Q - Q^T\right) \quad \text{[see equation 3 on Serie 3]} \tag{6}$$

$$\text{and } \|\mathbf{u}\| = \tan\left(\frac{\phi}{2}\right), \tag{7}$$

where ϕ is the rotation angle of Q about a unit rotation axis $\mathbf{n} = \frac{\mathbf{u}}{\|\mathbf{u}\|}$ (see Qu 2 in serie 2). Equation (5) above is another form of Euler-Rodrigues formula as already seen in Serie 3. Note that equations (5) & (6) give explicit forms of the mapping and its inverse between vectors $\mathbf{u} \in \mathbb{R}^3$ & $SO(3)$ rotation matrices through an angle $[0, \pi)$ (for which $(1 + trQ) > 0$).

The matrix exponential for any matrix is defined through the Taylor series in equation (4) of Qu3 Serie 2.

1. Show that for any matrix $N \in \mathbb{R}^{n\times n}$ for which $(I - N)$ is invertible

$$\exp(N) - CayTra(\frac{1}{2}N) = -\frac{1}{12}N^3 + \mathcal{O}(\|N\|^4) \tag{8}$$

where exp is defined in equation (4) of Serie 2, and $CayTra$ is defined in (4). Thus we see if the matrix N is small the Cayley transform of $N/2$ is very close to $\exp(N)$.

2. For any $\mathbf{v} \in \mathbb{R}^3$ with $\|\mathbf{v}\| < \pi$, prove that if $Q = \exp([\mathbf{v}\times])$,

$$[\mathbf{v}\times] = \frac{\theta}{2\sin(\theta)}\left(Q - Q^T\right), \tag{9}$$

where

$$\|\mathbf{v}\| = \theta, \tag{10}$$

so that (equation (5) of Serie 2), (9) and (10) are the analogues for exponential coordinates (sometimes called logarithms) to equations (5), (6) and (7) for the Cayley transformation and the Cayley vector. **Note:** One can prove directly that $\log(Q) := [\mathbf{v}\times]$, where log is the matrix logarithm.

Remark: This shows that exponential logarithmic coordinates $\mathbf{v}$ are parallel to the unit rotation axis $\mathbf{n}$ with $\|\mathbf{v}\| = \theta$ and the Cayley vectors $\mathbf{u}$ are also parallel to the unit rotation axis $\mathbf{n}$ with $\|\mathbf{u}\| = \tan\left(\frac{\phi}{2}\right)$.

5 Completing the square in vector quadratic forms (will use result in 6th lecture)

Suppose that $c_i \in \mathbb{R}, \boldsymbol{\mu}_i \in \mathbb{R}^n \& K_i \in \mathbb{R}^{n\times n}$ for $i = 1, 2, 3 \ldots, N$ with $K_i = K_i^T$ but K_i not necessarily invertible (and so not necessarily positive definite). What are the conditions such that a sum of shifted quadratic forms can be written as a single quadratic form? That is when does there exist $c \in \mathbb{R},\ \boldsymbol{\mu} \in \mathbb{R}^n \& K \in \mathbb{R}^{n\times n}$ such that

$$\sum_{i=1}^{N} \left[\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i) \cdot \mathbf{K}_i(\mathbf{x} - \boldsymbol{\mu}_i) + c_i\right] = \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{K}(\mathbf{x} - \boldsymbol{\mu}) + c. \quad \forall \quad \mathbf{x} \in \mathbb{R}^n. \tag{11}$$

Compute the expressions for $c, \boldsymbol{\mu} \& \mathbf{K}$ when possible.