

Genome folding and function: from DNA base pairs to nucleosome arrays and chromosomes

Ryan Ding | Mentor: Wilma Olson



Background

- DNA holds important genetic information about our bodies
 - Code comprised of four bases: adenine (A), thymine (T), cytosine (C), and guanine (G)
 - Represented via long sequences of characters
- Bases pair specifically with one another
 - Adenine pairs with thymine
 - Cytosine pairs with guanine
- **DNA folds!**
 - Important for packaging and functions

```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCTCTTTTCTTATCATTGACATTTAACTCTGGGGCAGTCTCTCGGTAGAACGGCGCTGTAGATCT
GCCACTTCCCCGCGAGCGCGGTGAGAAAGTGTGGAAACCGCGCTGCCAGGCTCACCTGCCTCCCCG
CCTCCGCTCCCAGGTAACGCCCCGGCTCCGGCCCCCGCCGCTCGGGCCCCGGGGGCTCTCCGCTG
CCAGCGACTGCTGTCCCAAATCAAAGCCCCGCCCAAGTGGCCCCGGGGCTGATTTTGGCTTTAAAG
GAGGCATACAAAGATGGAAGCGATTACTGAGGGCAGGATAGGAAGGGGGTGGAGGAGGACTGTCTT
TGCCGAGTGTGCTCTCTGCAAAAGTAGCAAAATGTTCCACTCTAAGAGTGGACTCCAGTCGCGCCCT
GAGCTGGGAGTAGGGGCGGAGTCTGCTGCTGCTGCTGCTAAAGCCACTCGCACCGCAAAAATGCA
GGAGTGGGGACGCATTTGCATCCAGACTCTCTGATCGCAGTTACAGACATCCACGCTTGGGAAG
TCCGTACCCGCGCTGGAGGCTTAAAGACACCTGCGCGGGTGGGGAGTGCAGCAGAGTTTCCC
GCGGTTGCAAAAGTGCAGATGGCTGGACCGCAAAAGTCTAGAGATGGGGTTCGTTTTCAGAAAGCCG
```

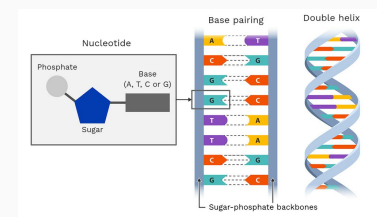


Figure 1 (top). Example of a DNA sequence file format, FASTA. From *Computational Genomics in R*, by Altuna Akalin, 2020. Retrieved from <https://compgenomr.github.io/book/fasta-and-fastq-formats.html>

Figure 2 (bottom). DNA structure broken down, with the backbone comprised of sugar phosphate and the base pairs represented in the bases described. From *Structure of DNA*. Retrieved from <https://theory.labster.com/structure-dna/>

Energy Optimization

- **DNA folds to fit into nuclei and fluctuates about a 'resting state'**
- Different sequences lead to different conformational structures
 - DNA folding is highly dependent on its stored elastic energy, which varies by sequence and environmental conditions; two sequences may end up having different conformations!
- Folding can be described in the orientation of base pairs

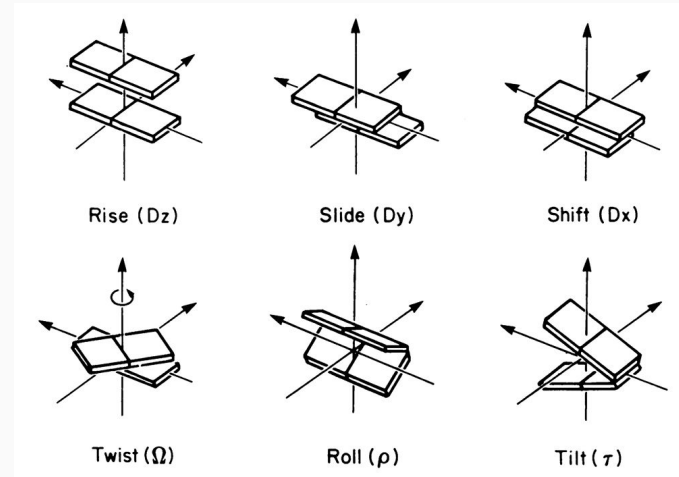


Figure 3. The six fundamental orientation parameters base pairs can be described in. From *Yildirim RNA Dynamics Lab*, 2017. Retrieved from <https://cosweb1.fau.edu/~iyildirim/DNAbending.html>

Existing Works

- **emDNA** software developed by the Olson Lab to minimize the energies of DNA and joint DNA-protein complexes
- **Tetrameric step energy optimization model** by Zoe Wefers (DIMACS 2021) in emDNA, which uses a sequence of 4 adjacent base pairings to perform energy calculations

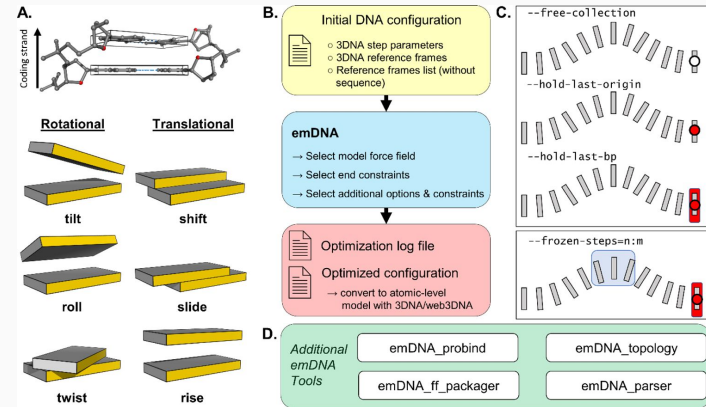


Figure 4. Diagram of emDNA configurations and workflows. From *emDNA - A Tool for Modeling Protein-decorated DNA Loops and Minicircles at the Base-pair Step Level*, by Young et. al., 2022. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9177622/>

Goals

1. Investigate sequence-dependency within loop-like DNA structures and see what types of conformations occur when two ends of a sequence are held in place using tetrameric energy calculations.
2. Integrate these energy calculations and findings into the emDNA software such that it can handle loop-like structure calculations.

References and Acknowledgements

Clauvelin & Olson, *Synergy between Protein Positioning and DNA Elasticity: Energy Minimization of Protein-Decorated DNA Minicircles*, Journal of Physical Chemistry

Young et. al., *Revisiting DNA Sequence-Dependent Deformability in High-Resolution Structures: Effects of Flanking Base Pairs on Dinucleotide Morphology and Global Chain Configuration*, Life 2022

This presentation was a part of the 2024 DIMACS REU program at Rutgers University, supported by NSF grant CNS-2150186.