# THE INFLUENCE OF DNA SEQUENCE ON THE ELASTICITY OF LAC REPRESSOR-BOUND LOOPED STRUCTURES

K. Grover, B. Saha, A. Sankaralingam, R. Shete, R.T. Young, PhD[1]

[1] Office of Undergraduate Research-First-Year Innovation & Research Experience
University of Maryland-College Park, MD, USA

## Introduction

- Lac Repressor Proteins (LacR) are proteins that regulate gene expression in the bacteria E.COLI. but they play an important role in helping us understand DNA function and structure relationships for all genomes, including humans.
- LacR binds to DNA at two different sites known as operators, forming loops that manipulate gene expression [1].
- The catch is that there are 8 possible configurations of these loops - 2 versions each of the A1, A2, P1, and P2 loop types - due to the various potential orientations of the binding sites.
- The focus of our project is to investigate how sequence influences these loop structures by making computational models with an updated, knowledge-based potential.
- Additionally, the position of the DNA base pairs are characterized by parameters like tilt, roll, twist, slide, and rise, so our project also explored the variation in these parameters caused by shifting the TAL sequence (a specific portion of the DNA sequence)

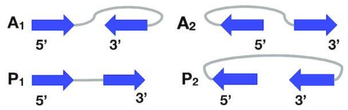


***Figure 1:*** *Shows difference in direction for the different loops [2]*


***Figure 2****: The image shows the movement of TAL sequence about the 133.5 bp as it moves closer to the O2 site. By incrementally adjusting the TAL-region's position, +1 along the DNA sequence, 57 DNA sequences will be created. Following this, optimizations will be conducted.*

## Methods

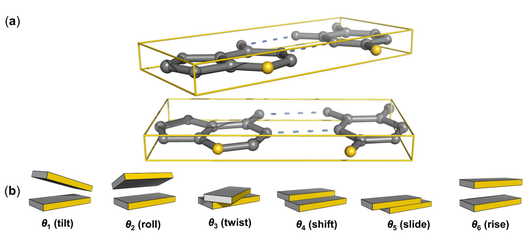


***Figure 3****: (a) Showcases an atomic model of DNA, emphasizing the purine and pyrimidine atoms situated along the DNA minor groove. (b) illustrates the orientation of DNA base pairs using parameters: tilt, roll, twist, shift, slide, and rise[3].*

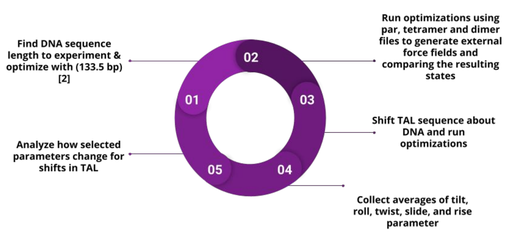


***Figure 4:*** *This figure illustrates the systematic steps undertaken in the study: Initially, each of the four loops is optimized with a new sequence, followed by the use of tetramer, and dimer files to generate external force fields[4]. Subsequent stages involve shifting the TAL sequence relative to the DNA and conducting further optimizations for new sequences, but also taking account for parameters.*

## Visualization of Loop Structure

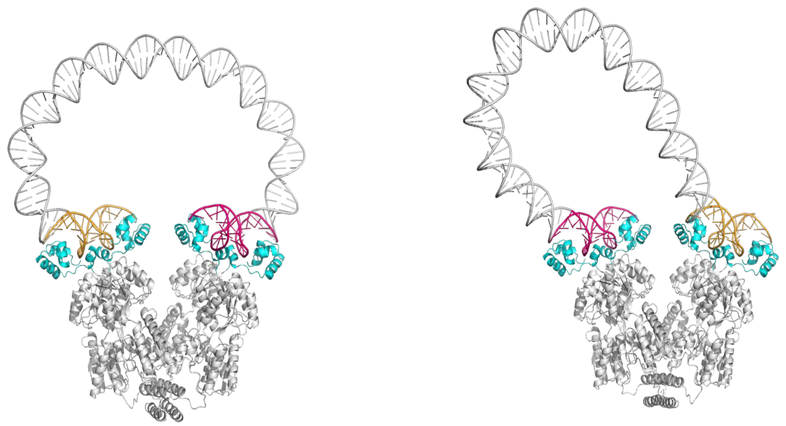

***Figure 5:*** *The top left image depicts the A2 loop, top right shows the P2 loop, bottom right shows the P1 loop and the bottom right shows the A1 loop, and their respective interactions with the lac repressor. The interaction between the DNA and protein is colored as follows: $O_{sym}$ as hot pink, $O_2$ as orange, and the protein as blue. The DNA have distinct shapes, due to their respective interactions with their lac repressor [5].*

FIRE: THE FIRST-YEAR INNOVATION & RESEARCH EXPERIENCE

## Results

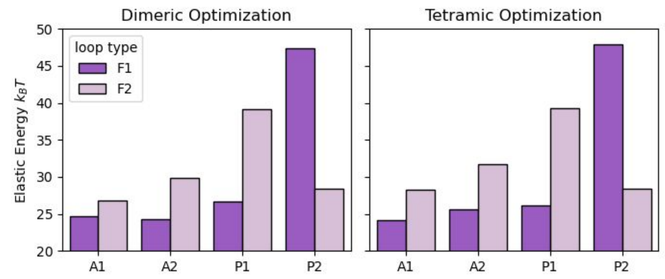


***Figure 6:*** *Compares dimeric and tetrameric optimization energy levels for F1 and F2 loops. Dimeric means we are optimizing by two base pair while tetrameric is with four base pairs. We can see the A1F1 and A1F2 have the lowest energy levels and the F2 family is higher in energy except for P2.*

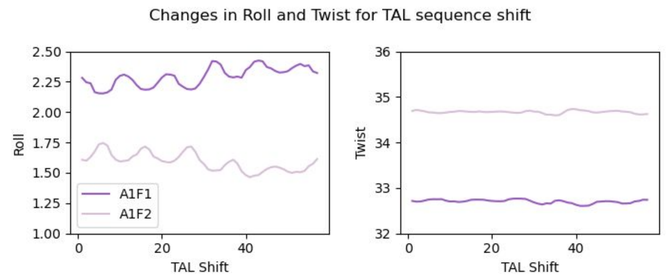


***Figure 7:*** *Shows difference in roll and twist for shifts in the TAL sequence for both families of the A1 loop. We see roll is higher for for F1 and twist is higher for F2. We also see that roll is periodic for every 10.5 base pairs.*

## Conclusion & Next Steps

- DNA biologically prefers lower energy states over higher energy states
- Thus, for A1, A2, and P1 loops, the F1 version is more likely to form than the F2 version, since the F1 energy values are lower.
- On the other hand, the F2 version is more likely to form than the F1 version for a P2 loop, since the F2 energy values are lower
- With Dimeric Optimization, DNA is most likely to form A2F1 lac loops. However, with Tetrameric Optimization, DNA is most likely to form A1F1 lac loops.
- As the TAL sequence gets shifted in the A1F1 AND A1F2 loops, the roll values change in a periodic manner every 10.5 base pairs, which signifies a full turn of DNA
- Furthermore, as the F2 version has higher twist and energy values then the F1 version for the A1 loop, this indicates that twist and energy values are likely directly proportional
- In the future, we plan to investigate the effects of increasing or decreasing the number of base pairs in DNA sequence being studied to understand their impact on DNA structure and function.
- We would also like to utilize machine learning for predictive analysis, focusing on sequence behaviors and energy state valuations.

## Works Cited

Garcia, H.G. et al. Biopolymers 85, 2, 2007.
Tse DH et al. Nucleic Acids Res, 2021.
Young RT *et al.* Life (Basel), 2022.
Clauvelin, N.; Olson, J. Phys. Chem. B 2021.
The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.