



BERKELEY LAB

Bringing Science Solutions to the World



U.S. DEPARTMENT OF
ENERGY

Office of Science

Analyzing Different Conformations of DNA

EX02 - Veer Mistry, Karthik Venuturimilli, Mariah Lillard-Richner, Anthony Zhou, Feranmi Olawoye, Thu Shun Lei

Mentor: Dr. Yang Ha

Alumni: Lachlan Mossman, Colin Zhang, Robert Chuang

Table of contents

Analyzing Different Conformations of DNA

01

Introduction: Why This Matters

02

Scientific Background

03

Methods & Tools Used

04

Results

05

Data Analysis

06

Conclusions & Next Steps

07

Acknowledgements

Why Predicting DNA Interactions Matters



Standard
B-DNA
Model

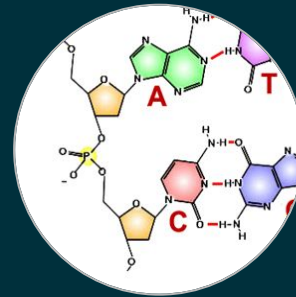
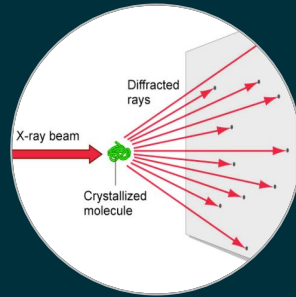
- ❑ DNA shape directs function
- ❑ Essential for **gene editing**
- ❑ Locating **protein binding** sites
- ❑ Crucial in **drug targeting**
- ❑ Enables better diagnostics



Barriers in DNA Structure Prediction

Existing methods for predicting **DNA shape** are limited and need improvement

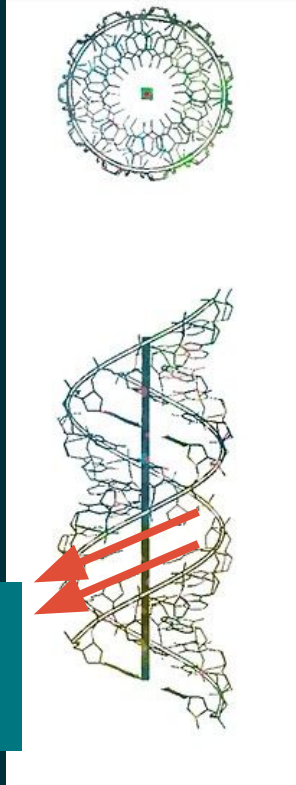
Genomic **data** often lacks clear structural **information**



DNA Conformations

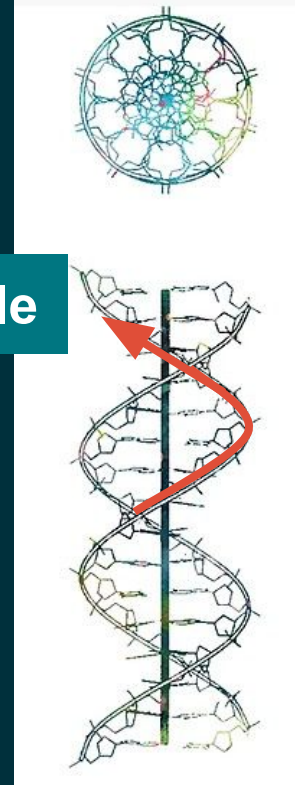
Key structural differences:
varying groove sizes, twist
angles, and propellor twists

A-DNA



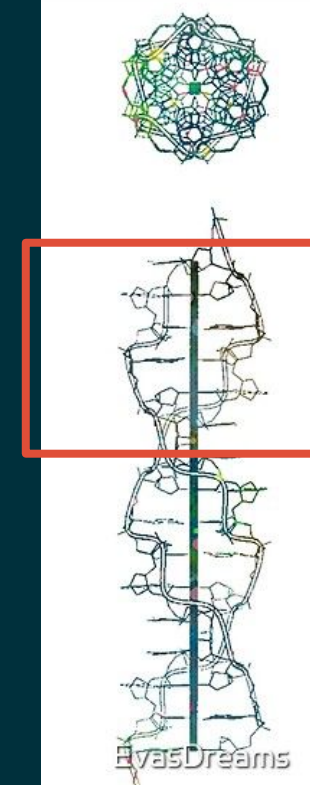
Propellor
twist

B-DNA



Twist angle

Z-DNA



Groove size

Methods

- Google Deepmind's **AlphaFold 3** AI model used to predict the structure of many macromolecules - the newest model allows the addition of predicting DNA structures
- 10 Base pair palindromes inputted into AlphaFold of the following types:
 - First position G/A/T/C
 - Second position G/A/T/C
- Results in 1024 different structures

Sample palindrome: GATAGCTATC

Each of us were assigned 256 structures to input and analyze through Alpha Fold and record our analyzation of structures in excel

AlphaFold Server allows you to model a structure consisting of many biological molecules



Entity type
DNA



Copies
2

>Paste sequence or fasta
GATAGCTATC

+ Add entity

Continue and preview job

GATAGCTATC

[← Back](#) [Download](#) [Clone and reuse](#) [Feedback on structure](#)

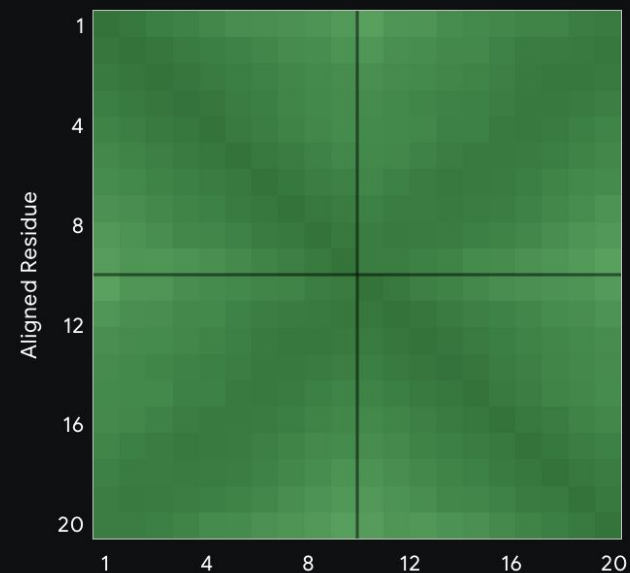
Very high (pLDDT > 90)

Confident (90 > pLDDT > 70)

Low (70 > pLDDT > 50)

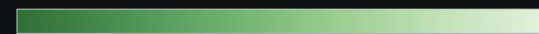
Very low (pLDDT < 50)

ipTM = 0.04 pTM = 0.05 [learn more](#)



Display Chain Borders

Scored Residue



0 5 10 15 20 25 30
Expected Position Error (Ångströms)

Methods

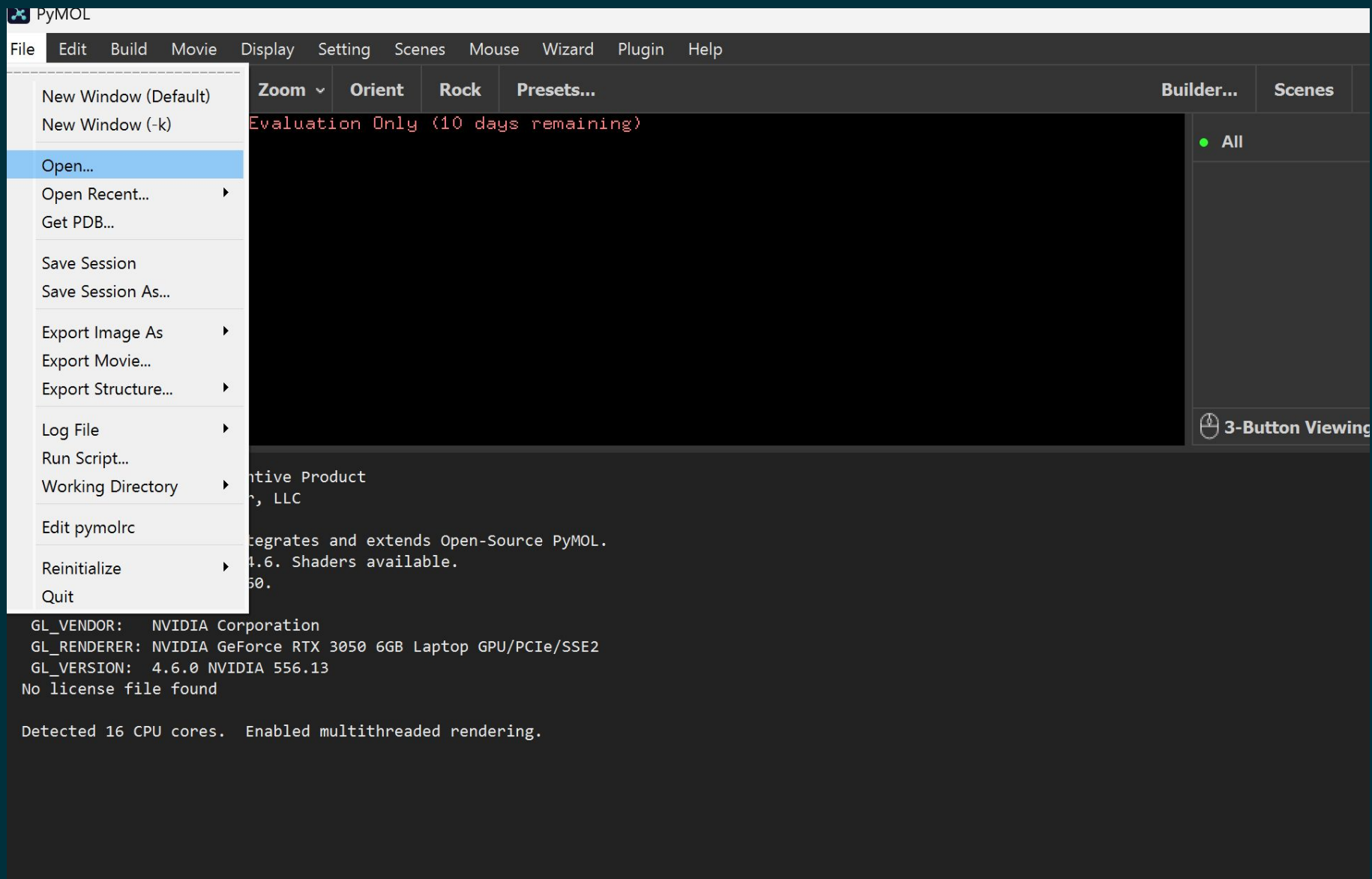
Pymol and Web 3DNA Fiber used to compare models and run calculations by aligning two models

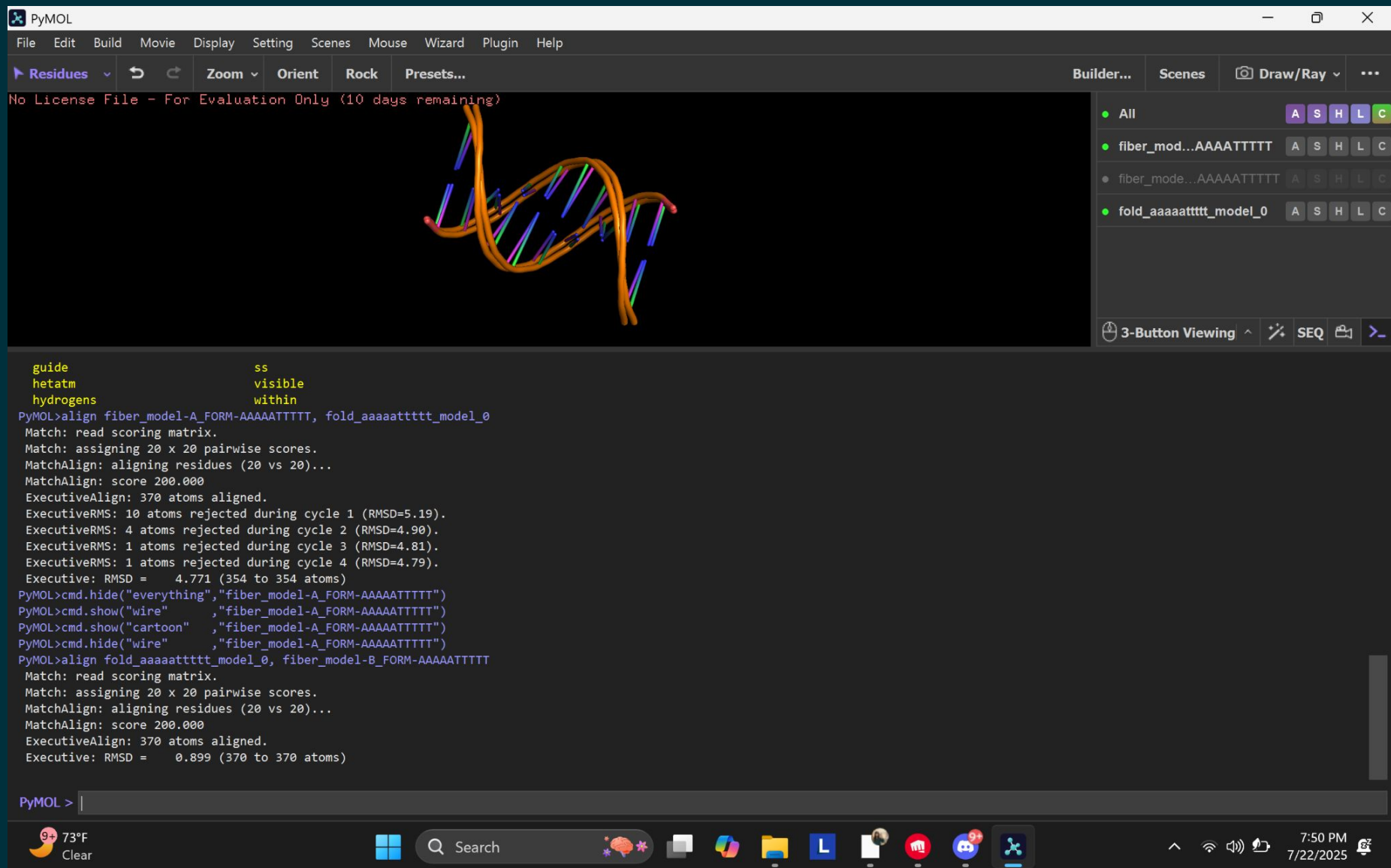
Commonly used fiber models

Short description	Twist (°)	Rise (Å)	
RNA (generic sequence: A, C, G, U)	32.7	2.548	use this model
A-DNA (generic sequence: A, C, G, T)	32.7	2.548	use this model
B-DNA (generic sequence: A, C, G, T)	36.0	3.375	use this model
C-DNA (generic sequence: A, C, G, T)	40.0	3.32	use this model
Pauling triplex model (generic sequence: A, C, G, T, U)	105.0	3.40	use this model
poly(l) : poly(l) : poly(l) : poly(l)	31.3	3.410	use this model

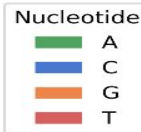
 fiber_model-B FORM-AAAAATTTTT.pdb	7/22/2025 7:42 PM	PDB File
 fiber_model-A FORM-AAAAATTTTT.pdb	7/22/2025 7:41 PM	PDB File
 fold_aaaaatddd	7/22/2025 7:44 PM	File folder

1. Downloaded the structure file from AlphaFold and downloaded both A form and B form fiber models from Web 3DNA

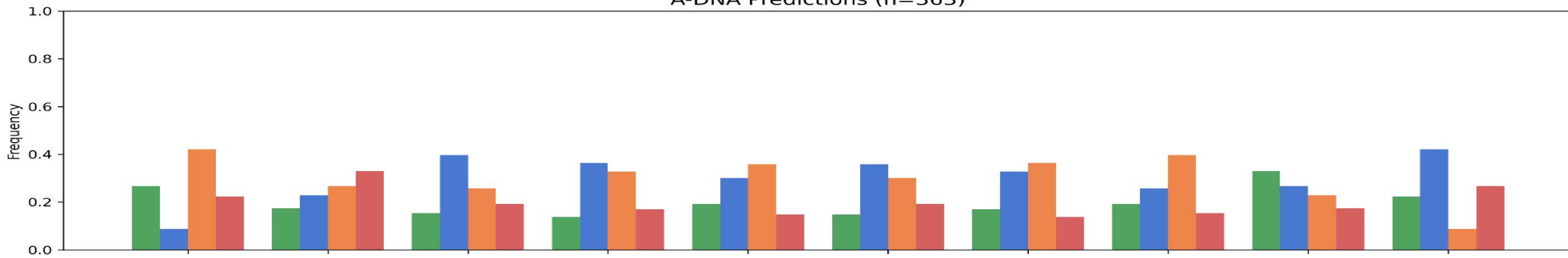




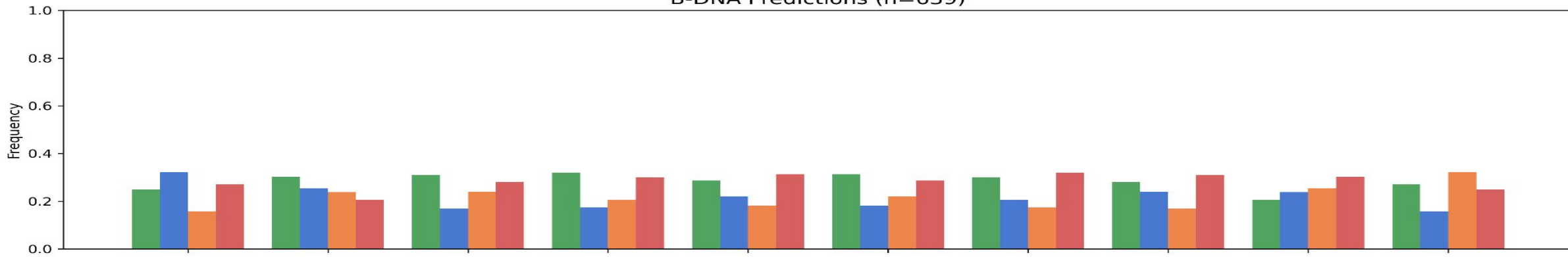
Nucleotide Frequency by Position for Predicted DNA Forms



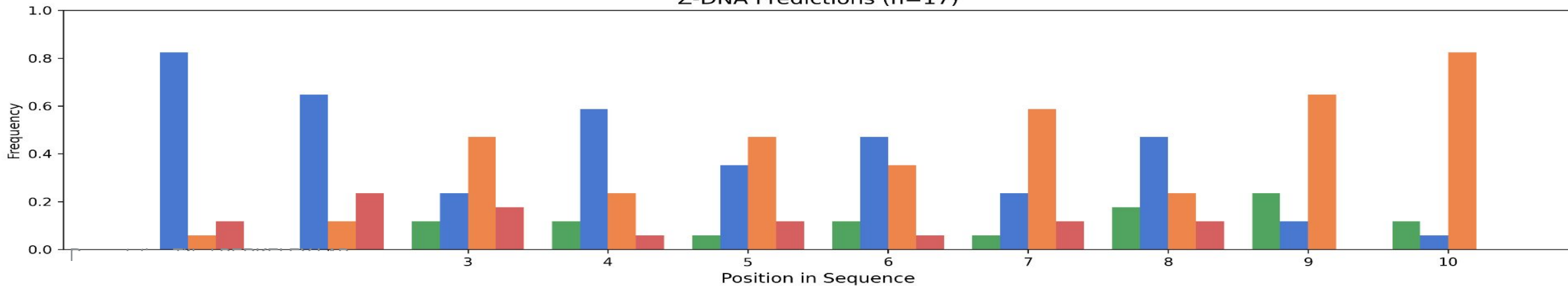
A-DNA Predictions (n=363)



B-DNA Predictions (n=639)

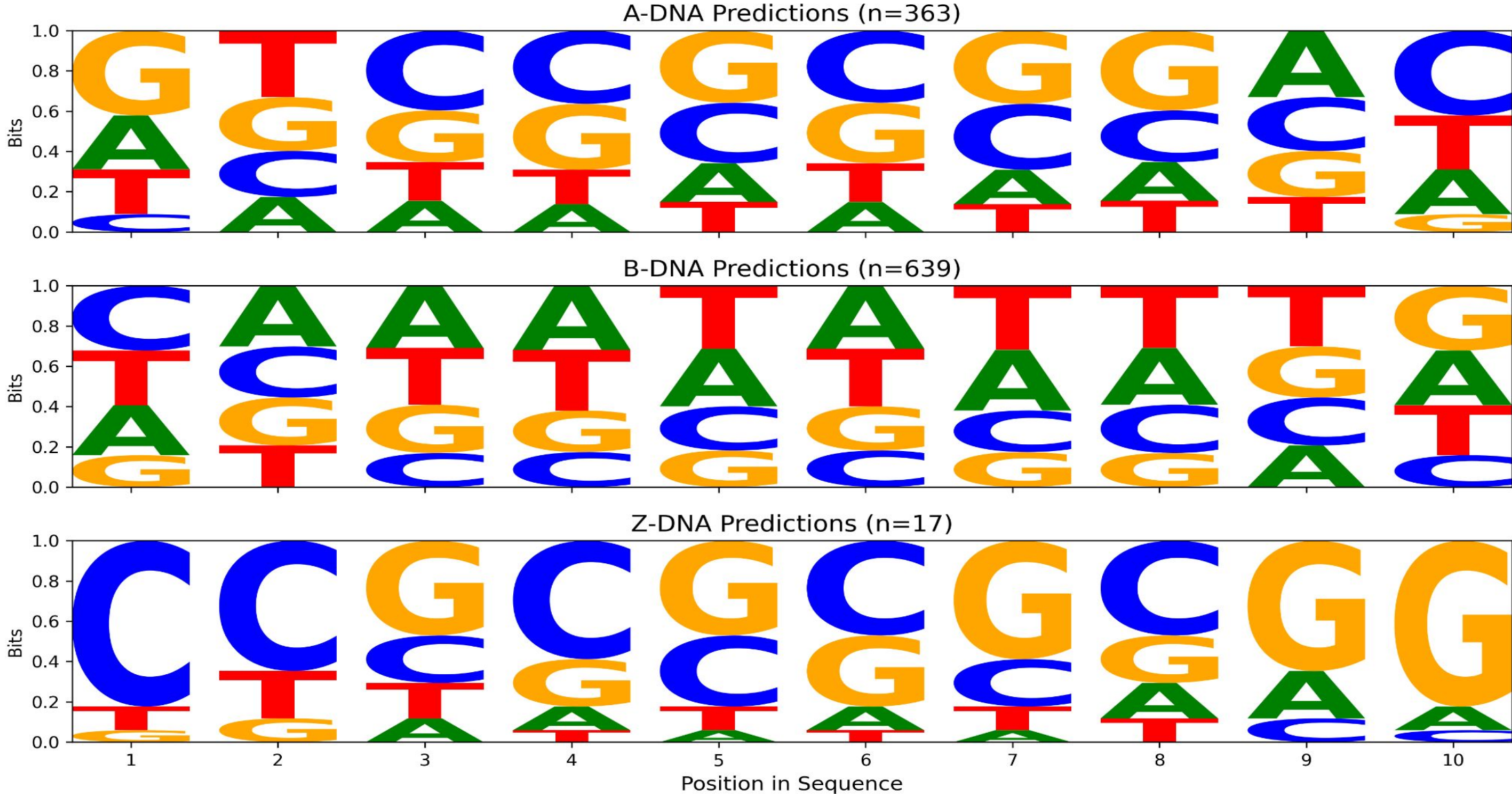


Z-DNA Predictions (n=17)



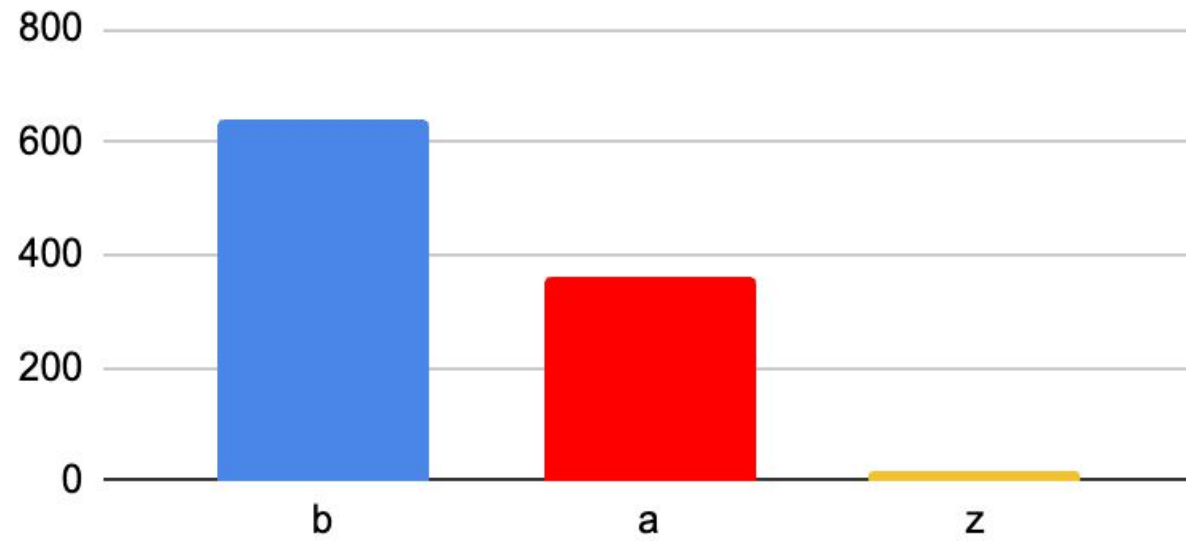
Data Analysis

Sequence Logos of DNA Forms Predicted by AlphaFold 3



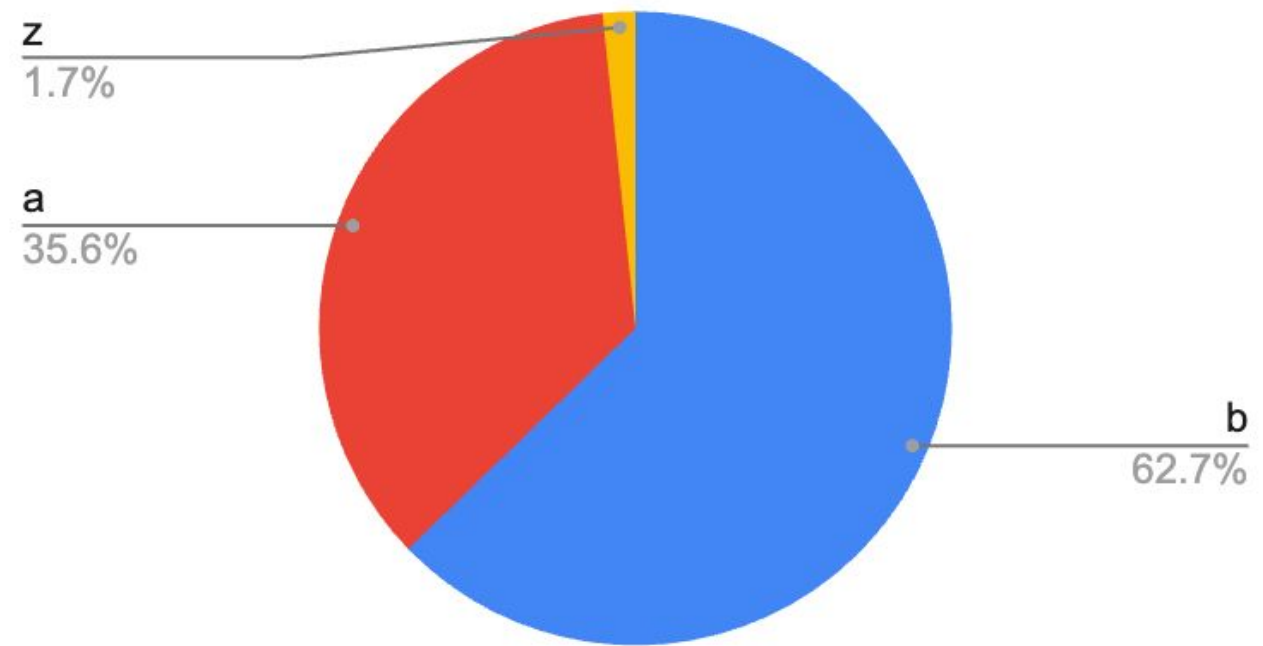
Results

Count of Forms



Count of Forms

Count of Forms



Results Analysis & Final Takeaways

A-DNA - 363 occurrences

- Displays a relatively balanced frequency of all nucleotides across positions.

B-DNA - 639 occurrences

- Frequencies are more uniform across all positions and nucleotides.
- Indicates no strong positional bias, suggests a stable nucleotide distribution.

Z-DNA - 17 occurrences

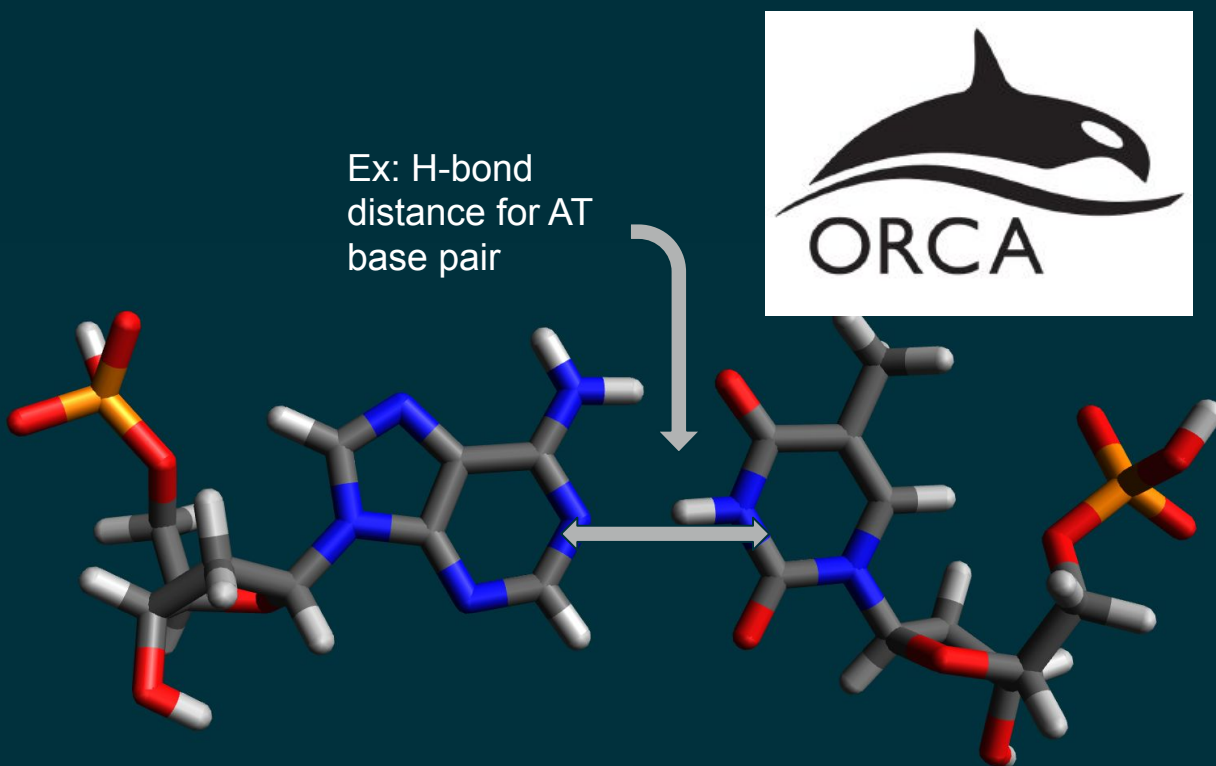
- Strong dominance of G (blue) at positions 1 and 2 (over 0.6–0.8 frequency).
- C (orange) increases sharply at positions 7 and 10 (up to 0.8).
- Small sample size (n=17) might exaggerate patterns.

The data for B-DNA shows a balanced nucleotide frequency across all positions, which aligns with the fact that B-DNA is the most stable and prevalent form in natural cellular conditions.

Future steps

Quantum Chemistry Calculations

- investigate physics behind DNA conformation trends
- Why does C/G lead to Z-form and not A/T?
- ORCA lets us examine molecular structure



Machine Learning

- Machine learning is all about creating algorithms to predict and deduce based on training
- How can we apply it?
 - detecting common patterns
 - help recognition with easier tasks

Conclusion

- Teamwork
 - Throughout this experience, we worked as a team collaborating which allowed us to share responsibilities, exchange ideas and learn from each other. The collective effort from each one of us improved the overall quality of our research
- Pymol - Alpha Fold
 - Using Pymol and Alpha Fold allowed us to visualize complex protein structures and better understand their configurations; these tools were essential for understanding the importance of our research.
- The importance of configuration of DNA
 - The configuration of DNA determines how it functions and interacts with other proteins. Understanding the structures of DNA is critical for advancements technology.

Acknowledgements

